> **Expression of Interest for an Integrated Project under Thematic Priority 1.1.6.3**
>
> # Assembling the Tree of Life - Europe (ATOL)
>
> **Call identifier: EOI.FP6.2002**
> **Date of submission: 6 June 2002**

List of Partners:

1. Royal Botanic Gardens, Kew, UK (Vincent Savolainen, Mark Chase, Mike Fay)
2. Imperial College, Univ of London, Ascot, UK (Tim Barraclough, Andy Purvis, Donald Quicke, Austin Burt)
3. The Natural History Museum, London, UK (Alfried Vogler, Mark Wilkinson, Richard Thomas, D.T.J. Littlewood)
4. Dept of Zoology, Univ of Oxford, Oxford, UK (Eddie Holmes, Andrew Rambaut)
5. Univ of Reading, Reading, UK (Mark Pagel, Paul-Michael Agapow, Julie Hawkins)
6. Univ of Glasgow, Glasgow, UK (Rod Page)
7. University Museum of Zoology, Cambridge, UK (Max Telford)
8. Natural History Museum of Geneva, Switzerland (Jean Mariaux)
9. Dept of Zoology, Univ of Geneva, Geneva, Switzerland (Jan Pawlowski)
10. Dept of Botany, Univ of Zurich, Zurich, Switzerland (Peter Linder)
11. Dept of Biology, National Univ of Ireland, Ireland (James McInerney)
12. Dept of Botany, Univ of Dublin, Ireland (Trevor Hodkinson)
13. Botanical Garden – Center for Biodiversity Conservation of the Polish Academy of Sciences, Warsaw, Poland (Piotr Bednarek)
14. Plant Breeding and Acclimatization Institute, Radzików, Poland (Wiesaw Podyma)
15. Univ of Uppsala, Uppsala, Sweden (Kåre Bremer, Fredrik Ronquist, Vincent Moulton, Katharina Huber, Mikael Thollesson)
16. Museum of Natural History, Stockholm, Sweden (Mari Källersjö, Steve Farris)
17. Dept of Evolutionary Botany, Univ of Copenhagen, Denmark (Gitte Petersen, Ole Seberg)
18. Zoological Museum, Univ of Copenhagen, Denmark (Niels Peder Kristensen, Rudolf Meier, Henrik Enghoff, Nikolaj Scharff, Nils Møller Andersen)
19. Zoological Institute, Univ of Copenhagen, Denmark (Jens Høeg, Henrik Glenner)
20. Novozymes A/S, Bagsvaerd, Denmark (Lene Lange)
21. ICAT, Lisboa, Portugal (Helena Cotrim)
22. Natural History Museum, Paris, France (Jean-Pierre Hugot, Simon Tillier, Thierry Bourgoin)
23. Univ Claude Bernard, Lyon, France (Manolo Gouy, Catherine Hänni)
24. Ecole Normale Superieure, Lyon, France (Vincent Laudet, Marc Robinson-Rechavi)
25. Univ of Montpellier, Montpellier, France (Olivier Gascuel, Vincent Berry, Emmanuel Douzery)
26. INSERM, Univ of Provence, Marseille, France (Pierre Pontarotti, André Gilles)
27. CIRAD, Montpellier, France (Xavier Perrier)
28. University of Jussieu, Paris, France (David Moreira, Purificación López-García)
29. Institute of Molecular Biology and Medecine, Free Univ of Brussels, Brussels, Belgium (Michel Millinkovitch)
30. National Botanic Garden, Meisse, Belgium (Elmar Robbrecht)
31. Dept of Zoology, Univ of Munich, Munich, Germany (Gerhard Haszprunar)
32. Dept. of Animal Sciences, Technical Univ. of Munich, Germany (Olaf Bininda-Emonds)
33. Genetics Institute, Univ of Koeln, Koeln, Germany (Diethard Tautz)
34. Zoological Research Institute and Museum Koenig, Bonn, Germany (Bernhard Misof)
35. German Collection of Microorganisms and Cell Cultures GmbH, Braunschweig, Germany (Erko Stackebrandt,)
36. Univ of Jena, Jena, Germany (Rolf Beutel)
37. Botanisches Museum Berlin-Dahlem, Freie Universität Berlin, Berlin, Germany (Christoph Oberprieler)
38. Palacky Univ, Czech Republic (Milada Bocakova, Ladislav Bocak)
39. Univ of Zaragoza, Spain (Pilar Catalan)
40. Natural History Museum, Madrid, Spain (Rafael Zardoya)
41. Dept of evolutionary biology, Univ of Siena, Siena, Italy (Francesco Frati)
42. Univ of Roma 2, Roma, Italy (Valerio Sbordoni)
43. Nationaal Herbarium Nederland, The Netherlands (Pieter Baas, Barbara Gravendeel, Lars Chatrou, Freek Bakker)
44. CBS. Fungal Biodiversity Center, Utrecht, The Netherlands (Dirk Van der Mei & Joost Stalpers, Pedro Crous)
45. Zoological Museum, Amsterdam, The Netherlands (Wouter Los)

