Table 5. Polymorphism in LTRs on chromosome III

Intergenic region	Europe						Far East					
	No.[1]	No. sites[2]	*S*	θ_π x1,000	θ_S x1,000	*D*	No.[1]	No. sites[2]	*S*	θ_π x1,000	θ_S x1,000	*D*
YCL069W - YCL068C	2	808	24	11.6	9.9	0.8	2	639	4	2.5	2.4	0.1
YCL024W - YCL021W-A	5	1075	10	3.8	3.4	0.6	5	952	12	5.0	5.8	-1.1
YCL021W-A - TCL018W	2	414	9	8.3	8.6	-0.2	1	182	3	8.5	7.2	1.0
YCL001W-B - TCR002C	1	373	0	0.0	0.0	NA	1	374	0	0.0	0.0	NA
YCR005C - YCR007C	1	311	6	4.4	6.4	-1.2	1	318	1	0.8	1.2	-1.1
YCR007C - YCR008W	2	287	2	3.8	2.4	1.8	2	386	5	6.0	5.4	0.6
YCR015C - YCR016W	3	899	8	2.8	3.2	-0.6	3	460	4	4.2	3.4	1.1
YCR018C - YCR019W	4	857	8	1.8	3.1	-1.7	4	1,076	7	2.8	2.6	0.4
YCR027C - YCR028C	2	316	1	0.5	1.0	-1.1	3	625	7	4.6	4.3	0.3
YCR061W - YCR063W	1	93	1	4.7	3.7	0.7	1	93	0	0.0	0.0	NA
YCR095C - YCR096C	1	369	0	0.0	0.0	NA	1	369	2	1.8	2.3	-0.9
Mean				3.8	3.8	-0.1				3.3	3.2	0.04
C.I.					2.1, 5.7	-0.8, 0.6				1.9, 4.5		-0.5, 0.5

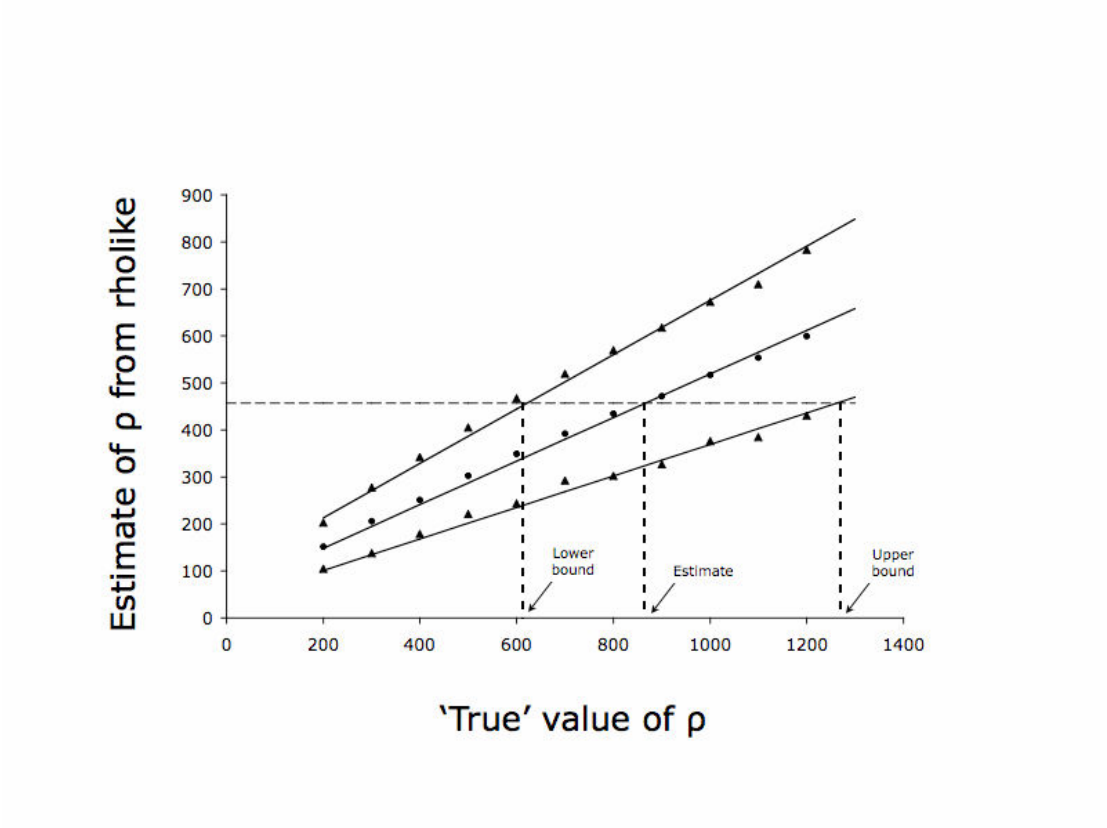
NA, any site at which any strain has a gap is excluded on the grounds that it may be an insertion and therefore more recent and less polymorphic.

