European Science Foundation Report
ESF Exploratory Workshop

Towards the Tree of Life....2004
June 8-10, 2004
Bergen, Norway

convened by David A. Liberles, Computational Biology Unit, BCCS, University of Bergen, 5020 Bergen, Norway

http://www.cbu.uib.no/etol/

Executive Summary

The exploratory workshop involved a number of researchers from across Europe who have an active interest in coalescing the tree of life. In addition, an American researcher active in the American effort to piece together the tree of life was also invited. Previous self-financed efforts to organize a European effort complementary to the American effort had taken place in Patras (Greece), Paris, London, and Brussels.

The meeting in Bergen was more focused on generating an assessment of ongoing scientific research in Europe, where researchers could collaborate on a tree of life effort. The first day was largely devoted to computational methods, while the second day was more devoted to sub-trees of life that were of particular interest to various European groups. Both days plus a third morning provided significant time for discussion about both computational and experimental solutions.

In addition to 21 invited speakers (3 local to University of Bergen), the meeting was opened to an additional 27 attendees. These included junior members of the laboratories of invited speakers, local Norwegian researchers, and others from Europe who applied to attend online. In addition to funding from ESF, FUGE, the Norwegian functional genomics research platform, also contributed money to make the workshop available to some interested Norwegian researchers.

The workshop was held on the University of Bergen campus, where the auditorium was donated for free. A poster session was held outside the lecture hall for non-invited attendees to present their work. A conference dinner was held after the second day, with a bus trip out to the tip of an off-coast island for a dinner of local species (e.g. halibut, cloudberries) coupled to a visit to a neighboring museum including an exhibit on local fauna and flora.

Scientifically, several important issues were raised. To begin with, several alternative algorithmic approaches were presented. One question that arose concerned the use of supertrees versus supermatrices versus other approaches as the best way forward. Further, especially in prokaryotes, it is unclear if a tree of life is the best representation, or if a network of life is more appropriate. The classic question of the proper use of morphological data and its role versus molecular data in this project emerged in several talks.

Ultimately, several speakers called for financial realism in the aims. A

general point that arose was the comparative lack of funding for the European effort compared with the American effort. Several strategies for ways forward were presented.

Overall, a productive scientific meeting was held. This presented both an overview of ongoing European tree of life research and a discussion towards a collaborative effort to build a tree of life.

Scientific Summary of Bioinformatics Discussion (written by James McInerney, National University of Ireland)
Database

This is the topic that occupied much of the discussion. We felt that we had identified a number of issues that might already have been addressed by other researchers and a number of desirable properties of an ideal database or set of databases.

The data that an appropriate database would contain will range from DNA and protein sequence data, Accession numbers of various kinds, photographs, DNA barcode data, general taxonomy data, pointers to locations of frozen tissues or DNA. These issues were thought to have been pretty much solved at this stage. The outstanding issues concerned whether or not there is a need for a new structure or datatype that would organize the information within this database in the format of a phylogenetic tree. How would we retrieve the information and what kind of information would we wish to retrieve from it? Could we retrieve data from different studies if we were just interested in some of the data and not all of it? It was generally decided that this was an issue that might warrant a meeting between researchers that had already developed relational databases for sequence data.

Visualization

It was generally agreed that WALRUS was an ideal platform for visualization of large trees. It offers a number of features that were considered desirable and in addition, there is the strong likelihood that it is going to become public domain software in the near future. Therefore, with further development, it should be possible to make visualization of large trees possible. It was further decided that a preliminary attempt to construct a phylogenetic tree using GenBank or a convenient subset of GenBank (or some comparable database) in the near future. Emmanuel Douzery to be contacted regarding the feasibility of such a study.

Computation

The amount of computation required for this project is undoubtedly going to be significant. Even with the development of excellent new algorithms (see next section), there will still be a requirement for a computing infrastructure that exceeds anything available in any single laboratory at the moment. Therefore, it has been suggested that we would explore the computing requirements and possible solutions to these requirements. The construction of a number of Virtual Organisations (VOs) using the GRID computing paradigm is going to be the first avenue to be explored. Alternatives include a special dedicated cluster system, although how this would be funded, maintained and used was not discussed.

Algorithms

This section, unsurprisingly, produced the greatest number of ideas and suggestions and required the shortest portion of the discussion. A number of subjects

were identified and a number of individuals expressed interest in becoming involved in these areas. One of the aspects that was considered to have a potential impact on how these methods would be developed was the issue of the kind of data that would come through from the experimental scientists and how it could be used. If the decision of the AToL community was to focus on the sequencing of a small number of universal genes, then there might be a different strategy for algorithm development than if there were group-specific genes being sequenced, etc. However, the major topics for discussion were super-tree and super-matrix methods, fast heuristics, parallel computing, super-network methods and methods for comparing trees. There were no shortage of ideas and opinions and it was generally felt that this would be a very productive area of the project.

Scientific abstracts from the meeting
Vincent Moulton (Uppsala University, Sweden)
Towards the network of life

We are now gaining access to data that is enabling us to make deep insights into how genomes are structured and evolved. However, the transition from gene analysis to genome analysis is presenting many challengeswithin phylogenetics. For example, evolutionary processes such as recombination, lateral gene transfer, and hybridization are all important in shaping the structure of genomes, but they do not always lend themselves well to traditional tree-based evolutionary analysis. In this talk we will present a brief overview of using networks as an alternative to evolutionary trees. As we shall see, this topic not only opens up many fascinating new directions for research, but has profound consequences for reconstructing the tree-of-life.