1. Need and Relevance

Assembling the Tree of Life (the phylogeny of all organisms; ToL) is a priority for describing and understanding biodiversity and its broad exploitation for human purposes. Biodiversity affects our lives at numerous levels: (i) natural world providing resources, agriculture, water management, eco-tourism; (ii) understanding and exploiting diversity at functional and genomic levels; and (iii) understanding the spread of disease agents and vectors. **A comprehensive ToL is essential for a predictive understanding of biodiversity and tightly linked to successful expansion of post-genomic knowledge.** Piecemeal systematic projects cannot achieve this aim; we need a major integrated effort to obtain data and tools needed for a comprehensive, world-leading resource.

We propose a bold collaboration bringing together diverse expertise in phylogenetic theory, software development, molecular systematics and taxonomic knowledge so that ToL knowledge is made available to end-users in an integrated manner. The ToL will become the ultimate taxonomic reference. **The objectives of Thematic priority 1.1.6.3 Biodiversity and Ecosystems can only be met if there is a solid platform on which to base the risk assessments of biodiversity and the understanding of ecosystems.**

Europe has abundant expertise in these areas, but an integrated approach is now needed: (i) to spread expertise among institutions, (ii) to co-ordinate research toward integrated goals for targeted end-users, and (iii) to produce vital hardware and software that enable data collection and analysis on an ambitious scale across Europe (e.g., via grid supercomputing).

Phylogenetics has been revolutionised by molecular methods and statistical analysis. DNA sequences are now routinely and successfully used to reconstruct trees, leading to a database of molecular phylogenetic information that is growing at an ever-increasing rate. Phylogenetic approaches underpin modern biological research.

GENOMICS   * large-scale genome changes and their effects on function
 * genomic diversity across life as a whole, the new field of phylogenomics
 * detecting the action of natural selection across genomes
 * EvoDevo: understanding the diversity of developmental mechanisms

---

BIODIVERSITY * predicting functional diversity within ecosystems and communities
 * evolution of biodiversity, speciation and adaptive radiation
 * setting priorities for species conservation, reserve locations, and identifying extinction risks
 * targeting biological control of invasive species and understanding pest-host evolution.

HEALTH * informed bioprospecting for pharmaceutical and agrochemical products
 * tracking the origin and spread of diseases and their vectors, disease-host coevolution
 * the evolution of DNA transposable elements and risk factors for GMOs

---

Molecular methods and DNA sequencing technology are now sufficiently advanced to make a concerted assault on the ToL. **The key challenge is to integrate data collection on a massive scale**, filling in major gaps, e.g. single-celled eukaryotes and microbes, and producing detailed data for a full range of potential applications. In addition, methods for phylogenetic analysis were typically developed with much smaller tasks in mind. **A vital step towards the ToL is development of new methods, software and strategies (e.g. parallel computing) for analysing large matrices and integrating data into a single resource.**

2. Scale of ambition and critical mass

The aim of ATOL will be to produce a world-leading European phylogenetic resource as a tool for biodiversity research and management.

Specific scientific objectives:

- *Phylogenetic framework: all organisms in Europe, from natives to collections*. As a framework for more detailed studies, ATOL will generate phylogenetic data for all higher groups of European organisms, including natives and those found in the major biodiversity collections. **The aim is to obtain complete representation for a global ToL, at least at the taxonomic level of the family but mostly in more detail depending on the taxon in question.** By focusing on Europe and its collections (museums, seed banks etc), the results will have maximum potential, by virtue of the massive amount of information that already exists for European organisms in terms of ecology, genetics, conservation status etc. We will co-ordinate sampling strategy with ToL projects in the USA and elsewhere.

