Heredity **75** (1995) 644–649 *Received 14 June 1995*

Haplotype frequencies of linked loci in backcross populations derived from inbred lines

P. M. VISSCHER* & R. THOMPSON
Roslin Institute (Edinburgh), Roslin, Midlothian EH25 9PS, U.K.

Linkage disequilibrium among loci is an important parameter in explaining genotypic means and variances in animal and plant breeding populations. Joint haplotype frequencies and their sampling covariance matrix for any number of linked loci were derived for backcross populations derived from inbred lines. The predicted frequencies can be used to test whether the linkage disequilibrium observed between (marker) loci in backcross populations is as expected.

Keywords: backcross population, gene frequency, inbred lines, linked loci.

Introduction

In animal and plant breeding populations, linkage disequilibrium among loci plays an important role in explaining observed genotypic means and variances. Understanding the behaviour of linked loci in breeding programmes is important because it may influence the selection decisions of breeders. In particular, with the advent of marker assisted selection and marker assisted introgression the prediction of joint frequencies of marker alleles and quantitative trait alleles partly determines the efficiency of such selection programmes. For example, when introgressing a gene based on flanking markers in a backcross breeding programme, it would be useful to know the expected frequency of the gene to be introgressed in the nth backcross population if parents in each generation are selected solely on the basis of their flanking marker genotypes.

For two linked loci, genotype frequencies have been documented for many plant and animal population structures (Kempthorne, 1957; Mather & Jinks, 1971; Bulmer, 1980). To our knowledge, results for three or more linked loci have not been documented. In this paper, we derive the genotype frequencies for any number of loci for a backcross population originating from inbred lines with random mating, which are required as a basis of studies of marker assisted selection and introgression programmes in these populations. Although alleles at the individual loci could be genes or markers, in our notation, at least for three loci, we use two flanking markers with a quantitative trait locus (QTL) somewhere in that interval. The reason for this is that we are interested in the frequency of marker haplotypes and in the average QTL value for different marker haplotypes because those determine the efficiency of marker assisted selection. For more than three loci, we present haplotype frequencies in a general way.

*Correspondence.

Methods

Gene frequencies for two and three loci

For the case of three loci, we consider a straightforward extension of methods presented for two loci by, for example, Kempthorne (1957), Mather & Jinks (1971) and Bulmer (1980).

Consider a single marker bracket of length L (Morgans) with a recombination rate of r between the markers (i.e. the loci flanking an interval with length L). r_1 is the distance between the first marker and the QTL and r_2 between the QTL and the second marker. Allele i ($i = 1,2$) is from the ith inbred line. Without loss of generality, assume that line 2 is the recurrent line. We wish to calculate the frequencies of all gametes ($M_iQ_jM_k$) in all generations. From those we can calculate average values for different marker genotypes (depending on the genetic model). Throughout, we assume the mapping function of Haldane (1919) without interference, i.e. $r = r_1 + r_2 - 2r_1r_2$.

Frequencies of marker haplotypes

Denote frequencies of marker haplotypes at generation t as $f_{11}(t)$ (M_1M_1), $f_{12}(t)$ (M_1M_2), $f_{21}(t)$ (M_2M_1), $f_{het}(t)$ (M_1M_2 and M_2M_1) and $f_{22}(t)$ (M_2M_2). Note that M_iM_j are the marker haplotypes received from the crossbred parent. (We only need to consider the contribution from the crossbred parent because the contribution from the recurrent inbred line is always $M_2Q_2M_2$.) The vector $\mathbf{f(t)}' = (f_{11}(t)\ f_{het}(t)\ f_{22}(t))$. In the absence of selection, $\mathbf{f(t+1)} = \mathbf{P}\ \mathbf{f(t)}$, with

$$\mathbf{P} = \begin{bmatrix} \frac{1}{2}(1-r) & 0 & 0 \\ r & \frac{1}{2} & 0 \\ \frac{1}{2}(1-r) & \frac{1}{2} & 1 \end{bmatrix}$$

and $\mathbf{f(0)}' = [1\ 0\ 0]$.

