

Scientific Report
For the Exploratory ESF Workshop
**“New tools for biodiversity conservation through the advancement of
phylogeographic methodologies”**

Reference: EW05-071

Title: New tools for biodiversity conservation through the advancement of
phylogeographic methodologies

Date: Monday 03 July 2006 – Thursday 06 July 2006

Location: University of East Anglia,
Norwich NR4 7TJ, UK

Telephone: (+44) (0)1603 456161
Fax: (+44) (0)1603 458553
Web page:
<http://www.cmp.uea.ac.uk/~vlm/cbg/workshops/esf/>

Convenors:

Professor Vincent Moulton,
School of Computing Sciences,
University of East Anglia,
Norwich NR4 7TJ, UK

Telephone: (+44) (0)1603 592607
Fax: (+44) (0)1603 593345
E-mail: Vincent.moulton@uea.ac.uk
Web site: <http://www.cmp.uea.ac.uk/people/vlm>

Dr Brent Emerson,
School of Computing Sciences,
University of East Anglia,
Norwich NR4 7TJ, UK

Telephone: (+44) (0)1603 592237
Fax: (+44) (0)1603 592250
E-mail: b.emerson@uea.ac.uk
Web site: <http://bioweb2.bio.uea.ac.uk/bioperson/facultyasp/EmersonB.aspx>

1. Executive summary

The aim of this exploratory workshop was to bring together a limited number of key researchers from across Europe in the fields of computational biology and phylogeography to explore the growing need for new methodologies and software for the analysis of the geographical distribution of genetic variation on regional and global scales. The Workshop was organised in such a way as to promote intensive discussion and critical appraisal of the state-of-the-art methodologies for analysing phylogeographic data, and to encourage future collaborations.

The general structure of the workshop included presentation sessions, in which speakers presented current challenges, and discussion sessions, in which participants were provided time to formulate possible new directions to address these. We now provide a short summary of the content of these sessions:

DAY 1

Presentation of the European Science Foundation (ESF). Since a representative of the ESF was unable to make it to the workshop, Prof. Moulton opened the workshop by providing an overview of the ESF, detailing the various organs of the foundation, its overall remit and membership, and the various functions it performs.

Presentation session I: Current problems in phylogeographic assessment of biodiversity. A current challenge in phylogeography is that the ease of obtaining genetic data has far outpaced the development of the analytical methods necessary to make full use of the data. This presentation session covered some of the current problems in the area of data analysis.

Discussion session I: Identification of themes and problems. In this session discussion points were identified for the rest of the conference. A lot of the discussion focused on the challenge in establishing a common language for phylogeographers and computational biologists to explore new methodological problems.

Presentation session II: Phylogenetic methods for the analysis of biodiversity.

It is becoming increasingly easier to obtain data from different DNA sequence regions from the same organism, and from different organisms with the same geographic distribution. This presentation session gave an opportunity for computational biologists to put forward various methodologies for analysing such data.

DAY 2

Presentation session III: Phylogeography in practice. The recent growth of phylogeography was fueled by the technical accessibility of mitochondrial (mt) DNA sequences in animal species. This growth appears to be exponential. Today the range of DNA techniques, combined with new analytical phylogeography methods and recent palaeoclimatic /geological studies, are providing considerable insight into the distribution and evolution of genetic diversity around the globe. This session served to highlight some of these.

Discussion session II: Identification of possible solutions to problems. In this session the discussion focused on formalising some of the current problems faced in phylogeographic data analysis so that they can be attacked using methods in mathematics, statistics and computer science.

Presentation session IV: Analysing population data. Genealogical networks are an example where the theory for phylogenetic trees has already been successfully generalized to the study of genetic diversity within an individual species. Although various methods have been proposed for constructing genealogical networks, their full potential to the analysis of biodiversity has still to be fully realised, and this presentation session explored some potential applications.

DAY 3

Discussion session III: Planning and follow up research activities and possible EU networking projects. In this final session, various intensive discussions were had concerning the possibility of establishing an EU network for developing new methodologies in phylogeography (e.g. ESF Research Networking Programs, and Framework 7).

