Ann. Hum. Genet. (1998), **62**, 453–464
Printed in Great Britain

European Mathematical Genetics Meeting

held at Reading, April 1998

Abstracts

An algorithm for exact LH calculation for pedigrees with close inbred loops. YU. S. AULCHENKO and T. I. AXENOVICH. *Institute of Cytology and Genetics, Novosibirsk, Russia.*

The maximum likelihood method is widely used in segregation and linkage analysis of pedigree data. Efficient algorithms for calculation of the likelihood for a pedigree without loops have been developed and applied in a set of computer programs. Although general algorithms to compute the likelihood of a pedigree with loops have also been elaborated, they require too much storage space or CPU time in the case of a large pedigrees with multiple loops (e.g. livestock pedigrees).

Approximate methods look rather promising for analysis in the case of complex pedigrees. One of the approaches to the approximate likelihood calculation is based on cutting the loops (Stricker *et al*. 1995; Wang *et al*. 1996). Cutting all loops results in a zero-loop pedigree, whose likelihood might be calculated by well-known efficient algorithms. The problem is that by cutting the loops one can lose genetic information. The more loops are cut and the shorter the loops are, the more information is lost.

The variety of short loops is small and therefore it is possible to introduce additional procedures for peeling short loops. Minimal loops which are common in animal pedigrees are those resulted from backcrosses and crosses between sibs. Introduction of peelings of these loops allows the saving of large amounts of information.

We constructed a peeling algorithm for calculation of the likelihood for a pedigree with short inbred loops. This algorithm retains intact the main idea of efficient algorithm for peeling a pedigree without loops, that is to peel a part of the pedigree onto a single individual. This allows easy insertion of the new peeling into a program which uses peeling technique. Insertion of the new peeling does not decrease the algorithm's efficiency. It is noteworthy that the problem of finding an optimal peeling sequence does not arise in the implementation of the algorithm.

The segregation analysis of simulated data was used to estimate the efficiency of the approach proposed. The structure of the pedigree under analysis was identical to the structure of a real livestock pedigree. Three methods of likelihood calculation were applied: (i) exact, (ii) approximation by cutting all loops and (iii) approximation which used additional peeling (i.e. long loops were cut while short inbred ones were peeled). We used transmission probabilities approach for testing the major-gene effect. The frequency of false rejections of major-gene effect was only 0.06 when we used (iii), which is close to the frequency of false rejection of 0.04 obtained with (i), while the method (ii) rejected true hypothesis with a frequency of 0.15.

REFERENCES

Stricker, C., Fernando, R. L. & Elston, R. C. (1995). An algorithm to approximate the likelihood for pedigree data with loops by cutting. *Theor. Appl. Genet.* **91**, 1054–1063.

Wang, T., Fernando, R. L., Stricker, C. & Elston, R. C. (1996). An approximation to the likelihood for a pedigree with loops. *Theor. Appl. Genet.* **93**, 1299–1309.

Segregation analysis of animal pedigree data. T. I. AXENOVICH. *Institute of Cytology and Genetics, Novosibirsk, Russia.*

The extension of mixed model is proposed for the analysis of pedigree data coming from crosses between genetically heterogeneous populations. The mixed model has been introduced (Morton & MacLean, 1974) and widely used for analysis of quantitative traits on the base of pedigrees coming from the single population. The model is based on the following assumptions: three factors (major-gene, polygene and environmental conditions) control the quantitative trait and these factors act independently.

For the hybrid pedigrees we have to introduce the additional assumptions: (i) the rule of phenotypic expression of genotype is the same for members of parental populations and their hybrids, (ii) the rule of genotype transmission from generation to generation is the same for members of parental populations and their hybrids. Formally it means that the effects of any major-genotype, effect of single non-zero allele of the polygene and the transmission probabilities do not depend on the origin of the individual. In this case the parental populations are different only in frequencies of major-gene alleles and of non-zero alleles in all loci of the polygene.

In the framework of additional assumptions the major gene effect is described via residual genotypic values m_g: $m_{aa} = 0$, $m_{aA} = td$ and $m_{AA} = t$, where t is a displacement and d is a dominance parameter.

