

Relations of the estimates in the text to maximum likelihood estimates

The observations on the different types of polymorphism can be partitioned as n_1 , n_2 and n_3 , corresponding to type 1, type 2/3 and type 4/5/6 events, respectively; their sum is n . Let P_a be the probability that a polymorphism is ancestral. The estimate of t given in the text is equivalent to equating the observed fraction of 1 polymorphisms among 1/2/3 polymorphisms to their expected value minus 0.5, which is independent of P_a . The estimate of P_a is obtained by comparing the overall frequency of 1/2/3 polymorphisms among all polymorphisms to its expected value, which is dependent on t . We can parameterize the likelihood model for the data in terms of θ_1 and θ_2 , where θ_1 is the expected fraction of type 1 polymorphisms among types 1/2/3, and θ_2 is the expected fraction of type 1/2/3 polymorphisms among the total. The likelihood of the data can be written as:

$$L = \Pr\{n_1 | (n_1 + n_2)\} \Pr\{(n_1 + n_2)\} \quad (\text{S1a})$$

so that

$$\ln L = \ln (\Pr\{n_1 | (n_1 + n_2)\}) + \ln (\Pr\{(n_1 + n_2)\}) \quad (\text{S1b})$$

where

$$\ln (\Pr\{n_1 | (n_1 + n_2)\}) = C_1 + n_1 \ln (\theta_1) + n_2 \ln (1 - \theta_1) \quad (\text{S2a})$$

$$\ln (\Pr\{(n_1 + n_2)\}) = C_2 + (n_1 + n_2) \ln (\theta_2) + n_3 \ln (1 - \theta_2) \quad (\text{S2b})$$

and $\theta_1 = P_1/P_{2+3}$, $\theta_2 = (P_1+P_{2+3})P_a$

Differentiating with respect to t and P_a , we have:

$$\frac{\partial \ln L}{\partial t} = \left(\frac{n_1}{q_1} - \frac{n_2}{(1 - q_1)} \right) \frac{\partial q_1}{\partial t} + \left(\frac{(n_1+n_2)}{q_2} - \frac{n_3}{(1 - q_2)} \right) \frac{\partial q_2}{\partial t} \quad (\text{S3a})$$

$$\frac{\partial \ln L}{\partial P_a} = \left(\frac{(n_1+n_2)}{q_2} - \frac{n_3}{(1 - q_2)} \right) \frac{\partial q_2}{\partial P_a} \quad (\text{S3b})$$

For a maximum of $\ln L$, both derivatives must be zero. It follows that the multiplicand of the partial derivative of θ_2 in equation (S3a) is zero, since this is required for equation (S3b) to equal zero. Hence, the multiplicand of the partial derivative of θ_1 in Eqn. (S3a) is also zero.

The last result is equivalent to equating $(1 - \theta_1)/\theta_1$ to n_2/n_1 ; this in turn is equivalent to 0.5 plus equation (2) of the text, so that this estimate of t is also the ML estimate.

Similarly, setting equation (S3a) to zero with this estimate of t yields the estimate of P_a as $r_T = (n_1 + n_2) / \{(n_1 + n_2 + n_3)(P_1 + P_{2+3})\}$, the expression that was used in the text .

Variances of the estimates

From eqn. (11b) of Charlesworth et al. (2005), the estimated *a priori* probability of an ancestral polymorphism is given approximately for a reasonably large sample size by:

$$\hat{P}_d = \frac{1}{3} + \frac{1}{2} \exp(-\hat{t}) + \frac{1}{6} \exp(-3\hat{t}) \quad (\text{S4})$$

so that

$$\begin{aligned} \frac{d\hat{P}_d}{dt} &= -\frac{1}{2} \{ \exp(-\hat{t}) + \exp(-3\hat{t}) \} \\ &= -\frac{1}{2} (\hat{x}^{-1} + \hat{x}^{-3}) \end{aligned} \quad (\text{S5})$$

where $x = \exp(t)$, $t = \ln(x)$.

