

**2 Postdoctoral Research Assistants (PDRA). Academic-related Research Staff Grade 1A:
Salary £19,460 - £29,128**

Two postdoctoral research positions of three years duration in Bioinformatics funded by BBSRC are available for working with Prof. Jotun Hein. This project will focus on models of sequence evolution that includes a process of insertion-deletion of nucleotides and their implementation. Furthermore, developed methods will be applied to complete genomes as they are determined in coming years. The work will include algorithm development, statistical modelling, software development and large scale comparative analysis of genomes. The research will take place at The Oxford Centre for Gene Function. Earliest starting date is 1st November 2004.

The Research Project: Practical Statistical Alignment.

Although bioinformatics perceived is a new discipline, certain parts have a long history and could be viewed as classical bioinformatics. For example, application of string comparison algorithms to sequence alignment has a history spanning the last three decades. The principle of choosing solutions by minimizing the amount of evolution is also called parsimony and has been widespread in phylogenetic analysis even if there is no alignment problem. Over the last two decades the parsimony method of phylogenetic reconstruction has been severely criticized and has lost terrain to methods based on stochastic modelling of nucleotides, codons or amino acids. The present project will develop, implement and simultaneously apply methods including insertion-deletions of nucleotides/amino acids to give a full model of sequence evolution.

At present there are only 2 techniques that can solve these problems – dynamic programming and MCMC – so that is where our focus lies. We are shifting emphasis from the former to the later; however, this does not mean dynamic programming will be without value. Dynamic programming provides exact solution and MCMC provides only approximation. The former can be used to test the validity of the later on small data sets.

The key goals of this project are:

i. Software package: A central aim of the project is to develop a package doing Bayesian statistical alignment for many sequences and performs a series of essential associated analyses: hypothesis testing for molecular clock, phylogeny, mutation rate, selection estimation, assessing ancestral sequences, and structure prediction. It should produce nice graphical outputs of different marginalisation (tree, alignment, structure, ancestral sequences, and posterior distribution of mutation rates) for easy visualization. Prior parameters will be allowed to be set by users. This package of easy to use, biologist-friendly software should be of wide use in the community. A web based server will be developed that researchers can submit jobs to.

ii. Development of better biological models and corresponding algorithms: The TKF91 model allows only single insertions and deletions which is biologically irrelevant. Modelling long insertions and deletions is necessary to make statistical alignment practical. Site heterogeneity and residue-dependent indel rates would also increase the strength of the method.

iii. Structure prediction: Most of the comparative structure prediction methods are based on alignments, which might be misleading when the alignment is wrong. Statistical alignment will improve this.

iv. Comparative genomics: At present the capabilities of our statistical alignment methods matches the growth of mammalian genomes. Several questions are of interest here. Since statistical alignment is based on a general model of molecular evolution including both insertions and deletions, issues related to the nature of these events can be addressed and parameterized. What is the nature of insertions and deletions, their length distribution and rates? There is no reason to assume that insertion process is the time reversed process of deletion as assumed by many models. Quantifying the departure from this

assumption would be of great interest. The removal of the intermediate alignment step makes statistical alignment very suited for large scale genomic analysis and question such as departure from the molecular clock and time reversibility of the underlying substitution process can be tested. A good model of neutral molecular evolution also increases the power to detect selection which is central in functional interpretation of individual genes. An especially interesting question at present is the enigma of the amount of selection outside genes.

The Appointments

Applications for the PDRA posts are invited from researchers interested in working on this high-impact project – the development of statistical models, software and the analysis of genomic data. Candidates should have relevant background and research experience in computer science, mathematics, statistical modelling and computational biology, preferably in more than one of these areas, and have solid experience of applying these skills in the biosciences. A PhD, or equivalent research experience, will be required.

Genomics and Bioinformatics are currently a particularly exciting area of modern science. There is an acute shortage of researchers trained on the analytical side of the subject, so that future job prospects for those moving into the field are very good, in both the academic and the private sectors.

The posts thus offer the opportunity of a substantial impact on one of the major international projects in human genetics, within the context of very strong academic groups.

The Department of Statistics

Professor John Gittins is currently Head of the Department of Statistics. The department has four established professorships (Professor Peter Donnelly, Professor Jotun Hein, Professor Brian Ripley and Professor S. Lauritzen), and an established readership (Dr Peter Clifford). Other members of the academic staff are Dr J F Bithell, Dr A M Etheridge, Professor R C Griffiths, Dr R W Hiorns, Dr C N Laws, Dr M Lunn, Dr A D Lunn, Professor C J H McDiarmid, Dr G McVean, Dr P Northrop and Dr G Reinert. Sir David Cox, although retired, is also based in the department.

In addition to the Bioinformatics group there are two other major research groups in the department: the Mathematical Genetics group led by Professor Donnelly, and the Computationally Intensive Statistics group led by Professor Ripley.

The department is located in two pleasant, adjacent, Victorian buildings, together with an annexe in the Peter Medawar Building for Pathogen Research and Oxford Centre of Gene Function (OCGF). There are well-equipped networks of Sun workstations, a cluster of fast Intel processor machines, and excellent computing support staff.

The department runs an MSc course in Applied Statistics and is in the process of setting up and recruiting staff to run a post-experience part-time MSc in Bioinformatics. The department is also extensively involved in undergraduate teaching.

General conditions

The current salary scales for the posts are given above, and the person appointed will progress by annual increments from a starting point, to be determined by qualifications and experience, to the maximum of the scale. Well-qualified successful applicants for the posts are likely to be appointed at or near the top of the respective salary scales. Each appointment will be for three years and will be subject to an initial probationary period of six months, during which the appointment may be terminated by one month's notice on either side.

The age of retirement for university appointments is 65. Staffs are eligible to join the Universities Superannuation Scheme, a contributory scheme to which members pay 6.35 per cent of annual salary. Subject to the Statement of Pensions Policy which will be issued to the successful candidate, the

appointee will be deemed to be in membership of the above pension scheme until such time as he or she gives notice in writing to exercise the right not to be a member of the scheme.

The appointee will be entitled to 38 days of annual leave (this figure of 38 includes 8 public holidays). The actual distribution of annual holidays, excluding the public holidays, is subject to the agreement of the Department.

The University has a generous maternity leave scheme, which goes well beyond the statutory provisions; and operates two childcare nurseries and a holiday programme for school age children.

Holders of postdoctoral positions (which would also include the scientific programming post) in the Department of Statistics are required to help with support teaching in the department, for an average of three hours per week over the three 8 week academic terms, for which no additional remuneration is provided. This teaching would typically take the form of supervising or running practical or exercise classes, or occasional lecturing, and every attempt is made to match the teaching to the postholders' interests.

Application Procedure

Informal enquiries about the post should be directed to hein@stats.ox.ac.uk.

Applications for the postdoctoral posts should comprise a *curriculum vitae*, details of relevant experience, and a list of publications together with the names, addresses, telephone, fax and e-mail details of three referees. Applicants should ask their referees to write directly to the department so that references arrive by the closing date.

Applications and references should be submitted to Mrs Sue Wood, Department of Statistics, 1 South Parks Road, Oxford, OX1 3TG. Applications faxed to +44 1865 272595 or e-mailed to wood@stats.ox.ac.uk are acceptable, provided they are followed by hard copy.

The closing date for applications is **Friday 24th September 2004**.