The polygene effect G is described via normal distributions with parameters z_1 and σ^2_{G1} for the first parental population and z_2 and σ^2_{G2} for the second parental population. The distributions of polygene for given hybrids may be described through the normal distribution with parameters:

$$z = (1-\epsilon)z_1 + \epsilon z_2$$

$$\sigma^2_G = (1-\epsilon)\sigma^2_{G1} + \epsilon\sigma^2_{G2} + \gamma[\epsilon_m(1-\epsilon_m) + \epsilon_f(1-\epsilon_f)],$$

where ϵ_m and ϵ_f are the blood-share of the second population for mother and father, respectively; $\epsilon = 0.5(\epsilon_m + \epsilon_f)$; γ is an additional parameter.

We introduce the heterotic effect $h = \eta[\epsilon_f(1-\epsilon_m) + \epsilon_m(1\text{-}\epsilon_f)]$, where η is the coefficient. It can be seen that $h = \eta$ for F1 hybrids ($\epsilon_m = 1$ and $\epsilon_f = 0$ or inverse); $h = 1/2\eta$ for F2 and backcross hybrids. In the presence of heterotic effect the penetrance function equals to the density of normal distribution $N(m_g + G + h, \sigma^2_e)$ in the point x:

$$p(x \boxtimes g, G) = \phi[(x - m_g - G - h), \sigma^2_e],$$

where σ^2_e is the variance due to environmental contribution.

The main difference between classical cross-breed analysis and method proposed here is that we do not assume that the initial parents are homozygous. The frequencies of alleles of the major-gene and polygene loci may be arbitrary and not the same in different populations. It allows us to analyse pedigree data coming from crossing between genetically heterogeneous populations, in particular from natural populations. However, the properties of tests used in segregation analysis (Elston & Stewart, 1971) are known for pedigrees from a single population only. Analysis of the properties of these tests in the case of hybrid pedigrees is the task of our future investigations.

REFERENCES

Elston, R. C. & Stewart, J. A. (1971). General model for the genetic analysis of pedigree data. *Hum. Hered.* **21**, 523–542.

Morton, N. E. & MacLean, C. J. (1974). Analysis of family resemblance. III. Complex segregation of quantitative traits. *Am. J. Hum. Gen.* **26**, 489–503.

MCMC methods for assessing departures from HW and gametic equilibrium. K. L. AYRES. *Department of Applied Statistics, University of Reading, Reading, UK.*

The assessment of departures from Hardy-Weinberg (HW) and gametic equilibria are important problems in statistical genetics, with increased interest due most recently to the use of linkage disequilibrium as a gene mapping tool (e.g. Jorde, 1995), and the study of population genetics issues surrounding the forensic use of DNA profiles. Traditional approaches to the investigation of disequilibrium have centred on hypothesis tests (e.g. Zaykin et al., 1995) or point estimation, due in part to their ease of implementation, a key factor at the time of their original development when computational power was limited. However, more information is available via probability density curves of disequilibrium parameters (Ayres & Balding, 1998).

Adopting a Bayesian approach, Markov Chain Monte Carlo methods are presented here for sampling from the posterior distributions of interest. The methods are flexible, and allow for the implementation of various models. First, a method is presented for sampling from the posterior distribution of the inbreeding coefficient (a measure of departure from HW), based on single-locus genotype data, and implemented via the inbreeding model. The more general 'fixation indices' model of Weir (1970) is also considered. The method is then extended to two loci, where the parameters of interest are additive gametic disequilibrium coefficients in addition to inbreeding measures. The methods are demonstrated via application to both simulated and published data. Implications for disequilibrium mapping are also discussed.

REFERENCES

Ayres, K. L. & Balding, D. J. (1998). Measuring departures from Hardy–Weinberg: a Markov chain Monte Carlo method for estimating the inbreeding coefficient. *Heredity* **80**, 769–778.

Jorde, L. B. (1995). Linkage disequilibrium as a gene-mapping tool. *Am. J. Hum. Genet.* **56**, 11–14.

Weir, B. S. (1970). Equilibria under inbreeding and selection. *Genetics* **65**, 371–378.

Zaykin, d., zhivotovsky, L. A. & Weir, B. S. (1995). Exact tests for association between alleles at arbitrary numbers of loci. *Genetica* **96**, 169–178.

