TAXONOMY SERVING SOCIETY

http://www.ucl.ac.uk/taxome/

Expression of Interest for
Integrated Project under
EU Framework VI
Call identifier: EOI.FP6.2002

presented by
The Lepidoptera Taxome Project
and other groups

James Mallet
Honorary Research Fellow, The Natural History Museum, London
&
Professor of Biological Diversity
University College London
The Galton Laboratory
4 Stephenson Way
LONDON NW1 2HE, UK
tel: (+44)-20-7679-7411

7 June 2002

Abstract

This project will develop large-scale complete taxonomy for a hyper-diverse group, the Lepidoptera, consisting of ~180,000 species, about 10% of all species known globally. The project will deliver the result to the widest possible audience online. It will integrate taxonomy with research in genomics, agricultural biology, modelling and prediction of global changes in biodiversity, and improve links between taxonomy and users of scientific information within industry, agriculture and conservation. The project will train a new generation of scientists in modern taxonomy and molecular genetics for a competitive Europe. The work will integrate scientists across Europe (including associated states), and encourage synergy and feedback between the disciplines of genomics, genetics, taxonomy, ecology, and applied sciences such as biotechnology, conservation and agricultural biology.

Need and relevance

Background. Taxonomy, the naming and classification of organisms, is fundamental to all biological science and biotechnology. Taxonomy is the original bioinformatics, and our modern binominal, Linnean nomenclature dates from the Age of Reason in Europe. In spite of revolutions in informatics for satellite imagery, genomics and other sciences, taxonomy informatics has remained virtually unchanged for 250 years.

State of the art in Europe. Taxonomy is a European invention, and Europe still contains the world's major taxonomic resources and among the best taxonomists. The majority of the world's type specimens probably occur in European museums ("types" are individual voucher specimens on which names of taxa, such as species or genera, are based). For example, in the Lepidoptera, over 80% of all types are housed in European museums, and for certain groups, well over 60% in one institute, The Natural History Museum, London[1]. Yet taxonomy is in decline, and the virtual lack of training in Western European countries is becoming critical[2,3]. Taxonomy is fundamental to all of biology, but is now virtually restricted to museums, so that integration with the rest of the scientific community has suffered. Because of very limited government funding or overheads for taxonomy, competitive and cash-strapped European universities and museums increasingly hire in other areas. In contrast, the USA recognizes the scientific and commercial value of taxonomy, and has repeatedly funded major programs in taxonomy and systematics[4].

Challenges and key gaps in current knowledge. Estimates of the total number of species on this planet vary widely over two orders of magnitude (3-100 million)[5], and species relationships remain obscure in even the best-known groups. Taxonomic information is scattered widely in small-circulation books and print journals, so that conservationists, biotechnologists, and other users of taxonomy cannot easily access data on or even count the ~1.5-2 million "known" species[6] estimated to have been described. The inadequacy of existing taxonomy informatics has triggered repeated calls from major scientific and public figures for modernization and increased funding[5,7]. Prof. C. Godfray is only the most recent supporter of a global web-based presentation of taxonomic knowledge[7], a concept we here term **"the Taxome Project"** (i.e. a complete taxonomy knowledge base for all of life, publicly available online).

A major demonstration for the taxonomy of 10% of the world's species is proposed here. This project will also integrate this taxonomic knowledge with other biological research and its user community in Europe. This project will serve society by delivering the means of monitoring impacts of global environmental change and ecosystem functioning, by researching new methods of sustainable crop protection in agriculture, and by training expertise for biotechnology[3] and other industries depending on taxonomy and molecular genetics. Such a project would stem the decline in taxonomy, develop innovative taxonomy across Europe, and demonstrate how this information can be put to work to foment European competitiveness in scientific excellence, biotechnology and agriculture.

