ARTICLE

Allelic association is increased by correlation of effective family size

Frédéric Austerlitz[1,3] and Evelyne Heyer[2]

[1]*Laboratoire Evolution et Systématique, CNRS Université Paris-Sud, Orsay;* [2]*CNRS, Laboratoire d'Anthropologie Biologique, Musée de l'Homme, Paris, France*

In a previous study, we showed that the observation of high frequencies of certain inherited disorders in the population of Saguenay Lac Saint Jean (SLSJ) in Quebec can be explained in terms of the variance and the auto-correlation of effective family size (EFS) across generations. Correlated EFS, across generations, allows alleles introduced as a single copy to reach very high frequencies in about 12 generations. Here, we investigate the impact of this same demographic process on allelic association between a disease locus and closely linked neutral markers. We model the fate of a disease allele, introduced as a single copy, and of its surrounding haplotype. We show that the auto-correlation of EFS across generations increases the expected proportion of individuals who carry the ancestral haplotype in the present generation. Thus, the length of a shared haplotype is longer, making this population useful for coarse mapping. But this autocorrelation decreases the estimated value; thus ignoring the auto-correlation in EFS leads to an underestimate of the recombination rate. This result is of importance, since socio-demographic processes such as auto-correlation of EFS across generations have been described in other human populations.
European Journal of Human Genetics (2000) **8**, 980–985.

Introduction

Allelic association, the excess frequency of a marker allele, in a sample of chromosomes carrying a disease allele, is the focus of numerous studies in human populations. As explained by Thompson and Neel,[1] the key factor in maintaining significant allelic association is the youth of a disease-causing variant. Many types of estimates of recombination rate, based on allelic association, have been developed. The first were moment estimators,[2,3] based on the Luria-Delbrück method.[4] Subsequently, maximum likelihood methods were developed.[5–10]

Such estimators have been used in various populations,[2,3,11–13] allowing the positional cloning of a disease locus in at least two cases.[14,15] All these models assume a Wright-Fisher model, usually with a constant population size or an exponentially growing population, but what happens if the assumptions are invalid?

In a previous paper,[16] we showed that the high frequencies of inherited disorders in the population of Saguenay-Lac Saint-Jean (SLSJ), Northeastern Quebec, can be explained in terms of the variance of effective family size (EFS), and in terms of the auto-correlation of EFS from one generation to the next. This demographic linkage is thought to exist in other populations as well.[17–20] In another paper,[21] we showed that the variance in EFS, without the correlation, has almost no impact on the expected level of allelic association.

The aim of this paper is, first, to design a model to study the joint impact of high EFS variance and of auto-correlation of EFS across generations on allelic association. We simulate a population with Poisson or geometric distribution of EFS, and given level of auto-correlation of EFS. Then we compute the average level of allelic association of founders' alleles with surrounding markers in a sample of 1000 individuals, as a function of the recombination rate and of the frequency of these founders' alleles in the sample. We also compare our results with the results of simulations on real genealogies of the SLSJ population.

Correspondence: Dr E Heyer, Laboratoire d'Anthropologie Biologique, Musée de l'Homme, 17 place du Trocadéro, 75116 Paris, France.
Tel: +33 1 44 05 72 52; Fax: +33 1 44 05 72 41;
E-mail: eheyer@mnhn.fr
[3]Current address: Department of Ecology, Evolution and Natural Resources, Rutgers University, New Brunswick, NJ 08901-8551, USA.

Second, we design a maximum likelihood method for estimating the recombination rate from population data, taking into account the peculiar demographic behaviour of the population. We use data from SLSJ, on a sample of 37 carriers of a disease gene, to show how this method can improve the estimation of recombination rate. We also investigate the impact of the assumed population carrier frequency of the mutant allele on the estimate.

Material and methods

Some of these data and simulation methods have been presented earlier.[16,21] We recall them briefly here to present the new results in a fashion consistent with those earlier efforts.

Demographic data

The population of SLSJ numbers approximately 300000 inhabitants, descended mostly from early founders, among the first 5000 settlers of Quebec, at the end of the seventeenth century. This implies an overall population growth rate of 1.41 per generation, taking into consideration only the children that remained in SLSJ.

