Plankton taxonomy in the computer age

Fabien JOUENNE, Ian PROBERT and Daniel VAULOT

UPMC Univ. Paris 06 et CNRS, UMR 7144, Station Biologique de Roscoff, Place Georges Teissier,
BP 74, 29682 ROSCOFF Cedex, France. Fax : +33 2.98.29.23.24. Email: jouenne@sb-roscoff.fr

Abstract: Preservation of biodiversity starts with knowledge and cataloguing of biodiversity, i.e. taxonomy. In this context, numerous web-based projects have developed over the last fifteen years, several with the ambitious aim of listing all living organisms described to date. Individual lists have been progressively incorporated into federative projects, such as Species 2000, the Global Biodiversity Information Facility (GBIF), or more recently the Encyclopedia of Life. We illustrate these changes by reviewing existing on line resources and presenting in some detail Plankton*Net, an interactive website dedicated to the taxonomy and images of plankton. We then provide some perspectives on the possible evolution of the links between taxonomy and the Internet.

Résumé : *La taxinomie du plancton à l'ère d'Internet.* La conservation de la biodiversité débute par la connaissance de cette biodiversité, c'est-à-dire par la taxinomie. Partant de ce principe, de nombreux projets Internet se sont développés ces quinze dernières années avec pour objectif de lister l'ensemble des organismes vivants. Les simples listes qui étaient disponibles au départ ont progressivement été incorporées dans des projets plus fédérateurs, tels que *Species 2000*, *Global Biodiversity Information Facility* (GBIF) ou plus récemment *Encyclopedia of Life*. Nous illustrons ces évolutions en synthétisant les différentes ressources disponibles en ligne pour la taxinomie et en présentant plus en détail Plankton*Net, un site web interactif et participatif dédié à la taxinomie du plancton. Nous concluons avec quelques réflexions sur l'avenir possible des liens entre Internet et taxinomie.

Keywords: Taxonomy ● Internet ● Plankton ● Databases ● Websites

Introduction

Global interest in environmental issues has increased in recent decades, with conservation of biodiversity at the centre of scientific, public and political concerns. Highly complex interactions take place between species in an ecosystem and perturbation of biodiversity has been shown to destabilize ecosystem productivity (Naeem & Li, 1997; Tilman et al., 1997), as well as to disrupt reservoirs of genetic variability (Heal, 2004). Changes in biodiversity have occurred for billions of years in response to "natural" environmental change and competition between species. During the Anthropocene, human activities have created threats to biodiversity either directly by converting,

degrading or fragmenting ecosystems, or indirectly by non-sustainable harvesting, introducing non-native species, or inducing climate change (Armsworth et al., 2004). Removal of keystone taxa (e.g. the passenger pigeon, Blockstein, 1998) or introduction of exotic species (e.g. the comb jelly *Mnemiopsis leidyi* or the macroalga *Caulerpa taxifolia*, Bax et al., 2001) have been repeatedly shown to lead to drastic changes in ecosystems, often with unpredictable and nefarious consequences. An illustration of the insurance role of biodiversity comes from the recent history of rice production. In the 1970s, a new virus threatened the Asian rice crop and thus the prosperity and food security of billions of people. To develop a form of rice resistant to this virus, scientists managed to locate a resistant variety of wild rice, only found in one location. Without the gene of this rice variety, apparently without any commercial value, the world's rice crop would have been seriously damaged (Heal, 2004). In light of the importance of biodiversity, the scientific community has come to recognise the need for biodiversity inventories in order to contribute to the conservation effort (Raven & Wilson, 1992).

Taxonomy, i.e. the science of describing, naming and classifying organisms, is obviously critical for assessing biodiversity. Hundreds of thousands of species have been described since the time of Linnaeus in the early 18th century, with international nomenclatural codes providing regulatory frameworks since the mid 19th century. Classical observation-based taxonomy is no longer, however, the main source of discovery of biodiversity. In the last two decades, molecular approaches have developed at an extremely rapid rate, providing methods for assessing the validity of existing taxonomy, but also revealing an ever increasing amount of 'hidden' (i.e. undescribed) biodiversity, particularly amongst unicellular prokaryotic and eukaryotic microbes (Giovannoni et al., 1990; Moon-van der Staay et al., 2001). Unfortunately, as molecular biology approaches have developed, recruitment of young taxonomists has fallen dramatically, preventing transmission of expertise between generations (the so called 'taxonomic impediment', Wheeler et al., 2004). During the last decades of the 20th century, taxonomy has been progressively considered as an ageing science pursued by ageing scientists. Paradoxically, however, a greater appreciation of our planet's changing biodiversity, revealed by molecular methods, has increased the need for traditional systematic expertise to identify organisms for which sequences have been acquired (Godfray, 2007). In the context of the recent explosion of genetic, genomic, proteomic and other molecular data (Godfray, 2002), there has been a pressing need to modernize taxonomic procedures. Computers were first used to hold taxonomic information in the 1960s (Sokal & Sneath, 1963 *in* Edwards

& Morse, 1995), but it is only in the last decade that a number of large-scale Internet-based taxonomy projects have been initiated, mostly focussing on metazoan and plant biodiversity.

Plankton, i.e. organisms that drift freely in the water column, plays key roles in the functioning of marine systems since it forms the base of most aquatic food webs and mediates many elemental cycles such as those of carbon, nitrogen and sulphur. In addition, a number of planktonic organisms have harmful effects, notably due to production of toxins. However, a very wide range of phylogenetic groups are represented in the plankton, especially in marine waters, and this heterogeneous group of organisms therefore offers specific challenges to taxonomy. Until quite recently very few web resources focusing on plankton were available, in part due to the fact that the smallest forms of plankton, i.e. protists and bacteria, often lack visible morphological features.

