Chambers, Susan (2012) Why does taxonomy take so long? Porcupine Marine Natural History Society Newsletter (32). pp. 16-21.

Why does taxonomy take so long?

Susan Chambers

National Museums Scotland, EH1 1JF, UK
s.chambers@nms.ac.uk

The aim here is to outline the taxonomic task when applied to the large numbers of living organisms and to describe some of the background steps required to identify and describe new taxa. Also, to explain why taxonomy takes so long but is not dull and only pursued by the old working in museums (Guerra-Garcia *et al*. 2008).

What is taxonomy?

Taxonomy is an artificial mechanism to identify, name and classify the living world in order to help us understand the complexity a little better. These are three separate processes and all require a different skill. The edges of these processes are fuzzy and overlap with phylogeny which is a classification based on evolution and molecular data. Both taxonomy and phylogeny are embraced by systematics which brings together the knowledge of the organisms and their relationships over time.

The scale of the problem? Or how many taxa?

Another way to make sense of the world is to record, assign a number and analyse the data. The numerical diversity of the biological world is still unknown. The Convention on Biodiversity, Rio Summit 1992, could not answer questions about the number of species and which ecosystems needed to be conserved. Politicians and policy makers need numbers to make decisions and informed choices about sustainable development. This task is particularly daunting for marine biologists as so much funding and research is concentrated on the terrestrial world. The Census of Marine Life took place between 2000-2010 to answer the question of the number of marine species and abundance of marine life. The Census was world wide and looked at all groups of animals and plants (Costello *et al*. 2010: www.plosone.org/article/info%3Adoi/10.1371/journal.pone.0012110).

The global estimate for eukaryotes is between 3 and 100 million taxa (May 2010). A recent application of statistics narrowed the number to approximately 8.7 million, of which 2.2 million are marine. Since Linnaeus' classification system was published 250 years ago, 1.2 million species have been catalogued, which breaks down as 14% terrestrial species and 9% of marine species. This means that 91% of marine species are still waiting to be described. At the current rate of 15,000 newly discovered species per year it will take 480 years to complete the task of identifying and naming all known taxa (Mora, *et al*. 2011).

Identification

Before you can name something you need to separate the specimen from the many others you see. Over the years many identification guides have been published but there are very few for marine life compared to guides for example, garden birds, or flowering plants. There are two main obstacles to the publication of marine guides:

- Knowledge of the fauna or flora
- How to write descriptions.

British Polychaetes are a good illustrative example and will be used here. Polychaetes are one of the most abundant groups in the benthic environment and there were 1,397 recorded in the Marine Directory of the British shallow water fauna (Howson & Picton 1997). My current estimate is nearer 2,000.

Knowledge of the British polychaete fauna: How do you start?

Literature: The first obstacle is that there is no single comprehensive polychaete publication for the British marine fauna. The standard texts for the last 100 years have included the British Annelid Monographs (McIntosh 1900, 1908, 1910, 1911, 1915 & 1923), Clare Island Survey (Southern 1914), Polychètes Errantes and Polychètes Sedentaires (Fauvel 1923, 1927) and Polychaeta, Tierwelt

Deutschlands (Hartmann-Schröder 1971, 1996) which are useful for some geographic areas of the British fauna. The drawback of these publications is that they are difficult to access for a variety of reasons including price, rarity and language. They are often poorly illustrated and difficult to interpret, so they demand a lot of time and skill. The recent series of the Linnaean Society Synopses of the British Fauna includes some polychaete families, but less than 10% of the total. (George & Hartmann-Schröder 1985; Pleijel & Dales 1991; Chambers & Muir 1997). The Marine Fauna of the British Isles and North West Europe (Hayward & Ryland 1990) has a polychaete chapter which only includes common species.