The research will address the priority theme of Global Change and Ecosystems (1.1.6.3), particularly to understand and predict global change and to develop strategies for prevention, mitigation and adaptation. The research will contribute towards preserving ecosystems and protecting biodiversity, and to sustainable use of land resources. Biodiversity monitoring and conservation strategies, are required by international agreements, such as the Convention on Biological Diversity, which has been ratified by most European countries. Europe suffers from a "taxonomic impediment" in biodiversity research[8] which can be ameliorated only by training more taxonomists and making taxonomic information more generally availabile. The research will thus also touch on issues of Knowledge-based Society and Social Cohesion (1.1.2 and 1.1.7.1).

Scale of ambition and critical mass

Main scientific objectives. The work focuses on the complete taxonomy of a large group and delivering this information directly over the internet and via key research groups to end users. We here choose the Lepidoptera, or butterflies and moths (other hyper-diverse groups might also be suitable as model taxomes, for example the flowering plants). With ~180,000 known species[9], the Lepidoptera form ~10% of the organic species diversity known on this planet, ~1.5 million described species[6] overall.

The research interface will consist of a number of units integrated as in Appendix 1. The composition of each unit is detailed in Appendix 3:

1) *The Lepidoptera Taxome Project*. This unit will develop a species level alpha-taxonomy for the entire Lepidoptera. A subset, the *Butterfly Taxome Project* will be completed to a very high accuracy and will incorporate morphological studies with molecular work (unit 2) into a phylogenetic classification.

2) *Genetics and Evolution.* (a) This unit will perform a "horizontal genome project" for the butterflies using a few standardized nuclear and mitochondrial genes. These molecular data will be used in combination with morphology to reconstruct phylogeny and provide a classification. (b) The unit will also be responsible for investigating and identifying genes for insecticide resistance and resistance in moths to crop plants genetically engineered with biocidal proteins, as well as (c) to understand the process of host race formation and speciation on natural and crop hosts.

3) *Prediction and Modelling of Biotic Change*. This unit will predict and model future biotic change and suggest strategies for amelioration or avoidance of damage.

4) *Agricultural Applications*. This unit will investigate economically important pests and their interactions with crops to design strategies for profitable and sustainable agriculture.

European strategic impact on competitiveness and social impact. As well as providing a model system for taxonomy in the service of society, Lepidoptera are important in their own right as economic pests of European agriculture and as biodiversity indicator taxa. Thus, Lepidoptera taxonomy will deliver research directly to industry and agriculture, and contribute towards assessing and forecasting changes in biodiversity, and investigating the structure, function and dynamics of ecosystems and their services. The work will monitor the biotic impacts of global climate change, and develop models enabling forecasting of changes in ecosystems' functioning, biodiversity and habitats. Perhaps even more importantly, the work will provide a major training program in strategies for sustainable agricultural and wild land management, and for expertise useful in gene discovery for the biotechnology industry[3].

Existing expertise of senior scientists within the consortium will allow implementation and critical mass to be found largely within Europe; these professionals will provide a strong background for training taxonomists, geneticists, molecular evolutionists, modellers, and agricultural biologists at postgraduate and postdoctoral level.

Summary of expected results and outputs. Overall, the project will integrate teams of museum taxonomists with molecular geneticists and evolutionary biologists, ecologists and biotic change modellers, and agricultural biologists. These groups will themselves interface closely with users in agriculture, conservation, biotechnology and education.

1) *The Lepidoptera Taxome Project*. This unit will deliver a species level alpha-taxonomy, consisting of a complete online synonymic catalogue of Lepidoptera. The *Butterfly Taxome Project* subset in particular will be completed to an extremely high level of accuracy (99.5%) and utility, including an online knowledge bank consisting of classification and lists of all known species-group and higher level

nomenclature including synonymy for the world's 17,500 species, a complete taxonomic literature reference database for all valid names and synonyms (99.5%), digitized original species descriptions for all taxa and photographs of type specimens, as well as distribution, ecological and life-history data.

The primary means of dissemination for this taxonomic information will be via a public online knowledge bank, from a single website (and mirror sites). The advantage of this mode of delivery is its widespread accessibility to scientists, not just in Europe, but worldwide as a form of data repatriation of information held in Europe's collections and systematic resources. Museums, of course, serve the public by means of education and outreach as well as research. Another advantage of online taxonomy is its flexibility of use. For example, museums will be able to provide public or schools outreach by developing simple and user-friendly access to the same taxonomic knowledge bank.