The year 2007 simultaneously saw the 300th anniversary of Linnaeus' birth and the launch of a very ambitious project, the Encyclopedia of Life, aimed at creating one web page for each extant species. It seems therefore an opportune time to review online taxonomic resources for plankton and to provide perspectives on future evolutions.

Online resources for taxonomy

General taxonomy

Internet taxonomy resources can be divided into two groups (Lilburn et al., 2006). The first group consists of databases and portals that provide information on species (names, classification, images, references, biogeographical data, etc.). Focus can be on a specific environment, a specific taxonomic group, or more general (Table 1). Initiated in 1996, the goal of the Tree of Life (ToL) project is to provide one web page for each group of organisms with images, text, and other information. Pages (almost 9000 at present) are linked in a phylogenetically hierarchical manner, pages on individual species being at the termination of the branches of the evolutionary tree. In 2001, the Catalogue of Life (CoL) was created from the association of Species 2000 and the Integrated Taxonomic Information System (ITIS), with the aim of becoming a comprehensive catalogue of all known species of organisms on Earth by 2011 (the 2008 edition of the Annual Checklist contains > 1.1 million species). The Global Biodiversity Information Facility (GBIF) was created in 2001 to facilitate digitalisation, global dissemination, and archiving of primary biodiversity data for many user groups (e.g. taxonomists, conservation biologists, molecular genetics researchers, administrative officers, school pupils, etc.). It gathers data

from more than 200 different data providers and contains more than 1.5 million records which can be explored by species, country, or dataset. The Encyclopedia of Life (EOL), launched in 2007, is a project aimed at giving life to E.O. Wilson's view: "Imagine an electronic page for each species of each organism on Earth, available everywhere by single access on command" (Wilson, 2003). It collects information from many different partners (including GBIF and Catalogue of Life).

A second group of electronic resources comprises tools that enable users to identify or classify an unknown specimen (Table 1). Among these, multi-access identification keys use a matrix of species and character combinations. A traditional dichotomous key requests precise information at each step about specific character states which can often only be determined with sophisticated techniques such as electron microscopy (e.g. inclination of striae on the valve of a diatom). In contrast, multi-access keys (e.g. photosynthetic euglenoids, http://euglena.msu.edu/lucid3/ Photosynthetic_Euglenoids. html, Uzwiak & Triemer, 2001) allow the user to choose characters freely. Species not displaying these characters are progressively eliminated, converging towards a few possibilities with final identification based on images provided for each species. Automatic identification systems are usually based on resident computer software rather than web applications (Gaston & O'Neill, 2004). They rely on the same principle as that used for fingerprint or human face recognition. Images can be acquired by a scanning system such as the Zooscan for zooplankton (Grosjean et al., 2004). During the initial training stage, species are manually identified by an expert in order to create a training set thereafter used as a reference. In the case of the Zooscan, the sample is poured directly into a scanning cell and a digital image taken. Morphometric measurements are subsequently performed on this scan. Each specimen is then identified based on reference data and the dimensions of individual organisms tabulated.

Plankton taxonomy

Several types of websites are devoted to plankton taxonomy (Table 2). Some originate from individual research teams (e.g. Tsukuba University website) or cover a limited geographic region (e.g. Phytocom dedicated to Mediterranean plankton). Other websites focus on specific taxonomic or functional groups (e.g. Extant Coccolithophore Taxonomy or Harmful Plankton Project). Finally, some databases cover all plankton groups and beyond (e.g. Algaebase, Micro*scope or Plankton*Net). Only a few sites offer taxonomic keys for plankton (Table 2, e.g. Uzwiak & Triemer, 2001; Shayler & Siver, 2006).

Most of the focus with respect to automatic identification has been on imaging mesoplankton and marine snow (Culverhouse et al., 2006; Benfield et al., 2007). A system called DiCANN (Dinoflagellate Categorisation by Artificial Neural Network) has been developed to analyse characters such as surface texture and shape of this group of microalgae (Culverhouse et al., 1996). The AIMS (Automatic Identification of Microbial populationS) project relied on artificial neural networks (ANNs) to combine data from flow cytometry and image acquisition for identification of specific populations (Jonker et al., 2000; Gaston & O'Neill, 2004 for other examples of ANNs applied to automated species identification). An automatic microscope image acquisition, evaluation and recognition system has been developed for the analysis of Utermöhl plankton chambers (PLASA, PLAnkton Structure Analysis, Rodenacker et al., 2006). However, these approaches have generally not been adopted by the wider plankton research community. The aim of the recently launched project RAPID (Research on Automated Plankton Identification, Benfield et al., 2007) is to increase the use of automatic identification for zoo- or phytoplankton samples and is part of the program of the current Scientific Committee on Oceanic Research (SCOR) Working Group 130 (http://www.scor-int.org/Working_Groups/wg130.htm). In contrast, web-based applications offering automatic identification are scarce. Moreover, automatic identification systems still face issues like lack of taxonomic expertise preventing accurate training of the system, limits on scanning resolution, morphological variability (intraspecific variability, sexual heteromorphism), orientation of organisms (tilted specimens), and contamination by detritus.