$\mathbf{P}$ has eigenvalues 1, $\frac{1}{2}$ and $\frac{1}{2}(1-r)$. If $\mathbf{P} = \mathbf{QDQ}^{-1}$, then

$$\mathbf{Q} = \begin{bmatrix} 0 & 0 & -1 \\ 0 & -1 & 2 \\ 1 & 1 & -1 \end{bmatrix}$$

and

$$\mathbf{Q}^{-1} = \begin{bmatrix} 1 & 1 & 1 \\ -2 & -1 & 0 \\ -1 & 0 & 0 \end{bmatrix}.$$

Following through gives expressions for the marker bracket frequencies at generation t,

$$f_{11}(t) = (\tfrac{1}{2})^t\ (1-r)^t$$

$$f_{het}(t) = 2(\tfrac{1}{2})^t\ [1-(1-r)^t]$$

$$f_{22}(t) = 1-(\tfrac{1}{2})^t\ [2-(1-r)^t].$$

In Table 1 the frequencies of marker gametes are shown for several values of r in backcross generations 1, 2, 5 and 10. Although tight linkage has a large effect on the frequencies in early backcross generations, frequencies after, say, five generations of backcrossing are either small (for f_{11} and f_{het}) or large (for f_{22}).

Frequencies of marker-QTL brackets

The same approach was taken to find the frequencies of marker-QTL brackets. The derivation is shown in Appendix A. Combining the frequencies for the same marker haplotypes gives the average value for the marker haplotypes at generation t. If the inbred lines are fixed for alternative alleles of the QTL, and the average effect of an allele substitution is α (i.e. the inbred lines differ by 2α), then the average values of marker haplotypes at generation t ($g(M_iM_j)$) are,

$$g(M_1M_1) = [1-2(1-r_1)^t(1-r_2)^t/(1-r)^t](\alpha/2)$$

$$g(M_1M_2) = [1-2(1-r_1)^t\{1-(1-r_2)^t\}/\{1-(1-r)^t\}](\alpha/2)$$

$$g(M_2M_1) = [1-2(1-r_2)^t\{1-(1-r_1)^t\}/\{1-(1-r)^t\}](\alpha/2)$$

$$g(M_2M_2) = [1-2\{1-(1-r_1)^t\}\{1-(1-r_2)^t\}/\{(2^t-1)-(1-(1-r)^t)\}](\alpha/2).$$

For example,

$$g(M_1M_1) = [\{h(M_1Q_2M_1)-h(M_1Q_1M_1)\}/\{h(M_1Q_2M_1)+h(M_1Q_1M_1)\}](\alpha/2)$$
$$= [1-2(1-r_1)^t(1-r_2)^t/(1-r)^t](\alpha/2).$$

In Table 2, average values for marker haplotypes received from the crossbred parent relative to the

Table 1 Haplotype frequencies ($\times 10{,}000$) for two linked loci in backcross generation t for recombination fraction of r between the loci

	f_{11}				f_{het}				f_{22}			
t	1	2	5	10	1	2	5	10	1	2	5	10
r												
0.5	2500	625	10	0	5000	3750	605	20	2500	5625	9385	9980
0.4	3000	900	24	0	4000	3200	576	19	3000	5900	9399	9981
0.3	3500	1225	53	0	3000	2550	520	19	3500	6225	9428	9981
0.2	4000	1600	102	1	2000	1800	420	17	4000	6600	9477	9982
0.1	4500	2025	185	3	1000	950	256	13	4500	7025	9560	9984

Table 2 Values (in %, relative to the mean of the recurrent inbred line) of marker haplotypes received from the crossbred parent in generation t for recombination fraction of r between the markers and a QTL in the centre of the marker bracket

	f_{11}				f_{het}				f_{22}			
t	1	2	5	10	1	2	5	10	1	2	5	10
r												
0.5	0.00	50.00	93.75	99.80	0.00	50.00	93.75	99.80	0.00	50.00	93.75	99.80
0.4	−74.54	−52.31	−1.23	48.76	0.00	22.05	65.51	92.39	74.54	80.77	95.73	99.82
0.3	−90.35	−81.17	−56.19	−21.98	0.00	12.80	44.46	76.53	90.35	91.05	97.30	99.85
0.2	−96.82	−93.70	−84.62	−70.42	0.00	6.97	26.37	52.73	96.82	96.57	98.66	99.90
0.1	−99.38	−98.76	−96.92	−93.89	0.00	2.92	11.56	25.27	99.38	99.25	99.63	99.97

value of the QTL allele from the recurrent line (= 100 per cent) are shown for various recombination fractions between the marker loci, assuming that the QTL is in the centre of the interval. If the QTL is marked by two flanking markers which are close together ($r<0.2$), the average value of marker haplotype M_1M_1 increases very slowly (because of double recombinants) whereas the value of M_2M_2 is nearly 100 per cent in all generations.

