

Supporting Information

Boeras et al. 10.1073/pnas.1103764108

SI Methods

Using the Test of Random Transmission analysis, given $n = 8$ transmission pairs, we address the question of whether or not the transmitted sequence is randomly distributed between rakes and nonrakes. All sequences were sampled from the GT. We pick a distance threshold, D , and define a cluster as any subset of sequences within distance D from one another. In other words, any two sequences s_1 and s_2 are within the same cluster if they differ by less than the distance D . We compute raw pairwise distances from all available sequences from any given donor and then assign sequences to clusters using the `cluster.dist` procedure in R.

For any given D , and for each transmission pair $i = 1, \dots, N$, we calculate the proportion, p_i , of sequences outside a cluster. Let $P_D(n)$ be the probability that n donors transmit a sequence outside a cluster. Then,

$$P_D(0) = \prod_{i=1}^N (1-p_i)$$
$$P_D(1) = \sum_{j=1}^N \prod_{i \neq j} p_i (1-p_i) = P_D(0) \sum_{i=1}^N \frac{p_i}{1-p_i}$$

and so on. Let g_D be the probability generating function defined by

$$g_D(x) = \sum_{i=1}^N P_D(i)x^i.$$

One can see that

$$g_D(x) = \prod_{i=1}^N [(1-p_i) + p_i x].$$

We can thus calculate $P_D(n)$ using the formula

$$P_D(n) = \frac{1}{n!} \frac{d^n}{dx^n} [g_D(x)]|_{x=0}$$

and Mathematica to compute $P_D(n)$ for $n = 1, \dots, 9$ and $D = 1, \dots, 10$ according to the above formula.

For each D , we call n_{obs} the number of donors that transmitted sequences outside a cluster and compute

$$p(D) = \sum_{n=n_{obs}}^N P_D(n).$$

The quantity $p(D)$ is the overall probability across eight donors that the observed number of sequences transmitted outside a cluster (i.e., not in a rake) is significantly different from what we would observe if transmission were equally likely among all sequences.

Table S2. Nucleotide differences between donor and recipient sequences

	D BL	D GT	BL source	GT source
F-DONORS				
RW36	10	10	PL	CA
ZM201	5	6	PL	SW
ZM216	4	3	PL	SW
ZM221	0	6	PB	SW
ZM238	4	3	PB	CA
ZM292	1	6	PL	CF
M-DONORS				
RW56	3	>15	PB	CA
ZM242	3	>15	PB	CA

D, donor; F, female; M, male; BL, blood; GT, genital tract.