*Plankton*Net: a portal for plankton taxonomy and observations*

Micro*scope (Table 2), launched in 2001, was the first website to try to cover the very wide taxonomic diversity of microbial groups. It contains images and descriptions of microbes arranged into collections. Micro*scope uses a taxonomically intelligent software written in PHP and linked to a MySQL database. Taxon names are organised within a unified classification called CU*STAR, covering all described organisms. CU*STAR communicates with the uBio NameBank which contains about 7 million records. Contributions to Micro*scope are made through a web interface. One key feature is that images are linked together through the underlying taxonomy. When visualizing a given image, a user can immediately access all available images for the same taxon and then navigate to related images at other taxonomical levels.

The Plankton*Net project focusing on plankton species was launched in 2004 based on the Micro*scope structure and received financial support in 2006 from the European Union allowing software development and data input.

Table 1. Selection of general web-based taxonomic resources. Websites rating (subjective) is based on richness, quality of information, update frequency, user-friendliness and template design.

Tableau 1. Sélection de ressources taxinomiques disponibles en ligne. Les sites web sont classés subjectivement selon la richesse, la qualité de l'information, la fréquence de mise à jour, la convivialité et la présentation.

Name	Web page	Description	Rating
PORTALS			
All Species Foundation	http://www.all-species.org/	Attempt to make an inventory of all species on Earth from 2000 to 2004	*
Barcode of Life Data Systems (BOLD)	http://www.barcodinglife.org/	Online workbench that aids collection, management, analysis, and use of DNA barcodes, consisting of 3 components: a repository for barcode records; a species identification tool; modules for bioinformaticians (Ratnasingham & Hebert, 2007)	****
Catalogue of Life	http://www.catalogueoflife.org/	Minimalist search engine based on species names. Systematic information on more than one million of species is compiled from Species 2000 and ITIS	****
Census of Marine Life (COML)	http://www.coml.org	Global network of researchers engaged in a scientific initiative to assess and explain the diversity, distribution, and abundance of life in the oceans (notably the project OBIS)	****
Encyclopedia of Life (EOL)	http://www.eol.org	Aims to create a single portal to authoritative and exhaustive web pages about all organisms: images, videos, references, descriptions, maps, classifications, etc.	****
ERMS (MARBEF)	http://www.marbef.org/data/erms.php	Authoritative taxonomic list of species occurring in the European marine environment	***
Global Biodiversity Information Facility (GBIF)	http://data.gbif.org/welcome.htm	International organisation that is working to make the world's biodiversity data accessible through a single portal. It collates information on species names from 220 data providers	****
ITIS	http://www.cbif.gc.ca/pls/itisca	The Integrated Taxonomic Information System is a taxon database offering a search engine on species names	***
MicrobeWiki	http://microbewiki.kenyon.edu/index.php/MicrobeWiki	Free wiki resource on microbes and microbiology, edited by students and monitored by microbiologists at Kenyon College. It refers to different domains: taxonomy, cell biology, genome structure, pathology	**
Species 2000	http://www.sp2000.org/	Goal is to create a validated checklist of all of the world's species (plants, animals, fungi and microbes)	***
The Taxonomicon	http://sn2000.taxonomy.nl/Taxonomicon	Biodiversity information system combining four major components: the Index of Life, the Tree of Life, the Web of Life, the Facts of Life. Classifications and references are the main assets	***
Tree of Life	http://tolweb.org/tree/phylogeny.html	Collection of information about biodiversity. Its goal is to provide one page with pictures, text, and other information for every species and for each group of organisms, living or extinct	**
uBio project	http://www.ubio.org/	Uses names and taxonomic intelligence to manage information about organisms	**

| Wikipedia | http://en.wikipedia.org/wiki/Main_Page | Multilingual, web-based, free content encyclopedia project. Wikipedia is written collaboratively by volunteers from all around the world. There are more than 75,000 active contributors working on more than 10,000,000 articles in more than 250 languages | *** |
| Wikispecies | http://species.wikimedia.org/wiki/Main_Page | Open, free directory of species. It covers Animalia, Plantae, Fungi, Bacteria, Archaea, Protista and all other forms of life | *** |

APPLICATIONS

Taxonomic keys

Linneaus II	http://www.eti.uva.nl/products/linnaeus.php	Multifunctional research tool for systematists and biodiversity researchers. It facilitates biodiversity documentation and species identification	
LucID	http://www.lucidcentral.org/	Software applications designed to help users with identification or diagnostic tasks	
X:ID	http://www.xidservices.com/	Allows authors to create their own "smart key" for the identification of any group of species, individuals, or objects, i.e., plants, animals, etc.	

Identification

DAISY	http://www.nhm.ac.uk/about-us/news/2005/august/news_6268.html	Digital Automated Identification Šystem - National History Museum of London (Gaston & O'Neill 2004). DAISY uses artificial intelligence and computer vision technologies to produce virtual collections of authoritatively identified specimens	
ABIS	http://www.informatik.uni-bonn.de/projects/ABIS/	The Automatic Bee Information System (Arbuckle et al., 2001) is designed to identify species of bees from images of their wings	
SPIDA-web	http://www.conbio.org/cip/article61WEB.cfm	The SPecies-IDentified-Automatically (Ness, 2005) uses artificial neural networks. It was first designed to identify Australian spiders and is now extended to other groups	
WEKA	http://www.cs.waikato.ac.nz/ml/weka/	The Waikato Environment for Knowledge Analysis is composed by a machine learning toolkit (artificial intelligence) associated with image analyser (ImageJ for example) (Mayo & Watson, 2007)	

Description

| DELTA | http://delta-intkey.com/ | Description Language for Taxonomy – Free program to record taxonomic descriptions for computer processing (generation and typesetting of descriptions and conventional keys, interactive identification and information retrieval, …). Online guides: Dallwitz, Paine & Zurcher 1993-2000 | |

Table 2. List of the web-based taxonomic resources for plankton. See Table 1 for rating criteria.
Tableau 2. Liste des ressources taxinomiques en ligne sur le plancton. Voir le Tableau 1 pour les critères de classement.