The effective family size (EFS) is the number of locally reproducing children per woman. The general shape of the EFS distribution resembles more a geometric than a Poisson distribution, and EFS is correlated from one generation to the next.[16] For the nineteenth century, this auto-correlation is 0.16, averaged between men and women. We also measured this correlation from eighteenth century individuals in the ascending genealogies and found an even higher correlation of 0.34. EFS distribution and correlation are almost the same for men and women.

Genealogical data

We used the genealogies of 978 individuals from the IREP ascending genealogies database, which contains all genealogical links back to 2631 individuals among the 5000 founders who settled in Nouvelle-France before 1700.[22] These 978 individuals are presumed Alzheimer cases, documented in the IMAGE project by Algène Biotechnologies. We have shown that the fact that they have been recruited as Alzheimer probands should not bias our results,[16] since the genetic contribution of the founders to these individuals is the same as their genetic contribution to a control group of people, not known to be carriers of any diseases.

Genetic data

As in previous work,[21] we used the haplotypic data of Labuda et al[3] on the Pseudovitamin D Deficient Rickets (PDDR). This is a recessive lethal condition, with a carrier frequency of about 1/26 in SLSJ. The haplotypic data consist of frequencies for the variants of a haplotype composed of nine microsatellite loci, adjacent to the locus of PDDR. The global recombination rate of the haplotype ($\theta = 0.07$) is independ-

ently known from Généthon linkage studies on CEPH families.[23] The PDDR mutant has recently been cloned,[24,25] confirming its monophyletic origin in the population and its location on the haplotype given in Labuda et al.[3]

Among 37 unrelated carrier chromosomes from SLSJ, 24 represented the same haplotypic variant, three haplotypes were seen twice and seven once each.[3] We consider here two-locus haplotypes made of the disease locus and a given marker locus. Table 1 gives, for each marker not in complete association with the disease gene, the number (x_f) carriers of the most frequent haplotype and the number (x_h) of different haplotypes in the sample, based on Figures 4 and 5 of Labuda et al.[3] We also computed these values for the total haplotype.

Demographic simulations

Using a branching process method, we simulated the genealogies of the population. We started with an initial population of 5000 individuals (2500 women and 2500 men), and let the population grow for 12 generations, with a growth rate λ of 1.41.

At each generation t, couples were formed at random. At $t = 0$, initial values for EFS were drawn from the distribution $D(2\lambda)$, where $D(2\lambda)$ is either a Poisson or geometric distribution, with mean 2λ. At each subsequent generation, the EFS of each couple was drawn according to the following distribution:

$$D\left(\alpha_t\left[(1-c)\lambda + c\left(\frac{n_{w_i} + n_{m_i}}{2}\right)\right]\right),$$

where n_{w_i} and n_{m_i} denote the EFS of the parents of a woman w_i and that of the parents of her husband m_i, respectively, and α_t is a normalisation parameter computed numerically so that growth rate remained equal to λ. The parameter α_t compensates the overall increase of EFS in the population that would otherwise occur due to the correlation of this value. The parameter c indicates the strength of the inter-generation dependency. For instance, a value of $c = 0$ corresponds to a correlation of $r = 0$, $c = 0.4$ corresponds to

Table 1 Number of carriers of the most common haplotype (x_f) and number of different haplotypes (x_h) in a sample of 37 individuals of the SLSJ population, for the total haplotype and for the haplotype consisting of the disease locus and each of the markers listed above, based on Figures 4 and 5 of Labuda et al.[3]

Marker	(x_f, x_h)
Total haplotype	(24, 11)
D12S325	(34, 3)
D12S103	(35, 2)
D12S359	(26, 9)
D12S312	(27, 8)
D12S355	(28, 6)
D12S83	(29, 5)

$r = 0.2$, and $c = 1$ to $r = 0.32$, when using the geometric distribution. In each simulation, we recorded all genealogies within the population. For more details on this model, see Austerlitz and Heyer.[16]

Allele dropping

Either on the simulated genealogies or on the real ascending genealogies of the 978 individuals, we adapted the method of allele dropping. Allele dropping[26,27] simulates the Mendelian transmission of alleles on genealogies. We extended it here to a two-locus haplotype, consisting of a disease locus and a marker locus, with recombination rate θ.