Using the delta-method formula for the variance of a function, we have:

$$V\{\hat{P}_d\} \approx \frac{1}{4} (\hat{x}^{-1} + \hat{x}^{-3})^2 V\{\hat{t}\} \quad (\text{S6})$$

By differentiating equation (S4) above, we find that:

$$\frac{d\hat{x}}{da} = \frac{\hat{x}}{(\hat{x} - 2a)} \quad (\text{S7})$$

so that

$$V\{\hat{x}\} \gg \frac{\hat{x}^2 V\{a\}}{(\hat{x} - 2a)^2} \quad (\text{S8})$$

where $a = (0.5f_1 + f_{2+3})/f_1$ and

$$V\{a\} = V\{f_{2+3}/f_1\} \quad (\text{S9a})$$

From equations (S4) and (S5) we find that:

$$V\{\hat{t}\} \gg \frac{1}{\hat{x}^2} V\{\hat{x}\} = \frac{V\{a\}}{(\hat{x} - 2a)^2} \quad (\text{S9b})$$

Again using the delta method, we have:

$$V\{f_{2+3}/f_1\} \gg \frac{V\{f_{2+3}\}}{f_1^2} + \frac{f_{2+3}^2 V\{f_1\}}{f_1^4} - \frac{2f_{2+3} \text{Cov}\{f_1, f_{2+3}\}}{f_1^3} \quad (\text{S10})$$

With k independent SNPs, the variances and covariances in equation (S10) are given by their multinomial expressions:

$$V\{f_1\} = f_1(1-f_1)/k, \quad V\{f_{2+3}\} = f_{2+3}(1-f_{2+3})/k, \quad \text{Cov} = -f_1 f_{2+3}/k \quad (\text{S11})$$

Substituting these into equations (S8) and (S9) thus gives us estimates of the variances of the estimates x and t .

The estimated proportion of ancestral polymorphisms is obtained as r_T , given by the ratio of $f_1 + f_{2+3}$ to the estimate of P_d .

The variance of r_T is thus given approximately by the delta method as:

$$\frac{V\{f_1\} + V\{f_{2+3}\} + 2\text{Cov}\{f_1, f_{2+3}\}}{\hat{P}_d^2} + \frac{(f_1 + f_{2+3})V\{\hat{P}_d\}}{\hat{P}_d^4} - \frac{2(f_1 + f_{2+3}) \text{Cov}\{(f_1 + f_{2+3}), \hat{P}_d\}}{\hat{P}_d^3} \quad (\text{S12})$$

where the variance of the estimate of P_d is obtained from equations (S4), (S9) and (S10).

Using equation (S4) and the delta method, we have

$$\text{Cov}\{f_1+f_{2+3}, \hat{t}\} \gg -\frac{1}{2}(\hat{x}^{-1} + \hat{x}^{-3}) \text{Cov}\{f_1+f_{2+3}, \hat{t}\} \quad (\text{S13a})$$

where

$$\text{Cov}\{f_1+f_{2+3}, \hat{t}\} \gg \frac{1}{(\hat{x}-2a)} \text{Cov}\{f_1+f_{2+3}, a\} \quad (\text{S13b})$$

and

$$\text{Cov}\{f_1+f_{2+3}, a\} = \frac{V\{f_{2+3}\} + \text{Cov}\{f_1, f_{2+3}\}}{f_1} - \frac{f_{2+3}[V\{f_1\} + \text{Cov}\{f_1, f_{2+3}\}]}{f_1^2} \quad (\text{S13c})$$

These expressions allow the approximate variances of the estimates t and P_a to be estimated.

Supplementary Table S1: Locus-by-locus breakdown of the (uncorrected) number of ancestral polymorphisms in 34 autosomal and 33 X-linked genes in *D. pseudoobscura* and *D. miranda*.