Name	Web page	Description	Rating
PORTALS			
Algaebase	http://www.algaebase.org/	Information on the world's algae that includes terrestrial, marine and freshwater organisms. It offers very exhaustive web pages on macro- and microalgae	*****
Algae Image Database	http://diatom.acnatsci.org/AlgaeImage/	Light micrograph images of diatom taxa from rivers throughout the USA	**
AlgaTerra	http://www.algaterra.net	Information system for micro algal biodiversity: a synthesis of taxonomic, molecular and ecological Information	**
Antarctic Freshwater Diatoms	http://huey.colorado.edu/diatoms/about/index.php	More than 500 pages of images and information about inland diatoms (mainly freshwater) from the McMurdo Dry Valleys region	***
Atlas of Dinoflagellates	http://dinos.anesc.u-tokyo.ac.jp/atlas_ver1_5/main.htm	Contains images of all the major bloom and toxin producing dinoflagellate species. Plates include images of whole cells, plate structure and cysts. Identification trees are also included	**
Copepods from the Bay of Villefranche	http://www.obs-vlfr.fr/%7Egaspari/copepods_guide/	A short description with pictures of each copepod species observed in the Bay. The ultimate goal of this site is to constitute a tool helping non-specialists to identify these organisms	***
Cyanobacterial Image Gallery	http://www-cyanosite.bio.purdue.edu/images/images.html	Simple list of images and a few videos on Cyanobacteria	*
Diatom Collection	http://www.calacademy.org/research/diatoms/names/index.asp	Relational database of diatom names and publications, compiled by staff at the California Academy of Sciences	****
East Sound Phytoplankton	http://thalassa.gso.uri.edu/ESphyto/index.html	Checklist of phytoplankton with a page for each taxon including photographs and notes on material from the San Juan Islands	**
Ehrenberg Collection	http://onlinedb.naturkundemuseum-berlin.de/v1/default.asp	The Ehrenberg Collection consists of several parts, including approximately 40,000 microscope preparations, 5,000 raw samples, 3,000 pencil and ink drawings, nearly 1,000 letters of correspondence, as well as index books and other documentation. Original drawings are downloadable	****
Extant Coccolithophore Taxonomy	http://www.nhm.ac.uk/hosted_sites/ina/CODENET/GuideImages/index.html	Images and classification to calcareous nanoplankton taxonomy	****
Friedrich Hustedt Study Centre For Diatoms	http://www.awi.de/en/research/research_divisions/biosciences/biological_oceanography/diatom_centre/	Diatoms database (material, samples and publications) hosted at the Alfred Wegener Institute (Bremerhaven)	***
Great Lakes Diatoms	http://www.umich.edu/%7Ephytolab/GreatLakesDiatomHomePage/top.html	Iconographic database on freshwater diatoms, University of Michigan	**
Harmful Plankton Project	http://www.liv.ac.uk/hab/intro.htm	Guide for identifying harmful marine dinoflagellates and diatoms	***

Index Nominum Algarum (INA)	http://ucjeps.berkeley.edu/INA.html	Card file containing bibliographic references pertaining to algal taxonomy (200,000 names of algae)	****
International Census of Marine Microbes	http://icomm.mbl.edu/	Cyberinfrastructure to index and organize what is known about microbes in oceans. Development of the Microbis database	****
Marine Planktonic Copepods	http://copepodes.obs-banyuls.fr/	One of the references on Copepods including taxonomic data and geographical distribution	***
Michael Martin's Phytoplankton Image Library	http://www.cedareden.com/phyto.html	Local database from Cedar Eden Environmental with mainly images arranged according to algal groups	**
Mediterranean Phytoplankton	http://193.205.231.4/botanica/data_en.htm	Taxonomic database dedicated to the Mediterranean Phytoplankton managed by the Stazione Zoologica from Naples	**
Micro*scope	http://starcentral.mbl.edu/microscope/portal.php	Communal website that promotes information on the biodiversity of microbes through an iconographic database on microbe taxonomy, supported by a powerful search engine	*****
Ostracod diversity	http://www.biology.missouristate.edu/ostracods/	Microphotographs of a wide variety of freshwater ostracods collected from North America	***
Phytocom	http://www.com.univ-mrs.fr/PHYTOCOM	Iconographic database devoted to Mediterranean phytoplankton with physico-chemical data and abundance of each species	**
Plankton*Net	http://planktonnet.awi.de/	European open-access iconographic and taxonomic databases on plankton. It offers images arranged in sets and associated to detailed information: description, classification, references, geolocalisation …	****
Planktonic Ciliate Project	http://www.liv.ac.uk/ciliate/intro.htm	Guide to identify marine coastal ciliates through exhaustive data sheets (PDF downloads)	***
Protist Image Data	http://megasun.bch.umontreal.ca/protists/protists.html	Pictures and short descriptions of selected protist genera, especially those genera whose species are frequently used as experimental organisms or are important in studies of organismal evolution	**
Protist Information Server	http://protist.i.hosei.ac.jp	Images and video clips of protists and other microorganisms as research and educational resources	***
Rotifers of Large Rivers	http://biology.missouristate.edu/rotifer	Photos from representative rotifer genera occurring in three large rivers in the central US (the Missouri, upper Mississippi, and Ohio rivers)	***
Skagerrak-Kattegat	http://www.smhi.se/oceanografi/oce_info_data/plankton_checklist/ssshome.htm	One of the first website dedicated to plankton (1996). Information on phytoplankton and heterotrophic protists found in the Skagerrak and the Kattegat, N. Atlantic Ocean	***
SilicaSecchiDisk	http://silicasecchidisk.conncoll.edu/Main_Menu_Frameset.html	Tools for scaled chrysophytes and diatoms, extensive limnological database and educational aids for the study of freshwater algae.	*
Tsukuba	http://www.biol.tsukuba.ac.jp/~inouye/ino/contents.html	(mainly in Japanese) Iconographic database on microalgae made by I. Inouye (College of Biological Sciences, University of Tsukuba)	**
Zooplankton project	http://www.cnas.missouristate.edu/zooplankton/Default.htm	Taxonomic records and photographs of freshwater zooplankton	**