In conclusion, all participants agreed that it was a very positive experience, resulting in a greater understanding of what the problems are in phylogeography, and what the potential for addressing these problems is. Talks were followed by constructive and useful discussions, which allowed all those involved to present open questions, understand the central methodological issues in phylogeography, and appreciate the complexities existing within the various disciplines involved.

Looking ahead to the future, it is envisaged that phylogeography will move toward whole genome scans, and there is a need to be thinking ahead to avoid a lag between this kind of data and the methods needed to analyse it. But even before that there is a need to improve existing methodology. There was general agreement that there is much room for fruitful collaboration within the workshop group and beyond, and in the closing discussion the potential for European projects to evolve from this workshop was tabled. All agreed that this would be both useful and beneficial, serving the wider scientific community as a whole.

2. Scientific content of the event

2.1 Scientific Background

Conservation and resource management agencies around the world face hard choices about where and how to conserve biodiversity with limited financial resources. If such choices are to be made in an informed way, there is the need for robust, quantitative methods for measuring, valuing and understanding the structure of biodiversity. The recently established and rapidly burgeoning field of phylogeography provides the framework required for developing such methods. Phylogeography is concerned with the analysis of the geographical distribution of genealogical lineages, with application to understanding the structure of biodiversity. While current tools for quantifying biodiversity typically measure the relative distributions of species within a region, phylogeography seeks to assess the geographical distribution of genetic variation within each species. Given the clear link between genetic variation and the potential for adaptive response, such assessments are critical for the long-term maintenance of individual species and species assemblages.

2.2 The Event

We now describe in detail the content of the workshop.

DAY 1

Presentation of the European Science Foundation (ESF).

A representative of the ESF was unable to make it to the workshop, so Prof. Moulton provided an overview of the ESF, detailing the various organs of the foundation, its overall remit and membership, and the various functions it performs.

Presentation session I: Current problems in phylogeographic assessment of biodiversity.

A current challenge in phylogeography is that the ease of obtaining genetic data has far outpaced the development of the analytical methods necessary to make full use of the data, and this presentation session covered some of the current problems in the area of data analysis.

Brent Emerson discussed the potential advantages of being able to identify ancestral haplotypes within an intraspecific phylogenetic network or phylogenetic tree, pointing out that this may facilitate the identification of ancestral areas from which a species has expanded its range out. Ideally outgroups can be used to identify the root location of a network, although the more distant the outgroup the more unlikely it is to identify the true root location. Thus, in the absence of good outgroup data, a challenge lies in identifying the root location of a network, although coalescent predictions regarding allele age and geography may offer some possibilities.

Pierre Taberlet identified some open questions regarding the importance of intraspecific biodiversity: (1) Is there a congruence between intra- and interspecific biodiversity?; (2) Do areas of high endemism, often coinciding with glacial refugia, harbour a greater degree of intraspecific diversity?; (3) Does habitat variation, characterised by environmental parameters, serve as a good surrogate for intraspecific diversity? He then focused attention on the emergence of population level genome scan data and the need for suitable analytical methods to analyse such data in order to answer these questions. He identified a need to analyse AFLP data in a non-trivial way for phylogeographic purposes, questioning how we might build reliable trees and how many lineages we might need to consider.

Jesus Gomez-Zurita brought attention to the complications arising below the species level through the population level processes of convergence, incomplete lineage sorting, recombination, paralogy, lateral gene transfer and hybridisation. Jesus focused on the issue of hybridisation and presented a method of detecting hybrid genomes using genotype networks in conjunction with the spatial information on the distribution of genetic variation. The advantages of this method are: (1) It is a statistical procedure to distinguish structured from randomly distributed homoplasy; (2) It pinpoints individual sequences responsible for incongruence; (3) It is a high resolution approach to identify the parental species of hybridisation events. However, as a caveat it was recognised that the method itself is not entirely unambiguous.

Petr Kotlik presented some worked examples to review a number of phylogeographic alternative methods to nested clade phylogeographic analysis (NCPA) to compare and contrast their relative merits. Using mammal and freshwater fish data for the geographic region of the black sea Peter concluded that both Bayesian and likelihood methods allow for the estimation of useful parameters without explicit reference to a “correct” tree topology, but that this is computationally demanding. Advantages of the gene-tree/population-tree approach is that multiple populations can be handled and that computation time is fast. Further to this likelihood methods were touted as probably being superior, but that what is needed is more complex but realistic models that incorporate geographic complexity.