Electronic keys: The National Marine Biological Analytical Quality Control (NMBAQC) scheme has published some keys electronically e.g. http://www.nmbaqcs.org/scheme-components/invertebrates/literature-and-taxonomic-keys.aspx. The Natural History Museum also publishes an electronic guide to polychaetes www.nhm.ac.uk/research-curation/research; the Marine Life Information Network www.marlin.ac.uk/phylumdetails.php?phylum=2448 describes some more common species and the environmental consultancy Thomson Unicomarine produces identification guides to some polychaete families.

These electronic guides are useful and interesting but again they are not always comprehensive or easy to use. To identify hundreds of samples you have to acquire a vast variety of reprints from various journals, monographs and the web which takes time to accumulate, requires knowledge of and access to the literature and becomes a major task in itself.

Specimens: To understand the range of characters, where one ends and another begins, you need to examine hundreds of specimens in good condition. Obtaining specimens in good condition is a challenge as collections of soft-bodied animals are usually obtained under pressure of the incoming tide or on research vessels where time is precious and cost is the main driver rather than the quality of the samples. Living material is the best way to see the characters in perfect condition but this is not always practical. Skill is required to prepare animals for narcotisation and preservation to ensure that essential characters are retained (Smaldon & Lee 1979; Mackie 1994; Pleijel & Rouse 2001).

A specimen without data is of no value to taxonomic research. Museum specimens are acquired from a range of sources, e.g. government agencies, commercial consultancies, government-funded environmental monitoring programmes such as the Strategic Environmental Assessment programme, Marine Laboratories such as Scottish Association for Marine Science and personal field-work. Specimens donated to museums are organised by the donor's research objective, e.g. monitoring sites, geographic location or chronological sequences such as yearly time series. The physical data needs to be matched to individual specimens before they are incorporated into the collection in systematic order. This is a time consuming task and is fundamental to the role of a curator (Chambers 2001; Mackie 2001).

As well as museums, many environmental consultancies, e.g. Unicomarine, Fugro and individual consultants, have established reference collections which are essential for their geographic area of work. A good example of why it is important to keep a reference collection is the cirratulid polychaete *Chaetozone setosa,* which was considered a common species for about 100 years with a world-wide distribution from the intertidal to the deep sea. In the 1980s a survey was completed that used multivariate analysis to examine the population dynamics of *C. setosa* in relation to changing organic enrichment (Hily 1987). It is unlikely that the species for this interpretation was *C. setosa* as we now know it does not occur in the area where the samples were collected. *'Chaetozone setosa'* is a complex of species and includes 2 intertidal species and 3 subtidal shallow water species around the UK (Chambers 2000). It is not known how many taxa were analysed for the population dynamics of the Bay of Brest as no samples were retained for future taxonomic work. If the samples had been deposited in a museum or university collection they could have been re-examined and checked to confirm their identification.

Good quality, data-rich specimens are invaluable for producing text and illustrations which become the foundations of taxonomy. All British National Museums have a statutory requirement to lend material to researchers all around the world to assist in systematic research. Once you have acquired good quality specimens and associated data, the next step is to begin the description.

How to write a new species description or re-description.

1. A group of animals is separated as distinct from known specimens. Ideally specimens should be from more than one locality and collected at different times of year.
2. A thorough search of the literature is made to check whether or not the entity has been described before. If it has been inadequately described then a re-description may be necessary.
3. Specimens from other surveys or type material from other museums are borrowed for comparison.

4. Most specimens need to be dissected and prepared for optical microscope slides or SEM so more than one specimen is required.

5. The morphological characters are described, and illustrated with line drawings, and photographs; internal structures and any morphological measurements are presented as a graphic, e.g. body length to segment ratio.

6. Other information where appropriate is also valuable to add to the data set, e.g. habitat, tubes/burrows, colour, smell, associated parasites/hosts, reproductive stages.

7. If the description is for a new species one specimen is selected as the Holotype (the first described specimen of that species) and assigned a name within a genus, e.g. *Chaetozone christiei*. The Holotype will always be associated with that name and cannot change. The specimen is only considered valid if it has the associated data attached, e.g. Northumberland coast, low shore, Low Newton-by-the Sea, 55^0 32'N 0.1^0 36'W, clean sand. It is good practice if this specimen is given a unique number from a museum, e.g. NMSZ.1988.122.