APPLICATIONS

Taxonomic keys

Benthic Dinoflagellates	http://names.mbl.edu/xid/index.php	This website provides five taxonomic keys, including one on benthic dinoflagellates through X: ID software
Ceratium	http://idbio.unice.fr/idbio/ flore_marine/index.htm	This website provides an iconographic taxonomic key to identify some of the *Ceratium* species (Dinoflagellates, Peridiniales), characteristic from Mediterranean phytoplankton (Villefranche-sur-Mer bay)
Euglenoids	http://euglena.msu.edu/lucid3/ Photosynthetic_Euglenoids.html	Taxonomic key on Euglenoids is provided through Lucid software (Uzwiak & Triemer, 2001)
Freswater Diatoms of Britain and Ireland	http://craticula.ncl.ac.uk/EADiatom Key/html/	Multiaccess key on freshwater Diatoms through Lucid software and detailed information on taxonomy of Diatoms
Freshwater microalgae	http://silicasecchidisk.conncoll.edu/ LucidKeys/Keys.html	Lucid keys on freshwater algae and notably Synurophyceae and the genus *Mallomonas* (Shayler & Siver 2006)
Image-Based Key To The Zooplankton Of The Northeast (USA)	http://cfb.unh.edu/CFBkey/html/ index.html	User-friendly taxonomic key on zooplankton. Videos available

Identification

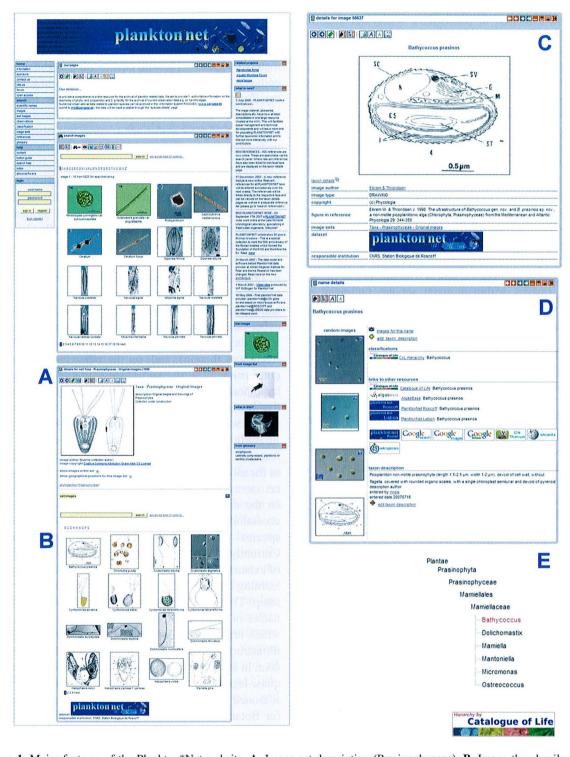|		
ADIAC Diatom Image Database	http://rbg-web2.rbge.org.uk/ ADIAC/db/adiacdb.htm	Automatic identification of diatoms using image information, i.e. both shape and ornamentation. The site also provides Diatom images
AIMS	http://scientiamarina.revistas.csic.es/ index.php/scientiamarina/article/ view/756	Automatic Identification and characterization of Microbial populationS. Integrating flow cytometric technology for identification of microbial cell populations and the determination of their cellular characteristics. (Jonker et al., 2000)
DIADIST	http://rbg-eb2.rbge.org.uk/DIADIST/ index.htm	Diatom and Desmid Identification by shape and texture. The aim of this project is to investigate methods to enable the visual indexing of images and drawings of biological specimens held in a database
DiCANN	http://www.dist.unige.it/dicann/	Dinoflagellate categorisation by artificial neural network. The website provides information on this tool and access to the image database
Plankton Identifier	http://www.obs-vlfr.fr/~gaspari/ Plankton_Identifier/index.php	*Automatic identification of objects (plankton and/or particles) from a set of images with their associated data. The software is freely downloadable on the website
PLASA	http://ibb.gsf.de/homepage/ burkhard.hense/Plasa/plasa.html	PLAnkton Structure Analysis (Rodenacker et al., 2006). Automatic archiving and analysis system composed of a fully automated image acquisition system using standard specimen holders (optical fixation from Utermöhl plankton chambers) and an also fully automated digital image analysis. The site provides a few information on PLASA system.

Figure 1. Major features of the Plankton*Net website. **A.** Image set description (Prasinophyceae). **B.** Image thumbnails. **C.** Image details (*Bathycoccus prasinos*). **D.** Taxon details. **E.** Taxon classification. In the image set panel (A), clicking on a thumbnail (B) (e.g. *B. prasinos*) opens up a new panel with more detailed information on this picture (C). Clicking on "taxon details" opens up a new panel (D) with information on *B. prasinos*, e.g. its taxonomical placement (E).

