Genet. Res., Camb. (1980), **35**, *pp.* 179–194
With 1 *text-figure*
Printed in Great Britain

The interpretation of gene conversion data from ordered eight-spored asci

By J. R. S. FINCHAM, W. G. HILL and E. C. R. REEVE
Department of Genetics, University of Edinburgh

(*Received* 2 *November* 1979)

SUMMARY

Currently favoured models postulate that gene conversion is due to the correction of mis-matches in heteroduplex DNA. If heteroduplex is formed reciprocally on both chromatids participating in recombination, the mis-matches due to a heterozygous site will be different on the two chromatids, and there will be four correction probabilities to be taken into account. It is shown that, given the frequencies of the five different kinds of aberrant ascus ratios, it is possible to calculate four alternative sets of values for the four correction probabilities and the total number of asci in which heteroduplex is formed. These four solutions reduce in effect to two when there are no other markers distinguishing the two chromatids. With the aid of flanking markers and the assumption that heteroduplex formation is chemically polarized, it is possible, in principle, to choose one best solution.

The method has been applied to the five one-point crosses in *Sordaria fimicola* from which most data are available. The data from four different mutants crossed to wild type are compatible with a restricted model in which the correction frequencies, from mutant to wild and from wild to mutant, are the same on both chromatids. In the case of the fifth mutant the data are not consistent with this restricted model, and indicate different correction frequencies in the two chromatids.

1. INTRODUCTION

Gene conversion and post-meiotic segregation, and the relation of these events to the crossing-over of flanking markers, have been particularly studied in the eight-spored Ascomycete species *Sordaria fimicola* (Kitani & Olive, 1967, 1969; Kitani & Whitehouse, 1974) and *Ascobolus immersus* (Stadler & Towe, 1971), mutants with altered ascospore colour being used in each case. The results have usually been interpreted in terms of the recombination model of Holliday (1964), a slightly elaborated version of which is shown in Fig. 1. The model proposes that the recombination events, responsible for both gene conversion and crossing-over, are initiated by a reciprocal exchange between chromatids of single DNA strands of like polarity, leading to the formation of duplex DNA of hybrid origin (heteroduplex) on both. If the region covered by heteroduplex includes a mutant site in one of the chromatids, both chromatids will have a wild/mutant mis-match at this site. This will be a mis-matched base pair, in the case of a base-pair

0016–6723/80/2828–7860 $01.00

substitution mutation, or a structural mis-match in the case of a deletion or frame-shift. As Fig. 1 shows, and as pointed out by Emerson (1966), the mis-matches will be different on the two chromatids.

If no correction of mis-matches occurs the result will be an aberrant 4:4 (ab4 : 4) segregation in the eight-spored ordered ascus, with two meiotic products showing segregation at the first post-meiotic mitosis. If correction occurs in opposite directions (wild-to-mutant and mutant-to-wild) on the two chromatids the result will be a normal 4:4 segregation indistinguishable from the great majority of asci in which no heteroduplex was formed over the site under observation. Correction in the same direction on both chromatids will give a 6:2 or 2:6 ascus depending upon which way the correction goes, while correction on one and no correction on the other will give either 5:3 or 3:5.

In Holliday's (1964) model the point of half-chromatid (single-stranded DNA) crossing-over is envisaged as being capable of resolution in either of two ways: by the cutting and rejoining of the two crossed strands restoring the parental linkages of flanking markers, and by cutting and rejoining of the 'outer' strands to make a whole-chromatid cross-over with recombination of flanking markers. Fig. 1 includes an assumption not included in Holliday's original model: namely, that the unwinding of DNA single strand from the breakpoint on one chromatid,