2) *Genetics and Genomics.* (a) This unit will deliver DNA sequence data for the 2000 genera of butterflies. It will collate existing sequence data and produce substantial new sequences of ~3-6 Kbp of key, standardized nuclear and mitochondrial genes for every butterfly genus group. The project will provide molecular characters for use in phylogeny reconstruction. The resulting estimated phylogenies will then be used in classification. While this may seem ambitious, 3 Kbp per genus group within the Nymphalidae, covering ~10% of butterfly genera is already done (Stockholm & USA), and much work on Papilionidae, Pieridae and Lycaenidae has also been performed (USA). Sequencing the total of 6-12 Mbp is trivial using current technology; the main challenges will be to obtain the DNA samples and develop new PCR primers to enable coverage of all genera. The unit will also interface with genome organizations such as the Lepidoptera genome projects and GENBANK. (b) The unit will also identify genes for resistance resistance in moths to insecticides and crops genetically engineered with biocidal proteins, and (c) provide data on host switching and host race formation for natural and crop hosts. Data from (b) and (c) will be used by the agricultural applications unit to design and implement sustainable pest management strategies.

3) *Prediction and Modelling of Biotic Change*. Butterflies are poikilotherms strongly dependent on solar heating, and are particularly sensitive to climate change and other environmental disturbances. Monitoring changes of their distribution in Europe has already led to an understanding of biotic impacts of habitat alteration and climate change[10,11]. Based on existing distribution databases and monitoring in Europe and worldwide, this unit will model and predict impacts of future climate and environmental change on geographic distribution of species and ecosystems. In collaboration with other monitoring and conservation organizations, this unit will provide data and suggest strategies for amelioration or avoidance of ecosystem damage.

4) *Agricultural Applications*. Information from units 1-3 will be applied to problems in agricultural ecosystems to promote sustainable, competitive crop protection. For example, recent work on *Heliothis* pests has elucidated the structure and function of alleles for resistance to *Bacillus thuringiensis* endotoxins used in GM crops[12]. This work will help in the race to design new endotoxins with different specificities. Work with European corn borer (*Ostrinia*) has demonstrated genetic differences between corn-infesting and native shrub-feeding populations, suggesting that wild hosts may harbour different, genetically incompatible strains of pests than crop plants[13]. Wild hosts are often assumed to be important reservoirs of moth susceptibility genes to insecticides and GM biocides, so management plans may now have to be altered. This work shows how genetics can interact with accurate taxonomy to design and improve more sustainable crop protection strategies.

Integrated activities and resources required

Development and demonstration of European competitiveness. This project on Lepidoptera will provide a first for Europe by demonstrating multidisciplinary research on large-scale taxonomy (10% of the world's species). It will integrate museum science with research in universities and other institutes,

and improve links to users of scientific information within industry, agriculture and conservation bodies. The project will train a new generation of scientists in modern taxonomy and molecular genetics to form a basis for European competitiveness. The work will integrate scientists working on Lepidoptera from across Europe (including associated states), and provide the basis for synergy and feedback between the disciplines of genomics, genetics, taxonomy, ecology, and applied science such as conservation, agricultural biology, and biotechnology.

Training, personnel and resources required. Most scientists to be hired will be trainees or technicians. We estimate at least 10 postdoctoral scientists, and similar numbers of postgraduate students and technicians are required per unit, while the majority of senior scientists (Appendix 3) will remain in existing long-term employments. The project is highly multidisciplinary: staff and trainees are required in disciplines as diverse as museum taxonomy, molecular and population genetics, population ecology, agricultural biology, biotechnology, and computer science.
Associated funding will be required for provision of servers and software to deliver the internet-based taxonomy reliably and securely, and for DNA sequencing facilities and consumables. Funds are also required for ecological work, travel, and basic taxonomy. Meanwhile, non-research costs for conservation and marketable agricultural applications will be funded via normal channels for conservation, biotechnology, and agriculture.