Figure 1. Caractéristiques principales du site Plankton*Net. **A.** Description de la collection (Prasinophyceae). **B.** Vignettes. **C.** Détails de l'image (*Bathycoccus prasinos*). **D.** Détails du taxon. **E.** Classification du taxon. Dans la fenêtre de la collection (A), cliquer sur une vignette (B) (e.g. *B. prasinos*) ouvre une nouvelle fenêtre avec une information plus détaillée sur cette image (C). Cliquer sur « taxon details » ouvre une nouvelle fenêtre (D) avec l'information concernant le taxon *B. prasinos*, par exemple sa classification (E).

Initially, a nodal structure had been implemented, each node accumulating data independently and communicating with the other nodes through XML protocols. During the initial phase of the project, four nodes were set up: Alfred Wegener Institute (AWI, Germany), Station Biologique de Roscoff (SBR, France), University of Lisboa (Portugal), Lake Kinneret Laboratory (Israël). However, coordination of databases between nodes proved difficult to implement and Plankton*Net moved in June 2008 to a centralized database and website located at the AWI in order to facilitate management and technical development.

Plankton*Net runs a content management system (CMS) called biodivDP (Biodiversity Data Provider) developed specifically for this project (Fig. 1). Like in Micro*scope, images are arranged in sets (Fig. 1A). Image sets can focus on a taxonomic group (e.g. coccolithophores), a location (e.g. the English Channel), a cruise or a facility (e.g. a culture collection). Images within a given set are presented as thumbnails (Fig. 1B). Clicking on a thumbnail opens a panel with more detailed information (Fig. 1C). Users can access information on the taxon itself such as its description (Fig. 1D), as well as other images available for the same taxon, its classification based on the Catalogue of Life (Table 1, Fig. 1E), and external resources (e.g. Google, Wikipedia). Registered users can contribute their own images and organize them in sets. They can also add references, original descriptions, definitions (glossary), and comment other images. A special section of the Aquatic Microbes Forum (http://aquatic_microbes.sb-roscoff.fr/) is dedicated to Plankton*Net where users can seek help and advice.

Plankton*Net contained in July 2008 5825 images grouped into 46 sets. Datasets contributed by AWI mainly contain phyto- and zooplankton images from the North Sea (e.g. Helgoland) or the Baltic Sea (e.g. IOW Phytoplankton monitoring). Datasets contributed by SBR are mainly dedicated to marine phytoplankton, notably picoplankton species from the Roscoff Culture Collection (Vaulot et al., 2004), local English Channel species, or species observed during cruises (e.g. the BIOSOPE cruise in the South East Pacific). Datasets assembled by Lisboa focus on harmful algae, notably dinoflagellates and their resting forms (cysts), while those from Israël are centred on freshwater species. Individual users have also contributed specific sets such as ciliate images from J. Dolan (Villefranche sur Mer, France).

Discussion

In recent years, genetic surveys of environmental samples have revealed many sequences that do not match any publicly available sequences from described species (for a recent review, see Epstein & López-García, 2008). A proportion of these sequences undoubtedly represent organisms that have been formally described, but for which genetic information has never been obtained. Obtaining sequence information for described species, using type material whenever possible or through expert designation of epitypes (currently permitted in the Botanical Code, but not the Zoological Code), should be a priority for molecular geneticists and taxonomists alike. However, a large proportion of these undefined sequences probably represent undescribed taxonomic entities. The recent development of metagenomics, i.e. the acquisition of large genome fragments directly from environmental samples (e.g. Béjà et al., 2000), and the advent of more powerful and cheaper sequencing technologies such as '454' pyrosequencing (Margulies et al., 2005) will undoubtedly lead to the discovery of even more hidden biodiversity in the near future.

In this context, an active and often animated exchange is ongoing in the scientific literature between, on the one hand, proponents of a stream-lined 'do-it-yourself' form of cyber-taxonomy with a stronger emphasis on genetic methods such as DNA barcoding (Godfray, 2007 and references therein), and, on the other hand, advocates of conserving 'traditional' systematics as a stand-alone hypothesis-driven science providing integrative knowledge that is essential to other areas of biology and conservation (de Carvalho et al., 2007 and references therein).

Traditionalists view the lack of investment in basic research, in collections, and in training of new taxonomists as the most important issues obstructing progress in modern taxonomy. For the pragmatist camp, the main issues centre on the need to accelerate the production of, and increase accessibility to taxonomic information in order to facilitate species identification for non-specialist end-users. Currently, describing new taxa or amending the description of existing taxa are complex and time-consuming tasks. The learning curve to the acquisition of expertise in taxonomy is steep. This stems in part from the unavoidably protracted nature of analytical methods such as electron microscopy, which are often essential for revealing morphological and ultrastructural characters of relevance for distinguishing taxa. In addition, the nomenclatural codes are presented in a quasi-legal style and are not readily accessible even to the seasoned taxonomist. For instance, the International Code for Botanical Nomenclature (ICBN, http://www.bgbm.org/iapt/nomenclature/code/SaintLouis/0000St.Luistitle.htm), designed to regulate the choice of taxon names, is composed of no less than 62 articles and 7 appendices. Among the elements of the codes that are increasingly viewed as outdated is the requirement to provide a latin translation of taxon diagnoses. There would appear to be considerable scope for simplifying and thus rendering more accessible the regulatory framework for taxonomy without impinging on the quality of taxonomic description.