8. Once accepted for publication the new species is then known as, *Chaetozone christiei* Chambers, 2000. The name, which includes the generic name, specific name, author and date are forever associated with the Holotype specimen.

9. Only after publication in a peer-reviewed journal and available in multiple copies is the name **valid** and **available** to the scientific community.

10. The name, description, figures, etc are the basis for comparison of future identifications.

It is good practice to donate type and non-type material to a museum and cite the location in the publication to avoid future confusion. If molecular data is available this can be included with the type material information. Publications often include new records or samples from locations which are expensive to re-visit, e.g. deep water sites, and these specimens are of great interest for zoogeographic information when trying to compose a guide. A large amount of valuable time is lost during taxonomic work by looking for type and non- type material that has not been donated to an institution with a good record of maintenance. It is not uncommon for a long time-series of research samples to be left under the marine biologist's desk and thrown away when they retire or leave this world. This leads to all kinds of problems including lost type specimens, which may then require a lectotype or neotype to be selected and assigned to the name. The change in status of the type material needs to be published and often requires the application of the International Code of Zoological Nomenclature rules (ICZN 1997). This is a separate process and leads to unnecessary delay of the original publication. There are a few ICZN guidance notes for publication of a new species.

- the description is published in a work that is obtainable in numerous identical copies, as a permanent scientific record.
- the scientific name must be spelt using the 26 letters of the Latin alphabet; binominal nomenclature must be consistently used; and new names must be used as valid when proposed.
- that names are consistently formed following certain rules; that original spellings can be established.
- that names are based on name-bearing types, the objective standard of reference for the application of zoological names.
- that general recommendations are followed for ethical behaviour.
- and that best practice should be used to give taxa names which are unique, unambiguous and universal.

Molecular data

More and more phylogenies are based on DNA sequence data; these are especially enticing for their potential to be automated and speed up the identification process. Eventually, this may change our understanding of evolutionary biology. Meanwhile a lack of taxonomic progress will not be solved by DNA systems alone, largely owing to costs and difficulty of practical applications, especially is less-developed regions of the world. We need both morphological and molecular techniques to construct phylogenies as DNA cannot be extracted from palaeontological material or rare specimens. There are too many mis-identifications in the invertebrate literature to produce clear results from sequencing. The vast majority of taxa has not been sequenced so adding another requirement to descriptions will slow the process down even more (Mallet & Wilmott 2003). Molecular techniques provide another set of information alongside the ecology, behaviour and reproductive strategies (Misof *et al*. 2005). At the moment the two systems are running in parallel; hopefully they will merge in the future.

Classification

One of the defining features of humans over millennia has been to find order and name it! This phenomenon has been found in many scientific disciplines, e.g. Physics: Newton's law of mechanics, Chemistry: Dalton's theory that matter is made of atoms and Mendeleyev arranged elements by atomic weight. Nature can be ordered as well, but the basic unit is harder to find. Aristotle began a classification of the living world approximately 300 BC, and later 18th and 19th century biologists such as Jussieu, Cuvier, Lamark, Haeckel (1866, introducing phylogeny), Banks and Darwin all looked for an underlying order. Most of these systems work on the principal of moving from the general to the specific and have been developed over centuries. The biological world uses a hierarchical system, e.g. *Chaetozone gibber* is classified as follows:

Domain: Eukaryota

Kingdom: Animalia

Phylum: Annelida

Class: Polychaeta

Order: Canalipalpata

Family: Cirratulidae

Genus: *Chaetozone*

Species: *gibber*

Biological classification itself has evolved and can be usefully divided into four phases (Tudge 2000).

Ancient: Aristotle (384 – 300 BC) demonstrated the need to choose characters carefully as some features gave unsatisfactory results. For example animals with two legs grouped birds and humans together whereas characters such as oviparous and viviparous were more helpful.