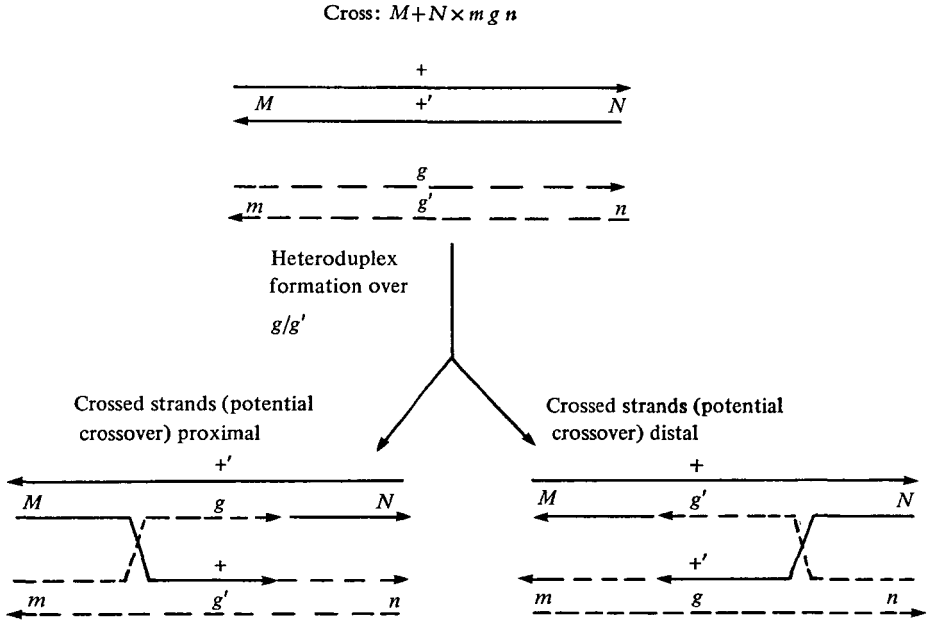

Fig. 1. Model for gene conversion and associated crossing-over. M/m, N/n are proximal and distal flanking markers. Only two of the four chromatids present at prophase I of meiosis are shown; the other two do not participate in the recombination event and retain the parental genotypes. Within each chromatid the opposite polarities of the DNA single strands are shown by arrows. $+/+'$ and g/g' are the complementary bases (or base sequences) of wild-type and mutant sites. It is assumed that heteroduplex spreads with a constant chemical polarity from its point of initiation. Correction probabilities: for $+/g'$ mismatch, $g' \rightarrow +' = p$, $+ \rightarrow g = q'$; for $+'/g$ mismatch, $g \rightarrow + = r$, $+' \rightarrow g' = s$.

https://doi.org/10.1017/S0016672300014038 Published online by Cambridge University Press

and its assimilation on to the other chromatid, proceeds with a constant chemical polarity. These seems quite plausible, especially in view of the properties of the 'unwindase' described from meiotic cells by Hotta and Stern (1978). This elaboration of the model has no effect on the reasoning in the following section, in which we describe the calculation of alternative sets of correction frequencies on the basis of the numbers of different ratios without taking flanking markers into account, but it becomes relevant when we consider how it may be possible, with the aid of flanking markers, to discriminate between the alternative solutions.

2. THE POSSIBILITY OF CALCULATING CORRECTION PROBABILITIES

The problem of calculating the probabilities of the four different kinds of mismatch correction in bilaterally formed heteroduplex covering a heterozygous site was first clearly formulated by Emerson (1966), but he did not arrive at an algebraic solution to the problem. Recently, Paquette & Rossignol (1978) have calculated correction probabilities in *Ascobolus immersus* using the assumption that the same two probabilities (of wild-to-mutant and mutant-to-wild correction respectively) applied to both chromatids. This assumption may be approximately true when wild and mutant sequences differ through deletion or insertion, but it can hardly be true in general and especially not when the mutation is a base-pair substitution. In this paper we continue Emerson's analysis and show how explicit solutions may be obtained.

In the *Sordaria fimicola* data (Kitani & Olive, 1967; Kitani & Whitehouse, 1974) on crosses of single ascospore colour mutants to wild type, there are five observed quantities: the respective numbers of 6:2, 2:6, 5:3, 3:5 and ab4:4 asci. What is not known is the total number of asci in which, because of hybrid DNA covering the gene under observation (*gray*, *g*), there was *opportunity* for gene conversion or post-meiotic segregation. This number would include the unobservable class of normal 4:4 asci due to conversion in opposite directions on the two chromatids. There are thus five parameters to be evaluated: the probabilities of correction from wild to mutant and from mutant to wild on the first chromatid (p and q respectively), the corresponding probabilities on the second chromatid (r and s respectively) and the total number of asci in which heteroduplex DNA is formed in the relevant region (N).

Then the expected and observed numbers of the various classes of aberrant asci are as shown in the following table:

+:m		Expected	Observed
6:2	Correction from mutant to wild on both chromatids	Npr	A
2:6	Correction from wild to mutant on both chromatids	Nqs	B
5:3	Correction from mutant to wild on one, no correction on the other	$N\{p(1-r-s)+r(1-p-q)\}$	C
3:5	Correction from wild to mutant on one, no correction on the other	$N\{q(1-r-s)+s(1-p-q)\}$	D
ab4:4	No correction on either chromatid	$N(1-p-q)\times(1-r-s)$	E

https://doi.org/10.1017/S0016672300014038 Published online by Cambridge University Press

It is shown in Appendix (i) that, by equating the expected and observed numbers of asci, two alternative sets of estimates of the five unknown parameters can be obtained.

$$\begin{aligned}(1)\ p &= Y/(X+Y+XY), & (2)\ p &= Y'/(X+Y'+XY'),\\ q &= X/(X+Y+XY), & q &= X/(X+Y'+XY'),\\ r &= Y'/(X'+Y'+X'Y'), & r &= Y/(X'+Y+X'Y),\\ s &= X'/(X'+Y'+X'Y'), & s &= X'/(X'+Y+X'Y),\end{aligned}$$

where

$$X = \frac{C+\sqrt{(C^2-4AE)}}{2A}, \quad X' = \frac{C-\sqrt{(C^2-4AE)}}{2A}$$

$$Y = \frac{D+\sqrt{(D^2-4BE)}}{2B}, \quad Y' = \frac{D-\sqrt{(D^2-4BE)}}{2B}.$$

Strictly, there are four alternative sets of solutions, but they reduce to two only when there is no means of distinguishing between the two chromatids. When such distinction can be made, because of the presence of distinguishing markers other than the one undergoing conversion, we can assign the values p and q to one chromatid and r and s to the other. In terms of Fig. 1, which defines p and q as the frequencies pertaining to the chromatid originally carrying the wild-type allele at the locus undergoing conversion, the values of (p, q) and (r, s) can be interchanged within each of the two solutions set out above.