- *Detailed phylogenetic studies for organisms with a major impact on human society and/or role in the environment*. Target groups include beetles, parasitic wasps, flies, mites (diverse components of terrestrial ecosystems, agricultural pests, biocontrol), monocots (comparative crop genomics and drylands), fish, mammals, birds (conservation), worms (soil and parasites), bacteria, viruses, and yeasts (ecosystem processes, disease and food safety). These will be sampled intensively and comprehensively. **We will develop new tools for specific uses, such as databasing and DNA-identification, forensics, marker assisted artificial selection.** This objective would take advantage of existing taxonomic networks and links with USA but focus on EU issues and taxa.

- *Gap-filling for poorly known organisms*. We will generate phylogenetic data for poorly known groups of organisms for which there is strong European taxonomic expertise. Target groups include protozoa, soil fungi, rotifers, parasitic crustaceans, urochordates, frogs, etc. all groups with dominant roles in ecosystems (including marine) but which are little studied. For example amoeboid protists are among the most abundant free-living microbial eukaryotes in terrestrial and aquatic ecosystems, playing a key role in nutrient cycling, but their diversity and phylogenetic relationships remain poorly known.

- *Analytical tools for ToL and user interface*. We will develop new methods and software for analysing large DNA matrices, genome-level data at a phylogenetic level (phylogenomics), and supertree approaches for combining data sets into a single ToL. **We will adopt a new conceptual approach, utilising computer networks for parallel computing. We will develop web-based resources for submitting, compiling and disseminating ToL information**, linked to major taxonomic collections, gene banks (EMBL), and other ToL-oriented web sites.

- *ATOL as a resource for biological research*. ATOL members will use the new data for a wide range of research applications. Key areas will include the function and evolution of diversity in ecosystems, phylogenomics, and the evolution of development (EvoDevo).

- *SMEs: bioprospecting, impact assessment, health care*. We will target SMEs as end-users, especially companies processing ToL-related information for applied use.

These objectives are timely and critical to European research in pure and applied biodiversity studies. Modern molecular technology and computer power now makes this an achievable aim, but the methods need scaling up. **Europe has played a leading**

role in this area but needs a major thrust to compete and collaborate with current USA initiatives, such as the 15-year NSF programme on Assembling the Tree of Life. Global collaboration will be essential for the formidable task of assembling a ToL.

Europe has world-class expertise in the theory and practise of phylogenetics, with many major taxonomic and biodiversity institutions. ATOL would take advantage of existing networks; it will bring together wide expertise, and researchers working on a wide range of organisms. **The onrush of the post-genomic age and primacy of environmental issues and emerging diseases make phylogenetics a central requirement for society**.

3. Description of Challenges and Scientific Strategy

Our challenges are to sample a significant portion of the Earth biodiversity in a clearly prioritised way, to produce massive phylogenetic data (mostly molecular) and new bioinformatics tools so that a large portion of the ToL can be assembled.

The ToL must have a predictive value, increase our knowledge in biology and medicine, and be particularly relevant to ecosystems and disease management. It must facilitate emerging scientific disciplines, and become a foundation for all life sciences.

- *Intensive sampling.* Target groups will be sampled from the field and from the immense collections (literally several billions of organisms) held by European museums, botanical gardens, and seed/tissue banks. Taxonomic expertise across Europe has built up over centuries and this guarantees the success of this initial phase of ATOL.

- *Data production*. ATOL will focus on assembling DNA sequence data. Existing sequences will be extracted from EMBL, but clearly massive data production is required for ATOL. **Targeted genes, genomes (plastid or mitochondrial), or nuclear loci (provided by cDNA libraries) will be sequenced by the participating teams, taking advantage of detailed genomic knowledge from model organisms.** Molecular expertise and infrastructure is considerable across the teams so there is no significant risk associated with this phase of ATOL. We will also compile data on morphology and fossils into a morphobank, vital sources of independent data and for dating with molecular clocks.

- *Bioinformatics.* Developing bioinformatics tools appropriate to the scale of ATOL is the most challenging part of our project. Current methods and software are unable to deal with very large data sets that will result from ATOL. **ATOL will develop new approaches for analysing large data matrices, including development of new heuristics and search algorithms, development of new frameworks for analysis e.g. Bayesian methods, and use of networked parallel computing.** Another key challenge is to combine trees obtained for different groups and DNA regions into a single estimate of the ToL, so-called supertree analysis. Supertrees have been reconstructed for a wide range of taxonomic groups, including parasitic worms, birds, mammals and plants. **ATOL will design algorithms for large-scale supertree reconstruction, implemented in a 'live' on-line application for submission and retrieval of phylogenetic information.** Expertise in bioinformatics is excellent and members of ATOL have written hundreds of phylogenetic applications. What is needed is to scale up these approaches and to integrate them in a user-friendly manner. One risk is the high computational demand needed for tree searching; we will explore radical approaches for optimizing computer resources across the network.