Emmanuel Douzery (University of Montpellier II, France)
Supertrees,supermatrices of characters, and impact of missing data on molecular phylogeny and divergence times

The recent increase in the number of DNA, RNA, and protein sequences available for a variety of organisms helps to understand complex questions of evolutionary biology in a phylogenetic and temporal framework. Usually, phylogeny reconstruction involves several genes for numerous species that might be analyzed in separate or combined ways. However, because some gene sequences are only available for subsamples of species, an important amount of missing data appears in data matrices. Two different approaches have been developed to circumvent this difficulty. (i) The "supertree" approach follows a combinatorial line; a phylogeny is reconstructed for each gene, and the resulting source trees are combined into a super-phylogeny. (ii) The "supermatrix of characters" approach involves analyzing all concatenated sequences, including missing data.

To illustrate these points, I will present a new supertree-building algorithm, based on the search of a super-dissimilarity distance matrix. I will compare it with the maximum likelihood (ML) analysis of the corresponding supermatrix of characters. Tree-building methods associated to a probabilistic principle (ML or Bayesian) are time consuming, but genetics and very fast ML algorithms (MetaPIGA, PHYML) now lowered computing times by one order of magnitude and allows the analysis of hundred of species with high topological accuracy. Biological examples will include mammals and eukaryotes. To conclude, I will evaluate the impact of missing data on topological accuracy. Because a growing number of phylogenies is used to infer molecular timelines for a variety of taxa, I will also discuss the influence

of missing entries on the accuracy and precision of divergence times estimated in a Bayesian relaxed clock framework.

This work is conducted on our Phyloinformatics platform in Montpellier, and actively involves the following people: Pascale Chevret, Frédéric Delsuc, Claudine Montgelard, and Emmanuel Paradis (ISE-M) ; Vincent Berry, Olivier Gascuel, Nicolas Lartillot (LIRM-Montpellier), Alexis Criscuolo (ISE-M / LIRM-M); François Chevenet (IRD); Nicolas Galtier (GPIA); and Hervé Philippe (University of Montreal, CA).

Michel Milinkovitch (Free University of Brussels, Belgium)
metaPIGA 2.0

MetaPIGA 2.0 is the new version of the metapopulation genetic algorithm for inference of very large phylogenetic trees (hundreds to thousands of taxa). This new version implements, beside many improvements such as GTR and protein models, multiple new stochastic approaches that are combined with consensus pruning.

David Liberles (University of Bergen, Norway)
The tree of life as a valuable tool for understanding gene function

The tree of life is a valuable tool in functional genomics. Various genomic processes can be studied in a phylogenetic context. These include gene content, gene sequence evolution, mRNA splice site usage and evolution, and the evolution of gene expression. Starting with the evolution of individual genes through duplication, speciation, mutation, drift, and selection, we have begun to catalog the evolution of genomes in a phylogenetic context in a resource called The Adaptive Evolution Database (TAED). Information from TAED can be used to understand both the global picture of genomic evolution and individual gene families in more detail.

Paul-Michael Agapow (University College London, UK)
Visualizing Large Phylogenies: A Curly Problem

Although traditional methods of visualizing phylogenetic trees are been adequate for current demands (i.e. 100s of nodes), the Tree of Life is a qualitatively different problem. Even the smallest fragment of the Tree (0.1% or tens of thousands of species) is too large to be comprehendible in current methods. This is not just a problem of aesthetics. Researchers need to be able to see the trees they are working on - if only for a quick visual confirmation of the state of tree - and to be able to edit the tree. However the size of the large phylogenies make it difficult to focus on features of interest. Here I present a portable software library for visualizing phylogenies that uses hyperbolic representation and selective emphasis of features to ease the task of representing phylogenies in acceptable detail at acceptable computational cost.

Matt Phillips (Oxford University, UK)
Integrating the tree of life with the tree of past life: a few concerns

By comparing trees inferred from standard (ACGT) and RY-coded mitochondrial protein-coding sequences it can be shown that even complex maximum-likelihood models can be poor branch-length estimators. This in turn

effects divergence estimates from molecular dating. Incorporating fossil calibration points from both shallow and deep nodes in trees increases the consistency of these divergence estimates, though depends on reliable fossil phylogenies from non-molecular (usually morphological/anatomical) data. I will briefly discuss some of the problems facing morphological phylogenetics and the implications of these problems for integrating the tree of life with the tree of past life.

Ziheng Yang (University College London, UK)
Maximum likelihood methods for estimating species divergence times

I will discuss recent developments in the use of maximum likelihood methodology to estimate species divergence times. My emphasis is on combining heterogeneous data sets and using multiple calibration points in an intergrated analysis.

James McInerney (National University of Ireland)
Supermatrices versus supertrees: some empirical observations

Phylogenies derived from concatenated multiple alignments are much more common than phylogenies derived using a supertree approach. The primary reason for this is the widespread belief that by concatenating datasets together any confounding signal from an individual gene will be overwhelmed by the one 'true' signal. A recent publication in Nature that employed this approach seemed to provide strong justification for the approach. Here I show that a supertree or consensus approach performs just as well as this supermatrix method and furthermore I show that in a situation in the bacterial genus Neisseria, where exchange of genes between different strains is very common, the supermatrix approach is vastly inferior, positively misleading and in fact, it behaves rather oddly.