Interpretation of linkage analysis with incomplete marker typing. M. C. BABRON[1], M. CHAVANCE[2], F. CLERGET-DARPOUX[1]. [1]*INSERM U155, Paris and* [2]*INSERM U472, Villejuif, France.*

The statistics used to test linkage, such as model-dependent (Morton, 1955), or model-free (Risch, 1993) lod scores, depend on two kinds of random fluctuations: one due to the drawing of the founders' genotypes, one due to the drawing of offspring' genotypes conditional on the genotypes of their parents. When the founders' typing are not available, the uncertainty on drawing of these genotypes may increase the variance of the statistics and thus affect both the test conclusion and the confidence interval of the estimated parameters (estimate of the recombination fraction in model-dependent methods, and estimate of the haplotype sharing vector in model-free analyses).

In the case of affected sib pair analysis, we investigated the variance due to random fluctuations of parental genotypes V(p) relative to that due to the overall fluctuations of both parents and offspring genotypes V(p + o). Under the null hypothesis of no linkage, we simulated 1000 replicates of nuclear families with 4 affected children and calculated the mean and variance of the MLS, assuming that parental typing were first available, and then not available. For one family, the ratio

V(p)/V(p+o) is negligible when the parents are known, whereas it reaches 22% when the parental genotypes are unknown. This ratio decreases as the number of families increases. It is largest when many affected children are issued from a small number of couples. In such a situation, results of linkage analyses should be interpreted with caution.

REFERENCES

Morton, N. E. (1955). Sequential tests for the detection of linkage. *Am. J. Hum. Genet.* **7**, 277–318.
Risch, N. (1990). Linkage strategies for genetically complex traits. III. the effect of marker polymorphism on analysis of affected relative pairs. *Am. J. Hum. Genet.* **46**, 242–253.

Evolutionarily stable stealing: game theory applied to kleptoparasitism. M. BROOM[1] and G. D. RUXTON[2]. [1]*School of Mathematical Sciences, University of Sussex, and* [2]*Division of Environmental and Evolutionary Biology, University of Glasgow.*

Food stealing (kleptoparasitism) occurs for many types of animals, and is especially common among birds, see Brockman &Taylor (1979). We present an individual-based model of a group of foraging animals. Individuals can obtain food either by discovering it themselves or by stealing it from others. Ruxton and Moody (1997) developed a mechanistic model where all opportunities to steal were taken. Given that challenging for a discovered food item costs time (which could otherwise be spent searching for an undiscovered item), attempting to steal from another may not always be efficient. We show that there is generally a unique strategy which maximises uptake rate, this is either to always or never challenge others. For any combination of parameter values, we can identify which strategy is appropriate. As a corollary to this, we predict that small changes in ecological conditions can under some circumstances (which we identify) cause a dramatic change in the aggressive behaviour of individuals. Further we investigate situations where searching for undiscovered food and searching for potential opportunities for stealing are mutually exclusive activities, i.e. that success at one can only be improved at the expense of the other. Using game theory, we are able to find the evolutionarily stable strategy for investment in these two activities, in terms of the ecological parameters of the model.

REFERENCES

Brockman, H. J. & Barnard, C. J. (1979). Kleptoparasitism in birds. *Anim. Behav.* **27**, 487–514.
Ruxton, G. D. & Moody, A. L. (1997). The ideal free distribution with kleptoparasitism. *J. Theor. Biol*, **186**, 449–458.

Applications of the Malecot model for allelic association. COLLINS, A. *Human Genetics, University of Southampton, Level G, Princess Anne Hospital, Southampton, UK.*

The Malecot model for isolation by distance, where distance here is between a marker and disease locus, has been implemented for mapping disease genes by allelic association and applied to published data sets. In addition to determining the location for the disease gene in a marker map other parameters reflect mono or polyphyletic origin, spurious association and the number of

generations since the original mutation(s). As an example of a monophyletic disease analysis of cystic fibrosis -F508 mutation haplotypes and a control sample placed the mutation to within 46 kb of its correct location (Collins & Morton, 1998). Applications to other diseases with polyphyletic origin and with heterogeneity in the genetic to physical map distances are encouraging. Using physical map distances the gene for Huntington's disease, a polyphyletic system, is localised to within its known location and the polyphyletic origin is reflected in a low value for parameter M = 0.282, (where M = 1 suggests monophyletic origin). The data for hereditary haemochromatosis (HFE) shows a poor fit to the physical map and a correspondingly poor location for the disease gene. However if the genetic map distances are used the fit is considerably improved and the error is only 0.035 cM. This reflects heterogeneity in the physical to genetic map distances due to a marked suppression of recombination proximal to HFE (the ratio of Mb/cM is 6.4).