A special problem arises if either A or B is zero. No such case occurs in the data considered in this paper (although $A = 1$ in one cross) but a procedure for dealing with the situation is given in Appendix (ii).

3. APPLICATION TO SORDARIA DATA

Table 1 shows data from Kitani & Olive (1967) and Kitani & Whitehouse (1974) for five of the most fully analysed one-point crosses of *g* mutants. The numbers of each type of aberrant segregation at the *g* locus are shown, and the asci are also classified with respect to segregation of the flanking markers. The flanking marker information becomes relevant in the following section.

Table 2 shows the values of the expressions C^2-4AE and D^2-4BE, the square roots of which are required for the calculation of the correction probabilities. A difficulty is at once apparent. In four of the five crosses the value of at least one of these expressions turns out to be negative, so that the equations for the calculation of correction probabilities have no real solutions. If these negative values were statistically significant they would have interesting implications. The obvious explanation would be that the assumption of independent correction on the two chromatids was false. If correction on one chromatid were correlated with correction on the other, so that correction tended to occur either on both or on neither, the values of A, B and E would obviously be inflated in comparison with C and D. Complete correlation would, in fact, reduce C and D to zero. Any substantial deficiency in ascus classes C and D due to correlated correction would tend to

Table 1. *Classification of aberrant asci from four* Sordaria fimicola *crosses of the form* $(M+N)/(m\,g\,n)$, *where g is an ascospore colour mutant at the gray locus and m and n are proximal and distal flanking markers*

(Only the genotypes of the two meiotic products which are presumed to have been involved in recombination at the gray locus are shown; the other two are the two parental types in each case.)

Ascus ratio + : *g*, and genotypes of relevant chromatids*

	6:2		2:6		5:3				3:5				ab4:4	
Cross	P	R	P	R	P1	P2	R1	R2	P1	P2	R1	R2	P	R
	$M+N$	$M+n$	MgN	Mgn	$M+N$	$M\pm_g N$	$m+N$	$M+n$	$M\pm_g N$	MgN	$m\pm_g N$	$M\pm_g n$	$M\pm_g N$	$M\pm_g n$
	$m+n$	$m+N$	mgn	mgN	$m\pm_g n$	$m+n$	$M\pm_g n$	$m\pm_g N$	mgn	$m\pm_g n$	Mgn	mgN	$m\pm_g n$	$m\pm_g N$
$+\times g_1$	237	163	15	13	99	73	84	66	23	4	13	10	14	9
	400		28		322				50				23	
$+\times g_6$	33	27	3	2	31	17	18	22	9	4	9	9	9	8
	60		5		88				31				17	
$+\times h_2$	22	18	13	7	68	29	36	51	63	23	19	42	147	138
	40		20		184				147				285	
$+\times h_3$	1	0	6	4	9	6	1	5	36	3	8	12	20	28
	1		10		21				59				48	
$+\times g_7$	46	25	6	3	42	15	24	29	7	7	12	12	26	17
	71		9		110				38				43	

From Table 5 of Kitani & Whitehouse (1974), which includes the data of Kitani & Olive (1967).

* The symbol $\pm_g$ indicates the presence of both + and *g* on the complementary strands of the DNA duplex.

Table 2. *Calculated correction frequencies from the data of Table 1*

Cross	*A*	*B*	*C*	*D*	*E*	Total	C^2-4*AE*	D^2-4*BE*	First solution *p*	*q*	*r*	*s*	Est. N4:4†	second solution *p'*	*q'*	*r'*	*s'*	Est. N4:4†
+ × g_1	400	28	322	50	23	823	66884	−76	No real solutions									
+ × g_6	60	5	88	31	17	201	3664	621	0·41	0·09	0·62	0·23	36	0·23	0·48	0·79	0·03	125
+ × h_2	40	20	184	147	285	676	−11744	−1191	No real solutions									
+ × h_3	1	10	21	59	48	139	249	1561	0·04	0·16	0·16	0·43	6	0·03	0·49	0·24	0·13	19
+ × g_7	71	9	110	38	43	271	−112	−104	No real solutions									

* Note that formally there are two other solutions, with *p*, *q* interchanged with *r*, *s* and *p'*, *q'* interchanged with *r'*, *s'*. These are only distinct from the two solutions given if there is some independent means of distinguishing between the *pq* and *rs* chromatids.

† Estimated number of normal 4:4 due to conversion in opposite directions.

result in C^2-4AE and D^2-4BE becoming negative. If the assumption of independent correction holds, however, there is no way these expressions could become negative other than by sampling error. A situation where negative values could readily occur would be if there were the same correction frequencies on each chromosome, i.e. $p = r$ and $q = s$. The expected values of C^2-4AE and D^2-4BE would both then be zero, so negative values of one or both of these quantities would happen in many samples by chance.

To test whether the negative values shown in Table 2 were significant and to enable estimation of the correction frequencies for these crosses, models with fewer than four parameters were fitted by maximum likelihood. Details are given in Appendix 3. The important model is that in which both r and s are constrained so that $p = r$ and $q = s$. The fit of the restricted models to the data was tested by comparing doubled log likelihoods to χ^2. Results are shown in Table 3, together with the expected numbers of each type of aberrant ascus, summing to the same total as the expected numbers.