- *Use as predictive tool*. Use of phylogenetics in biological research has exploded over the last decade, leading to greater predictive power and statistical rigour for

comparative biology. Theory for the statistical analysis of biological diversity on phylogenetic trees is extensive. **A key challenge for ATOL will be to foster use of the vastly increased potential offered by ToL-scale data.** We will develop new tools for phylogenetic analysis, particularly in phylogenomics, biodiversity, disease evolution, and EvoDevo.

4. Integrated Activities

The ATOL programme will focus on data production for targeted groups of organisms and the development of software for building and analysing the ToL. A major component will be providing advanced training at both pre- and postdoctoral levels. Teams will produce new data for each target group of organisms, and identify specific analytical challenges that require concerted and integrated effort. A team of about 10 bioinformatics scientists will work on developing the necessary software, including a core team located centrally, linked to computer experts and to the major bioinformatics and genomic centres in Europe (e.g. EBI/EMBL), and additional members interacting on a day-to-day basis with the biologist teams.

Building on biotechnological and advanced computer resources, ATOL will train approximately 80 young scientists recruited across Europe. There is a clearly uneven distribution of phylogenetic activities across Europe and special attention will be given to allocating resources to countries with expertise but fewer resources to conduct ToL research. The development of open web-based tools for accessing, analysing and utilising ATOL information will also contribute to making ATOL benefits available throughout Europe.

ATOL has such a broad scope that it is multidisciplinary in essence. Biological specialists must be involved in sample collection. Advanced biotechnology is required to achieve massive molecular data production. Bioinformatics, phylogenomic, mathematical and biological experts must be involved in the analytical phases before finished products can be delivered to the broader scientific community (ecology, medicine, etc.) and other end-users in society including SMEs (ecosystem managers, conservationists, pharmacologists, bioprospectors etc).

Computing resources will be concentrated on one site and provide a bioinformatics platform where members of the teams will work for 3-6-month periods of time. The aim will be to provide a core network accessible by all members of ATOL and optimising computer resources across participating institutions. Communication among the teams will be facilitated by an intranet web site and meetings. An external web site will enhance networking across the world.

Dissemination. ATOL will be very pro-active in disseminating its results and tools through available means and particular attention will be given to free electronic dissemination.

Management. There will be a Project Steering Group made up of core members working with the scientific co-ordinator and a full time project co-ordinator.

Budget: 30 MEURO over four years.

Overview of leading industrial and academic research centres and other major participants
(Gov=Government; Univ=University)

