Сборник трудов Зоологического музея МГУ им. М.В. Ломоносова
Archives of Zoological Museum of Lomonosov Moscow State University
Том / Vol. 54 Стр. / Pp. 116–133

INTEGRATIVE TAXONOMY: TEN YEARS OF PRACTICE AND LOOKING INTO THE FUTURE

Tricia C. Goulding, Benoît Dayrat[1]

Department of Biology, The Pennsylvania State University; [1]*bdayrat@gmail.com*

With a global crisis in species extinction, the field of taxonomy has been under pressure to speed-up the description and naming of new species before they disappear. This led to passionate debates in the early 2000's about the utility and limitations of DNA barcodes for taxonomy. In response to this debate, Integrative Taxonomy was developed as a way to incorporate DNA sequences into species discovery and species identifications, while maintaining that "traditional" taxonomic research is fundamental to delineating and understanding species. In the ten years since Integrative Taxonomy was proposed, DNA sequences have become prevalent in species description and have been widely used in biodiversity surveys. DNA sequences have become a valuable tool for species discovery but cannot greatly speed up species description which require careful work by trained experts. With the proliferation of DNA barcodes, there is increased need to link newly discovered species with proper names and describe their natural history. High-throughput sequencing has revolutionized many fields of biology, and will undoubtedly be incorporated into alpha-taxonomy in the future. However, in order to move forward with Integrative Taxonomy, it is fundamental that DNA sequences can be linked to individual specimens, and the use of multiplex identifiers with high-throughput sequencing has not yet matched the low cost of uni-directional Sanger sequencing. Only by linking DNA sequences to specimens can individual variation within species be described and can fundamental questions of "which species have been described?" and "which species are new for science?" be answered.

ИНТЕГРАТИВНАЯ СИСТЕМАТИКА: ДЕСЯТЬ ЛЕТ ПРИМЕНЕНИЯ И ВЗГЛЯД В БУДУЩЕЕ

Триша Гулдинг, Бенуа Дейра

В связи с глобальной угрозой вымирания видов систематика оказалась перед необходимостью ускорения описания и именования новых видов прежде их исчезновения. В начале 2000-х гг. это привело

к горячим дебатам по поводу возможностей и ограничений использования ДНК-баркодинга в систематике. В качестве реакции на эти дебаты возникла интегративная систематика как способ, сочетающий использование ДНК-сиквенсов в выявлении и идентификации видов с признанием фундаментальности «традиционных» систематических исследований для понимания границ и природы видов. Спустя 10 лет после формулировки идей интегративной систематики использование ДНК стало превалировать в описании видов и получило широкое применение в изучении биоразнообразия. Секвенирование ДНК стало важным инструментом для выявления видов, но оно не может существенно ускорить описание видов, которое требует внимательного исследования квалифицированных специалистов. Широкое распространение ДНК-баркодинга приводит к необходимости «связывания» вновь открываемых видов с соответствующими названиями и описания их биологии. Высокопроизводительное секвенирование ДНК революционизировало многие разделы биологии и в будущем несомненно будет включено в альфа-систематику. Однако для развития интегративной систематики крайне важно, чтобы ДНК-сиквенсы могли быть «привязаны» к конкретным экземплярам, а использование множественных идентификаторов на основе высокопроизводительного секвенирования пока не сравнялось по доступности с простым секвенированием по Сэнгеру. Только «связывание» ДНК-сиквенсов с экземплярами может позволить описывать внутривидовую изменчивость и получать ответы на ключевые вопросы: «что за виды были описаны?» и «какие виды — новые для науки?».

1. Introduction

Taxonomy has deeply changed in the first decade of the 21st century. A major change is that DNA sequences are now routinely included in papers on species diversity, which was not the case ten years ago. This radical change cannot be understood out of a historical context which is worth briefly presenting here.

In a visionary paper entitled “The Biological Diversity Crisis” (the shorter “biodiversity” had not come yet to existence), Wilson (1985) expressed a series of ideas that laid the foundation for most of the discussions that were to come on species diversity and taxonomy. Those ideas include the urgent need to explore the earth’s species diversity so that attempts could be made to conserve it, the fact that millions of species are still undescribed and unknown, and, importantly, the lack of taxonomists who could describe species. After the Convention on Biological Diversity, in 1992, by the United Nations, the interest and concern for biodiversity continued to grow internationally. For instance, the US National Science Foundation launched its PEET (Partnership for Enhancing Expertise in Taxonomy) program in 1995 (Rodman, 2007), as an attempt to support taxonomic studies of poorly known taxa and train a new generation of taxonomists.