Discussion session I: Identification of themes and problems.

The bringing together of both phylogeographers and computational biologists at this workshop highlighted that there are differences in language between the two groups that can lead to misunderstanding. In this session, it became clear that while in phylogeography the term network has a fairly broad meaning, in computational biology (specifically, phylogenetics) this term refers to a tree with cycles. Thus there are gaps that need to be bridged between the two disciplines with regard to terminology and definition to achieve a clear understanding by both.

The problem was raised about how to deal with uncertainty in data – both in terms of sampling effects and data quality. As we are often dealing with sequences/other data forms with low levels of natural divergence between them, errors accidentally incorporated into data may lead to significant downstream implications for results and interpretation. This point was emphasized by both Pierre Taberlet and Peter Forster, both advocating the need to better controls. Pierre pointed out that studies in his laboratory had critically evaluated error in AFLP data and found that numerous controls are necessary, even if all the work is done in the same lab.

Perhaps one of the fundamental problems in phylogeography is how to relate trees/networks from genetic data with geographic data. At present there are a limited number of approaches, with many being contentious. Additional to this there is the problem of how to root a network when there are no suitably close outgroups for this purpose. Given that coalescent theory makes implicit assumptions about allele age within a network, can this somehow be utilised? Methods for identifying geographic clusters are also currently lacking. It should be possible to estimate an optimal number (from one up) of geographic evolutionary clusters within a data set, but there seems to have been no focus upon this issue. Perhaps something analogous to the software Structure?

Another area of difficulty brought to the attention of the workshop group was exactly when to use networks or phylogenetic trees. Beyond this there seems to be little information on how to evaluate the performance of different network approaches against each other, or for a specific data set. Also, when we have more than one network (say for two unlinked gene regions sampled from the same set of samples), how do we assess to what extent they are either similar or different?

While much of the discussion was focused on the analysis of DNA sequences attention was also drawn to the emergence of studies analysing other genetic markers (e.g. AFLPs, RFLPs, microsatellites) for studies of phylogeography. How can these be best analysed to take the most from the data without going too far? Can we develop appropriate models of evolution for AFLPs that could assist with the construction of networks? Are there some questions that are better addressed with AFLPs, and others better addressed with DNA sequence data?

Presentation session II: Phylogenetic methods for the analysis of biodiversity.

It is becoming increasingly easier to obtain data from different DNA sequence regions from the same organism, and from different organisms with the same geographic distribution. However, to analyze them in concert it is now time to follow the lead taken by the field of molecular phylogenetics that has progressed from single marker studies to more rigorous studies of multiple independent genetic markers. In particular, some specific gaps in our analytical ability include: (1) How can we critically assess the degree to which two or more gene genealogies are concordant with one another? (2) How can we critically assess the degree to which demographic inferences from a gene genealogy/genealogies are concordant with non-DNA sequence genetic data (e.g. microsatellites, AFLPs, SNPs)? (3) How do we bridge the gap between phylogenetic trees and genealogical networks to deal with data sets spanning both temporal frames?

It seems probable that the theoretical and quantitative problems arising by studying questions such as (1)-(3) can be addressed using adaptations of current methods for inferring phylogenetic trees, and this presentation session was an opportunity for computational biologists to put forward various methodologies that may suit this purpose.

Arndt von Haeseler presented an approach for estimating phylogenetic diversity (PD) – a measure of the diversity of a group of taxa. The problem posed is that from a tree with n taxa, one wants to identify k taxa such that the resulting subset W maximised phylogenetic diversity. Typically this approach has been applied above the species level, but we can consider this method as transferable

to intraspecific genetic variation in order to identify areas that may contain the maximum amount of genetic variation within a species.

Stuart Baird talked on individual based methods and phylogeography, as opposed to population based methods. A recurrent problem in species level phylogenetics, and one that similarly affects intraspecific phylogenetic analysis and phylogeography, is when two DNA sequence data sets give conflicting genealogies. A useful tool for quantifying this is the quartet distance – the number of quartets (four names taxa in an unrooted tree) that don't have the same topology in two trees. Christian Storm Pederson presented the tools Qdist and QuartetDist for computing the quartet distance between both fully and partially resolved trees, and suggested this may be extended to the comparison of network structures, with the possibility of identifying local similarities.