In conclusion, this approach is effective for fine-scale mapping of both monophyletic and polyphyletic diseases given a reliable marker map and an understanding of the extent of variation in recombination rates in the candidate region.

REFERENCES

Collins, A. & Morton, N. E. (1998). Mapping of disease locus by allelic association. *Proc. Natl. Acad. Sci. USA* **95**, 1741–1745.

Fisher, Haldane and Wright in the decade 1922–1932. A. W. F. EDWARDS. *Department of Community Medicine, University of Cambridge, Cambridge, UK.*

The development of mathematical population genetics in the decade following the publication of R. A. Fisher's 1922 paper introducing stochastic models was traced in detail, with special reference to the contributions of, and Fisher's interactions with, J. B. S. Haldane and Sewall Wright. Particular attention was paid to Haldane's sequence of papers *A mathematical theory of natural and artificial selection* and his 1932 book *The Causes of Evolution*, to Wright's influential paper of 1931 *Evolution in Mendelian populations* and his 1932 Congress of Genetics paper, and to Fisher's 1930 book *The Genetical Theory of Natural Selection*.

Variation of the Maximum Likelihood Score. S. EICHENBAUM-VOLINE[1,2], M-C. BABRON[1], B. PRUM[2] and F. CLERGET-DARPOUX[1]. [1] *INSERM U155, Paris, France and* [2] *URA CNRS 1323, University of Paris V, France.*

The Maximum Likelihood score (MLS) (Risch, 1990) is a model-free method of linkage detection widely used in sib pairs analysis. The MLS enables the estimation of the probability of sharing alleles identical by descent (IBD) given observations on marker genotypes and measures its deviation from the expected IBD under no linkage. Under the hypothesis of no linkage, a null MLS is expected while under the alternative, positive values are expected. However, there is a great variation of the MLS values depending of the underlying genetic model. Furthermore, for a given alternative, the range of the possible values of the MLS may be huge and thus complicate the interpretation of linkage analysis and their replications. Indeed, given this great variation, once a significant result has been obtained in a first study, what can be called a replication when analysing a second sample ?

We computed the exact distribution of the MLS statistic at a completely informative marker locus in order to determine the 95% interval of variation of the MLS under different alternatives. As an example, for an IBD distribution at the marker locus equal to (0.18, 0.43, 0.39), the possible values for the MLS vary between 0.39 and 6.14 for a sample of 100 affected sib pairs. Thus, the analysis of two different samples could lead to two extreme MLS values, one being significatively different from 0 and the other not. A preliminary replication study, in terms of power, could enable us to avoid such contradictory results.

REFERENCES

Risch, N. (1990). Linkage strategies for genetically complex traits. III. The effect of Marker polymorphism on analysis of affected relative pairs. *Am J Hum Genet* **46**, 242–253.

Cluster analysis of the polygene system of determination of a quantitative trait. D. FURMAN, Yu. MATUSHKIN and P. MOROZOV. *Institute of Cytology and Genetics, Siberian Branch of the RAS, Russia.*

Results of applying cluster analysis to reveal polygenes in *achaete-scute*, a polygene system of determination of a quantitative trait in *D. melanogaster*. The system consists of the major gene *achaete-scute* located on the Õ chromosome and polygenes whose number and positions were unknown till recently. Major-gene mutations cause the fly body to lose macrochaetae, while polygenes control the frequency of the event (penetrance). The trait of interest is penetrance with respect to loss of macrochaetae in the *scute* mutant lines. The likely position of the polygenes were labelled with mobile elements (ME). To treat the problem, a correlation was to be found between the penetrance value and the availability of characteristic sites of localisation of ME in the lines with different genotypes.

The array of experimental data was a total of 168 localization sites of four MEs determined by *in situ* hybridization in 33 isogenic *scute* mutant lines with different penetrance (Furman *et al.*1995) Primary data processing included ranking the lines with respect to penetrance and breaking them into non-overlapping classes (penetrance group, PG). The breaking criterion was the significance of the differences ($p = 0.95$) between the maximum penetrance in the previous PG line and minimum penetrance in the next. Four PGs resulted, PG1 with 16 lines (mean penetrance 11.84%, s.d. 18.8); PG2 with 4 (penetrance 41.3%, s.d. 8.0); PG3 with 11 (penetrance 69.1%, s.d. 13.3); PG4 with 2 (penetrance 95.4%, s.d. 3.8).