A potential solution could be to reduce the emphasis on morphological comparison in taxonomy and even, by adopting the logic that a genetic sequence is a variable character that can be assigned a unique identifier (a name), to base taxonomy and identification of biodiversity solely on sequence data. While the taxonomy of prokaryotes, which are relatively devoid of morphological characters, is heading in this direction, the overwhelming majority of eukaryote biologists seem to concur that such a reductionist approach would significantly weaken the scientific foundation of organismal biology and in fact hinder rather than enhance efforts to understand ecosystem functioning and biological responses to global change. As more sophisticated technologies for morphological, genetic, biochemical, and physiological characterisation are becoming accessible, it would make sense to include a larger number and a greater diversity of characters in taxonomy. In fact, many of the web projects reviewed above are already pushing in this direction by using the integrating power of technologies such as XML to draw together dispersed resources such as images, genetic data, and taxonomic descriptions.

Taxonomists already benefit greatly from the increasing on-line availability of information. The legacy of more than 200 years of taxonomic literature underpins modern taxonomy, but identifying and obtaining relevant scientific literature and type material has traditionally been a difficult and time-consuming process. Many taxa have changed names, sometimes several times, and it is often difficult to establish which name is the currently accepted one. As these changes often occur without widespread diffusion of the information, data become frequently outdated or unreliable (Wheeler et al., 2004). The Internet is changing this through widespread access to current scientific journals, together with large initiatives such as the Biodiversity Heritage Library (http://www.biodiversity-library.org/) which aims to digitize and provide on-line access to over 2 million volumes of biodiversity literature going back 200 years, and smaller more focussed initiatives such as the *AlgaTerra* Information System (http://www.algaterra.org) which will provide scans of original documents from the collection of Ehrenberg (1795-1876), one of the founders of microbiology and micropalaeontology.

Future directions

Taxonomy can gain much from the development of the Internet, but in our opinion rapid progress is needed in some key directions:

- **Online species description**. Taxonomists should move towards the use of open-source web-based systems to report the description of new species and revision of existing taxonomy. Publication in peer-reviewed journals could be replaced by publication on permanent websites. The peer-review system must be retained, but moving to a web system would offer many advantages, in particular in terms of reducing the turnover time for publication and increasing accessibility.

- **Image depository**. A GenBank-like system could be developed for the deposit of taxon diagnoses and images linked to taxonomic descriptions. A formal requirement for this in the international nomenclatural codes would probably be needed to establish this as a standard protocol, as well as long-term national or international funding for the actual web-based database system.

- **Unitary project.** Taxonomists have historically found it difficult to forge united initiatives. However, web-based biodiversity guides need unitary and interoperable projects (Godfray, 2002) as much as they need consensus on taxonomic names and classification (Chavan et al., 2005; Patterson et al., 2006). Existing web-based taxonomy initatives should try to converge to a few high profile projects that would attract the mass of taxonomists. EOL could be one good candidate because of its ambitious goals, secure funding, and its integrating power.

Much has changed in the last ten years since the appearance of the first websites dedicated to plankton listing, taxonomy, and identification. It is highly likely that the pace of change will continue to increase and that, in less than five years, the web will become the primary information source for taxonomy in this area. This will benefit not only dedicated plankton taxonomists, but will also be extremely useful in applied contexts, e.g. in harmful algae monitoring, and will largely facilitate research for our colleagues from developing countries who until now have had tremendous difficulties to access reference books and primary literature.

Acknowledgements

References

Arbuckle T., Schröder S., Steinhage V. & Wittmann D. 2001. Biodiversity informatics in action: identification and monitoring of bee species using ABIS. *15th International*

Symposium Informatics for Environmental Protection, Zürich. 425-430.

Armsworth P.R., Kendall B.E. & Davis F.W. 2004. An introduction to biodiversity concepts for environmental economists. *Resource and Energy Economics*, **26**: 115-136.

Bax N., Carlton J.T., Mathews-Amos A., Haedrich R.L., Howarth F.G., Purcell J.E., Rieser A. & Gray A. 2001. The control of biological invasions in the world's oceans. *Conservation Biology*, **15**: 1234-1246.

Béjà O., Aravind L., Koonin E.V. & Suzuki M.T. 2000. Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science*, **289**: 1902-1906.

Benfield M., Grosjean P., Culverhouse P.F., Irigoien X., Sieracki M.E., Lopez-Urrutia A., Dam H., Hu Q., Davis C.S., Hansen A., Pilskaln C., Riseman E.M., Schultz H., Utgoff P.E. & Gorsky G. 2007. RAPID Research on automated plankton identification. *Oceanography*, **20**: 172-187.

Blockstein D.E. 1998. Lyme disease and the passenger pigeon? *Science*, **279**: 1831.

Chavan V., Rane N., Watve A. & Ruggiero M. 2005. Resolving taxonomic discrepancies: role of electronic catalogues of known organisms. *Biodiversity Informatics*, **2**: 70-78.

Culverhouse P.F., Simpson R.G., Ellis R.E., Lindley J.A., Williams R., Parisini T., Reguera B., Bravo I., Zoppoli R., Earnshaw G., McCall H. & Smith G. 1996. Automatic classification of field-collected dinoflagellates by artificial neural network. *Marine Ecology Progress Series*, **139**: 281-287.