Classical: The 16th, 17th and 18th centuries were dominated by the practical needs of commercial policies, e.g. timber trade, plants for pharmacy. In 1758, Linnaeus combined a hierarchy from a kingdom to species with a bi-nomial method of naming. This was very easy to use and reduced the need to re-state the characters.

Immediate post Darwin: Darwin proposed a dynamic process of natural selection leading to the theory of evolution. It is a mechanism to describe evolution but it has fundamentally changed the way the world thought about biology. This had immense consequences for classification.

Cladistics: Hennig (1966) introduced rules to distinguish between primitive and derived characters. This led to an entire new philosophy of classification based on natural relationships which reflected an evolutionary history.

Nomenclature

Linnaeus, in 1758, published the 10th edition of his hierarchical classification system of plants and animals and gave them all two names, a binomial system (Linnaeus 1758). This is considered as the starting point for current biological classification. It led to the need for objective rules and in 1895 a committee was formed to produce guidance on the rules for zoological names. This is known as the International Code of Zoological Nomenclature. The aim of ICZN is to regulate the application of zoological names to ensure each name is unique and universally available. The code is international and has evolved over more than 100 years. An electronic version of the 4th edition of the code is available: www.iczn.org. One of the main issues for the committee is the need to revise the plans for the 5th edition of the code to include electronic publication and registration of names. Consequently an interactive discussion forum has been established to enable a wide involvement of all concerned www.iczn.ansp.org/wiki.

Hennig's development of phylogeny and the application of cladistics has led to the proposal for a new system of classification. It is based on clades which do not recognise the standard naming of ranks. This new system would require a new set of rules or code, to be known as a Phylocode, The adoption of this proposal has not yet received universal acceptance. (www.ohiou.edu/phylocode).

Catalogues

As the classification of zoological specimens increased in size and complexity after 1758, cataloguing this information became the obvious next step. Over the next 250 years there have been numerous catalogues published to suit particular commercial and aesthetic needs, e.g. types of timber, agricultural pests, toxic organisims. In response to the Convention of Biodiversity held in Rio, 1992, the Global Taxonomic Initiative was launched to improve data capture for conservation policies and planning. There were specific European initiatives such as Fauna Europea, and European Register of Marine Species which have been subsequently linked to the Species 2000 framework. At about the same time computers became widely available and software to produce bigger catalogues was developed eg. (www.sp2000.org; www.gbif.org; www.eol.org). There are numerous ways to catalogue by species in systematic order, alphabetical, geographical area etc. There are also combinations of some of these factors, e.g. the Marine Directory of British fauna and flora (Howson & Picton 1997), Fish catalogue (www.fishbase.org), Checklist of European Marine Mollusca (http://www.somali.asso.fr/clemam/index.clemam.html).

Recent developments include catalogues of DNA sequences commonly known as bar-codes. DNA bar-coding uses a specific mitochondrial genome (CO1) to provide a single fingerprint tag (Marshall 2005).

The sequences are registered in a gene bank and there are three main International GenBanks, the European Molecular Biology Laboratory, the Data bank of Japan and the USA International Nucleotide Sequence Database which all have publically available DNA sequences. Each GenBank includes concise descriptions of the sequence, scientific name and taxonomy and other sites of biological significance. The Barcode of Life Database (BOLD) manages the use of DNA barcodes and best practice includes vouchered specimens. DNA barcodes have a standardised method for non-experts to identify species using a DNA sequence.

Electronic developments

Web-based applications have allowed the taxonomic community to share and access data in imaginative and various ways; this is known as Biodiversity Informatics. A new name proposed by Quentin Wheeler is Cybertaxonomy (www.v.smith.info/cybertaxonomy.) which he defined as a fusion of taxonomy, computer science and engineering. Electronic tools have helped to create, store and share large amounts of data to produce electronic descriptions and guides, e.g. the National Museum of Wales produced a British Bivalve web-based guide; http://naturalhistory.museumwales.ac.uk/britishbivalves. There is also the development of *Scratchpads* which had been funded by ViBRANT, an EU project. The aim was to increase collaboration in an electronic framework and accelerate the pace of biodiversity research. http://scratchpads.eu/.