Jens Lagergren (Royal Institute of Technology, Sweden)
Probabilistic and combinatorial analysis of gene families w.r.t. gene duplications and lateral gene transfers

Comparative genomics in general and orthology analysis in particular are becoming increasingly important parts of gene function prediction. Previously, orhtology analysis and reconciliation has been performed only with respect to the parsimony model. This discards many plausible solutions and sometimes precludes finding the correct one.

In many areas in bioinformatics probabilistic models have proven to be both more realistic and powerful than parsimony models. We introduce a probabilistic gene evolution model based on a birth-death process in which a gene tree evolves "inside" a species tree. Based on this model, we develop a tool with the capacity to perform practical orthology analysis, based on Fitch's original definition, and more generally for reconciling pairs of gene and species trees w.r.t duplications and losses.

We develop a Bayesian analysis based on MCMC which facilitates approximation of an a posteriori distribution for reconciliations. That is, we can find the most probable reconciliations and estimate the probability of any reconciliation, given the observed gene tree. This also gives a way to estimate the probability that a pair of genes are orthologs. The algorithm performs very well on synthetic as well as biological data. Using standard correspondences, our results carry over to allele trees as well as biogeography.

When also lateral transfers are considered reconciliation is much harder.

We give a combinatorial model and parsimony algorithms for gene duplications and lateral gene transfers. These algorithms detect lateral gene transfers with very low error rates.

Petter Bjørstad (University of Bergen, Norway)
Computational Resources for The Tree of Life

This talk will briefly review the state of art computer resources that are available for BioInformatics research, in particular, for the Tree of Life project. We focus on inexpensive, high latency cluster computing and some of the issues and challenges when extending this to a grid computing environment.

Michael Whiting (Brigham Young University, Utah, USA)
Distributed computing and phylogenetic reconstruction: Searching the insect tree of life

Recent advances in DNA sequencing technology have created large data sets upon which phylogenetic inference can be performed. However, current research is limited by the prohibitive time necessary to perform tree search on even a reasonably sized data set. This difficulty is particularly apparent in insect phylogeny, where a vast amount of data is required to accurately represent insect diversity. This presentation will focus on our ongoing efforts to reconstruct a tree of life for insects based on DNA sequence data, and describe logistical challenges in obtaining the pertinent data. I will also describe the DOGMA parallel processing system which allows for efficient distributed computing using existing, trusted phylogenetic software across a variety of computational platforms.

Olaf Bininda-Emonds (Technical University of Munich, Germany)
The rebirth of supertree construction in the genomic age

In combining source trees rather than the primary character data underlying those trees, supertree construction currently represents a tradeoff between an inherent loss of information and the ability to combine all available phylogenetic hypotheses to achieve the most complete phylogeny possible. To date, the lack of sufficient compatible character data for most groups have justified this tradeoff. By combining existing phylogenetic hypotheses from the literature, supertree construction has been able to provide complete phylogenetic estimates of even very large clades that have yet to be approached using conventional phylogenetic techniques. However, the ever-growing wealth of sequence data means that comparable taxonomic coverage for many groups will soon be achievable using supermatrix approaches. But rather than fade away, I argue that supertrees will continue to play an essential, but different role in phyloinformatics for the efficient analysis of very large sequence matrices as part of a divide-and-conquer strategy. Under such a framework, the single large supermatrix is broken down into many smaller and therefore computationally simpler subproblems for analysis. The global answer is then achieved by using supertree construction to combine the results of the subproblems. Thus, it will only be through the compatible strengths of the supertree and supermatrix approaches that a successful attempt to reconstruct the Tree of Life can be made.

Victor Albert (University of Oslo, Norway)

The Floral Genome Project and the genetic architecture of floral diversification

The Tree of Life is largely meant to address phylogenetic patterns. However, the time is also ripe to use emerging phylogenetic knowledge to begin to understand the processes that have led to these patterns. Among the evolutionary mechanisms of interest to the plant community are the developmental genetics of flower evolution. The Floral Genome Project (FGP; see www.floralgenome.org) was funded by a $7.4 million grant from the United States National Science Foundation to bridge the evolutionary gap between the most broadly studied plant model systems. *Arabidopis* and rice, although now completely sequenced and under intensive comparative genomic investigation, last shared a common ancestor ca. 125 million years ago and cannot in isolation provide a comprehensive perspective on structural and functional aspects of flowering plant genome dynamics. We discuss new genomic resources comprising cDNA libraries and EST sequences for a suite of phylogenetically basal angiosperms specifically selected to interconnect the evolutionarily divergent model plants. Initial comparisons illustrate the utility of the EST data sets toward discovery of the basic floral transcriptome. Progress has included a completed target-sized EST library (ca. 10,000 ESTs, 6000 estimated genes; see www.pgn.cornell.edu) for the basal eudicot California poppy (*Eschscholzia*), a nearly complete EST library for *Amborella* (the most basal angiosperm), and an overview of florally expressed gene families sampled among poppy, *Amborella*, avocado (*Persea*), tulip tree (*Liriodendron*), water lily (*Nuphar*), and *Welwitschia* (a gymnosperm). The overall picture that is emerging is that both gymnosperms and basal angiosperms had the basic 'tool kit' for flower development, as viewed from the perspective of more apomorphic plants like *Arabidopsis*. These first findings and the genetic resources available will afford opportunities to address conspicuous evolutionary genomic issues, including genome-wide duplication history, duplication and functional divergence of gene lineages, and adaptive molecular evolution.