We started the simulations with all founders carrying a different allele at both disease and marker loci. The assumption was straightforward for the disease locus: we assumed that the mutant allele had been introduced by a single founder, thus we treated each founder in turn as the source of the mutant allele in the population.

Regarding the marker locus, the bias introduced by a minimally polymorphic marker would be an underestimation of recombination events. We are dealing, here, however, with nine highly polymorphic markers. Thus, we could assume, following Labuda *et al*,[3] that all recombination events in the population are detectable. The real situation was therefore equivalent to the approximation that we make here, where all founders carry a different allele at the marker locus.

For each individual of the next generation, we randomly chose one of the two alleles at the disease locus carried by his/her mother and one of the two carried by his/her father. Then, for each of the two alleles, we drew a random number from the U[0,1] distribution to decide whether recombination occurred between the disease and marker loci.

This model of genetic transmission was propagated, starting from the founders and proceeding along the genealogical paths. All alleles were assumed to be neutral. We have previously shown[27] that lethality has very little influence on changes in allele frequencies for a rare recessive allele.

Allelic association

The result of one simulation consisted of all haplotypes, either ancestral or recombinant, in the final population. To estimate the allelic association, we sampled 1000 individuals in the final population at random, to be consistent with the simulations on the 978 ascending genealogies. For each founder's disease allele, we counted the number of copies (z) of this allele in the sample. Then we calculated the fraction (p_f) of the most frequent haplotype among these z carriers of the founder's allele.

We repeated this simulation 1000 times, to compute the empiric expectation of p_f conditional on z. Each simulation yielded many (z, p_f) pairs. Each pair with the same z was treated as a separate repetition. The average value $\overline{p_f}(z)$ of p_f conditional on z was calculated by averaging over all repetitions of all simulations. Thus, we obtained several

repetitions, for any value of z. An average value of p_f calculated over too small a number of repetitions would have been meaningless. So we calculated $\overline{p_f}(z)$ only for the value of z for which we had at least 50 repetitions.

Maximum likelihood estimation

Using the data listed in Table 1, we wanted to estimate the recombination rate θ, for a given value of c. Thus, for input parameters, c, λ and θ, we did repeated simulations. In each simulation, we kept only the haplotypes containing the founder's allele the frequency of which had reached a given target interval in the population. Each simulation yielded, for each of these founder's alleles, the number of carriers of this founder gene (n) in the population, and the number of carriers of the most frequent haplotypes (n_f) among these n carriers.

We then computed the probability of observing the data (x_f, x_h) in a sample of size s, drawn from these n carriers. Therefore, unlike the above, where we sampled chromosomes at random in the population, we sampled here from the chromosomes carrying a given founder's disease allele. Let X_f and X_h be the random variables representing the number of carriers of the most frequent haplotype and the number of different haplotypes in the sample. The probability p_{x_f} that $X_f = x_f$ is given by a hypergeometric distribution:

$$p_{x_f} = P(X_f = x_f) = \frac{\binom{n_f}{x_f}\binom{n-n_f}{s-x_f}}{\binom{n}{s}}.$$

Regarding x_h, the conditional probability $p_{x_h} = P(X_h = x_h / X_f = x_f)$ of observing exactly ($x_h - 1$) haplotypes among the ($s - x_f$) carriers of other than the most frequent haplotype in the sample was then estimated empirically by doing repeated sampling. The joint distribution of X_f and X_h was then $P_{x_f x_h} = P(X_f = x_f, X_h = x_h) = p_{x_f} p_{x_h}$.