For the two crosses, $+ \times h_2$ and $+ \times g_7$, where both $C^2 < 4AE$ and $D^2 < 4BE$, the restricted model with $p = q$ and $r = s$ fit the data very well. Furthermore, both crosses, $+ \times g_6$ and $+ \times h_3$, which gave real solutions on the general model are easily compatible with the restricted model. We may thus conclude that the results from four out of the five crosses here analysed are consistent with the same two frequencies of correction (from mutant to wild and from wild to mutant respectively) on both chromatids. Paquette & Rossignol (1978) found that their data from *Ascobolus immersus* were consistent with this restricted model. To the extent that it is found to be applicable it suggests that the mutations under study are recognized by the correction mechanism(s) as physical features (such as small deletions or insertions) which will be the same on the two chromatids, rather than as specific mis-matches of single base pairs, which, on the basis of the currently accepted models of the Holliday type (Fig. 1), are not expected to be the same on the two chromatids.

On the other hand, cross $+ \times g_1$ from which the largest number of asci was analysed and in which $D^2 < 4BE$, gave results which were *not* compatible with $p = r$ and $q = s$ ($P < 0{\cdot}01$). The data were, however, well fitted by a less restricted model constructed such that the expected value of D^2-4BE, but not that of C^2-4AE, was zero. This implies only one set of solutions (or two if the 'mirror image' solutions obtained by interchanging p and r and q and s are counted) with the restriction $s = q(1-r)/(1-p)$. This restriction has no apparent implications as regards correction mechanism or mode of recognition of mis-matches; the four correction frequencies are still all different from each other as in the general model. The data fit this restricted model very well, consistent with independent correction. But when the further restriction of $p = r$ and $q = s$ is tried we find that the results from $+ \times g_1$ are not compatible with correction frequencies being the same on the two chromatids. This could be explained if g_1 were a single base-pair substitution and if different mis-matched base pairs in a Holliday structure were recognized with different efficiencies by a correction enzyme.

It is not clear to what extent the fit of the data from the other four crosses to the model with $p = q$ and $r = s$ was due just to small numbers of observations. We do not present standard errors of the estimates of correction frequencies because they are not very useful. The estimates of the parameters, p, q, r and s are so highly correlated that, providing they are changed together rather than singly, a good fit to the data can be obtained over a wide range of values.

4. POSSIBILITIES OF DISCRIMINATING BETWEEN ALTERNATIVE SOLUTIONS

Although the existing Sordaria data are not adequate for the evaluation of meaningful correction probabilities, it is still perhaps of some interest to consider whether, given larger numbers, one might be able to choose between the alternative sets of solutions derived in Appendix 1. In fact, on the basis of the model shown in Fig. 1, it is possible to deduce the correct set, provided that closely placed flanking markers are scored. It has been pointed out to us by Dr Sterling Emerson that the model predicts that whichever side – whether proximal or distal – the DNA crossed-strands lie in relation to the site at which conversion is observed, completion of crossing-over will always associate a particular flanking marker recombinant type with a particular mis-match in the heteroduplex. Thus, in terms of the conventions of Fig. 1, *Mn* recombinant chromatids will carry the mis-match $+/g'$ and *mN* recombinant chromatids the mis-match $+'/g$. This is a necessary consequence of the assumption, embodied in the Figure, that heteroduplex spreads with a constant chemical polarity. The cases where heteroduplex formation is not associated with crossing-over do not give comparable information. The *MN* products, for example, will carry the $+/g'$ mis-match when the crossed strands are distal and the $+'/g$ mis-match when they are proximal. There is no way of resolving the distal/proximal ambiguity in the case of those conversion asci which are not recombined for the flanking markers, and so only the cross-over asci are used in the following argument.

From inspection of Fig. 1 it can be seen that 5:3 asci can arise in two ways:

(R2) No correction on the *Mn* chromatid and correction $g \rightarrow +$ on the *mN* chromatid. The frequency will be proportional to $(1-p-q)\,r$.

(R1) No correction on the *mN* chromatid and correction $g' \rightarrow +'$ on the *Mn* chromatid. The frequency will be proportional to $(1-r-s)\,p$.

3:5 asci will arise from:

(R2) No correction on *Mn*, correction $+ \rightarrow g$ on *Mn*; frequency proportional to $(1-r-s)\,q$.

(R1) No correction on *mN*, correction $+' \rightarrow g'$ on *Mn*; frequency proportional to $(1-p-q)\,s$.

Thus

$$\frac{5:3\,(\mathrm{R1})}{3:5\,(\mathrm{R2})} = \frac{p}{q} \quad \text{and} \quad \frac{5:3\,(\mathrm{R2})}{3:5\,(\mathrm{R1})} = \frac{r}{s}.$$

https://doi.org/10.1017/S0016672300014038 Published online by Cambridge University Press

Table 3. *Goodness of fit of the data of Table 1 to restricted models*

(Model 1: $p=r$, $q=s$ so expected values of C^2-4AE and of D^2-4BE are zero. Model 2: $s=q(1-r)/(1-p)$, so expected value of D^2-4BE is zero.)