[1] No. of segments with LTR homology.

[2] No. of sites analyzed (Variscan, option numnuc = 4).

Calculation of Rho Using Rholike

The algorithm implemented in rholike can give significantly biased estimates of ρ, and Li and Stephens (1) suggest using simulations to get unbiased estimates. Accordingly, we generated simulated datasets by using ms (2) and the observed values for number of alleles (*n*), number of segregating sites (*S*), number of base pairs (l), and ρ ranging from 100 to 1,200 in steps of 100. Five hundred datasets were generated for each value of ρ, and each one was used as input to rholike. As an example, Fig. 5 below shows the median estimated ρ as a function of the true input ρ, and also the 2.5% and 97.5% quantiles of the estimated ρs, using *n*, *S*, and l from the European population. Using the actual data for the European sample, rholike gave an uncorrected estimate of ρ = 457 (horizontal dashed line). We estimated the true ρ for the population and 95% confidence limits as indicated on the plot. These are then converted to values per kb by dividing by the length of the alignment (282 kb).



For consistency, estimates and confidence limits using the Wakeley and Pairwise methods were obtained in the same way.

Estimating the Frequency of Intratetrad Mating

Consider a population in which a fraction t of zygotes are derived by random outcrossing, s_h are derived from haplo-selfing, and s_i are derived by intratetrad mating ($t + s_h + s_i = 1$). Among the outcrossed zygotes the heterozygosity (relative to Hardy-Weinberg proportions) will be 1. Among the haplo-selfed zygotes heterozygosity will be 0. Finally, the effect of intratetrad mating is to reduce heterozygosity such that it will be $(2 + e^{-3x})/3$ times what it was in the previous generation, where x is the map distance (in Morgans) between *MAT* and the locus in question (3, 4). Therefore, across all zygotes, the heterozygosity (H_1) in any particular generation will be

$$H_1 = t\,(1) + s_h\,(0) + s_i\,H_0\,(2 + e^{-3x})/3,$$

where H_0 is the heterozygosity in the previous generation. At equilibrium,

$$H_1 = H_0 = H = \frac{3t}{3-\left(2+e^{-3x}\right)s_i}$$

Thus heterozygosity is expected to decline as a function of distance from *MAT*. For a region of the genome to one side of *MAT*, starting a Morgans away and ending b Morgans away, the average equilibrium heterozygosity within that region will be

$$\overline{H}_{ab} = \frac{\int_a^b H dx}{b-a} = \frac{t\left(\mathrm{Ln}\left[s_i - e^{3b}\left(3-2s_i\right)\right] - \mathrm{Ln}\left[s_i - e^{3a}\left(3-2s_i\right)\right]\right)}{\left(b-a\right)\left(3-2s_i\right)}$$

In particular, the ratio of average heterozygosity in the 20-kb regions on each side of *MAT* to average heterozygosity for the whole chromosome (which runs from 188 kb to the left of *MAT* to 91 kb to the right) is expected to be:

$$\frac{\overline{H}_{MAT}}{\overline{H}_{chrom}} = \frac{\frac{\int_0^{0.096} Hdx}{0.096}}{\frac{\int_0^{0.9024} Hdx + \int_0^{0.4368} Hdx}{0.9024 + 0.4368}} = \frac{13.95(Ln[4.00127 - 3.66751s_i] - Ln[3 - 3s_i])}{Ln[44.9618 - 30.9745s_i] + Ln[11.123 - 8.41531s_i] - 2Ln[3 - 3s_i]}$$

In deriving this expression we have converted physical distances to genetic distances by using the chromosome average of 0.0048 Morgans/kb (5). Note that this ratio is a function of just one variable, s_i. We do not have information on relative heterozygosity near MAT, but we do have information on ρ, which is a linear function of H (recall that $\rho = 4Nr(1 - F) = 4NrH$). Thus we can estimate the ratio of heterozygosity near the MAT to that for the whole chromosome by estimating the same ratio for ρ:

$$\frac{\overline{\rho}_{MAT}}{\overline{\rho}_{chrom}} = \frac{4Nr\overline{H}_{MAT}}{4Nr\overline{H}_{chrom}} = \frac{\overline{H}_{MAT}}{\overline{H}_{chrom}}$$

In our datasets the ratio of ρ near *MAT* to that for the whole chromosome does not differ significantly between Europe [1.9 (95% C.I.: 0.9-3.8)] and Far East [2.7 (0.3-15.2)], and combining data from the two populations we estimate the ratio to be 2.1 (1.1-4.2). We can then calculate s_i to be 0.94, with bounds of 0.41 and 0.999.

1. Li N, Stephens M (2003) *Genetics* 165:2213-2233.

2. Hudson RR (2002) *Bioinformatics* 18:337-338.

3. Kirby GC (1984) *Heredity* 52:35-41.

4. Zakharov IA (2005) *Russ J Genet* 41:402-411.

5. Cherry JM, *et al.* (1997) *Nature* 387:67-73.