Averaging $p_{x_f x_h}$ over 1000 replicates gave the estimated joint probability of observing x_f and x_h in the sample. Then, holding c and λ fixed, this process was repeated for values of θ ranging from 0.001 to 0.12, with a step interval of 0.002. This yielded a maximum likelihood estimate of θ. The confidence interval of the estimates was defined with the classical (max – 2) criterion of the log likelihood, as described in Kaplan *et al*.[6]

In all cases, we used the geometric distribution and values of c (strength of inter-generation dependency) ranging from 0.4 to 1.0. In order to assess the effect of the assumed population frequency of mutant allele carriers on the estimate of recombination rate, we computed these likelihood estimates for three different frequency target intervals, [1/120, 1/60], [1/60, 1/40] and [1/30, 1/20]. The last interval corresponds to the real data of 1/26, allowing for sampling uncertainty. As we showed in Austerlitz and Heyer,[16] it is not possible for a disease allele introduced by a single founder to

reach the target frequencies given above, if c 0.4 for the geometric distribution or if c takes any value for the Poisson distribution.

Results
Allelic association
In the random sample of 1000 individuals in the population, the average fraction $\overline{p_f}(z)$ of most frequent haplotype carriers among the z carriers of a founder's allele decreased with z (Figure 1), and reached an asymptotic value as z became large. When the geometric distribution was used, association increased with the level of cultural inheritance (c). For example, when $\theta = 0.07$, about 60% of individuals still carried the ancestral haplotype when $c = 1$ ($r = 0.32$), this number fell to 45% when $c = r = 0$. In the Poisson case, the c parameter had very limited impact. The results of the real ascending genealogies were similar to the result obtained with the simulated geometric distribution and $r = 0.32$. Although $\overline{p_f}(z)$ declined more slowly with z for the real genealogies than for the simulated geometric distribution, with $r = 0.32$, a similar asymptotic value was reached.

Estimation of recombination rate using real data
Estimation, using the real haplotypic data (Table 1), of the recombination rate of the haplotype surrounding the disease locus was clearly affected by the value chosen for the c parameter. The estimator $\hat{\theta}$ increased with c (Figure 2). The increase was rather large, sometimes reaching 1.5 cM. This effect held for the estimated recombination rate of the
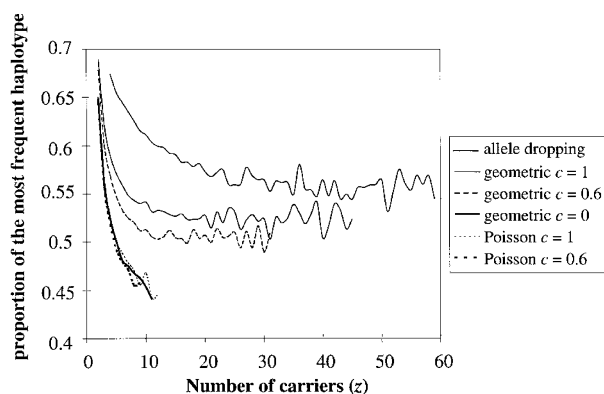
disease locus with each surrounding marker (see Table 2). The assumed carrier frequency of the mutant allele in the population also had an impact; the higher the estimated carrier frequency, the lower the estimator $\hat{\theta}$. This bias was also high, approximately 0.5 cM. The confidence interval was always wide (more than 8 cM for the total haplotype with $c = 1$) and it increased with c (see Table 2).
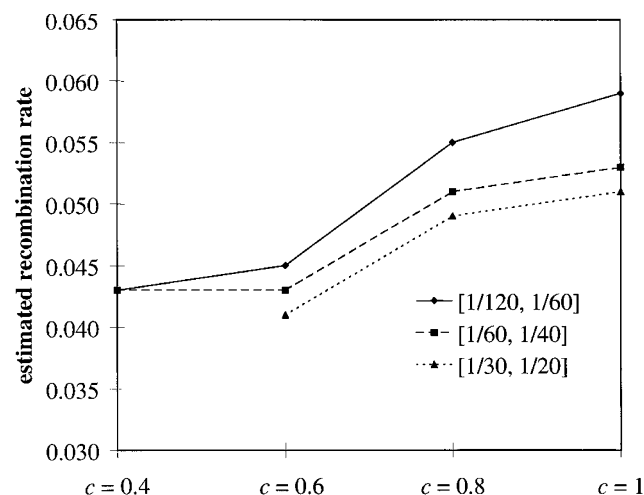
Discussion
Impact on allelic association

Figure 2 Impact of the c value on the maximum likelihood estimator of recombination rate (θ) for several target frequencies of carriers of the mutant allele (see Material and methods for details). The estimated value corresponding to a c value of 0.4 and a target interval of [1/30, 1/20] could not be calculated, since it is almost impossible to reach this target interval with a c value of 0.4.[16] The estimated value increased with c, whereas it decreased with the target frequency of carriers.