Locus	Type 1 (Shared)	<i>D. pseudoobscura</i>			<i>D. miranda</i>		
		<i>n</i>	Type 2/3 (Ancestral)	Type 4/5 & 6 (‘ <i>de novo</i> ’)	<i>n</i>	Type 2/3 (Ancestral)	Type 4/5 & 6 (‘ <i>de novo</i> ’)
Autosomal							
<i>ade3</i>	0	16	2	7	15	0	4
<i>crinkled</i>	0	15	3	14	15	1	0
<i>GA10135</i>	0	16	0	10	15	0	3
<i>GA10344</i>	0	16	0	7	15	0	0
<i>GA12147</i>	1	15	0	10	16	0	4
<i>GA12512</i>	0	16	1	10	15	0	0
<i>GA12664</i>	0	16	1	13	16	0	3
<i>GA12722</i>	0	16	1	9	16	0	1
<i>GA13578</i>	0	15	1	6	15	0	1
<i>GA14694</i>	0	15	1	6	16	0	0
<i>GA14715</i>	0	15	0	2	14	1	1
<i>GA15377</i>	0	16	0	8	16	0	3
<i>GA16473</i>	0	15	1	7	15	0	2
<i>GA17300</i>	0	15	3	10	15	0	1
<i>GA17553</i>	0	16	1	7	16	0	2
<i>GA17997</i>	1	16	1	7	15	0	1
<i>GA17998</i>	0	16	1	5	16	0	1
<i>GA18219</i>	0	15	5	17	16	0	1
<i>GA18470</i>	0	16	0	4	15	0	1
<i>GA18654</i>	0	16	0	7	15	1	2
<i>GA19119</i>	0	16	3	7	15	0	3
<i>GA19169</i>	0	16	0	8	15	0	2
<i>GA19326</i>	0	16	1	1	15	0	1
<i>GA19427</i>	0	16	1	4	15	0	3
<i>GA19649</i>	0	15	2	10	16	0	5
<i>GA19678</i>	0	16	0	5	16	0	3
<i>GA20117</i>	0	16	2	7	16	0	1
<i>GA20218</i>	0	16	0	9	16	0	1
<i>GA20407</i>	0	16	0	4	15	0	0
<i>GA22152</i>	1	15	0	9	15	1	2
<i>GA25303</i>	1	16	3	11	14	0	5
<i>GA25341</i>	0	14	0	3	14	1	1
<i>GA25997</i>	0	16	1	9	16	0	2
<i>lamin</i>	0	16	1	5	16	0	1
Sub-total	4	532	36	258	521	5	61
X-linked							
<i>GA10596</i>	0	16	0	8	14	0	3
<i>GA10819</i>	1	16	1	11	13	0	0
<i>GA11209</i>	1	16	2	10	16	0	1
<i>GA11389</i>	0	16	3	6	16	1	6
<i>GA12234</i>	0	16	0	7	16	0	1
<i>GA12817</i>	0	16	1	6	16	0	2
<i>GA12834</i>	0	15	3	5	14	1	1

GA12844	0	15	3	16	14	0	0
GA13242	0	16	0	1	16	0	0
GA13913	0	15	0	0	16	0	0
GA14058	1	15	1	4	16	1	3
GA14176	0	16	0	10	14	0	1
GA14572	0	16	1	4	15	0	3
GA14705	1	16	0	4	16	1	0
GA15328	0	16	0	8	12	0	5
GA15393	0	16	0	0	12	1	1
GA15436	0	15	0	10	16	0	0
GA15687	0	16	0	6	16	0	0
GA16943	0	16	3	3	14	0	0
GA17354	0	16	0	1	16	1	1
GA17564	0	16	0	1	15	0	1
GA17590	1	16	1	3	16	2	0
GA17594	0	16	0	1	15	0	1
GA17599	0	16	0	1	16	1	0
GA17782	0	16	0	4	16	0	1
GA18107	0	16	0	11	16	0	0
GA19139	0	16	1	2	16	0	0
GA19265	1	16	1	9	16	0	4
GA19408	0	16	1	13	13	1	0
GA20067	0	16	2	3	16	0	1
GA20648	0	16	0	4	16	0	0
GA20842	0	16	0	5	16	0	5
GA21237	0	16	1	9	16	0	1
GA21946	0	15	1	12	14	0	0
Sub-total	6	538	26	198	515	10	42
TOTAL	10	15.74*	62	456	15.24*	15	103

(*) denotes an *average*, rather than a total.

Supplementary Table S2: Counts of transitions and transversions across both species for X-linked (X) and autosomal (A) datasets. The ratio of transitions to transversions did not differ significantly between X and A ($\chi^2_{[1]} = 0.246, p > 0.05$); thus, the datasets were combined to give empirical rates of mutation for transitions and transversions.

	<i>D. pseudoobscura</i>		<i>D. miranda</i>		TOTAL
	X	A	X	A	
Transitions	170	221	46	59	496
Transversions	79	113	17	22	231
TOTAL	249	334	63	81	727