Zookeys (www.pensoft.net/journals/zookeys) is a peer-reviewed, open-access, rapidly disseminated journal launched to accelerate research and free information exchange in taxonomy, phylogeny, biogeography and evolution of animals. Zookeys will publish and give priority to manuscripts with large keys, new descriptions and identifications which many standard journals find a challenge (Smith & Penev 2011).

The future for taxonomy?

Acquiring identification skills takes time and patience and is often a lonely occupation as there are fewer and fewer people to pass on their knowledge. Most people start their identification career from a very low knowledge level and can only proceed slowly due to lack of basic literature and skill transfer. Also, informatics is not a substitute for science and this includes taxonomy (Knapp *et al*. 2002). The Census of Marine Life summary (Costello *et al*. 2010) found that there was a positive relationship between availability of taxonomic guides and knowledge of biodiversity. More than 80% of phyla are found in the sea which is a good reason for taxonomists to turn their attention from the land (May 1992). Species are complex but taxonomy is a mature and stimulating science. It is a dynamic process and by no means static, dull or only for the elderly.

References

Chambers, S. J. 2000. A redescription of *Chaetozone setosa* Malmgren, 1867 including a definition of the genus, and a description of a new species of *Chaetozone* (Polchaeta: Cirratulidae) from the Northeast Atlantic. *Bulletin of Marine Science* **67**: 587-596.

Chambers, S. J. 2001. The Atlantic Frontier Environmental Network surveys – A good example of how to develop sample collections. *In:* Rothwell, R.G. (ed.) pp. 20-21 *Marine sample collections: their value, use and future*. IACMST Information Document No. 8.

Chambers, S. J. & Muir, A. I. 1997. Polychaetes: British Chrysopetaloidea, Pisionoidea and Aphroditoidea. *Synopses of the British Fauna (New Series) No 54* . Linnaean Society of London & The Estuarine and Coastal Sciences Association, London 1-202.

Costello, M. J., & Coll, M., Danovaro, R., Halpin,P., Ojaveer, H., Miloslavich, P. 2010. A census of marine biodiversity, knowledge, resources and future challenges. *PLoS*. 5(8) e 12110 (http://www.plosone.org/article/info%3Adoi/10.1371/journal.pone.0012110).

Fauvel, P. 1923. Polychètes errantes. *Faune de France*. **5**: 1-488.

Fauvel, P. 1927. Polychètes sedentaires. *Faune de France*. **16**: 1-494.

George, J. D. & Hartmann-Schröder, G.1985. Polychaetes: British Amphinomida, Spintherida and Eunicida. *Synopses of the British Fauna (New Series) No 32* . Linnaean Society of London & The Estuarine and Coastal Sciences Association, London 1-221.

Guerra-Garcia, J. M., Espinosa, F. & Garcia-Gomez, J. C. 2008. Trends in taxonomy today: an overview about the main topics in taxonomy. *Zoologica Baetica* **19**: 15-49.

Hartmann-Schröder, G. 1971. Annelida, Borstenwürmer, Polychaeta. *Tierwelt Deutschland*. **58**: 1-594.

Hartmann-Schröder, G. 1996. Annelida, Borstenwürmer, Polychaeta.2 neubearbeitete Auflage. *Tierwelt Deutschland*. **58**: 1-648.

Hayward, P.J. & Ryland, J.S. 1990. Volume 1. Introduction and Protozoans to Arthropods. *The Marine Fauna of the British Isles and North West Europe*. Clarendon Press, Oxford 1-627.

Hily, C. 1987. Spatio-temporal variability of *Chaetozone setosa* Malmgren populations on an organic gradient in the Bay of Brest, France. *Journal of Experimental Marine Biology and Ecology* **112**: 201-216

Hennig, W. 1966. *Phylogenetic Systematics*. University of Illinois Press, Urbana.

