obtained. Control genotypes were based on these frequencies assuming Hardy-Weinberg equilibrium. Case genotypes were simulated under two contrasting models, one where disease risk is influenced by only one SNP and one where a combination of SNPs is important. The four selection methods compared were R2 (Chapman et al, 2003. Hum Hered 56:18-31), an entropy-based method (Butler et al, in press. BMC Genetics) and two simple methods, one that picks equally spaced SNPs and one that chooses the SNPs with the highest minor allele frequencies. Results will be presented comparing the (equally sized) subsets of SNPs selected by the four methods. The power to detect association between the gene and disease will be compared between the four methods under the different disease models.

The Length of External Branches in Coalescent Trees

A. Caliebe[1], R. Neininger[2], M. Krawczak[1] and U. Rösler[3]
[1] *Institute of Medical Informatics and Statistics, University of Kiel*
[2] *Department of Mathematics, University of Frankfurt*
[3] *Department of Mathematics, University of Kiel*

Keywords: Coalescence, external branches, weak convergence, neutral evolution

We study the asymptotic distribution of the length of a typical external branch in a coalescent tree. If Z(n) denotes the length of a randomly chosen external branch in a coalescent tree of n individuals, the following recursion applies in distribution

$$Z(n) = B(n)Z(n-1) + T(n),$$

where B(n) is Bernoulli distributed with parameter $1-2/n$ and T(n) is exponential distributed with parameter n choose 2.

We show that in distribution nZ(n) converges to Z, where Z has Lebesgue density

$$f(x) = 8/((2+x)^3)$$

for x in [0,infinity).

This asymptotic distribution allows the determination of the time to the first common ancestor for a randomly chosen individual and any of the remaining individuals in the population or sample. It describes the relatedness of a randomly chosen individual to the rest of the population.

Caution in the interpretation of association in presence of missing data

P. Croiseau, F. Clerget-Darpoux, M. C. Babron and E. Génin
INSERM U535, Villejuif, France

Keywords: Missing data, association studies, Multiple Sclerosis, CTLA4

Association studies between a disease and a set of markers are often performed to identify variants which could be involved in the disease susceptibility. This was the case for a recent study of 450 multiple sclerosis family trios in which an association was detected with one SNP in the promotor of the CTLA4 gene. To determine if this SNP was the most likely involved in the disease susceptibility, eighteen others SNPs selected to ensure a good coverage of the CD28-CTLA4-ICOS gene cluster were genotyped and analyzed using the program Unphased [F. Dudbridge, 2003]. This program performs an association analysis of haplotypes from unphased genotype data. In presence of missing data, all possible phased genotypes are considered and weighted by the appropriate haplotype frequencies estimated from the available data.

We propose here to evaluate by simulation the performance of this approach in the situation of missing data. We assume the effect of one of the 19 SNPs, and simulate the genotyping of all the SNPs in 450 family trios with the same linkage disequilibrium and proportion of missing data as in the real sample.

We show that if the variant involved in the disease susceptibility is poorly typed, association may be found stronger with another variant. Caution is therefore needed in the interpretation of association studies in the situation of missing data and poorly typed markers should not be discarded from further studies.

A spatial probit model for fine-scale mapping of disease genes

M. De Iorio and C. Verzilli
Dept. of Epidemiology and Public Health, Imperial College London, London, UK

Keywords: association studies, Bayesian probit regression, retrospective likelihood

We present a novel statistical method for disease gene association mapping by linkage disequilibrium which leads to markers physically close to each other along the chromosome tending to be correlated. The pattern of linkage disequilibrium is influenced by many factors such as recombination, mutation, genetic drift and the full extent of variation in linkage disequilibrium patterns is not yet well characterised.

We propose a Bayesian multivariate probit model for phase-known marker data that can take into account the spatial correlation between markers. In a case-control setting, we use a retrospective likelihood that can accurately reflect the study design enabling us to identify regions where marker frequencies differ significantly across cases and controls and, in addition, formally quantify these differences. Full Bayesian inference is implemented through Gibbs sampling. We demonstrate our approach on simulated and real data, comparing the results with those obtained from models which ignore the apparent spatial correlation in the data.

Space and temporal analysis to estimate the evolution of the yield by balanced lethal systems

M. Delucis[1], R. Boggio[1], D. Almorza[2] and J. C. Salerno[3]
[1] *National University of La Plata (Argentina)*
[2] *University of Cádiz (Spain)*
[3] *Genetic Institute INTA-Castelar (Argentina)*

Keywords: inbred lines, backcross, heterosis, grain yield trials, genetic load

Several researches carried in the Genetic Inst. E. Favret, about the relation between the genetic load and the yield in different populations of maice, show a positive regression with a very high level of significance between the yield and the frequency of lethals. We have realized several essays in space and time testing lines with and without lethals to evaluate the yield and the transfer of balanced systems to public lines and to estimate the evolution of the yield. In this work we prove that the balanced lines showed a yield higher than the ones obtained with the non balanced lines and the public lines through the years.

Material and Methods. Data from essays were analized to test the behaviour of the line BLS1, that includes one chlorophyll lethal and one gametic gen. From this line is derived another that includes only the chlorophyll gen. They were tested with four public lines. In every essay we used randomized blocks design with 3 replications. With the same criteria in other comparative grain yield trial, another line with chlorophyll lethals, and another one without lethals, were tested for 4 years. Essays to test the transfer of balanced systems to public lines were made through backcrossing.

Data from essays for two years carried in Argentina were analyzed to test the behaviour of simple hibrids wich include in their genetic load both parental lines with lethal balanced, another with only one of with lethal balanced, and another group without lethals in both parental lines.

Result. Balanced lines showed a higher yield than their derivated non balanced lines and than the public lines.

Reference

Boggio, R., Sorarrain, O. M., Salerno, J. C. & Favret, E. A. (1997) Theoretical Analysis of Lethal Factors in Plant Populations. *Math Bios* **140**, 85–99.