Table 2 Impact of the c parameter on the maximum likelihood estimator $\hat{\theta}$ of the recombination rate θ, for the actual data, and on its confidence interval, using both number of non-recombinant haplotypes x_f and number of different haplotypes x_h. The results are given for the recombination rate of the whole haplotype and for that of the disease locus with the different surrounding markers, using the values given in Table 1. In all cases, a geometric distribution of EFS was used, with a target range for the frequency of carriers of mutant allele of [1/30, 1/20]

	$c=0.6$	$c=1$
Total haplotype	0.041 (0.023–0.089)	0.051 (0.023–0.103)
D12S325	0.039 (0.019–0.079)	0.043 (0.021–0.089)
D12S103	0.033 (0.017–0.071)	0.039 (0.017–0.081)
D12S359	0.029 (0.013–0.063)	0.035 (0.013–0.073)
D12S312	0.025 (0.011–0.055)	0.027 (0.011–0.061)
D12S355	0.007 (0.003–0.023)	0.009 (0.003–0.027)
D12S83	0.011 (0.003–0.031)	0.013 (0.003–0.035)

Figure 1 Fraction $\overline{p_f}(z)$ of carriers of the most frequent haplotype among the carriers of a given founder's allele against the number of carriers (z) of this allele in a sample of 1000 individuals. Results are given for Poisson and geometric distribution and various level of cultural inheritance (c), and a recombination rate of 0.07. More points are shown when c increases, because it becomes easier to reach a given value of z. Results for the allele dropping method on the sample of genealogies of 978 individuals are also given (upper curve). Allelic association increased with c, and results are similar between the geometric distribution with high level of correlation and the allele dropping.

This study confirms that specific socio-demographic behaviour can have a strong impact on the genetics of human populations. We have shown that variance of EFS and correlation between generations together influence the level of allelic association, just as they have been shown to affect the frequency of initially rare variants.[16] The similarity between the results of our simulated populations and the real genealogies is good evidence of the validity of our method, which can thus be used in populations where genealogies are unknown.

Our first result is that the cultural inheritance (c) of EFS increases allelic association rather strongly (see Figure 1). Our results can be related to previous results, in a population with no correlation of EFS between generations.[21] For a given range of carrier frequency in the present population, an increase of population growth rate (λ) yields an increase of allelic association. Our explanation is that for low values of λ, the founder's allele can never reach a high carrier frequency, unless it has increased strongly during the stochastic growth that occurs during the first few generations. In this case, the coalescent tree of the chromosomes carrying the allele is starlike.[28] Thus, the total size of this coalescent tree is larger, more recombination events occur, and the decay of allelic association is stronger.

Here, an increase in c has the same impact as an increase in λ in the previous study, in the sense that an increase in c yields an increase of the growth rate of the number of copies of the few founder's alleles that succeed by chance in the process. Thus a high number of copies of the allele is reached with a smaller coalescent tree.

Estimating recombination rate

Because the maximum likelihood estimate ($\hat\theta$) of the recombination rate (θ) increases with c, not taking into account the level of cultural inheritance would yield an underestimate of the recombination rate. With a c value of 1, the estimate of θ is closer to the Généthon estimate of 7 cM for the haplotype than when lower values are used. It is even closer to the genemap estimate of 5.9 cM (http://www.ncbi.nlm.nih.gov/genemap/map.cgi). Thus, though one might suspect large errors to occur for estimates based either on familial or population studies, the result seem reasonably consistent in both cases. This c value of 1 ($r = 0.32$) is realistic, both in terms of demography and of changes in carrier frequency inferred in our population.[16] A more elaborate statistical study will be needed in the future to evaluate possible biases of this maximum likelihood estimate.