The sites were classified according to which PG any of the four mobile elements was specific for entirely. The sites were compared by pairs and their similarity with respect to the criterion was evaluated using various measures, including Nei's estimate (Nei & Miller, 1990), Euclidean distance, modified Euclidean measure with weights, etc. Results were gathered into a matrix in which the sites were automatically clustered using UPGMA implemented in VOSTORG software (Zharkikh *et al.* 1991). The comparison of the contents of the clusters that followed from applying all these measures showed that the clustering remains the same no matter which measure of similarity. Twenty-four significant sites have been pinpointed: eight in PG1; one in PG2, two in GP4; the other 13 in more than one PG at a time. Note the number of characteristic sites does not depend on the number of lines in the PG. Data obtained allow the genome regions labelled with sites referring to different PGs to be associated with the positions of the polygenes that modify penetrance.

REFERENCES

Furman, D. P., Rodin, S. N. & Kozhemyakina T. A. (1995). Transposable elements and the penetrance of quantitative characters in *D. melanogaster*. *Theor. Appl. Gen.* **91**, 1095–1100.

Nei, M. & Miller, J. C. (1990) A simple method for estimating average number of nucleotide substitutions within and between populations from restriction data. *Genetics* **125**, 873–879.

Zharkikh, A. A., Rzhetsky, A. Yu., Morozov, P. S. *et al.* (1991). VOSTORG: a package of microcomputer programs for sequence analysis and construction of phylogenetic trees. *Gene* **101**, 251–254.

Non-independence of parental transmissions in the transmission disequilibrium test (TDT). A. P. MORRIS, J. C. WHITTAKER and R. N. CURNOW. *Department of Applied Statistics, University of Reading, Reading, UK.*

The transmission disequilibrium test was developed originally as a test of linkage of an unknown disease locus to a bi-allelic marker locus (Spielman *et al.* 1993). The test is based on the observation of marker alleles transmitted from parents to affected offspring. In the TDT analysis, parental transmissions of marker alleles within a nuclear family are treated under the null hypothesis of no linkage. In general, however, parental transmissions are not independent. It can be shown, asymptotically, that the form of the test statistic and the symmetry of allelic transmissions under the null hypothesis of no linkage removes any covariance between parental contributions (Whittaker *et al.* 1998). Thus parents can be treated as independent in a TDT analysis.

REFERENCES

Spielman, R. S., McGinnis, R. E. & Ewens, W. J. (1993). Transmission test for linkage disequilibrium: the insulin gene region and insulin dependent diabetes mellitus (IDDM). *Am. J. Hum. Genet.* **52**, 506–516.

Whittaker, J. C., Morris, A. P. & Curnow, R. N. (1998). Using information from both parents when testing for association between marker and disease loci. *Genet. Epidemiol.* **15**, 193–200.

Meta-analysis for complex inheritance. MORTON, N. E. *Human Genetics, University of Southampton, Level G, Princess Anne Hospital, Southampton, UK.*

Evidence from linkage and allelic association on the location of oligogenes is commonly reported graphically (which defies meta-analysis), or as a statistic giving significance as χ^2 with n d.f. or a significance level P. Methods to synthesize this evidence transform χ^2_n to P, P to $\chi^2_{(1)}$ and $\chi^2_{(1)}$ to a lod. Analysis of such lods assumes that each is based on the same amount of information and so $\chi^2_{(1)} = t^2$. While grossly approximate, this leads to acceptable statistics for main effect as $(\mathrm{St})^2/\mathrm{n}$ and for homogeneity as the residual. The latter is distributed as expected under $\mathrm{H_0}$ except for genuine heterogeneity in the tail of the distribution. This analysis defines candidate regions for definitive analysis of the more informative but unfortunately rare case where effect and location are simultaneously estimated and combined efficiently with other studies presented in the same way, although not necessarily with the same markers, phenotypes, or ascertainment or from the same ethnic group.