General case of n *linked loci*

Although the same approach can be taken to derive the haplotype frequencies for more than three loci, the matrices become rather large and the algebra is tedious. Fortunately, the method can be generalized for any number of loci if joint frequencies of marker (loci) haplotypes are related to marginal frequencies. We find it simpler to predict partially marginal frequencies. The joint frequencies can be written as $f_u(t)$ with $\mathbf{u}$ an n-vector with each element $u_i (i = 1, \ldots, n)$ 1 or 2. If we sum over a set of loci then we will get a marginal frequency. We use $\mathbf{s}$, an n-vector with elements 0 or 1, with $s_i = 0$ indicating summation over the alleles of the ith loci, so that

$$f_s^0(t) = \sum_{i=1}^{n} \left(\sum_{u_i=1}^{2-s_i} f_u(t) \right). \qquad (1)$$

Marginal frequencies are calculated if $s_i = 0$ and conditional frequencies are calculated if $s_i = 1$. Therefore, with two loci ($n = 2$),

$$f_{11}^0(t) = f_{11}(t)$$

$$f_{01}^0(t) = f_{11}(t) + f_{21}(t)$$

$$f_{10}^0(t) = f_{11}(t) + f_{12}(t)$$

$$f_{00}^0(t) = f_{11}(t) + f_{12}(t) + f_{21}(t) + f_{22}(t).$$

Then, using the equations previously derived for marker haplotypes,

$$f_{11}^0(t) = (\tfrac{1}{2})^t(1-r)^t$$

$$f_{01}^0(t) = f_{10}^0(t) = (\tfrac{1}{2})^t$$

$$f_{00}^0(t) = 1.$$

Similarly for three loci,

$$f_{111}^0(t) = (\tfrac{1}{2})^t(1-r_1)^t(1-r_2)^t$$

$$f_{110}^0(t) = (\tfrac{1}{2})^t(1-r_1)^t$$

$$f_{101}^0(t) = (\tfrac{1}{2})^t(1-r)^t$$

$$f_{100}^0(t) = (\tfrac{1}{2})^t$$

$$f_{011}^0(t) = (\tfrac{1}{2})^t(1-r_2)^t$$

$$f_{010}^0(t) = (\tfrac{1}{2})^t$$

$$f_{001}^0(t) = (\tfrac{1}{2})^t$$

$$f_{000}^0(t) = 1.$$

Hence, each of these partially marginal frequencies changes at a rate of g_s per generation. The formula for g_s depends on the loci that $f_u(t)$ is summed over to form $f_s^0(t)$. This number of loci is $b = \text{bits}(\mathbf{s})$, where bits($\mathbf{s}$) is the number of nonzero elements of $\mathbf{s}$. If $b = 0$, then $g_s = 1$, and if $b = 1$, then $g_s = \frac{1}{2}$. Finally, if $b > 1$ then $f_s^0(t)$ is a marginal frequency for b loci with $u_i (i = 1, \ldots, b)$ including the b elements of s_i which are zero, and,

$$g_s = \tfrac{1}{2} \prod_{i=1}^{b-1} (1 - r_{u_i\, u_{i+1}})$$

with $r_{i,j}$ the recombination rate between the ith and jth loci. For all marginal frequencies,

$$f_s^0(t+1) = f_s^0(t) \times g_s, \text{ or}$$

$$f_s^0(t) = (g_s)^t.$$

These results generalize for any number of loci. Hence, given recombination rates between loci, the marginal frequencies can be calculated. In matrix notation, let $\mathbf{f(t)}$ be the vector of elements $f_u(t)$ and $\mathbf{f^0(t)}$ the vector of the $f_s^0(t)$. As before, $\mathbf{f(t+1) = Pf(t)}$. If $\mathbf{T}$ is the transformation taking $\mathbf{f(t)}$ to $\mathbf{f^0(t)}$, i.e. $\mathbf{f^0(t) = Tf(t)}$, and $\mathbf{f^0(t+1) = Df^0(t)}$, then we see that $\mathbf{f(t+1) = T^{-1}DTf(t)}$, so that elements of the vector $\mathbf{g}_s$ correspond to the eigenvalues of matrix $\mathbf{P}$, and $\mathbf{T}$ gives the eigenvectors of $\mathbf{P}$.