Patrick Mardulyn (presenting on behalf of Insa Cassens who could not attend at the last minute because of a family bereavement) tackled the issue of intraspecific network construction techniques, presenting the results of simulations for comparing construction methods. At the heart of phylogeographic inference from DNA sequence data is the assembly of DNA sequence haplotypes into a single tree or network representing the best estimate of the evolutionary history of those sequences. A comparison of four methods (minimum spanning networks, median joining networks, statistical parsimony networks, and the union of most parsimonious trees) reveals that different methods can generate different networks, at least with moderately distant sequences, and that this appears to be due to differing abilities to minimise the number of loops (ambiguities). While it is not clear that there is a single best method, it is recommended that users should compare the results obtained with different methods.

DAY 2

Presentation session III: Phylogeography in practice.

The recent growth of phylogeography was fuelled by the technical accessibility of mitochondrial (mt) DNA sequences in animal species. This growth appears to be exponential, and an ISI search listed some 800 papers referring to the topic in 2003-4. The backbone of the majority of these publications is the reconstruction of gene genealogies (networks) from DNA sequence data that allows the determination of, for example, areas of biodiversity importance, gene flow across the landscape and source populations, and transmission dynamics of pathogens. Today the range of DNA techniques, combined with new analytical phylogeography methods and recent palaeoclimatic /geological studies are providing considerable insight into the distribution and evolution of genetic diversity around the globe, and this session served to highlight some of these.

Patrick Mardulyn has over recent years assembled a growing database of DNA sequence data for the leaf beetle *Gonioctena pallida*, a specialist feeder on willow or hazel leaves. The principle questions of phylogeographic interest for this species are (1) for how long have the populations on different mountain ranges been isolated, and (2) how has the climate history of Eurasia influences the current geographic distribution of genetic diversity for *G. pallida*. Using coalescent models appears to be an important approach for the inference of the ages of most recent common ancestors (mrca). However there are a number of obstacles to doing this, namely the need for more complex models and uncertainty about how much demographic information is included in a genealogy, given the stochastic nature of the evolutionary process.

Walter Salzburger presented a series of phylogeographic analyses to assess ice age mediated effects in old world fishes, and the origin of the superflock of cichlid fishes in Lake Victoria. For both these studies constructing haplotypes networks has been at the core of inferring evolutionary history and origins, and the strength of conclusions would seem to be a function of sampling, both in terms of the number of geographic locations and the number of samples within each location.

Toomas Kivisild shifted the focus to human evolutionary history, a species for which sampling is becoming increasingly comprehensive. His focus was to ask if it is possible and reasonable to distinguish structural units in human populations, and his presentation discussed the relative merits of phylogenetic approaches using linked markers and statistical approaches using independent markers. Peter Forster continued with the theme of human phylogeography detailing the mitochondrial DNA chronology of prehistoric human migrations. This presentation highlighted the integration of different data sets for generating a general picture for the understanding of the phylogeographic history of an organism. In the case of humans one has access to archaeological data, climate data, and fossil DNA.

One important cautionary note, also highlighted by Pierre Taberlet, is that extreme caution is necessary in checking for data errors, as these have the potential to result in incorrect historical inferences. Bengt Oxelman discussed the problem of reticulation in the analyses of plant phylogenies, a problem analogous to the problem of recombination in intraspecific gene genealogies. Essentially what is needed is more efficient ways to generate multiple gene phylogenies, the ability to distinguish among processes causing discordant gene phylogenies, and development of methodology to make consensus trees from multi-trees.

Thomas Schmitt then shifted the emphasis to multi-species phylogeographic analysis, presenting data from his long-term analysis of European alpine butterfly species. This also highlighted the need for the development of methodologies for the analysis of unlinked genetic markers (in this case allozymes) for population analyses, and suitable methods for testing for congruence of pattern across multiple species. Importantly such methods need to be able to accommodate missing data, because species ranges are nearly never fully overlapping.