Block updating in constrained Markov chain Monte carlo sampling with application to pedigree analysis. N. A. SHEEHAN. *Department of Mathematical Sciences, Loughborough University, UK.*

Probability and likelihood calculations are central to the analyses of genetic data on groups of related individuals, or pedigrees. Such calculations become computationally intensive when the pedigree, viewed as a graph, has too many cycles, or loops. The *peeling* method for computing probabilities on pedigrees, generalised by Cannings *et al* (1978) from the algorithm of (Elston & Stewart, 1971) to include arbitrarily complex pedigrees and genetic models, has limitations in practice because of the enormous storage requirements involved. Thus, exact calculations are often not possible, even for simple genetic systems, on large complex pedigrees unless some structural information is discarded. Computational problems can also arise on small relatively simple pedigrees, especially in the area of linkage analysis, with the availability of more and more polymorphic markers creating difficulties for exact multipoint linkage calculations.

Markov chain Monte Carlo methods provide an alternative to interfering with the pedigree structure (see Thompson (1994), for example). Unlike many other Bayesian applications of these methods, the problem of simulating realisations from the posterior distribution of genotypes given phenotypic data on a pedigree requires simulation from a distribution subject to constraints. The individual by individual Gibbs sampler (Geman & Geman, 1984) was shown to provide good estimates of posterior ancestral genotype probabilities on a highly complex pedigree of Greenland Eskimos (Gilberg *et al*. 1978) which cannot be peeled for any genetic system, however simple, by Sheehan (1992), for example. However, once the genetic system of interest has more than two alleles, the underlying Markov chain is not guaranteed to be irreducible and the Gibbs sampler does not necessarily converge to the true posterior distribution of genotype given phenotype. The resulting estimates are thus unreliable. Reducibility is data-dependent, but, to date, there is no deterministic algorithm for detecting the problem on a large complex pedigree (Jensen & Sheehan, 1998).

Several ways of getting around this problem have been proposed over the last few years, most of which involve relaxing the genetic model and augmenting the statespace of the Markov chain. The method of importance sampling with weights of zero and one, proposed by Sheehan & Thomas (1993), is one example. Although this gets around the reducibility problem, it may necessitate very high rejection rates on a large complex pedigree and is thus inefficient. Slow mixing is always a concern with these methods. Even when the chain is irreducible, it may not converge to the true posterior distribution in 'finite' time (Geyer, 1992).

Joint updating of individuals, or blocking, is one approach to improve mixing. The Blocking Gibbs sampler of Jensen *et al*. (1995), although not guaranteed to be irreducible if the genetic system has more than two alleles, has been shown to be very efficient in many applications. The proposal here is to apply the method of Hurn *et al*. (1998) to a pedigree problem, by which a random block updating approach is suggested which combines the ideas of blocking with importance sampling. The resulting sampler can be shown to be irreducible but, with careful choice of the blocking scheme, should also mix quickly. The method is currently being tested on the Greenland Eskimo pedigree.

REFERENCES

Cannings, C., Thompson, E. A. & Skolnick, M. (1978). Probability functions on complex pedigrees. *Advances in Applied Probability*, **10**, 26–61.

Elston, R. C. & Stewart, J. (1971). A general model for the genetic analysis of pedigree data. *Human Heredity* **21**, 523–542.

Geman S. & Geman D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **45**, 721–741.

Geyer, C. J. (1992). Practical Markov chain Monte Carlo (with discussion). *Statistical Science* **7**, 473–511.

Gilberg, A., Gilberg, L., Gilberg, R. & Holm, M. (1978). Polar Eskimo genealogy. Meddelelser on Gronland, Nyt Nordisk Forlag Arnold Busck, Kobenhavn 203 (4).

Hurn, M. A., Rue, H. & Sheehan, N. A. (1998). Block updating in constrained Markov chain Monte Carlo sampling. *Statistics and Probability Letters* In press.

Jensen, C. S., Kong, A. & Kj Ae Rulff, U. (1995). Gibbs sampling in very large probabilistic expert systems. *International Journal of Human-Computer Studies* **42**, 647–666.

Jensen, C. S. & Sheehan, N. (1998). Problems with determination of non-communicating classes for Markov chain Monte Carlo applications in pedigree analysis. *Biometrics* **54** (2), 253–262.

Sheehan, N. (1992). Sampling genotypes on complex pedigrees with phenotypic constraints: the origin of the B allele among the Polar Eskimos. *IMA Journal of Mathematics applied in Medicine and Biology* **9**, 1–18.