Interestingly, in the exact same decade, i. e., the 1990s, Sanger sequencing became increasingly affordable. It is the second piece of that historical puzzle. In the early 1990s, getting DNA sequences was time-consuming and expensive, reserved to some rare labs of molecular systematics. By 2000, Sanger sequencing had become fast, common, and inexpensive.

Finally, one should not forget a third and essential piece of the puzzle, the conceptual development of phylogeography, which also took place during that same period (Avise et al., 1987; Avise, Ball, 1990; Avise, 2000). In the 1990s, DNA sequences (and, for the most part, mitochondrial COI sequences) were already part of phylogeographic studies in exploring intraspecific variation. In that regard, phylogeography played a critical pioneering role. Other important concepts emerged during the last decade of the 20th century as well, such as Moritz's (1994) Evolutionary Significant Unit (ESU), which also directly depends on the use of DNA sequences for delineating species-level units.

With the new century came new discussions around the general theme of "what taxonomy for the 21st century?" or "what future for taxonomy?". By 2000, the time was ripe to incorporate modern scientific advances into taxonomic practice. A conceptual framework was in place, which had already demonstrated that species-level units could be delineated and species variation explored using DNA sequences. A technique was routinely available and cheap. And, obviously, huge needs existed because most of the Earth's species diversity was — and still is — unknown and that we were — and still are — lacking the taxonomists to study it. As often in science, but most especially in taxonomy, discussions quickly became heated. Any new proposal was seen by taxonomists as a direct attack coming from outsiders who did not know the field of taxonomy. There were a few years (mostly from 2002 to 2005) during which the exchanges were quite lively, to say the least. That was the context in which Integrative Taxonomy was defined (Dayrat, 2005).

The idea of integrative taxonomy came up mostly as an attempt to provide a constructive response to the unfruitful debate generated by the first publications on DNA barcoding and DNA taxonomy (e. g., Baker et al., 2003; Hebert et al., 2003; Tautz et al., 2003). Indeed, as soon as the use of short fragments of DNA sequences for identifying species was proposed, quarrels arose over taxonomy between — with only a slight exaggeration — those supporting molecules and those supporting morphology. A main issue in that debate was that the different tasks of taxonomy were not distinguished clearly enough. Taxonomists discover, describe, name, classify, and identify species. Those different tasks are all connected but they each present distinct challenges.

The fundamental idea on which Integrative Taxonomy is based is that, because species delineation is difficult, it should be based on as many tools available as possible. It is pointless to argue whether molecules or morphology should be used to delineate species. Instead, the variation of different kinds of data (DNA sequences, morphology, behavior, etc…) should be studied and integrated to obtain a broad view on species boundaries. Since 2005, many authors have adopted Integrative Taxonomy as a sound approach to species delineation — Dayrat's (2005) paper has been cited more than 400 times so far.

In the present contribution, we review the changes that took place in the last ten years and led to DNA sequences now playing a major role in alpha-taxonomy. We also discuss some challenges that remain regarding integrative taxonomy. Looking into the future, we discuss the use of high-throughput next-generation sequencing in alpha-taxonomy and the challenges associated with it.

2. The major role of DNA sequences in biodiversity research

DNA barcoding (Hebert et al., 2003) and DNA taxonomy (Tautz et al., 2003) claimed that short fragments of DNA had the prom-

ise to speed up taxonomy and help solve the biodiversity impediment. The idea behind was simple. There are not enough taxonomists given the huge diversity that remains to be discovered and traditional taxonomy is too slow. One of the strongest criticisms expressed about those DNA-based approaches was that they neglected the link with traditional species names and, through them, a goldmine of information accumulated over more than two centuries about species (Dayrat, 2005; Teletchea, 2010; and references therein). Species diversity could be explored, but the link with the past literature would be lacking. Obviously, DNA sequences would need taxonomists because only they could go through the taxonomic literature to find species names and retrieve valuable biological information.

However, the most important issue in the original papers on DNA barcoding and DNA taxonomy probably was that species delineation and species identification were confused. Differences between species identification and species delineation are deep and it is essential to understand them. Species identification should be as easy, fast, and as reliable as possible. A distinctive, external color trait may be enough or DNA sequences may be needed, as in the case of cryptic species. It should not matter which data are best or necessary for identification. Species identification is what comes at the very end, after the real hard work has been done. Indeed, species units should first be delineated before they can be identified, regardless of whether it is based on DNA sequences or morphology.