A common criticism leveled at phylogeographic analyses is that sampling is often not sufficient. But the reality is that often the species we are interested in are either globally or locally rare, or for political or financial reasons, thorough sampling is impossible. This is a topic Christophe Thebaud focused on with his talk entitled “Is a minimalist phylogeographic approach better than nothing?” While it is clear that more is always better, Christophe pointed out quite clearly that the issue is not one of limited sampling. One must accept that the questions we can answer will to a large part be dictated by the sampling of individuals within the species and genes within the genome.

Discussion session II: Identification of possible solutions to problems.

Discussion began with a focus on the question of how one might best infer ancestral geographic areas. There is a substantial body of literature on this with regard to inferring ancestral areas for species level phylogenies, but an absence of literature on the treatment of the topic below the species level. Given a network, geographic localities, and extant individuals, can we infer geographic localities of ancestral individuals?

It was suggested that one could simply apply parsimony analysis, but as is done in cladistic biogeography, while the argument was also made for the need to introduce population level models and statistical tests. For the greatest utility both suggestions would need to be applicable to networks/trees that incorporate information on ambiguity (cycles) and this might be best done through a maximum likelihood or Bayesian approach.

The problem of formulating and testing alternative phylogeographic hypotheses was addressed. Consideration was first given to a network with geographic localities, and whether it might be possible to determine the “optimal” number of geographic “units” in the network. One potential solution put forward was to employ a cluster analysis, analogous to the software Structure, to group samples into geographic units. However the issue was then raised by Stuart Baird as to whether it

would not be more correct to use geography with DNA (or an alternative set of genetic markers) sequence data for the coestimation of a genealogy/network. Typically workers tend to treat these two sets of data independently – construct a network and then map on the geographic coordinates of DNA sequences to make some historical inference. Stuart made a compelling argument for not separating these two data sets.

Arndt von Haseler then put forward the general question of what can or cannot be inferred from genetic data to analyse geographical questions, and there was some agreement that the major determinant of this is the mutation rate and model of the DNA sequence being utilised. Summary statistics on networks were seen as a potential solution to some of the problems identified in discussion session I. It was also felt that splits may be a useful measure of both geographic groupings and genetic groupings. With regard to the comparison of networks constructed from independent genetic loci, this may represent a useful measure of matching between the two networks.

Presentation session IV: Analysing population data.

Genealogical networks are an example where the theory for phylogenetic trees has already been successfully generalized to the study of genetic diversity within an individual species. These networks permit the representation of phenomena such as living ancestors and their multiple descendants in a single diagram. Although various methods have been proposed for constructing genealogical networks, their full potential to the analysis of biodiversity has still to be fully realised, and this presentation session explored some potential application.

Katharina Huber presented an overview of median networks and their application to phylogeography. These networks and their relatives have been successfully applied in human population studies and also in biodiversity studies, but work is needed to make them more biologically realistic. In addition there needs to be an assessment of the relationship between median networks and the networks produced by statistical parsimony (NCPA) and the union of most parsimonious trees (UMP). Can NCPA be used to develop a statistical framework for median networks?

There is increasing interest in the incorporation of heterochronous (sampled from different time points) DNA sequences in phylogeographic analysis, and some inroads have been made to the analysis of DNA sequences from different time points. Miguel Navascues reviewed some of the methods available and presented a critique of these approaches by asking “how ancient are ancient DNA sequences?”

Finally, in this session Oliver Pybus looked at integrating the demographic, spatial and evolutionary dynamics of populations. The underlying premise is that the demographic, spatial and evolutionary dynamics of non-recombining gene sequences are underlined by a single, true, unknown, bifurcating genealogy. From this it is then suggested that hypothesis testing should be carried out using a fully parametric genealogical statistical model, treating haplotype networks, median networks etc. not as an end in themselves, but as model selection and hypothesis generation tools.

DAY 3

Discussion session III: Planning and follow up research activities and possible EU networking projects

In this final session, various intensive discussions were had concerning the possibility of establishing an EU network for developing new methodologies in phylogeography (e.g. ESF Research Networking Programs, and Framework 7.) In addition, Peter Forster suggested the

possibility of establishing an annual workshop for the objective testing of phylogeographic methods.

In general, it was agreed that it would be more appropriate to look into the possibility to apply for a network that would include a large data gathering effort (e.g. Pierre Taberlet suggested the use of AFLP data, that could become important in the future), so that methods could be developed in tandem with new data types. In particular, various participants agreed to look into future possibilities in this direction.