Recent reviews from the FGP:

Soltis, D. E., Soltis, P. S., Albert, V. A., Oppenheimer, D. G., dePamphilis, C.W., Ma, H., Frohlich, M. W. and Theissen, G. (2002) Missing links: the genetic architecture of flower and floral diversification. *Trends in Plant Science* 7: 22-31.

Albert, V.A., Oppenheimer, D.G. and Lindqvist, C. (2002) Pleiotropy, redundancy, and the evolution of flowers. *Trends in Plant Science* 7: 297-300.

Soltis, D. E. and Soltis, P. S. (2004) The role of phylogenetics in comparative genetics. *Plant Physiology* 132: 1790-1800.

Buzgo, M., Soltis, D. E., Soltis, P. S., and Ma, H. (2004) The role of developmental morphology in molecular developmental genetics. *Trends in Plant Science* 9: 164-173.

Vincent Savolainen (Royal Botanic Gardens, Kew, UK)
Towards building a complete generic-level phylogenetic tree for flowering plants

As part of the international efforts on reconstructing the Tree of Life, botanists have aimed at obtaining in the very near future comprehensive phylogenetic hypotheses of relationships for all ca. 13,000 genera of angiosperms. Here we present

simulation studies that show that an accurate tree could be reconstructed for such a large number of taxa using limited sequencing effort. We also evaluate empirically how rbcL and other genes archived in EBI/GenBank could be used to infer the generic angiosperm tree and finally we explore how 3D hyperbolic space could be used to represent such a large tree.

Bernd Schierwater (Hannover School of Veterinary Medicine, Germany)
In the beginning there were diploblasts

For more than a century the question of the ancestral metazoan bauplan, the "urmetazoan", has been subject of controversial hypotheses/speculations, and there has been little hope to ever resolve the issue. Recent advances in molecular systematics and EvoDevo research have created new hope to test some of the hypotheses. We have been (i) sequencing mitochondrial and nuclear genes, (ii) predicting secondary structures of ribosomal RNA morphological, (iii) characterized the structure, genomic organization and function of Hox genes, and (iv) analysed the sum of data in supertrees and supermatrices. The results are rewarding.

Daniel Chourrout (University of Bergen, Norway)
Derived and ancestral features of the Oikopleura (Urochordata) genome
Urochordates (Tunicates) occupy a basal position in the phylogenetic tree of chordates. Their body plan is far less complex than that of vertebrates. Among Urochordates, the free-swimming larvaceans represent the most basal class and they keep the chordate tail complex during the entire life. The larvacean *Oikopleura* has a particularly small cell number, a very short generation time, and the smallest genome size identified in the animal kingdom. Urochordates are often considered as the best approximation of the chordate ancestor.

We have examined several features of the newly sequenced *Oikopleura* genome, including the retrotransposon content, the intron-exon organisation and the *Hox* gene complement. Our observations concur to propose that, even though urochordates were the earliest group to diverge from other chordates, they have also evolved faster than them. Their simple anatomy is likely to be the result of a drastic simplification of ancient chordates.

Jaume Baguñà (University of Barcelona, Spain)
A multigenic approach further supports acoelomorph plathelminthes as the earliest basal bilaterians

Sequences from the 18S rDNA, 28S rDNA, myosin II, tropomyosin and mitochondrial protein genes have been used to further test the placement of the Acoelomorpha (Acoela+Nemertodermatida) plathelminthes as the earliest extant basal bilaterian taxa. Under different evolutionary assumptions and using different statistical tests and methods of phylogenetic inference, acoelomorphs appear to be the earliest branching Bilateria. This is further supported from analyses of the number and types of Hox and ParaHox genes from both acoels and nemertodermatids which show they bear an intermediate (Anterior-group 3-Central-Posterior) number of Hox genes, two ParaHox genes (*Xlox*-like and *Cdx*-like), and no molecular signature specific of Hox genes from the three big bilaterian superclades (Deuterostomia, Lophotrochozoa and Ecdysozoa). This positioning may have far-reaching implications for understanding the evolution of body axes, life-cycles, and the Hox/ParaHox clusters,

for the origin of mesoderm, and for our perception of the so-called Cambrian explosion. To that aim EST collections and BAC libraries are currently being developed from selected acoels and nemertodermatids.

Rafael Zardoya (National Museum of Natural History, Spain)
Molecular systematics of Mollusks

Molluscs are a large and diverse group of soft-bodied unsegmented animals that includes more than 100,000 extant and 35,000 fossil species. Most molluscs are marine organisms, but some (notably Pulmonates) have radiated on land and freshwater habitats. There are eight main classes currently recognized within Mollusca: Solenogastres, Caudofoveata, Polyplacophora, Monoplacophora, Cephalopoda, Scaphopoda, Bivalvia, and Gastropoda. These groups comprise structurally diversified organisms with very distinct body plans. Nevertheless, all share several morphological features i.e. the mantle, the ventral foot, the shell, and the radula. A wealth of information on alpha-taxonomy of molluscs is available, but the systematics of the group is still poorly understood. Within the European ATOL framework, we propose to establish a robust phylogenetic hypothesis of Molluscs relationships using a genomic approach. Representatives of the eight main groups of molluscs will be analyzed with a particular emphasis (and more thorough coverage) on gastropods, bivalves, and cephalopods. Two kinds of molecular markers will be used to determine the backbone of the mollusc tree: complete mitochondrial genomes will be sequenced for a representative of each of the eight main groups of molluscs, as well as for a representative of each of the main lineages of gastropods, bivalves, and cephalopods. Nuclear markers (mainly ESTs-Expression Sequence Tags) derived from cDNA libraries of a representative of each of the eigth main groups of molluscs will be determined anew.