Species discovery starts in the field, when individuals are collected, but new species are not necessarily recognized as new in the field. Hypotheses are often formed in the field based on observations of morphology, habitat, and behavior. Evaluation of these hypotheses with taxonomic research is then needed and often reveals some taxa with high morphological variation, as well as unseen cryptic diversity. Yet, field collecting may not be done by experts, which can make describing species more difficult if notes and photographs do not describe relevant natural history information. The core of taxonomic research is the delineation of species boundaries, which addresses the following question: given a number of specimens, what species are there? It seems simple, but in reality it is complicated because of individual and eventually other kinds of within-species variation. Everything in life varies and variation is the essence of biodiversity but it also is the reason why studying it is so difficult.

As the core of taxonomic research, delineating species boundaries should not be done as fast as possible, because fast research rarely rhymes with good research. It is unfortunate because the need for taxonomy is immense, i. e., considerable amounts of new species wait to be discovered and described. Anybody with any bit of taxonomic experience will agree that working on a taxon in which new species have been described (and new names created) recklessly and carelessly in the past is basically a taxonomic nightmare. So, poor-quality work is not an option. Some would argue that current species descriptions include character information that is not useful in diagnosing species, and that more concise species descriptions are needed. This has led to the recent creation of the term “turbo-taxonomy”, which has been applied to recent descriptions of species in hyper-diverse groups of invertebrates (Butcher et al., 2012; Riedel et al., 2013). As the expertise of taxonomists is greatly needed to address undescribed biodiversity, there is certainly a need to be efficient in producing species descriptions. At the same time, this methodology may be less applicable in less diverse groups or those with complicated tax-

onomic history. A lesson from the past tells us that the best service to future taxonomists and future users of taxonomic data is to produce good work, i. e., accurate species units. It may take time, which unfortunately we do not have a lot of given the multiple threats to biodiversity and a lack of experts, but there is no such thing as fast and good science, at least not in the realm of taxonomy.

When standardized DNA-based approaches to taxonomy were first proposed (Baker et al., 2003; Hebert et al., 2003; Tautz et al., 2003), they generated vigorous debate. While some were ardently opposed to using DNA sequences for taxonomy (Ebach, Holdrege, 2005) or believed it would be too inefficient (Hołyński, 2010), others were enthusiastic about how it could revolutionize taxonomy (Miller, 2007) and the proposal gained many supporters (Blaxter et al., 2005; Valentini et al., 2009; Teletchea, 2010; Mitchell, 2011). For animals, a region of the mitochondrial cytochrome oxidase I gene (COI) was selected as a barcode region due to the ease of PCR amplification in a broad variety of taxa, and a mutation rate that enables many closely related species to be distinguished. In plants and other gross taxonomic groups it has been debated which locus should be used as a DNA barcode and in some cases multiple loci are used (Kress, Erickson, 2008; Zimmermann et al., 2014).

The choice of a mitochondrial marker for the animal DNA barcode was controversial because the mode of mitochondrial inheritance and the evolution of the mitochondrial genome differ from the nuclear genome and therefore the two genomes may differ in how they estimate evolutionary relationships (Rubinoff, 2006). Also, gene trees may not represent species trees and the analysis of a single molecular marker should be interpreted cautiously regarding the evolutionary history of a group (Degnan, Rosenberg, 2009). Others have pointed out the potential difficulty in determining a "barcode gap", or a universal threshold separating intraspecific and interspecific diversity (Meyer, Paulay, 2005; Hickerson et al., 2006). The potential limitations and debate about using a mitochondrial marker have been discussed in many reviews of DNA barcoding (Waugh, 2007; Frézal, Leblois, 2008; DeWalt, 2011) and do not need to be repeated here, but suffice it to say that the limitations have not prevented the adoption of DNA barcoding.

As the number of groups in which DNA barcoding has been applied has grown, it has become clear that levels of genetic divergence representative of multiple species will differ between taxonomic groups, and therefore a standard and universal threshold for delineating species is not practical. However, that does not interfere with the ability of individual studies to delineate species within their respective taxonomic groups. The identification of a "barcode gap" as a threshold for delineating species within a group has not be found in all taxonomic groups and may be complicated in species-poor data sets; in these cases, other loci and data are needed to delineate species (Meyer, Paulay, 2005; Meier et al., 2006; Boyer et al., 2007; Piganeau et al., 2011; Dellicour, Flot, 2015). There is also likely to be a few taxa in any group that present additional challenges for delineation (van Velzen et al., 2012). Yet in many taxa there is a barcode gap that enables reliable species-level delineation which can be tested using different analytical methods (Lefébure et al., 2006; Puillandre et al., 2012; Fujisawa, Barraclough, 2013).