		Numbers minimally adjusted to fit alternative models (observed numbers in parentheses)						Correction frequency calculated on adjusted numbers				
Cross	Model	*A*	*B*	*C*	*D*	*E*	χ^2(D.F.*)	*p*	*q*	*r*	*s*	Est. N4:4‡
$+\times g_1$	2	400·0	27·8	322·0	50·4	22·8	0·01(1)	0·40	0·32	0·86	0·074†	355
	1	425·8	22·2	270·4	61·7	42·9	25·80(2)	0·65	0·15	0·65	0·15	194
		(400)	(28)	(322)	(50)	(23)						
$+\times g_6$	1	64·7	7·3	78·7	26·4	23·9	5·21(2)	0·51	0·17	0·51	0·17	43
		(60)	(5)	(88)	(31)	(17)						
$+\times h_2$	1	34·3	19·6	195·5	147·8	278·8	1·75(2)	0·22	0·16	0·22	0·16	52
		(40)	(20)	(184)	(147)	(285)						
$+\times h_3$	1	1·8	13·2	19·4	52·5	52·0	2·52(2)	0·11	0·29	0·11	0·29	10
		(1)	(10)	(21)	(59)	(48)						
$+\times g_7$	1	70·9	8·7	110·1	38·6	42·7	0·02(2)	0·47	0·16	0·47	0·16	50
		(71)	(9)	(110)	(38)	(43)						

* χ^2 approximated by calculation of twice log likelihood; degrees of freedom in parentheses. All the data fit Model 1 satisfactorily, except those for $+\times g_1$ which fit Model 2.

† There is formally a second solution obtained by interchanging the values of p and r and of q and s.

‡ Estimated number of normal 4:4 due to conversion in opposite directions.

With the aid of these relationships it is possible, in prinicple, to choose the best values for the four correction probabilities and also to determine which pair of values pertains to which chromatid.

To illustrate the reasoning we will take the data for the cross of wild type $\times h_3$, ignoring for the moment the small numbers of observations and the unreliability of the estimates.

From Table 2 we obtain:

$$\text{(Solution 1)} \quad \frac{p}{q} = 0{\cdot}27, \quad \frac{r}{s} = 0{\cdot}37$$

$$\text{(Solution 2)} \quad \frac{p'}{q'} = 0{\cdot}053, \quad \frac{r'}{s'} = 1{\cdot}89.$$

In each solution the two ratios may equally well be reversed since, without using the flanking marker information, we can not distinguish one chromatid from the other.

From Table 1 we obtain:

$$\frac{p}{q} = \frac{5:3\,(\text{R1})}{3:5\,(\text{R2})} = \frac{1}{12} = 0{\cdot}083$$

and

$$\frac{r}{s} = \frac{5:3\,(\text{R2})}{3:5\,(\text{R1})} = \frac{5}{8} = 0{\cdot}625.$$

Solution 2 agrees rather better with the observed ratios in the odd-numbered conversion asci than solution 1. Accepting the ratios of solution 2, the indications from the flanking marker analysis ($p/q \ll 1$, $r/s \sim 1$) suggest that the assignment of the ratios to chromatids was correct as first written. We thus arrive at best estimates:

$$p = 0{\cdot}03, \quad q = 0{\cdot}49, \quad r = 0{\cdot}24 \quad \text{and} \quad s = 0{\cdot}13.$$

In fact, due to the small numbers, the data are not inconsistent with $p = r$ and $q = s$ (Table 3).

As we noted above, the subclasses P1 and P2 of the 5:3 and 3:5 asci are more complex in origin (if the model of Fig. 1 is accepted) since a given type of correction can have different consequences depending on whether the crosses strands are proximal or distal. However, having determined the best values of p, q, r and s on the basis of the R1 and R2 asci, we can apply these to the interpretation of the P1 and P2 asci and draw conclusions about the relative frequencies of proximal and distal crossed strands (potential crossover positions).

5. THE CONTRIBUTION OF UNILATERAL HETERODUPLEX

A very serious limitation of the model shown in Fig. 1 is its assumption that heteroduplex is always formed symmetrically on both chromatids (bilaterally). In fact there are strong indications that heteroduplex may be formed unilaterally on one chromatid only. In such a case the donor chromatid, which must be presumed to have contributed a single DNA strand to the chromatid with the hetero-

duplex, must be repaired not by reciprocal DNA transfer but by complementary copying of its own remaining strand. In Saccharomyces, indeed, the extremely low frequency of ab4:4 asci (Fogel & Mortimer, 1978) suggests strongly that nearly all gene conversion and post-meiotic segregation in this organism is due to unilateral heteroduplex. In the eight-spored Ascomycetes that have been studied the indications are that heteroduplex formation is sometimes unilateral and sometimes bilateral. Paquette & Rossignol (1978) have recently presented evidence that in *Ascobolus immersus* the relative contributions of the two modes may change systematically from one side of the gene map to the other. Whitehouse (1974) suggested a substantial contribution from unilateral heteroduplex in *Sordaria fimicola* on the basis of the consistent excess of P1 over P2 odd-numbered conversion asci (termed by him tritype and tetratype asci respectively). As he pointed out (and the same point was made in the case of *Ascobolus immersus* by Stadler & Towe, 1971), bilateral heteroduplex formation would be expected, overall, to yield P1 and P2 asci in equal numbers; unilateral heteroduplex, on the other hand, can only lead to P1 and never to P2.