Culverhouse P.F., Williams R., Benfield M., Flood P.R. & Sell A.F. 2006. Automatic image analysis of plankton: future perspectives. *Marine Ecology Progress Series*, **312**: 297-309.

de Carvalho M.R., Bockmann F.A., Amorim D.S., Brandao C.R.F., de Vivo M., de Figueirido J.L., Britski H.A., de Pinna M.C.C., Menezes N.A., Marques F.P.L., Papavero N., Cancello E.M., Crisci J.V., McEachran J.D., Schelly R.C., Lundberg J.G., Gill A.C., Britz R., Wheeler Q.D., Stiassny M.L. J., Parenti L.R., Page L.M., Wheeler W.C., Faivovich J., Vari R.P., Grande L., Humphries C.J., DeSalle R., Ebach M.C. & Nelson G.J. 2007. Taxonomic impediment or impediment to taxonomy? A commentary on systematics and the cybertaxonomic-automation paradigm. *Evolutionnary Biology*, **34**: 140-143.

Edwards M. & Morse D.R. 1995. The potential for computer-aided identification in biodiversity research. *Trends in Ecology and Evolution*, **10**: 153-158.

Epstein S. & López-García P. 2008. "Missing" protists: a molecular prospective. *Biodiversity and Conservation*, **17**: 261-276.

Gaston K.J. & O'Neill M.A. 2004. Automated species identification: why not? *Philosophical Transactions of the Royal Society of London B*, **359**: 655-667.

Giovannoni S.J., Britschgi T.B., Moyer C.L. & Field K.G. 1990. Genetic diversity in Sargasso Sea bacterioplankton. *Nature*, **345**: 60-63.

Godfray H.C. J. 2002. Challenges for taxonomy. *Nature*, **417**: 17-19.

Godfray H.C. J. 2007. Linnaeus in the information age. *Nature*, **446**: 259-260.

Grosjean P., Picheral M., Warembourg C. & Gorsky G. 2004. Enumeration, measurement, and identification of net zooplankton samples using the ZOOSCAN digital imaging system. *Journal of Marine Science*, **61**: 518-525.

Heal G. 2004. Economics of biodiversity: an introduction. *Resource and Energy Economics*, **26**: 105-114.

Jonker R., Groben R., Tarran G., Medlin L.K., Wilkins M., Garcia L., Zabala L. & Boddy L. 2000. Automated identification and characterization of microbial populations using flow cytometry: the AIMS project. *Scientia Marina*, **64**: 225-234.

Lilburn T.G., Harrison S.H., Cole J.R. & Garrity G.M. 2006. Computational aspects of systematic biology. *Briefings in Bioinformatics Advance*, **7**: 186-195.

Margulies M., Egholm M., Altman W.E., Attiya S., Bader J.S., Bemben L.A., Berka J., Braverman M.S., Chen Y.-J., Chen Z., Dewell S.B., Du L., Fierro J.M., Gomes X.V., Goodwin B.C., He W., Helgesen S., Heen Ho C., Irzyk G.P., Jando S.C., Alenquer M.L.I., Jarvie T.P., Jirage K.B., Kim J.-B., Knight J.R., Lanza J.R., Leamon J.H., Lefkowitz S.M., Lei M., Li J., Lohman K.L., Lu H., Makhijani V.B., McDade K.E., McKenna M.P., Myers E.W., Nickerson E., Nobile J.R., Plant R., Puc B.P., Ronan M.T., Roth G.T., Sarkis G.J., Simons J.F., Simpson J.W., Srinivasan M., Tartaro K.R., Tomasz A., Vogt K.A., Volkmer G.A., Wang S.H., Wang Y., Weiner M.P., Yu P., Begley R.F. & Rothberg J.M. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, **437**: 376-380.

Mayo M. & Watson A. T. 2007. Automatic species identification of live moths. *AI 2006, The 26th SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence*, **20**: 195-202.

Moon-van der Staay S.Y., De Wachter R. & Vaulot D. 2001. Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature*, **409**: 607-610.

Naeem S. & Li S. 1997. Biodiversity enhances ecosystem reliability. *Nature*, **390**: 507-509.

Ness E. 2005. SPIDA-web. Artificial neural networks dill in for taxonomists. *Conservation in Practice*, **6**: 35-36.

Patterson D.J., Remsen D., Marino W.A. & Norton C. 2006. Taxonomic indexing - Extending the role of taxonomy. *Systematics Biology*, **55**: 367-373.

Ratnasingham S. & Hebert P.D.N. 2007. BOLD: the barcode of life data system (www.barcodinglife.org). *Molecular Ecology Notes*, **7**: 355-364.

Raven P.H. & Wilson E.O. 1992. A fifty-year plan for biodiversity surveys. *Science*, **258**: 1099-1100.

Rodenacker K., Hense B., Jütting U. & Gais P. 2006. Automatic analysis of aqueous specimens for phytoplankton structure recognition and population estimation. *Microscopy Research and Technique*, **69**: 708-720.

Shayler H.A. & Siver P.A. 2006. Key to Freshwater Algae : a web-based tool to enhance understanding of microscopic biodiversity. *Journal of Science Education and Technology*, **15**: 298-303.

Tilman D., Knops J., Wedin D., Reich P., Ritchie M. & Siemann E. 1997. The influence of functional diversity and composition on ecosystem processes. *Science*, **277**: 1300-1302.

Uzwiak D.J. & Triemer R.E. 2001. A web-based digital key to

euglenoids. *Journal of Phycology*, **37**: 49-50.

Vaulot D., Le Gall F., Marie D., Guillou L. & Partensky F. **2004.** The Roscoff Culture Collection (RCC): a collection dedicated to marine picoplankton. *Nova Hedwigia*, **79**: 49-70.

Wheeler Q.D., Raven P.H. & Wilson E.O. **2004.** Taxonomy: impediment or expedient. *Science*, **303**: 285.

Wilson E.O. **2003.** The encyclopedia of life. *Trends in Ecology and Evolution*, **18**: 77-80.