DNA sequences are now frequently used to examine intraspecific genetic diversity within morphospecies, and this has resulted in many discoveries of morphologically cryptic lineages that need further study (Bickford et al., 2007). DNA sequences have

the additional benefit of standardizing biodiversity surveys of unknown groups across systems so that data can be compared, and the completeness of sampling can be analyzed with DNA sequence accumulation curves (Smith et al., 2009). While there will be a degree of error in initial estimations using mitochondrial DNA, DNA sequences have been undoubtedly useful in biodiversity surveys, especially in groups in which morphological identifications are difficult or time-consuming (Bhadury et al., 2006; Hajibabaei et al., 2011). DNA sequencing has also been invaluable in revealing cryptic microbial biodiversity from environmental samples via metagenomics, which is now being done with high-throughput sequencing (Roesch et al., 2007; Rajendhran, Gunasekaran, 2011). It has been more difficult to define "species" in many of the microbial taxa compared to metazoans due to frequent horizontal gene transfer. Thus, the taxonomic status of many molecular operational taxonomic units of microbes remains uncertain, although different strategies are being developed to produce molecular operational taxonomic units that approximate species (Caron et al., 2009; Bruno et al., 2015). Bibliographic searches reveal that DNA sequences are now pervasive in studies of biodiversity (Teletchea, 2010). Interestingly, many studies use *COI* sequences to estimate biodiversity or cryptic diversity without explicitly characterizing it as a barcoding approach (Faurby et al., 2011; O'Loughlin et al., 2011; Liu et al., 2013), and many others use the same locus (*COI*) to study phylogeography within species (Abellán et al., 2009; DeBoer et al., 2014; Einfeldt et al., 2014).

It is pertinent to point out here that there are fundamental differences in approaches between the use of *COI* sequences in biodiversity surveys (with a reference or not to DNA barcoding) and their use in phylogeographic studies or broader systematic studies. While information about the origin of species is relevant to understanding species diversity, it is not strictly needed in order to delineate and describe species. It often requires research at the population level which is beyond the scope of biodiversity surveys. For phylogenetic studies, there is a clear need to utilize more than one marker in order to better reconstruct species relationships and resolve geographic or ecological patterns in speciation. Using both nuclear and mitochondrial DNA sequences enables taxa to be more confidently supported and more accurately represent species diversification. It has therefore become expected that phylogenetic studies should utilize multiple genetic loci to investigate the evolutionary history of a particular clade and this has become common practice (e. g., Frey, Vermeij, 2008; Leavitt et al., 2013; Tamar et al., 2015). Ideally, biodiversity surveys could also incorporate multilocus data to delineate species (Fujita et al., 2012), but large numbers of individuals are usually sequenced in order to get an estimate of species diversity, and there is often not enough time or resources to sequence multiple loci from a large number of samples in order to comprehensively investigate phylogenetic relationships at multiple levels. With many undescribed species needing the attention of taxonomists, obtaining multilocus data for each species description will often not be the best use of resources.

Efforts recently emerged to incorporate DNA barcoding into broader studies in biodiversity and ecology (Kerr et al., 2007; Janzen et al., 2009; Ward et al., 2009; Leray et al., 2012; Dincă et al., 2015) and a large amount of sequence data is continually being added to databases (e. g., Genbank, CBoL). A major challenge that we already face is the curation of those sequences, from the permanent storage of an actual specimen associated with a

particular sequence to the species binominal name attached to that sequence. In that regard, Integrative Taxonomy provides us with a framework for linking the new and massive acquisition of COI sequences with other data on species diversity, such as natural history, historical literature, past museum collections, etc… Needless to say, challenges remain. Some are discussed below.

3. Current challenges in Integrative Taxonomy

The past ten years have shown that the scientific community has broadly and enthusiastically adopted the use of DNA sequences in biodiversity studies. Regardless of the heated exchanges that took place in the early years after the emergence of DNA barcoding, DNA sequences are now routinely used in species diversity studies, often without any reference to barcodes or barcoding. However, important challenges remain with respect to the integration of the data from DNA sequences with all the rest of taxonomic information (Padial et al., 2010; Tan et al., 2010; Jörger et al., 2014).