In *Sordaria fimicola* data reviewed by Whitehouse there were 453 P1 and 226 P2 5:3 and 3:5 asci in data pooled from a number of crosses involving different mutants of the *gray* series. As he pointed out, the excess of 227 P1 asci suggested that about one third of the odd-numbered conversion asci arose from unilateral heteroduplex (all P1) and about two-thirds from bilateral (equal numbers of P1 and P2). This is an uncertain line of argument applied to a cross of any one mutant to wild type, since in a single case the heteroduplex correction probabilities could easily happen to be such as to tend to restore one or both of the original chromatid genotypes, a situation indistinguishable from heteroduplex not always being formed at all. In other words, in a particular case, the relationships $p > r$ and $s > q$ might well hold. Over a number of mutant × wild crosses, however, there is no reason on the basis of the model of Fig. 1 why these inequalities should apply consistently; yet the excess of P1 over P2 does seem to be consistent over all mutants. The estimate of one third of 5:3 and 3:5 asci arising from unilateral heteroduplex thus seems very plausible.

These considerations do not prevent the calculation of values for p, q, r and s, but they do call into question the interpretation of these parameters. If unilateral heteroduplex is significant, then a part of p and of s will be due to restoration of parental chromatid constitution in cases where no heteroduplex was ever formed. Only where there is independent evidence that heteroduplex is nearly always bilateral can the values obtained be taken as good estimates of mis-match correction probabilities.

We are indebted to Sterling Emerson, whose ideas initiated this study and with whom we have had many useful discussions.

REFERENCES

ELANDT-JOHNSON, R. C. (1971). *Probability Models and Statistical Methods in Genetics*. New York: Wiley.

EMERSON, S. (1966). Quantitative implications of the DNA-repair model of gene conversion. *Genetics* **53**, 475–485.

FOGEL, S., MORTIMOR, R., LUSNAK, K. & TAVARES, S. Meiotic gene conversion: a signal of the basic recombination event in yeast. *Cold Spring Harbor Symposia on Quantitative Biology* XLIII, 1325–1342.

HOLLIDAY, R. (1964). A mechanism for gene conversion in fungi. *Genetical Research* **5**, 282–304.

HOTTA, Y. & STERN, H. (1978). DNA unwinding protein from meiotic cells of *Lilium*. *Biochemistry* **17**, 1872–1880.

KITANI, Y. & OLIVE, L. S. (1967). Genetics of *Sordaria fimicola*. VI. Gene conversion at the *g* locus in mutant × wild type crosses. *Genetics* **57**, 767–782.

KITANI, Y. & OLIVE, L. S. (1969). Genetics of *Sordaria fimicola*. VII. Gene conversion at the *g* locus in interallelic crosses. *Genetics* **62**, 23–66.

KITANI, Y. & WHITEHOUSE, H. L. K. (1974). Aberrant ascus genotypes from crosses involving mutants at the *g* locus in *Sordaria fimicola*. *Genetical Research* **24**, 229–250.

PAQUETTE, N. & ROSSIGNOL, J.-L. (1978). Gene conversion spectrum of 15 mutants giving post-meiotic segregants in the *b2* locus of *Ascobolus immersus*. *Molecular and General Genetics* **163**, 313–326.

STADLER, D. R. & TOWE, A. M. (1971). Evidence for meiotic recombination in Ascobolus involving only one member of a tetrad. *Genetics* **68**, 401–413.

WHITEHOUSE, H. L. K. (1974). Genetic analysis of recombination at the *g* locus in *Sordaria fimicola*. *Genetical Research* **24**, 251–279.

APPENDIX: ESTIMATION OF CORRECTION PROBABILITIES

(i) *General case*

The observed numbers (A–E) and their total (M) of the five classes of abnormal asci in terms of the four correction probabilities (p, q, r, s) and the number of asci at risk (N) are:

Class	Observed	Expected
(6:2)	A	Npr
(2:6)	B	Nqs
(5:3)	C	$N[r(1-p-q)+p(1-r-s)]$
(3:5)	D	$N[s(1-p-q)+q(1-r-s)]$
(ab4:4)	E	$N(1-p-q)(1-r-s)$
Total	M	$N(1-ps-qr)$

The numbers in each class are multinomially distributed, so the log-likelihood, conditional on the numbers observed, is

$$L = \text{constants} + A \ln(pr) + \ldots + E \ln[(1-p-q)(1-r-s)] - M \ln(1-ps-qr).$$