Alfried Vogler (Imperial College, UK)
The use of Genbank data for building the Tree-of-Life

The increasing taxonomic representation of public databases provides a great resource for compiling comprehensively sampled Trees. A reiterative procedure is proposed for data extraction, tree building and inclusion of multiple gene datasets and periodical update of trees. The Blast algorithm can be used to for the initial placement of newly added taxa, providing starting trees for more extensive searches. A preliminary demonstration of the approach is provided using large data sets for Coleoptera (beetles).

Tim Barraclough (Imperial College, UK)
Interpreting the Tree of Life: dynamics of lineage branching

The Tree of Life is a potent resource for uncovering evolutionary processes behind diversification. An array of statistical methods are available to study speciation and extinction from patterns of lineage branching, but these typically make untested and stringent assumptions about the accuracy of the timing of branching events and the completeness of sampling at the species level. The talk will discuss the application of these methods to 'Tree of Life'-type data, i.e. high-throughput approaches, supertrees and haphazard sampling. In particular, it will focus on the species-boundary and future applications based on individual-based sampling schemes. Finally, it will outline some general issues for the utility of current developments in systematics for studies of diversification.

Assessment and Future Work

As previous attempts to obtain funding from EU funding sources have not succeeded, several strategies were discussed for a way forward. An ongoing discussion has been established with CETAF (http://www.cetaf.org) to collaborate on a tree of life/biodiversity proposal. The tree of life organization would take a key role in work packages in phylogenetics and bioinformatics, two areas of relative strength. A tree of life Europe website (http://www.cbu.uib.no/atol) was also established to collect information and facilitate collaboration.

Scientifically as a way forward, mining Genbank and generating an automated pipeline to build the tree of life from known gene sequences was seen as a starting point. Alfried Vogler's group has done this with beetles. Groups in Bergen and Lyon have gene family databases that can be mined for this type of purpose. Emmanuel Douzery's group was given the coordination lead on this project. A tentative decision to coordinate any progress towards this goal involved a meeting of the bioinformatics groups in conjunction with the Marseille Evolutionary Biology meeting in September.

In assessing significant European contributions to an international effort, one goal was to compare Genbank sequence data with biodiversity data to identify relatively virgin areas of research that Europe can contribute to with little competition, given its reduced funding level. However, one worry with this is that such taxa may not be areas of scientific strength in Europe and taxa available to research may not be so flexible (researchers in a botanical garden can't start systematically sequencing archaea).

At the level of standards, some discussion included a suggestion for journals to require that all published trees are deposited into treebase. The desire for Genbank to impose a uniform standard for gene names was seen as desirable. Further guidelines for DNA banking and data curation were recommended. However, ultimately these decisions require coordination with others, including non-European scientists.

xxxbioinformatics efforts and CETAF coordination

Program

Tuesday, June 8

8:45-9:00 **Meeting Introduction** David Liberles (University of Bergen, Norway) and Joerg Ott, (Standing Committee for Life and Environmental Sciences, European Science Foundation (ESF))

Phylogeny Reconstruction

9:00- 9:20 Vincent Moulton (Uppsala University, Sweden) *Towards the network of life*

9:25-9:45 Emmanuel Douzery (University of Montpellier II, France) *Supertrees,supermatrices of characters, and impact of missing data on molecular phylogeny and divergence times*

9:50-10:10 Michel Milinkovitch (Free University of Brussels, Belgium) *metaPIGA 2.0*

10:15-10:35 **break**

Analyzing Phylogeny

10:35-10:55 David Liberles (University of Bergen, Norway) *The tree of life as a valuable tool for understanding gene function*

11:00-11:20 Paul-Michael Agapow (University College London, UK) *Visualizing Large Phylogenies: A Curly Problem*

11:25-11:45 Matt Phillips (Oxford University, UK) *Integrating the tree of life with the tree of past life: a few concerns*

11:50-12:10 Ziheng Yang (University College London, UK) *Maximum likelihood methods for estimating species divergence times*

Lunch 12:15-13:30

13:30-13:50 James McInerney (National University of Ireland) *Supermatrices versus supertrees: some empirical observations*

13:55-14:15 Jens Lagergren (Royal Institute of Technology, Sweden) *Probabilistic and combinatorial analysis of gene families w.r.t. gene duplications and lateral gene transfers*

Computational Resources for The Tree of Life

14:20-14:40 Petter Bjørstad (University of Bergen, Norway) *Computational Resources for The Tree of Life*

14:45-15:15 **break**

15:15-17:15 ESF roundtable open discussion after computational presentations on visions for actualizing a computational resource for the tree of life