Many DNA sequences are being accumulated, mostly because it is technically easy and fast — little training is required. There is no question that it is a great thing for biodiversity and taxonomy because molecular data are clearly invaluable. DNA sequences can provide quick and amazingly interesting biodiversity estimates. So, if nothing else, they say how many species there are within a particular region or taxon, given a number of individuals. For people interested in some biogeographic or ecological questions, a species count may be enough. However, trying to answer the question "what species are there?" is far more informative that the question "how many species are there?" Far more informative but far more complicated as well, because it requires to deal with taxonomic data with which only well-trained experts are normally familiar, such as the taxonomic literature, nomenclatural history, natural history, behavior, anatomy, and within-species variation of traits. So, in other words, as insightful as getting DNA sequences is to answer the question "how many species are there", it only is a beginning and not an end in itself. And it is the beginning of studies that can take a long time and sophisticated taxonomic training. DNA sequences will not speed-up taxonomic research because the core of Integrative Taxonomy means dealing with the question "what species are there?" which requires considerable training, additional data, and time.

How cryptic species have been dealt with in the recent years illustrates well the current reality of biodiversity studies. DNA sequence data has revealed the presence of morphologically cryptic species across the tree of life, but it often is unclear to the users of this sequence data whether species names already exist and could be applied to those units, or if new names are needed. This is especially prevalent in invertebrates, in which biodiversity is high but still poorly known. Often, a complex of cryptic species is referred to by the name of a known species and cryptic species are given a numeric designation (e. g., Hebert et al., 2004; García-Morales, Elías-Gutiérrez, 2013; Dayrat et al., 2014). The naming of new species can be left for subsequent publications, but often these cryptic species end up left in limbo without a subsequent effort to address naming. This data on the presence of cryptic species is being left for future study by other taxonomists, but in some cases there may be no other expert working on the group who can name the species. If the species names are not addressed at the time species are being studied, it could be a long time before the names are addressed.

So, at the core of the integration of DNA sequences with other taxonomic data are the species names. Suppose a series of species that are well delineated. Which species have been named in the past? Which species are new to science? Answering these two questions means that one has to deal with the nomenclatural history of a taxon. There is no way out. In that regard, it is worth reiterating here again the critical importance of checklists of species names. These are frequently not considered of high scientific importance but should be valued for their essential contribution to the biodiversity studies. Without checklists of species names, the biodiversity of a group cannot be investigated, at least not within a context of Integrative Taxonomy. Identifying all the species names created within a given group, those available for use and those not available, makes it possible for investigators to apply names — old or new — to species being delineated. And, as surprising as it might seem to some people, a fact is that putting together a species checklist is far more time consuming and difficult than delineating species using COI sequences. Nomenclature is not a mere formality that can be overlooked for biological studies. Without names, DNA sequences are disconnected from historical data on natural history and geographic distribution. It is important that we do not let this divide between traditional taxonomy and DNA sequence data continue to grow and potentially generate a gap between these types of data.

Another key aspect of the Integrative Taxonomy which requires enormous time efforts and expertise is the actual description of the species that need to be described, i. e., all the species new to science and all the species that have been named in the past but remain poorly known and need to be redescribed properly. Indeed, beyond names, closely-related species are entities that differ in ways that need to be described. Differences in nucleotide sequences certainly are part of biodiversity, but other differences should not be neglected. Morphological, anatomical, behavioral, and ecological differences between closely related species need to be described for species diversity to take all its deep and real meaning. Biodiversity goes well beyond counting species in a location or in a group. Biodiversity is about the characteristics of each of those closely-related species. Unfortunately, there remains a bottleneck in the number of trained experts who can describe species once they are delineated. And, unfortunately, describing species is time consuming, but it is necessary in order to provide useful information to all scientists that seek to study species and their communities.

Finally, it may be worth reinforcing here the critical importance of natural history data directly collected in the field. Experts should be present in the field when surveying biodiversity whenever possible. Taxonomists are above all naturalists, and one has to be an expert in a particular group to know what traits to pay attention to, what natural history data to collect, etc… Collectors bringing all kinds of specimens from the field play an immense role in biodiversity studies, but the eye of an expert is irreplaceable. Regardless of how they are collected in the field and by whom, natural history data are critical for the species diversity. It needs to be included in papers and attached to voucher specimens being deposited in museums or somewhere accessible to other researchers (Martin, 1990; Funk et al., 2005).