Sheehan, N. & Thomas, A. (1993). On the irreducibility of a Markov chain defined on a space of genotypic configurations by a sampling scheme. *Biometrics* **49**, 163–175.

Thompson E. A. (1994). Monte Carlo Likelihood in Genetic Mapping. *Statistical Science* **9** (3), 355–366.

Genetic Anticipation in Familial Hodgkin's disease. Y. SHUGART. *Department of Family Medicine and Clinical Epidemiology, University City of Pittsburg, USA.*

Anticipation is a phenomenon that disease severity increases or age at onset decreases in successive generations. This hypothesis was challenged by Penrose (1948). In his view, so called 'anticipation' was causes by a number of ascertainment bias including (1) overrecruitment of parents with late disease onset due to limited reproductive capability in the HD parents with early age of onset; (2) preferential selection of offspring with early onset due to rarity or severity; and (3) selection of cases with simultaneous onset in parent and child resulting in increased physician awareness of both individuals being affected and increased reporting by physicians. Even though Penrose's point of view is theoretically sound, it suffers from lack of evidence in practice. The discovery in rapid succession of several diseases caused by expansion of CTA and CAG triplet repeats has provided a biological basis for anticipation, even if it is unlikely to account for all reported disease which exhibit anticipation. In this paper, we will describe the observation that age at onset decreases with each generation in Hodgkin's disease (HD) and discuss a few notions put forwarded by Penrose regarding ascertainment bias.

Because the detection of HD parent-child pairs is a rare event, we pooled several published data sets to test whether there is a difference in cancer age of onset between parents and children who are affected with HD. Thirty parent-child pairs with confirmed diagnosis and well-documented age of diagnosis were included in this study. In all pairs except one, HD children reveal a younger age of onset. The mean age of onset is 46 in parents and 22 in children. This significant difference between the age of diagnosis of parents and children was detected using the Mann–Whitney test ($n = 30$, $U = 40.5$, $p < 0.0001$). One may argue that the smaller number of parents of relatively young age among the pairs reported in the 50's may be due to reduced fitness as a consequence of poorer treatment. To address this issue, the analysis was repeated after removing these pairs. The difference

in age of onset between the two generations still remained significant (Mann-Whitney, $n = 19$, $U = 12.5$, $p < 0.0001$) and the mean age of onset is 44 in parents and 20 in children. Therefore, our results support the hypothesis of anticipation in familial HD. Nevertheless, as pointed out by Penrose (1948), false claims of genetic anticipation may be the result of various selection biases. A more optimal study design should be based on prospectively selected cases as discussed by Horwotz *et al* (1996). In addition, infectious agents such as the Epstein–Barr Virus (EBV) have been implicated in the etiology of familial HD. The observed 'anticipation' may also be related to simultaneous parent-child exposure to viral infection. To unfold this intriguing relation, our further study will focus on population based ascertainment design in which we can possibly correct for bias. We are also planing to examine the EBV status of relative pairs through a collaboration.

REFERENCES

Cimino, G., LoCoco, F., Cartoni, C., Callerano, T., Luciani, M., Lopez, M., de Rossi, G. (1988). Immune-deficiency in Hodgkin's disease (HL): a study of patients and healthy relatives in families with multiple cases. *Eur. J. Cancer Clin. Oncol.* **24**, 1595–1601.

Devore, J. W. & Doan, C. A. (1957) Studies in Hodgkin's syndrome. XII. Hereditary and epidemiological aspects. *Ann. Intern. Med.* **47**, 300–316.

Haim, N., Cohen, Y. & Robinson, E. (1982) Malignant lymphoma in first-degree blood relatives. *Cancer* **49**, 2197–2200.

Hors, J. & Dausset, J. (1983) HLA and susceptibility to Hodgkin's disease. *Immu. rev.* **70**, 167–192.

Hors, J., Steinberg, G., Andrieu, J. M., Jacquillat, C., Minev, M., Messerschmitt, J., Malinvaud, G. *et al.* (1980). HLA genotypes in familial Hodgkin's disease. Excess of HLA identical affected sibs. *Eur. J. Cancer* **16**, 809–812

Horwitz, M., Goode, E. L., Jarvik, G. P. (1996). Anticipation in familial leukemia. *Am. J. Hum. Genet.* **59**, 990–998.