Posters 17:15-19:00

Wednesday, June 9

Subtrees of Life

9:00- 9:20 Michael Whiting (Brigham Young University, Utah, USA) *Distributed computing and phylogenetic reconstruction: Searching the insect tree of life*

9:25- 9:45 Mark Chase (Royal Botanic Gardens, Kew, UK) *Overview of the Tree of Life-EU Process*

9:50-10:10 Olaf Bininda-Emonds (Technical University of Munich, Germany) *The rebirth of supertree construction in the genomic age*

10:15-10:35 Victor Albert (University of Oslo, Norway) *The Floral Genome Project and the genetic architecture of floral diversification*

10:40-11:00 **break**

11:00-11:20 Vincent Savolainen (Royal Botanic Gardens, Kew, UK) *Towards building a complete generic-level phylogenetic tree for flowering plants*

11:25-11:45 Bernd Schierwater (Hannover School of Veterinary Medicine, Germany) *In the beginning there were diploblasts*

11:50-12:10 Daniel Chourrout (University of Bergen, Norway) *Derived and ancestral features of the Oikopleura (Urochordata) genome*

12:15-12:35 Jaume Baguña (University of Barcelona, Spain) *A multigenic approach further supports acoelomorph plathelminthes as the earliest basal bilaterians*

Lunch 12:40-14:00

14:10-14:30 Rafael Zardoya (National Museum of Natural History, Spain) *Molecular systematics of Mollusks*

14:35-14:55 Alfried Vogler (Imperial College, UK) *The use of Genbank data for building the Tree-of-Life*

15:00-15:20 Tim Barraclough (Imperial College, UK) *Interpreting the Tree of Life: dynamics of lineage branching*

15:25-15:55 **break**

Working Group Discussions

16:00-18:00 Experimental Approaches, including sequencing, led by Alfried Vogler (Imperial College, UK);

16:00-18:00 Bioinformatics, led by James McInerney (National University of Ireland);

Conference Dinner at Herdla Museum, leaving at 18:30 by bus. The return bus left Herdla at 22:30. Time to explore the museum or the area after dinner.

Thursday, June 10

Discussions and Progress

9:00-12:00 discussions led by Tim Barraclough (Imperial College, UK), Mark Chase (Royal Botanic Gardens, Kew, UK), Vincent Savolainen (Royal Botanic Gardens, Kew, UK), and Alfried Vogler (Imperial College, UK); **break** at 10:40;

12:00-12:30 closing remarks by David Liberles and Alfried Vogler

Lunch 12:35-13:45

List of Participants
1. Paul-Michael Agapow (UK); p.agapow@ucl.ac.uk
University College London
Department of Biology
Darwin Building, Gower Street, London WC1E 6BT
UK

2. Victor Albert (Norway); victor.albert@nhm.uio.no
Natural History Museums & Botanical Garden
University of Oslo
Box 1172 Blindern, 0318 Oslo
Norway

3. Heidi Lie Andersen (Norway); heidi.andersen@bio.uib.no
UiB, Department of Biology
Allégaten 41
5007 Bergen
Norway

4. Nadia Anwar (UK); n.anwar@udcf.gla.ac.uk
University of Glasgow
314C Graham Kerr Building
G128QQ
UK

5. Himanshu Ardawatia (Norway); himanshu@ii.uib.no
Computational Biology Unit, BCCS
University of Bergen
5020 Bergen
Norway

6. Miquel Arnedo (Spain); marnedo@ub.edu
Dept. Biologia Animal, Universitat de Barcelona
Fac. Biologia, Av. Diagonal 645, 08028 Barcelona
Spain

7. Jaume Baguña (Spain); bagunya@bio.ub.es
Departament de Genètica, Facultat de Biologia, Universitat de Barcelona
Diagonal 645, 08028 Barcelona
SPAIN

8. Tim Barraclough (UK); t.barraclough@imperial.ac.uk

Department of Biological Sciences
Imperial College London
Silwood Park Campus
Ascot, Berkshire
UK

9. Matthew Betts (Norway); matthewb@ii.uib.no
Computational Biology Unit
Bergen Centre for Computational Science
University of Bergen
N-5020 Bergen
Norway

10.Olaf Bininda-Emonds (Germany); Olaf.Bininda@tierzucht.tum.de
Lehrstuhl für Tierzucht
Technical University of Munich
Alte Akademie 12
85354 Freising-Weihenstephan
Germany

11.Petter Bjørstad (Norway); petter@ii.uib.no
Department of Informatics
University of Bergen
5020 Bergen
Norway

12.Mark Chase (UK); m.chase@kew.org
Royal Botanic Gardens, Kew
Jodrell Laboratory
Richmond, Surrey
TW9 3DS
United Kingdom

13.Daniel Chourrout (Norway); daniel.chourrout@sars.uib.no
SARS Institute for Marine Molecular Biology
University of Bergen
5020 Bergen
Norway

14.Pascal DalFarra (Belgium); pdalfarra@ibelgique.com
Laboratory of Evolutionary Genetics
Institute of Molecular Biology & Medicine
12 rue Jeener & Brachet
B-6041 Gosselies
Belgium

15.Katie Davis (UK); k.davis@udcf.gla.ac.uk
University of Glasgow
DEEB, Graham Kerr Building,
University of Glasgow, Glasgow, G12 8QQ
UK