Partner	Type	Expertise	Role (with indication of team leaders)
Royal Botanic Gardens, Kew, UK	Gov	Botany	Assemble plant trees (Vincent Savolainen, Mark Chase, Mike Fay)
Imperial College, Univ of London, Ascot, UK	Univ	Evolutionary biology, bioinformatics	Comparative analyses and trees (Tim Barraclough, Andy Purvis, Donald Quicke, Austin Burt)
The Natural History Museum, London, UK	Gov	Bioinformatics, zoology	Assemble animal trees, supertree (Alfred Vogler, Mark Wilkinson, Richard Thomas, D.T.J. Littlewood)
Dept of Zoology, Univ of Oxford, UK	Univ	Bioinformatics	Assemble bacterial and viral trees, software (Eddie Holmes, Andrew Rambaut)
Univ of Reading, UK	Univ	Bioinformatics,	Develop new bioinformatic tools (Mark Pagel, Paul-Michael Agapow, Julie Hawkins)
Univ of Glasgow, UK	Univ	Bioinformatics	Develop new bioinformatic tools (Rod Page)
Univ Cambridge, UK	Univ	Zoology	Assemble animal trees (Max Telford)
Natural History Museum, Switzerland	Gov	Zoology	Assemble trees for parasitic worms (Jean Mariaux)
Dept of Zoology, Univ of Geneva, Switzerland	Univ	Zoology	Assemble trees for free-living protists (Jan Pawlowski)
Univ of Zurich, Switzerland	Univ	Botany	Assemble plant trees (Peter Linder)
National Univ of Ireland, Ireland	Univ	Bioinformatics	Develop software for supertrees (James McInerney)
Dept of Botany, Trinity College, Ireland	Univ	Botany	Assemble trees for grasses (Trevor Hodkinson)
Polish Academy of Sciences, Poland	Gov	Germplasm	Assemble plant trees (Piotr Bednarek)
Plant Breeding Institute, Poland	Gov	Botany	Assemble plant trees (Wiesaw Podyma)
Univ of Uppsala + Museum Stockholm Sweden	Univ	Bioinformatics, biogeography	Assemble trees, algorithms (Bremer, Ronquist, Moulton, Huber, Kallersjo, Farris, Thollesson)
Univ of Copenhagen, Denmark	Univ	Botany	Assemble plant trees (Gitte Petersen, Ole Seberg)
Zoological Museum, Univ. of Copenhagen, Denmark	Univ	Zoology	Assemble Arthropod trees (N. Kristensen, R. Meier, H. Enghoff, N. Scharff, N. Andersen)
Zoological Institute, Denmark	Univ	Zoology	Assemble animal trees (Jens Høeg, Henrik Glenner)
Novozymes A/S, Bagsvaerd, Denmark	SME	Mycology	Assemble fungal trees (Lene Lange)
ICAT, Portugal		Botany	Assemble plant trees (Helena Cotrim)
Natural History Museum, France	Gov	Zoology Paleontology	Assemble animal trees (Jean-Pierre Hugot, Simon Tillier, Thierry Bourgoin)
Univ Claude Bernard, France	Univ	Bioinformatics	Develop software in molecular evolution (Manolo Gouy, Catherine Hänni)
Ecole Normale Superieure, France	Acad	Bionformatics	Assemble animal trees (Vincent Laudet, Marc Robinson-Rechavi)
Univ of Montpellier, France	Univ	Bioinformatics, zoology	Develop algorithms, assemble animal trees (Olivier Gascuel, Vincent Berry, Emmanuel Douzery)
INSERM, Univ of Marseille, France	Gov	Marine Biology	Develop phylogenomics of marine invertebrates (Pierre Pontarotti, André Gilles)
CIRAD, France	Gov	Bioinformatics	Develop supertrees (Xavier Perrier)
University of Jussieu, France	Univ	Bioinformatics, Protistology	Develop new algortithm, Assemble trees for marine protists (David Mareira, Purificación López-García)
Free Univ of Brussels, Belgium	Univ	Bioinformatics	Assemble animal trees (Michel Millinkovitch)
National Botanic Garden, Belgium	Gov	Botany	Assemble plant trees (Elmar Robbrecht)
Univs, Munich, Germany	Univ	Bioinformatics, zoology	Develop supertrees (Olaf Bininda-Emonds, Gerhard Haszprunar)
Genetics Institute, Univ of Koeln, Germany	Univ	Zoology	Assemble animal trees (Diethard Tautz)
Zoological Research Institute, Germany	Gov	Zoology	Assemble arthropod trees (Bernhard Misof)
German Collection of Microorganisms and Cell Cultures GmbH, Germany	Gov	Mycology, bacteriology	Assemble micro-organism trees (Erko Stackebrandt,)
Univ of Jena, Germany	Univ	Zoology	Assemble arthropod trees (Rolf Beutel)
Botanisches Museum Berlin-Dahlem, Freie Universität Berlin, Germany	Univ	Botany	Assemble plant trees (Christoph Oberprieler)
Palacky Univ, Czech Republic	Univ	Zoology	Animal trees (Milada Bocakova, Ladislav Bocak)
Univ of Zaragoza, Spain	Univ	Botany	Assemble plant trees (Pilar Catalan)
Natural History Museum, Madrid, Spain	Gov	Zoology	Assemble insect trees (Rafael Zardoya)
Univ of Siena, Italy	Univ	Zoology	Assemble arthropod trees (Francesco Frati)
Univ of Roma 2, Italy	Univ	Zoology	Assemble arthropod trees (Valerio Sbordoni)
Nationaal Herbarium Nederland	Gov	Botany	Assemble plant trees (Pieter Baas & coll.)
CBS. Fungal Biodiversity Center, The Netherlands	Gov	Mycology, bacteriology	Assemble micro-organismal trees (Dirk Van der Mei, Joost Stalpers, Pedro Crous)
Zoological Museum, Amsterdam, The Netherlands	Gov	Zoology	Assemble animal trees (Wouter Los)