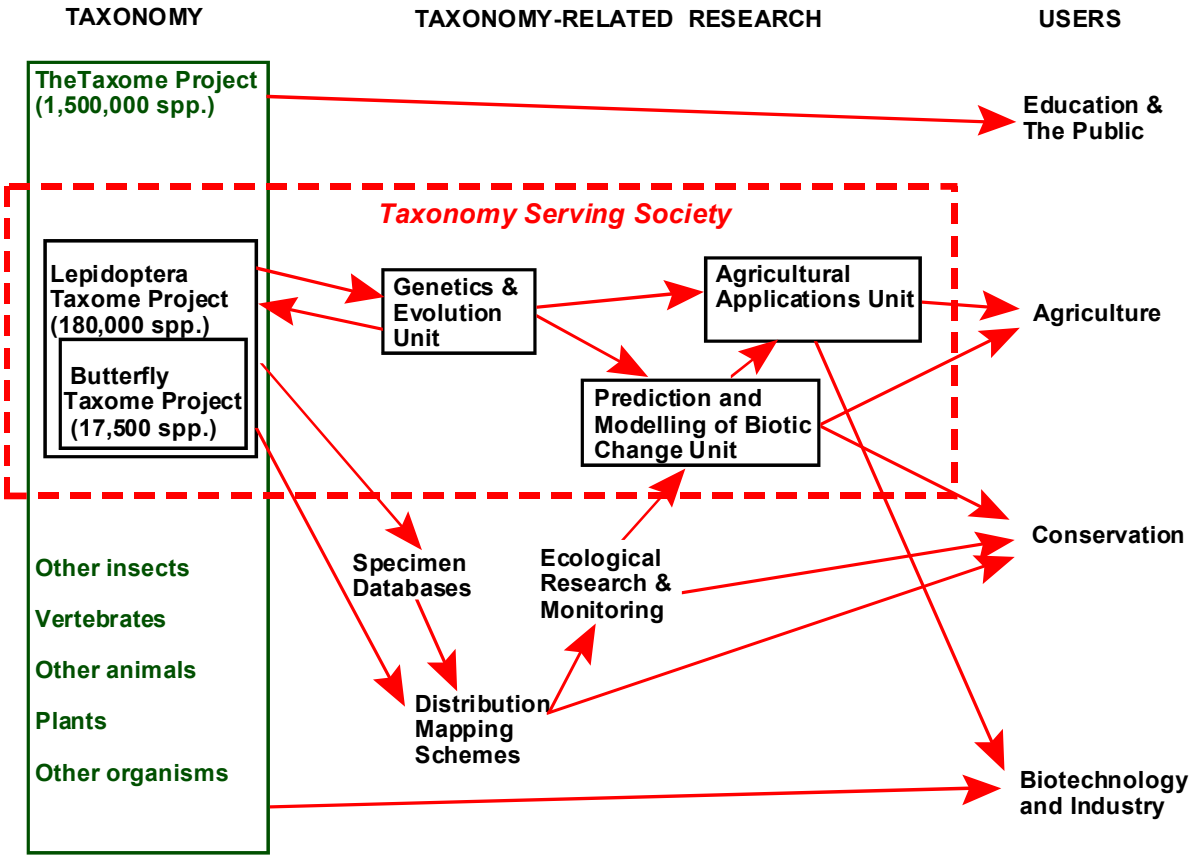
Organisation and management. This Integrated Project will be coordinated by a small overall management team probably based in London, while management of each semi-autonomous unit will be accomplished by team leaders from across Europe. Each unit will itself consist of interdisciplinary multi-national consortia.

Collaboration with groups outside the IP. The project depends on collaboration with conservation organizations, ecological monitoring groups, distribution mapping schemes, industrial groups (particularly those involved with biotechnology and crop protection) and online specimen databases. Some of these could well become partners with our IP before proposal submission; other collaboration will be outside the project.

A major long-lasting output of this Integrated Project will be the production of >99% complete online taxonomy for 10% of the world's described species, available to all users. Other projects with similar aims exist[14,15], such as Species 2000 and All Species, that aim to make available complete taxonomic information on the web. However, those projects are all in very early stages. This project aims to tackle a major hyper-diverse group, which can be completed within 5 years. This will be a European first, but as with any such global informatics project, work by our teams must complement and cooperate with similar projects worldwide.