The workshop concluded with an executive summary of the proceedings from the key-note speaker Mike Steel (a New Zealand researcher in the field of phylogenetics). In addition to summarizing some of the main mathematical problems that arose in the conference, he pointed out that many of the problems arising at the interface between phylogeography and computational biology are of the same nature as those arising in any interdisciplinary area (e.g. use of language and differences in training). Mike Steel ended by giving a talk concerning novel algorithms and developments in tools for understanding and assessing biodiversity.

3. Assessment of the results, contribution to the direction of the field and outcome

The discipline of phylogeography is constrained by the limited methodologies that have been implemented for general use. It does appear that there is much potential for new algorithms to be developed, and perhaps even whole new methodological approaches. It was clear from the meeting that there are questions being asked by phylogeographers that do have potential solutions from computational biologists, but the problem is often that these questions are not clearly formulated, or easily transferable to computational biologists. This may in part be due to differences in “language” between these two groups, but equally it may be due to a lack of clearly formulated problems that phylogeographers would like to be able to solve.

All involved in the meeting agreed that it was a very positive experience, resulting in a greater understanding of what the problems are in phylogeography, and what the potential for addressing these problems is. What is needed is a more explicit understanding of what different network construction techniques are really doing, and an evaluation of when each of the different techniques might be more appropriate to use. Additionally, support indices on networks are needed, particularly when networks involve a number of cycles. These all seem to be issues that can potentially be addressed one way or another, and this meeting has provided the stimulus for moving toward these answers. For example, the splits method would seem to have application for assessing both the geographic clustering of haplotypes and assessing the degree of concordance between independent networks.

Looking ahead to the future, it is envisaged that phylogeography will move toward whole genome scans, and there is a need to be thinking ahead to avoid a lag between this kind of data and the methods needed to analyse it. But even before that there is a need to improve existing methodology. There was general agreement that there is much room for fruitful collaboration within the workshop group and beyond, and in the closing discussion the potential for European projects to evolve from this workshop was tabled. All agreed that this would be both useful and beneficial, serving the wider scientific community as a whole.

4. Final programme

PROGRAMME

Tuesday 4 July 2006

- 09:30 – 10:00 **Opening**
- Presentation of the European Science Foundation (ESF), Vincent Moulton (University of East Anglia, UK)**
- Presentation session I: Current problems in phylogeographic assessment of biodiversity**
- 10:00 -10:20 **Brent Emerson (University of East Anglia, UK).** Inferring ancestral haplotypes and ancestral areas.
- 10:20 -10:40 **Pierre Taberlet (CNRS, France).** Ne data: new challenges in comparative phylogeography.
- 10:40 – 11:20 *Coffee break and discussions*
- 11:20 – 11:40 **Jesus Gomez-Zurita (Zoologische Staatssammlung München, Germany).** Statistical analysis of mosaic genomes: detection of hybrids using genotype networks.
- 11:40 – 12:00 **Petr Kotlik (Academy of the Sciences of the Czech Republic).** A statistical phylogeographical approach to inferring glacial refugia.
- 12:00 – 13:30 *Lunch*
- 13:30 – 15:00 **Discussion session I: Identification of themes and problems**
- 15:00 -15:30 *Coffee*
- Presentation session II: Phylogenetic methods for analysis of biodiversity**
- 15:30 -15:50 **Arndt Von Haeseler (Center for Integrative Bioinformatics, Austria).** Efficiently computing phylogenetic diversity.
- 15:50 -16:10 **Stuart Baird (University of Montpellier, France).** Individual based methods and phylogeography.
- 16:10 -16:30 **James McInerney (National University of Ireland, Ireland)** A tree-like phylogeny only exists at the tips in the prokaryotes.
- 16:30 -16:50 **Christian Storm Pedersen (University of Aarhus, Denmark).** Tools for quartet based comparison of trees.
- 16:50 -17:10 **Insa Cassens (Universite Libre de Bruxelles, Belgium).** Intraspecific networks: a comparison of construction methods using simulated and empirical data. *Due to a death in the family of Insa Cassens shortly before the workshop, this presentation was given by Patrick Mardulyn.*

Wednesday 5 July 2006

- Presentation session II: Phylogeography in practice**
- 09:00 – 09:20 **Patrick Mardulyn (Universite Libre de Bruxelles, Belgium).** Phylogeography of *Gonioctena pallida*.