An increase of c yields an increase of the confidence interval around the estimator. Since a higher c value relaxes the constraints on the shape of the coalescent tree, more topologies are probable. The increase of both the length of the shared haplotype and the size of the confidence interval makes SLSJ unpractical for fine mapping. On the other hand, it makes it more suitable for broad scale mapping, since allelic association can be expected on a wider scale. Thus, this population becomes particularly interesting for genome screening of disease genes. This result is of great interest in the current debate on the most useful populations, as well as on the number of necessary markers loci, for such kinds of study.[29–32]

This study has shown that it is important to take demographic factors into account when estimating a genetic parameter like recombination rate. It adds another nuisance parameter (and not the least) to those already described by Terwilliger et al.[33] As we have already pointed out, cross-generational correlation of EFS is not unique to the SLSJ population.[17–20]

If no demographic data are available for the population under study, it would be useful to design a method to infer both recombination rate and demographic parameters from genetic data. On the other hand, if demographic data are available, it is important to integrate them into the estimation method for recombination rate, as we have done here. Currently, disequilibrium estimation methods are being applied to monogenic and oligogenic diseases, and application to complex diseases[34,35] is being considered. Allelic association can also be used as a 'genetic clock' to infer the time of foundation of a population.[36] Such estimates will also be affected by cultural inheritance of EFS. We must keep in mind that various elements in the socio-demographic history of human populations can affect genetic processes strongly, and that those elements need to be factored into estimation procedures.

References
1 Thompson EA, Neel JV: Allelic disequilibrium and allelic frequency distribution as a function of social and demographic history. *Am J Hum Genet* 1997; **60**: 197–204.
2 Hästbacka J, de la Chapelle A, Kaitila I, Sistonen P, Weaver A, Lander E: Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. *Nat Genet* 1992; **2**: 204–211.
3 Labuda M, Labuda D, Korab-Laskowska M *et al*: Linkage disequilibrium analysis in young populations: pseudo-vitamin D-deficiency rickets and the founder effect in French Canadians. *Am J Hum Genet* 1996; **59**: 633–643.
4 Luria SE, Delbrück M: Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* 1943; **28**: 491–511.
5 Kaplan NL, Weir BS: Are moment bounds on the recombination fraction between a marker and a disease locus too good to be true? Allelic association mapping revisited for simple genetic diseases in the Finnish population. *Am J Hum Genet* 1995; **57**: 1486–1498.
6 Kaplan NL, Hill WG, Weir BS: Likelihood methods for locating disease genes in nonequilibrium populations. *Am J Hum Genet* 1995; **56**: 18–32.

7 Rannala B, Slatkin M: Likelihood analysis of disequilibrium mapping, and related problems. *Am J Hum Genet* 1998; **62**: 459–473.

8 Slatkin M, Rannala B: The sampling distribution of disease-associated alleles. *Genetics* 1997; **147**: 1855–1861.

9 Graham J, Thompson EA: Disequilibrium likelihoods for fine-scale mapping of a rare allele. *Am J Hum Genet* 1998; **63**: 1517–1530.

10 Collins A, Morton NE: Mapping a disease locus by allelic association. *Proc Natl Acad Sci USA* 1998; **95**: 1741–1745.

11 Levy EN, Shen Y, Kupelian A *et al*: Linkage disequilibrium mapping places the gene causing familial Mediterranean fever close to D16S246. *Am J Hum Genet* 1996; **58**: 523–534.

12 Lehesjoki AE, Koskiniemi M, Norio R *et al*: Localization of the EPM1 gene for progressive myoclonus epilepsy on chromosome 21: linkage disequilibrium allows high resolution mapping. *Hum Mol Genet* 1993; **2**: 1229–1234.

13 Suzuki K, Bustos T, Spritz RA: Linkage disequilibrium mapping of the gene for Margarita Island ectodermal dysplasia (ED4) to 11q23. *Am J Hum Genet* 1998; **63**: 1102–1107.

14 Hästbacka J, de la Chapelle A, Mahtani MM *et al*: The diastrophic dysplasia gene encodes a novel sulfate transporter: positional cloning by fine-structure linkage disequilibrium mapping. *Cell* 1994; **78**: 1073–1087.