Since there are only five classes and four parameters to be estimated, the maximum-likelihood estimates of the correction probabilities are obtained by equating the observed to expected numbers. Thus

$$\frac{C}{A} = \frac{1-p-q}{p} + \frac{1-r-s}{r}, \tag{1}$$

$$\frac{D}{B} = \frac{1-p-q}{q} + \frac{1-r-s}{s}, \tag{2}$$

https://doi.org/10.1017/S0016672300014038 Published online by Cambridge University Press

$$\frac{E}{A} = \frac{(1-p-q)(1-r-s)}{pr}, \tag{3}$$

$$\frac{E}{B} = \frac{(1-p-q)(1-r-s)}{qs}. \tag{4}$$

Squaring (1), subtracting (4) × (3) and taking the square root:

$$\frac{1-p-q}{p} - \frac{1-r-s}{r} = \frac{\sqrt{(C^2-4AE)}}{A}. \tag{5}$$

Taking the positive root and adding (1) and (5):

$$\frac{1-p-q}{p} = \frac{C+\sqrt{(C^2-4AE)}}{2A} = X. \tag{6}$$

Similarly, from (2) and (4),

$$\frac{1-p-q}{q} = \frac{D+\sqrt{(D^2-4BE)}}{2B} = Y. \tag{7}$$

From (6) $(1-q)/p = 1+X$, and from (7), $(1-p)/q = 1+Y$, so the estimates are

$$p = \frac{Y}{X+Y+XY} \quad \text{and} \quad q = \frac{X}{X+Y+XY}.$$

Subtracting (5) from (1)

$$\frac{1-r-s}{r} = \frac{C-\sqrt{(C^2-4AE)}}{2A} = X', \tag{8}$$

and subtracting (7) from (2):

$$\frac{1-r-s}{s} = \frac{D-\sqrt{(D^2-4BE)}}{2B} = Y'. \tag{9}$$

From (8), $(1-s)/r = 1+X'$, and from (9), $(1-r)/s = 1+Y'$, to give the estimates

$$r = \frac{Y'}{X'+Y'+X'Y'} \quad \text{and} \quad s = \frac{X'}{X'+Y'+X'Y'}.$$

Finally, from $Npr = A$, the estimate of N is

$$\frac{A(X+Y+XY)(X'+Y'+X'Y')}{YY'}.$$

Alternative solution(*s*). If, instead of taking the positive roots in both (5) and (7), we take the positive root in (5) and the negative in (7) we obtain from (6), $(1-q)/p = 1+X$, and from (7), $(1-p)/q = 1+Y$. Hence alternative estimates are:

$$p = \frac{Y'}{X+Y'+XY'}, \quad r = \frac{Y}{X'+Y+X'Y'},$$

$$q = \frac{X}{X+Y+XY'}, \quad s = \frac{X'}{X'+Y+XY}.$$

https://doi.org/10.1017/S0016672300014038 Published online by Cambridge University Press

Or, if we take the negative root in (5) and the positive in (7) we obtain from (6) $(1-q)/p = 1+X'$, and from (7), $(1-p)/q = 1+Y$. Hence

$$p = \frac{Y}{X'+Y+X'Y}, \quad r = \frac{Y'}{X+Y'+XY'},$$

$$q = \frac{X}{X'+Y+X'Y}, \quad s = \frac{X'}{X+Y'+XY'},$$

i.e. we obtain the same four expressions, but with p and q converted into r and s, respectively and vice versa. Taking negative roots in both (5) and (7) clearly gives the same four expressions as obtained with both roots positive, but with p,q and r,s again interchanged. Hence, unless we have some independent means of distinguishing between the 'p,q' and 'r,s' chromatids, there are two distinguishable sets of estimates:

$$\begin{aligned}
(1)\ p &= Y/(X+Y+XY), & (2)\ p &= Y'/(X+Y'+XY')\\
q &= X/(X+Y+XY), & q &= X/(X+Y'+XY'),\\
r &= Y'/(X'+Y'+X'Y'), & r &= Y/(X'+Y+X'Y),\\
s &= X'/(X'+Y'+X'Y'). & s &= X'/(X'+Y+X'Y).
\end{aligned}$$

where

$$X = \frac{C+\sqrt{(C^2-4AE)}}{2A}, \quad X' = \frac{C-\sqrt{(C^2-4AE)}}{2A},$$

$$Y = \frac{D+\sqrt{(D^2-4BE)}}{2B}, \quad Y' = \frac{D-\sqrt{(D^2-4BE)}}{2B}.$$

(ii) *Solutions when* $A = 0$ *or* $B = 0$

If no individuals are observed in either the A (6:2) or B (2:6) classes, the above solutions cannot be used since equations (6) and (8) or (7) and (9) involve division by zero. Note that in cross $+ \times h_3$, where $B = 1$, this nearly happened. The solutions are now obtained by setting the estimate of one of the correction probabilities to zero, for example, if $A = 0$ either $p = 0$ or $r = 0$.

$A = 0$. Setting $p = 0$, from (7), $(1-q)/q = Y$, and

$$\frac{1-r-s}{r} = \frac{E}{C} = W.$$

Thus W is used instead of X' in (8). Using (9) and previous results the solution follows, as do the alternative distinguishable solutions.

$$\begin{aligned}
(1)\ p &= 0, & (2)\ p &= 0,\\
q &= 1/(1+Y), & q &= 1/(1+Y'),\\
r &= Y'/(W+Y'+WY'), & r &= Y/(W+Y+WY),\\
s &= W/(W+Y'+WY'). & s &= W/(W+Y+WY).
\end{aligned}$$

Two other solutions are obtained by converting p and q to r and s, respectively.

https://doi.org/10.1017/S0016672300014038 Published online by Cambridge University Press

$B = 0$. Similarly, the two alternative distinguishable solutions are given by

$$\begin{aligned} (1)\ p &= 1/(1+X), & (2)\ p &= 1/(1+X'), \\ q &= 0, & q &= 0, \\ r &= Z/(X'+Z+X'Z), & r &= Z/(X+Z+XZ), \\ s &= X'/(X'+Z+X'Z). & s &= X/(X+Z+XZ). \end{aligned}$$

where $E/D = Z$.