Consensus and collaboration with the worldwide Lepidoptera taxonomy community is necessary will ensure that the work is respected and serviceable worldwide. We are in touch with all the main world experts on Lepidoptera, and will involve them fully in decision-making about most units in this project, particularly unit 1. The project can and will operate within existing International Commission for Zoological Nomenclature guidelines, rather than adopting or promoting new standards as suggested elsewhere[7].

Appendix 1. Information flow in "Taxonomy Serving Society" Integrated Project.
Relationships among units in and currently outside the project also shown.



Appendix 2. Notes and cited literature

1 Estimates are based on existing data collated for Riodinidae, Nymphalidae and Geometridae held at the NHM, London. The lack of total information makes even these mild claims hard to assess; Taxome Projects outlined here will, among other achievements, enable an assessment of the global value of European and other important world type collections for the first time.

2 Disney RHL 1998 Nature 394: 120
3 Jaspars M 1998 Nature 394: 413
4 e.g. NSF 2002 http://www.nsf.gov/pubsys/ods/getpub.cfm?nsf02074
5 Gaston KJ, May RM 1992 Nature 356: 281-282
6 Wilson EO 1992 The Diversity of Life
7 Godfray HCJ 2002 Challenges for taxonomy. Nature 417: 17-19
8 Hoagland KE 1996 http://www.ascoll.org/Newsletter/taxImp.htm
9 Kristensen NP (Ed.) 1999 Lepidoptera: Moths and Butterflies. Handbuch der Zoologie
10 Parmesan C et al. 1999 Nature 399: 579-583
11 Thomas CD et al. 2001 Nature 411: 577-581
12 Gahan LJ, Gould F, Heckel DG 2001 Science 293, 857-860
13 Bourguet D et al. 2000 Proc. Roy. Soc. Lond. B 267: 1177-1184
14 In Europe: http://www.cetaf.org/, http://www.biocase.org/, http://www.insects-online.de/, http://www.faunaeur.org/enbi/info.html, http://www.sp2000.org/, http://viadocs.essex.ac.uk/html/
15 Outside Europe: http://www.all-species.org/, http://www.gbif.org/, http://www.ento.csiro.au/globis/, http://www.sis.agr.gc.ca/pls/itisca/taxaget, http://tolweb.org/tree/phylogeny.html

Appendix 3. Taxonomy Serving Society: suggested principal personnel.
Tentatively proposed leading organizations stand at the head of each unit.

The Lepidoptera Taxome Project
Niels Kristensen, University of Copenhagen
Malcolm Scoble, Brian Pitkin, George Beccaloni, Natural History Museum, London
Joel Minet, Muséum National d'Histoire Naturelle, Paris
Ugo Dall'Asta, Royal Museum for Central Africa, Brussels
& others

The Butterfly Taxome Project
Dick Vane-Wright, Phil Ackery, Campbell Smith &c, The Natural History Museum London
Rienk de Jong, National Museum of Natural History, Leiden
Christoph L. Häuser, Staatliches Museum für Naturkunde, Stuttgart (Chair of GBIF)
Christian Köppel, Verlag für interaktive Medien Gbr
Henri Descimon, Université de Marseille
Zsolt Bálint, Hungarian Natural History Museum, Budapest
& others

Genetics and Evolution Unit
James Mallet, University College London
Soren Nylin, Niklas Wahlberg, Stockholm University
Valerio Sbordoni, Tor Vergata University, Rome
David Heckel, University of Melbourne, Australia
François Michel, CNRS, Gif-sur-Yvette
Steph Menken, University of Amsterdam
Frantisek Marec, Institute of Entomology ASCR, Ceske Budejovice
& others

Agricultural Applications Unit
Denis Bourguet, INRA, Guyancourt
Ian Denholm, Hugh Loxdale, IACR, Rothamsted
& others

Monitoring, Modelling and Prediction of Biotic Change Unit
Brian Huntley, University of Durham
& others