Joint frequencies, $f_u(t)$, can be calculated from the marginal frequencies by a simple modification of Yates' algorithm (Yates, 1937). This procedure is used to work out effects associated with an analysis of variance in a 2^n table, and involves n rounds of replacing a 2^n-vector by sums and differences of elements of this vector. To get joint frequencies from partially marginal frequencies we need the reverse of the operation defined by eqn (1). It is convenient to form the frequency in $(n+1)$ steps. In the ith step $(i = 1, \ldots, n)$ we form

$$f_s^{(i)}(t) = f_s^{(i-1)}(t) - f_v^{(i-1)}(t)$$

for $s_i = 0$ and $s_j = 0,1$ $(i \neq j)$, and

$v_i = 1$ and $v_j = s_j$ $(i \neq j)$.

This operation forms joint frequencies for the ith locus. For example,

$$f_{00}^{(1)}(t) = f_{21}(t) + f_{22}(t), \quad f_{00}^{(2)}(t) = f_{22}(t)$$

$$f_{01}^{(1)}(t) = f_{21}(t), \quad f_{01}^{(2)}(t) = f_{21}(t)$$

$$f_{10}^{(1)}(t) = f_{11}(t) + f_{12}(t), \quad f_{10}^{(2)}(t) = f_{12}(t)$$

$$f_{11}^{(1)}(t) = f_{11}(t), \quad f_{11}^{(2)}(t) = f_{11}(t).$$

In the $(n+1)$th stage we relate $f_s^n(t)$ to $f_u(t)$ using $u_i = 2 - s_i$, changing identifiers of $f^n(0,1)$ to the indicators of the alleles (1 and 2). For $n = 2$,

$$f_{22}(t) = f_{00}^{(2)}(t)$$

$$f_{21}(t) = f_{01}^{(2)}(t)$$

$$f_{12}(t) = f_{10}^{(2)}(t)$$

$$f_{11}(t) = f_{11}^{(2)}(t).$$

By ordering the elements in the initial f^0 vector in a standard way, the elements to be changed in the ith round can be simply expressed in terms of powers of 2. Fortran algorithms which calculate the geometric factors g_s and joint frequencies from marginal ones can be obtained from the authors.

As an example, we calculate the joint frequencies of four evenly spaced linked (marker) loci which are 20 cM apart for backcross generations 1, 5 and 10. Results are presented in Table 3. The marginal frequencies $f_s^0(1)$, i.e. marginal frequencies at the first generation and the corresponding $\mathbf{s}$ vectors are,

$\mathbf{s}$ = (0000, 1000, 0100, 1100, 0010, 1010, 0110, 1110, 0001, 1001, 0101, 1101, 0011, 1011, 0111, 1111)

$f_s^0(1)$ = (1.0000, 0.5000, 0.5000, 0.3784, 0.5000, 0.3159, 0.3784, 0.2863, 0.5000, 0.2838, 0.3159, 0.2390, 0.3784, 0.2390, 0.2863, 0.2166).

For example, for $\mathbf{s} = (0111) = \frac{1}{2}(1-r_{23})(1-r_{34}) = \frac{1}{2}(1-0.2433)^2 = 0.2863$.

In Appendix B we show that the variances of haplotype frequencies are,

$$\text{var}(\mathbf{f(t)}) = \text{diag}(\mathbf{f(t)}) - \mathbf{f(t)}\ \mathbf{f(t)}'. \qquad (2)$$

An algorithm to calculate the joint frequencies $\mathbf{f(t)}$ has already been presented. Therefore, calculating

Table 3 Joint haplotype frequencies ($\times 10{,}000$) for four equally spaced (marker) loci on a chromosome of length 100 cM for backcross generations 1, 5 and 10