16.Emmanuel Douzery (France); douzery@isem.univ-montp2.fr
Institute of Evolutionary Sciences of Montpellier
Molecular Phylogeny Team
University Montpellier II - CC064
Place E. Bataillon ; 34095 Montpellier Cedex 5
France

17.Torbjørn Ekrem (Norway); Torbjorn.Ekrem@zoo.uib.no
Bergen Museum, UiB, Norway
Vitenskapsmuseet, NTNU, Trondheim, Norway
Muséplass 3
5007 Bergen
Norway

18.Carmela Gissi (Italy); carmela.gissi@unimi.it
Dipartimento di Scienze Biomolecolari e Biotecnologie
Universita di Milano
Via Celoria, 26 - 20133 Milano
Italy

19.Heike Hadrys (Germany); heike.hadrys@ecolevol.de
Insitut of Ecology and Evolution
Veterinary School of Hannover
Buenteweg 17d
D-30559 Hannover
Germany

20.Katharina Huber (Sweden); katharina.huber@lcb.uu.se
Department of Biometry and Engineering
The Swedish University of Agricultural Sciences
Box 7013
Uppsala
Sweden

21.Joseph Hughes (UK); j.hughes1@imperial.ac.uk
Imperial College London
Natural History Museum London
Department of BIological Sciences
Silwood Park, Ascot, Berkshire
UK

22.Timothy Hughes (Norway); tim@cbu.uib.no
Computational Biology Unit
University of Bergen
High Technology Center
N-5020 Bergen
Norway

23.Bjarte Jordal (Norway); bjarte.jordal@vm.ntnu.no
Norwgian University of Science and Technology

Vitenskapsmuseet
Erling Skakkesgt 47
7491 Trondheim
Norway

24.Lawrence Kirkendall (Norway); lawrence.kirkendall@bio.uib.no
Univ. Bergen, Dept. Biology
Allegaten 41
5020 Bergen
Norway

25.Jens Lagergren (Sweden); jensl@nada.kth.se
KTH
Albanova
Stockholm Bioinformatics Center
S - 100 44 Stockholm
Sweden

26.David Liberles (Norway); liberles@cbu.uib.no
Computational Biology Unit, BCCS
University of Bergen
5020 Bergen
Norway

27.Christoph Mayer (Germany); cm@tp4.rub.de
Spezielle Zoologie
Ruhr Universitaet Bochum
Ruhr Universitaet Bochum, ND 05 788
44780 Bochum
Germany

28.James McInerney (Ireland); james.o.mcinerney@may.ie
NUI Maynooth
Maynooth
Co. Kildare
Ireland

29.Harald Meier (Germany); meier@informatik.tu-muenchen.de
Angewandte Bioinformatik
Lehrstuhl für Rechnertechnik und Rechnerorganisation
Institut für Informatik
Technische Universität München
Office: 02.13.061, Boltzmannstr.3, 85748 Garching bei München
Germany

30.Michel Milinkovitch (Belgium); mcmilink@ulb.ac.be
Laboratory of Evolutionary Genetics
Institute of Molecular Biology & Medicine
12 rue Jeener & Brachet
B-6041 Gosselies
Belgium

31.Vincent Moulton (Sweden); vincent.moulton@lcb.uu.se
The Linnaeus Center for Bioinformatics
Uppsala University
Box 598
751 24 Uppsala
Sweden

32.Matt Phillips (UK); mattphillips73@hotmail.com
Oxford University
Zoology Department, Oxford University
South Parks Road, Oxford
United Kingdom

33.Carles Ribera (Spain); cribera@ub.edu
Department of Biologia Animal
Universitat de Barcelona
Facultat de Biologia
Av. Diagonal 645
Barcelona - 08028
Spain

34.Marta Riutort (Spain); mriutort@ub.edu
Dpt. de Genètica, Fac. de Biologia
Universitat de Barcelona
Av. Diagonal, 645
08028 Barcelona
Spain

35.Christian Roth (Norway); chregu@cbu.uib.no
CBU
BCCS, UIB
5020 Bergen
Norway

36.Nicolas Salamin (Ireland); salamin@tcd.ie
Botany Department
Trinity College
College Green
Dublin 2
Ireland

37.Francesco Santini (France); fsantini@mnhn.fr
UMR 7138 CNRS Systematique, Adaptation, Evolution and
Service de Systematique Moleculaire
Department Systematique et Evolution
Museum National d'Histoire Naturelle
43 rue Cuvier
75231 Paris cedex 05
France

38.Vincent Savolainen (UK); v.savolainen@kew.org
Jodrell Laboratory, Royal Botanic Gardens, Kew
TW9 3DS Richmond
UK

39.Bernd Schierwater (Germany); bernd.schierwater@ecolevol.de
Insitute of Ecology and Evolution
Veterinary School of Hannover
Buenteweg 17d
30559 Hannover
Germany

40.Marie Skovgaard (Norway); maries@ii.uib.no
Computational Biology Unit
BCCS
University of Bergen
5020 Bergen
Norway

41.Åsa Tellgren (Norway); aasat@ii.uib.no
Computational Biology Unit
Centre for Computational Science
N-5020 Bergen
Norway

42.Alfried Vogler (UK); apv@nhm.ac.uk
Imperial College London, and
The Natural History Museum
Cromwell Road
London, SW7 5 BD
U.K.