Howson, C. M. & Picton, B. E. (eds). 1997. *The Species Directory of the Marine Fauna and Flora of the British Isles and Surrounding Seas*. Ulster Museum and The Marine Conservation Society, Belfast and Ross-on-Wye.

International Commission on Zoological Nomenclature, 1999. *International code of zoological nomenclature. Forth edition*. International Trust for Zoological Nomenclature, London.

Knapp, S., Bateman, R. M., Chalmers, N. R., Humphries, C. J., Rainbow, P. S., Smith, A. B., Taylor, P. D., Vane-Wright, R. I. & Wilkinson, M. 2002. Taxonomy needs evolution, not revolution. *Nature* **419**:559

Linnaeus, C. 1758. Systema naturae per regna tria naturae: secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis. 10th edition. Laurentii Salvii, Holmiae.

Mackie, A. S. Y. 1994. Collecting and preserving polychaetes. *Polychaete Research* **16**: 7-10.

Mackie, A. S. Y. 2001. Marine invertebrate collections in the National Museum of Wales. *In:* Rothwell, R.G. (ed.) pp. 24-25 *Marine sample collections: their value, use and future*. IACMST Information Document No. 8.

Mallet, J. & Wilmott, K. 2003. Taxonomy: renaissance or Tower of Babel? *Trends in Ecology and Evolution* **18**: 57-59

Marshall, E. 2005. Will DNA bar codes breathe life into classification? *Science* **307**: 1037.

May, R. 1992. Bottoms up for the oceans. *Nature* **357**: 278-279.

May, R. 2010. Tropical Arthropod species, more or less? *Science* **329**: 41-42

McIntosh, W. C. 1900. *A Monograph of British Annelids*. Vol. 1, Part 2*: Polychaeta Amphinomidae to Sigalionidae.* Ray Society, London.

McIntosh, W. C. 1908. *A Monograph of British Annelids*. Vol. 2, Part 1: *Polychaeta Nephtydidae to Syllidae.* Ray Society, London.

McIntosh, W. C. 1910. *A Monograph of British Annelids*. Vol.2 Part 2: *Polychaeta Syllidae to Ariciidae*. Ray Society, London.

McIntosh, W. C. 1911. Notes from the Gatty Marine Laboratory, St Andrews. No. XXXII . *Annals and Magazine of Natural History* **7**: 145–173.

McIntosh, W. C. 1915. *A Monograph of British Annelids*. Vol. 3 Part 1:*Polychaeta, Ophelidae to Ammocharidae.* Ray Society, London.

McIntosh, W. C. 1923. *A Monograph of British Annelids*. Vol. 4 Part 2: *Polychaeta, Sabellidae to Serpulidae.* Ray Society, London.

Misof, B., Klütsch, C. F. C., Niehuis, O., & Patt, A. 2005. Of phenotypes and genotypes: two sides of one coin in taxonomy? *Bonner zoologische Beiträge* **53**: 121-133.

Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G. B. & Worm, B. 2011. How many species are there on Earth and in the ocean? *PLoS Biol* **9**: e101127.

Pleijel, F. & Dales, R.P. 1991. Polychaetes: British Phyllocoideans, Typhloscolecoideans. *Synopses of the British Fauna (New Series) No 45*. Linnaean Society of London & The Estuarine and Coastal Sciences Association, London 1-202.

Rouse, G. W. & Pleijel, F. 2001. *Polychaetes*. Oxford University Press, UK, pp. 354.

Smaldon, G. & Lee, E. W. 1979. A synopsis of methods for the narcotisation of marine invertebrates. *Information Series Natural History Royal Scottish Museum* **6**: 1-96.

Smith, V & Penev, L. (eds) 2011. e-Infrastructures for data publishing in biodiversity science**.** *ZooKeys* **150:** 1–417**.**

Southern, R. 1914. Clare Island Survey. Part 47. Archiannelida and Polychaeta. *Proceedings of the Royal Irish Academy* **31**: 1-160

Tudge, C. 2000. *The Variety of Life*. Oxford University Press, UK. pp.684