15 Engert JC, Berube P, Mercier J *et al*: ARSACS, a spastic ataxia common in northeastern Quebec, is caused by mutations in a new gene encoding an 11.5-kb ORF. *Nat Genet* 2000; **24**: 120–125.

16 Austerlitz F, Heyer E: Social transmission of reproductive behavior increases frequency of inherited disorders in a young-expanding population. *Proc Natl Acad Sci USA* 1998; **95**: 15140–15144.

17 Neel JV: Lessons from a 'primitive' people. *Science* 1970; **170**: 815–822.

18 MacCluer JW, Neel JV, Chagnon NA: Demographic structure of a primitive population: a simulation. *Am J Phys Anthropol* 1971; **35**: 193–207.

19 Murray-McIntosh RP, Scrimshaw BJ, Hatfield PJ, Penny D: Testing migration patterns and estimating founding population size in Polynesia using human mitochondrial DNA sequences. *Proc Natl Acad Sci USA* 1998; **95**: 9047–9052.

20 Motulsky AG: Jewish diseases and origins. *Nat Genet* 1995; **9**: 99–101.

21 Austerlitz F, Heyer E: Impact of demographic distribution and population growth rate on haplotypic diversity linked to a disease gene and their consequences for the estimation of recombination rate: example of a French Canadian population. *Genet Epidemiol* 1999; **16**: 2–14.

22 Heyer E, Tremblay M, Desjardins B: The 17th century European origins of hereditary diseases in the Saguenay population (Quebec, Canada). *Hum Biol* 1997; **69**: 209–225.

23 Gyapay G, Morisette J, Vignal A *et al*: The 1993–94 Généthon human genetic linkage map. *Nat Genet* 1994; **7**: 246–339.

24 Fu GK, Lin D, Zhang MY *et al*: Cloning of human 25-hydroxyvitamin D-1 alpha-hydroxylase and mutations causing vitamin D-dependent rickets type 1. *Mol Endocrinol* 1997; **11**: 1961–1970.

25 Wang JT, Lin CJ, Burridge SM *et al*: Genetics of vitamin D 1alpha-hydroxylase deficiency in 17 families. *Am J Hum Genet* 1998; **63**: 1694–1702.

26 MacCluer JW, VandeBerg JL, Read B, Ryder OA: Pedigree analysis by computer simulation. *Zoo Biol* 1986; **5**: 147–160.

27 Heyer E: One founder-one gene hypothesis in a new expanding population. The case of the Saguenay (Québec, Canada). *Hum Biol* 1999; **71**: 91–101.

28 Slatkin M, Hudson R: Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* 1991; **129**: 555–562.

29 Jorde LB, Watkins WS, Kere J, Nyman D, Eriksson AW: Gene mapping in isolated populations: new roles for old friends? *Hum Hered* 2000; **50**: 57–65.

30 Ott J: Predicting the range of linkage disequilibrium. *Proc Natl Acad Sci USA* 2000; **97**: 2–3.

31 Collins A, Lonjou C, Morton NE: Genetic epidemiology of single-nucleotide polymorphisms. *Proc Natl Acad Sci USA* 1999; **96**: 15173–15177.

32 Kruglyak L: Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat Genet* 1999; **22**: 139–144.

33 Terwilliger JD, Zollner S, Laan M, Paabo S: Mapping genes through the use of linkage disequilibrium generated by genetic drift: 'drift mapping' in small populations with no demographic expansion. *Hum Hered* 1998; **48**: 138–154.

34 Slatkin M: Disequilibrium mapping of a quantitative-trait locus in an expanding population. *Am J Hum Genet* 1999; **64**: 1764–1772.

35 Terwilliger JD, Weiss KM: Linkage disequilibrium mapping of complex disease: fantasy or reality? *Curr Opin Biotechnol* 1998; **9**: 578–594.

36 Labuda D, Zietkiewicz E, Labuda M: The genetic clock and the age of the founder effect in growing populations: a lesson from French Canadians and Ashkenazim. *Am J Hum Genet* 1997; **61**: 768–771.