Lin, A. Y., Kingma, D. W., Lennette, E. T., Fears, T. R., Whitehouse, M., Ambinder, R. F. *et al.* Epstein–Barr Virus and Familial Hodgkin's Disease. *Blood* **88**, 3160–3165.

Penrose (1948). The problem of anticipation in pedigrees of dystrophia myotonica. *Ann. Eugen.* **14**, 124–132.

Razis, D. V., Diamond, H. D. & Craver, L. F. (1959). Familial Hodgkin's disease: its significance and implications. *Ann. Intern. Med.* **51**, 933–971.

Vianna, N. J., Davies, J. N. P., Polan, A. K. & Wolfgang, P. (1974). Familial Hodgkin's disease: an environmental and genetic disorder. *Lancet* **2**, 854–857.

Finite sample properties of tests for allelic association. J. C. WHITTAKER[1] and D. THOMPSON[2]. [1]*Department of Applied Statistics, University of Reading, UK and* [2]*CRC Genetic Epidemiology Unit, University of Cambridge, UK.*

There has been much recent interest in the use of family-based association tests to detect linkage between marker and disease loci. A number of test statistics have been proposed, and the strengths and weaknesses of these statistics have been much debated. Investigation of test statistic properties has usually been done by simulation or by reliance on standard asymptotic results. In this talk we

discuss relationships between some of the commonly used association tests, concentrating in particular on biallelic markers. For biallelic markers it is easy to calculate the finite sample properties of the test statistics under consideration, and this gives some interesting results. In particular, the standard chi-squared approximation used to assess significance is found to be poor if we require very low Type I error rates. Such very low Type I error rates are necessary if association tests are to be used in genome scans, as has recently been advocated by Risch & Merikangas (1996).

Risch, N. & Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science* **273**, 1516–1517.

Inference for Ancestral Populations. I. WILSON. *QMW, London* (*Visiting Research Fellow, Department of Applied Statistics, University of Reading*).

Patterns of similarity and difference in genes sampled from natural populations carry information about both the evolutionary history of the genes and evolutionary processes. Genes sampled from different subpopulations within a species can tell us about the historical relationships between subpopulations, and the process of gene flow between them. Here we develop a Markov-chain based inferential method for 'local' evolutionary parameters, for example gene flow and population sizes – which removes much of the computational complexity compared to full genealogical methods, and also allows us to separate recent events from more distant population histories and hence allow discrimination between models for evolution. These methods are demonstrated on data for microsatellite variation from introduced red foxes (*Vulpes vulpes*) surveyed from a small island and neighbouring mainland populations in Australia, described by Lade *et al.* (1996).

REFERENCE

Lade, J. A., Murray, N. D., Marks, G. A.. & Robinson, N. A. (1996). Microssatellite differentiation between Philip Island and mainland Australian populations of the red fox *Vulpes vulpes*. *Molecular Ecology* **5**, 81–87.

Bayesian estimation of molecular evolutionary trees. ZILHENG YANG[1] and BRUCE RANNALA[2]. [1]*Department of Biology, University College London and* [2]*Department of Integrative Biology, University of California at Berkeley.*

Since a phylogenetic tree representing the relationships among species is different from a conventional statistical parameter, use of the maximum likelihood methodology for phylogeny estimation involves complexities (Yang *et al.* 1995). The Bayesian approach appears more natural. We use a birth-death process with species sampling to specify the prior distribution of phylogenies and ancestral speciation times and a Markov process model of nucleotide substitution to calculate the probability of the data given the prior. The posterior probabilities of trees are then calculated for tree selection. Monte Carlo integration was used to integrate over the ancestral speciation times, and a Markov chain Monte Carlo algorithm was used to generate trees with the highest posterior probabilities. Application of the method to a data set of DNA sequences from nine primate species generated reasonable results (Rannala & Yang, 1996; Yang & Rannala, 1997).

REFERENCES

Rannala, B. & Yang, Z. (1996). Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. *J. Mol. Evol.* **43**, 304–311.

Yang, Z. & Rannala, B. (1997). Bayesian phylogenetic inference using DNA sequences: a Markov chain Monte Carlo method. *Mol. Biol. Evol.* **14**, 717–724.

Yang, Z., Goldman N. & Friday A. E. (1995). Maximum likelihood trees from DNA sequences: a peculiar statistical estimation problem. *Syst. Biol.* **44**, 384–399.