09:20 – 09:40	Walter Salzburger (University of Konstanz, Germany). Phylogeography and ice-age mediated effects in Old World fishes.
09:40 – 10:00	Toomas Kivisild (Tartu University and Estonian Biocentre, Estonia). Revealing population structures from genetic data in humans.
10:00 – 10:40	<i>Coffee break</i> and discussions
10:40 – 11:00	Peter Forster (University of Cambridge, UK). The mitochondrial DNA chronology of prehistoric human migrations.
11:00 – 11:20	Bengt Oxelman (Uppsala University, Sweden). The hierarchical model and reticulate plant phylogenies.
11:20 – 11:40	Thomas Schmitt (University of Trier, Germany). Molecular biogeography of alpine and continental butterfly species.
11:40 – 12:00	Christophe Thebaud (University of Toulouse, France). Is a minimalist phylogeographic approach better than nothing?
12:00 – 13:30	<i>Lunch</i>
13:30 – 15:00	Discussion session II: Identification of possible solutions to problems
15:00 -15:30	<i>Coffee</i>
	Presentation session III: Analysing population data
15:30 -15:50	Kathi Huber (University of East Anglia, Norwich, UK). Median networks and phylogeography.
15:50 -16:10	Miguel Navascues (Centre of Forest Research (CIFOR-INIA), Spain). How ancient are ancient DNA sequences?
16:10 -16:30	Oliver Pybus (University of Oxford, UK). Integrating the demographic, spatial and evolutionary dynamics of populations.

Thursday 6 July 2006

09:00 – 10:15	Discussion session III: Planning of follow up research activities and possible EU networking projects
10:15 – 10:30	Concluding remarks on discussion sessions: Mike Steel (University of Canterbury, Christchurch, New Zealand)
10:30 – 11:00	<i>Coffee Break</i>
11:00 – 11:45	Keynote speaker: Mike Steel (University of Canterbury, Christchurch, New Zealand). Phylogenetic diversity: from combinatorics to conservation.
11:45 – 12:00	Closing remarks
12:00 – 13:00	<i>Lunch</i>
Afternoon	<i>Departure</i>

5. List of participants

Stuart Baird (University of Montpellier, France)
Brent Emerson (University of East Anglia, UK)
Peter Forster (University of Cambridge, UK)
Jesus Gomez-Zurita (Zoologische Staatssammlung München, Germany)
Toomas Kivisild (Tartu University and Estonian Biocentre, Estonia)
Katharina Huber (University of East Anglia, UK)
Petr Kotlik (Academy of the Sciences of the Czech Republic, Czech Republic)
Patrick Mardulyn (Universite Libre de Bruxelles, Belgium)
James McInerney (National University of Ireland, Ireland)
Vincent Moulton (University of East Anglia, UK)
Miguel Navascues (Centre of Forest Research, Spain)
Bengt Oxelman (Uppsala University, Sweden)
Oliver Pybus (University of Oxford, UK)
Walter Salzburger (University of Konstanz, Germany)
Thomas Schmitt (University of Trier, Germany)
Mike Steel (University of Canterbury, New Zealand)
Christian Storm Pedersen (University of Aarhus, Denmark)
Pierre Taberlet (CNRS, France)
Christophe Thebaud (University of Toulouse, France)
Arndt Von Haeseler (Center for Integrative Bioinformatics, Austria)

An additional invitee, *Insa Cassens* (Universite Libre de Bruxelles, Belgium), was unable to attend because of a family bereavement just before the workshop, but a colleague, *Patrick Mardulyn*, was able to make her presentation on her behalf.

6. Statistical information on participants

Overall, 20 scientists attended the workshop. One invitee, Insa Cassens from Brussels was unable to attend at the last minute because of a death in her family. Institutes and Universities from 11 European countries were represented (United Kingdom, Ireland, Denmark, Czech Republic, Spain, France, Austria, Germany, Estonia, Belgium, Sweden). Approximately half of the participants were computational biologists, and half were phylogeographers. One of the participants was female (there were originally two, but see above). The youngest participants were in their mid to late twenties, and the oldest in their late forties to early fifties.