	Backcross generation				Backcross generation		
Haplotype	1	5	10	Haplotype	1	5	10
$M_2M_2M_2M_2$	2166	9015	9963	$M_1M_2M_2M_2$	697	215	9
$M_2M_2M_2M_1$	697	215	9	$M_1M_2M_2M_1$	224	8	0
$M_2M_2M_1M_2$	224	167	9	$M_1M_2M_1M_2$	72	9	0
$M_2M_2M_1M_1$	697	55	1	$M_1M_2M_1M_1$	224	3	0
$M_2M_1M_2M_2$	224	167	9	$M_1M_1M_2M_2$	697	55	1
$M_2M_1M_2M_1$	72	9	0	$M_1M_1M_2M_1$	224	3	0
$M_2M_1M_1M_2$	224	44	1	$M_1M_1M_1M_2$	697	14	0
$M_2M_1M_1M_1$	697	14	0	$M_1M_1M_1M_1$	2166	5	0

the covariance matrix of $\mathbf{f(t)}$ is straightforward using eqn 2. In the case of independent loci, the variance of haplotype frequency i simply reduces to $\mathbf{f(t)}_i(1-\mathbf{f(t)}_i)$ which is the expected variance using binomial errors.

Discussion

We have shown that if there is no selection the joint haplotype frequency is relatively easily calculated for any number of linked loci. For three loci (two marker loci and a QTL), the value of a particular marker bracket quickly changes because of recombination. The covariance matrix of the haplotype frequencies is easy to calculate and requires knowledge only of the expected frequencies at any generation.

In practice, selection may operate in that genes may be introgressed (from a donor line) and at the same time the rest of the genome of the donor line may be selected against, so that the predictions of frequencies will not be accurate. The presented algorithm can be used to test whether linkage disequilibrium is as expected from random mating in the absence of selection.

Acknowledgements

P.M.V. was funded by the Marker Assisted Selection Consortium of the U.K. pig industry (Cotswold Pig Development Company, J.S.R. Farms, National Pig Development Company, Newsham Hybrid Pigs, Pig Improvement Company, and the Meat and Livestock Commission) and by MAFF, DTI and the BBSRC. R.T. acknowlegdes support from MAFF and BBSRC. We thank Chris Haley for helpful comments and discussions, Bill Hill for critical comments and John Whittaker for useful comments on an earlier version of the manuscript.

References

BULMER, M. G. 1980. *The Mathematical Theory of Quantitative Genetics*. Clarendon Press, Oxford.

HALDANE, J. B. S. 1919. The combination of linkage values, and the calculation of distance between the loci of linked factors. *J. Genet.*, **8**, 299–309.

KEMPTHORNE, O. 1957. *An Introduction to Genetic Statistics*. John Wiley, New York.

MATHER, K. AND JINKS, J. L. 1971. *Biometrical Genetics*, 2nd edn. Chapman and Hall, London.

YATES, F. 1937. *The Design and Analysis of Factorial Experiments*. Imperial Bureau of Soil Science, Harpenden.

Appendix A

Joint frequencies of two marker loci and a single QTL

Let $\mathbf{h(t)}$ be a vector with frequencies of gametes $[M_1Q_1M_1\ M_1Q_2M_1\ M_1Q_1M_2\ M_1Q_2M_2\ M_2Q_1M_1\ M_2Q_2M_1\ M_2Q_1M_2\ M_2Q_2M_2]$ at generation t. As for the case of two loci, let,

$\mathbf{h(t+1) = H\,h(t)}$, with

$$\mathbf{H} = \begin{bmatrix}
\frac{1}{2}(1-r_1)(1-r_2) & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\frac{1}{2}r_1r_2 & \frac{1}{2}(1-r) & 0 & 0 & 0 & 0 & 0 & 0 \\
\frac{1}{2}(1-r_1)r_2 & 0 & \frac{1}{2}(1-r_1) & 0 & 0 & 0 & 0 & 0 \\
\frac{1}{2}(1-r_2)r_1 & \frac{1}{2}r & \frac{1}{2}r_1 & \frac{1}{2} & 0 & 0 & 0 & 0 \\
\frac{1}{2}(1-r_2)r_1 & 0 & 0 & 0 & \frac{1}{2}(1-r_2) & 0 & 0 & 0 \\
\frac{1}{2}(1-r_1)r_2 & \frac{1}{2}r & 0 & 0 & \frac{1}{2}r_2 & \frac{1}{2} & 0 & 0 \\
\frac{1}{2}r_1r_2 & 0 & \frac{1}{2}r_1 & 0 & \frac{1}{2}r_2 & 0 & \frac{1}{2} & 0 \\
\frac{1}{2}(1-r_1)(1-r_2) & \frac{1}{2}(1-r) & \frac{1}{2}(1-r_1) & \frac{1}{2} & \frac{1}{2}(1-r_2) & \frac{1}{2} & \frac{1}{2} & 1
\end{bmatrix}$$

and $\mathbf{h(0)}' = [1\ 0\ 0\ 0\ 0\ 0\ 0\ 0]$.