$A = 0$ and $B = 0$. In this pathological case, the distinguishable solutions are $p = 0, q = 0, r = C/(C+D+E), s = D/(C+D+E)$, and $p = 0, q = D/(D+E), r = C/(C+E), s = 0$.

(iii) *Solutions when $C^2 < 4AE$ or $D^2 < 4BE$: reduced models*

If the observations are such that $C^2 < 4AE$ or $D^2 < 4BE$ the estimation procedure given in (i) cannot be used since negative roots are obtained. This occurred in three of the five sets of data. To obtain real solutions, the parameter estimates have to be constrained in some way, and it seemed reasonable that when, for example, $C^2 < 4AE$, the parameters should be constrained such that the expected value of this quantity is exactly zero, to give three independent parameters.

$C^2 < 4AE$, $D^2 \geqslant 4BE$. Setting the expected values of $C^2 = 4AE$ gives, from Appendix 1:

$$[r(1-p-q)+p(1-r-s)]^2 = 4pr(1-p-q)(1-r-s),$$

which reduces to $r(1-q) = p(1-s)$, so s can be replaced by $s = 1-r(1-q)/p$ in each of the numbers of A–E expected. Since there is now one less parameter than observation, the likelihood cannot be maximized by equating observation to expectation. No analytic expressions for the parameter values at the maximum has been obtained, so a numerical 'hill-climbing' procedure on the likelihood shown in Appendix 1 was used. The number of equivalent maxima (solutions) has not been deduced analytically, but numerical checks suggest that there are two when $C^2 < 4AE$. The computer program was set to find these two by taking as initial values the solutions from Appendix 1 obtained by assuming $C^2 = 4AE$, i.e. $X = X' = C/2A$.

Since only three rather than four parameters are fitted, a goodness-of-fit test of the reduced model to the data can be practised. This was done by comparing twice the difference in log likelihood from fitting observed frequencies equal to expectation (i.e. assuming expected frequency of class A (6:2) is A/M and so on) and the log likelihood fitting the model with $s = 1-r(1-q)/p$. If the model fits, the doubled difference in log likelihood has an approximate chi-square distribution with one degree of freedom.

$D^2 < 4BE$, $C^2 \geqslant 4AE$. The same procedure is used, but the restraint on the parameters, obtained by setting expectations of $D^2 = 4BE$, is

$$s = q(1-r)/(1-p).$$

The initial values used for the hill climbing were obtained by setting $Y = Y' = D/2B$.

https://doi.org/10.1017/S0016672300014038 Published online by Cambridge University Press

$C^2 < 4AE$, $D^2 < 4BF$. Setting both expectations $C^2 = 4AE$ and $D^2 = 4BE$ implies $p = r$ and $q = s$. The maximum likelihood parameter estimates can now be obtained explicitly. From Appendix (i), the log likelihood is now:

$$L = \text{constants} + A\ln(p^2) + B\ln(q^2) + C\ln[2p(1-p-q)] + D\ln[2q(1-p-q)] + E\ln[(1-p-q)^2] - M\ln(1-2pq).$$

It is convenient to set

$$F = 2A + C, \quad G = 2B + C, \quad H = C + D + 2E.$$

The log likelihood now becomes

$$L = \text{constants} + F\ln p + G\ln q + H\ln(1-p-q) - M\ln(1-2pq).$$

Differentiating L with respect to p and q, and setting the derivatives to zero gives the following solution for p:

$$p = \frac{2F + H \pm \sqrt{(4FG + H^2)}}{2(F - G + H)}$$

and for q by substituting the solutions for p into

$$q = \frac{pG}{F - p(F - G)}.$$

Although there are two solutions to the quadratic equations, in all examples checked so far there has only been one solution in which both p and q lie in the acceptable range $0 \leqslant p, q \leqslant 1$.

Fit of this reduced model to the data can also be checked using a likelihood ratio test, i.e. computing the difference in doubled log likelihood and comparing to chi-square. Since there are now only two parameters fitted, there are 2 D.F. for chi-square against the full model with p, q, r and s fitted, or 1 D.F. for chi-square against the reduced models where s was constrained to take account of either $C^2 < 4AE$ or $D^2 < 4BE$. Notice that these likelihood ratio tests can be conducted whether or not these inequalities are satisfied; so, for example, the simple two parameter model with p and q only can be examined directly.

(iv) *Sampling errors of estimates*

Because the estimates of the correction probabilities involve ratios and square roots of the observations or, in the case of some of the reduced models, no explicit solution at all, exact formulae for variances of the estimates cannot be obtained. Since these are maximum likelihood estimates, Cramer–Rao lower bounds for the variances can be obtained (see, for example, Elandt-Johnson, 1971).