43.Michael Whiting (USA); Michael_Whiting@byu.edu
Department of Integrative Biology
Brigham Young University
Provo, UT 84602
USA

44.Endre Willassen (Norway); endre.willassen@zmb.uib.no
University of Bergen
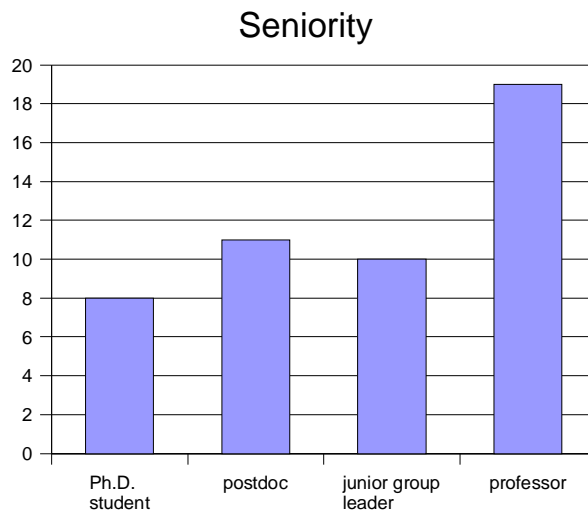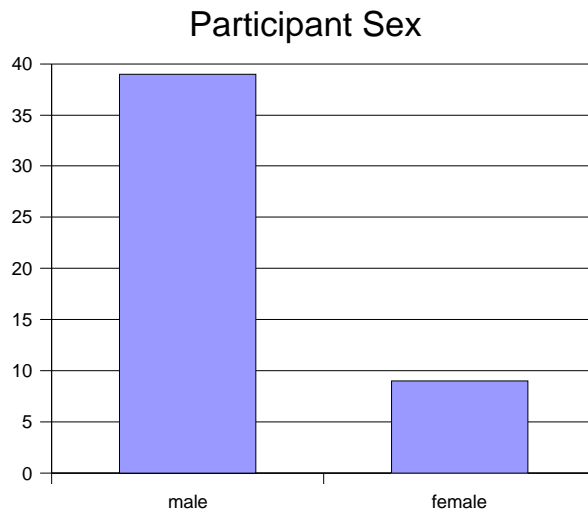Museum of Zoology
Muséplass 3
N-5007 Bergen
Norway

45.Gert Wörheide (Germany); gert.woerheide@geo.uni-goettingen.de
Geoscience Centre Göttingen
Dept. of Geobiology
Goldschmidtstr.3
37077 Göttingen
Germany

46.Ziheng Yang (UK); z.yang@ucl.ac.uk
University College London
Department of Biology
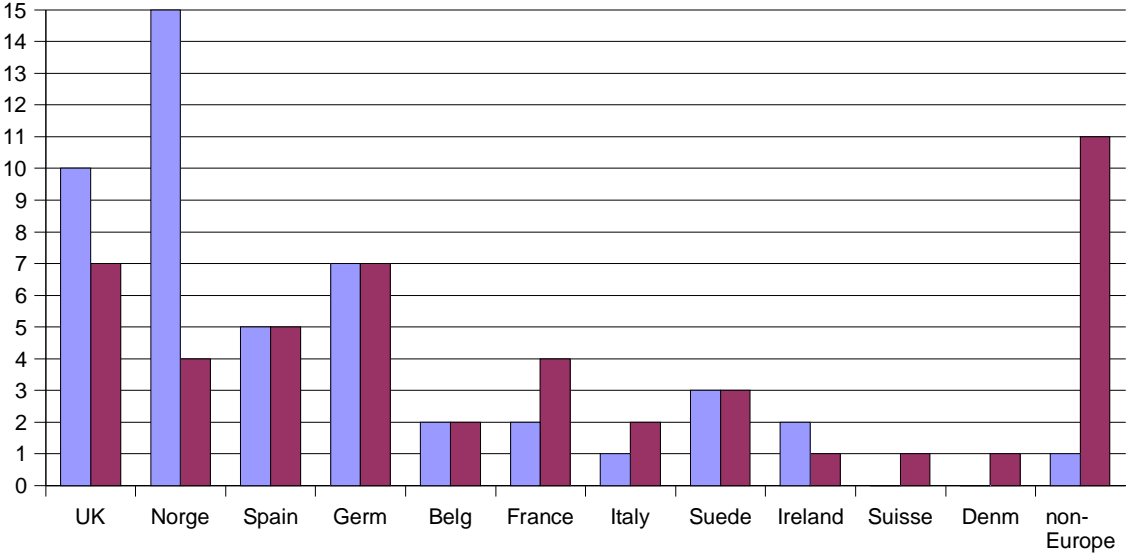Darwin Building, Gower Street, London WC1E 6BT
UK

47.Rafael Zardoya (Spain); rafaz@mncn.csic.es
Museo Nacional de Ciencias Naturales
Jose Gutierrez Abascal, 2
28006 Madrid
Spain

48.Mingli Zhang (Germany); minglizhang@hotmail.com
Special Institute of Botany
Mainz University
Bentzelweg 9a

Mainz 55099

Germany

Statistical Information on Participants

Participant Sex

	male	female
value	39	9

Seniority

Ph.D. student	postdoc	junior group leader	professor
8	11	10	19

country associations

blue= country of work

red=country of citizenship