Decompose as $\mathbf{H}' = \mathbf{U\ D\ U}^{-1}$; This can be performed using standard statistical software. $\mathbf{H}$ has eigenvalues $\frac{1}{2}(1-r_1)(1-r_2)$, $\frac{1}{2}(1-r)$, $\frac{1}{2}(1-r_1)$, $\frac{1}{2}$, $\frac{1}{2}(1-r_2)$, $\frac{1}{2}$, $\frac{1}{2}$, 1 and the matrices $\mathbf{U}$ and $\mathbf{U}^{-1}$ can be determined easily. Now, $\mathbf{h(t)}' = \mathbf{h(0)}'\ (\mathbf{H}')^t = \mathbf{h(0)}'\ \mathbf{U\ D}^t\ \mathbf{U}^{-1}$.

This gives the frequencies of the marker-QTL gametes at generation t as,

$$h(M_1Q_1M_1) = (\tfrac{1}{2})^t[(1-r_1)^t(1-r_2)^t]$$

$$h(M_1Q_2M_1) = (\tfrac{1}{2})^t[(1-r)^t-(1-r_1)^t(1-r_2)^t]$$

$$h(M_1Q_1M_2) = (\tfrac{1}{2})^t[(1-r_1)^t\{1-(1-r_2)^t\}]$$

$$h(M_1Q_2M_2) = (\tfrac{1}{2})^t[\{1-(1-r)^t\}-(1-r_1)^t\{1-(1-r_2)^t\}]$$

$$h(M_2Q_1M_1) = (\tfrac{1}{2})^t[(1-r_2)^t\{1-(1-r_1)^t\}]$$

$$h(M_2Q_2M_1) = (\tfrac{1}{2})^t[\{1-(1-r)^t\}-(1-r_2)^t\{1-(1-r_1)^t\}]$$

$$h(M_2Q_1M_2) = (\tfrac{1}{2})^t[\{1-(1-r_1)^t\}\{1-(1-r_2)^t\}]$$

$$h(M_2Q_2M_2) = (\tfrac{1}{2})^t[(2^t-1)-\{1-(1-r)^t\}-\{1-(1-r_1)^t\}\{1-(1-r_2)^t\}]$$
$$= 1-\tfrac{1}{2}{}^t[1+\{1-(1-r)^t\}+\{1-(1-r_1)^t\}\{1-(1-r_2)^t\}].$$

Appendix B

Variances of haplotype frequencies

Using the notation from the main text, the (co)variance matrix of $\mathbf{f(t)}$ can be derived easily. In general, using the joint frequencies,

$$\mathbf{f(t)} = \mathbf{P}\mathbf{f(t-1)} + \mathbf{e}, \text{ and} \tag{B1}$$

$$\text{var}(\mathbf{f(t)}) = \text{diag}(\mathbf{f(t)}) - \mathbf{P}^t \,\text{diag}(\mathbf{f(0)})\, \mathbf{P}'^t, \tag{B2}$$

with diag($\mathbf{f}$) indicating a diagonal matrix with elements corresponding to the vector $\mathbf{f}$. We assume that each individual has one offspring only and that sampling of frequencies is multinomial.

In our case, the vector $\mathbf{f(0)}$ has a special form, with its first element unity, and all other elements zero. Then eqn B2 simplifies, because

$$\mathbf{P}^t \,\text{diag}(\mathbf{f(0)})\, \mathbf{P}'^t = \mathbf{P}^t \,\text{diag}(\mathbf{f(0)})\, \text{diag}\,(\mathbf{f(0)})\, \mathbf{P}'^t = (\mathbf{P}^t)_1\, (\mathbf{f(0)})_1\, (\mathbf{f(0)})_1\, (\mathbf{p}'^t)_1$$
$$= \mathbf{P}^t\mathbf{f(0)} = \mathbf{f(t)}\, \mathbf{f(t)}'.$$

Hence,

$$\text{var}(\mathbf{f(t)}) = \text{diag}(\mathbf{f(t)}) - \mathbf{f(t)}\, \mathbf{f(t)}'. \tag{B3}$$