In cases where organisms were fixed in formalin for study, there are two types of vouchers which need to be deposited for species identifications: specimens preserved for anatomy and those which have been used for DNA sequencing. Molecular sequences

need to be directly linked in papers to catalogued voucher specimens so that they can be referenced or evaluated by others. Many DNA barcode sequences in databases like GenBank are misidentified, and when questions arise about the identity of the organisms they often cannot be checked because they lack voucher specimens. In addition, taxonomic identifications may need to be updated as additional data is collected, and therefore the original specimens need to be available. Currently, determining the reliability of a taxonomic identifications used for a DNA barcode requires searching through the original papers to determine which specimens were used for sequencing. To address this problem, Chakrabarty and colleagues have proposed a revised system for authors to rank the reliability of taxonomic identifications in sequence databases based on the material used for sequencing, i. e., type and non-type sources (Chakrabarty et al., 2013). If the GenSeq system were to be used in future submissions, it would help those using DNA barcodes for species identification to recognize the reliability of the taxonomic identifications.

We also wish to emphasize again the importance of within-species variation. The perils of describing a species from one specimen have been acknowledged for a long time; an obvious risk is that the same species is liable to be described multiple times with different names, causing a great deal of confusion. The same principle applies to using DNA sequences for species diversity. In particular, sampling species from across their geographic range is important as it provides information about intraspecific genetic variability, and makes future species identification more accurate (Bergsten et al., 2012).

Ultimately, as was originally emphasized (Dayrat, 2005), Integrative Taxonomy should take place within a revisionary framework. Unfortunately, like checklists of species names, revisions are highly regarded only by a small fraction of researchers. Regardless of that, taxonomic revisions allow comprehensive studies in which everything that matters in taxonomy (individual variation, nomenclatural history, field data, effective species delineation, etc…) can be integrated properly. However, taxonomic revisions take time and expertise.

4. Taxonomy in the era of high-throughput sequencing

Rapid advances in high-throughput sequencing have revolutionized many fields in biology with applications from human health to ecosystem diversity. Genomic studies have gone beyond sequencing the human genome to sequencing the human microbiome (Gill et al., 2006; Turnbaugh et al., 2006; De Filippo et al., 2010) and have great potential to improve our understanding of genetic diseases. As the cost for high-throughput sequencing has declined in recent years, it has become increasingly feasible to pursue genomic studies of non-model organisms. The ability to sequence a genome or analyze large numbers of genes from a species at lower cost opened up new investigative directions for answering evolutionary questions. While the analysis of such large amounts of data presents many computational challenges, phylogenomics has begun to be applied to understanding the deep evolutionary relationships across taxa (Prasad et al., 2008; Finet et al., 2010; Meusemann et al., 2010; Kocot et al., 2011).

Analyzing entire genomes or using transcriptomes to analyze large numbers of loci is considered ideal for answering challenging evolutionary questions: more data should resolve problems in identifying the relationships between organisms. A variety of next-generation sequencing techniques have begun to be applied to systematics and

phylogenetics such as restriction-site-associated DNA sequencing (RAD-seq), RNA-seq and hybrid enrichment, which all involve sequencing many regions of the genome without PCR. Those methods may be used to address various biological questions, but there are also various limitations: RAD-seq is a useful technique for sequencing the same regions from multiple individuals without a lot of time invested, but is most appropriate for investigating shallowly diverged groups; RNA-seq cannot be used for many samples previously collected because high-quality RNA is required, and hybrid enrichment techniques require significant time and monetary investment to design capture probes (Lemmon, Lemmon, 2013). Sequencing large genomic regions from a many specimens becomes less practical when evolutionary relationships are being examined across a large diversity of species. In these cases, one approach is to modify standard PCR protocols and then pool the PCR products into a high-throughput sequencing run.

Pooling PCR products in a high-throughput sequencing run is already common protocol in microbial genomics (Roesch et al., 2007; Stoeck et al., 2010; Webster et al., 2010). For microbes, the 16S ribosomal DNA is the standard genetic marker used to distinguish taxonomic units. To analyze diversity in a sample, the 16S region is amplified from DNA in the sample and multiplex identifiers (MIDs) can be attached to different environmental samples (Kozich et al., 2013) in order to pool them all into a sequencing run. Bioinformatics programs can later demultiplex the sequences using the unique barcode tags. With high sequencing coverage from new high-throughput sequencing platforms, it is now possible to multiplex more samples while retaining adequate sequencing depth, thus reducing the cost per sample. Genetic sequencing has proved especially useful for quantifying taxonomic units for bacterial communities and microbiota difficult to identify morphologically. High-throughput sequencing has also proved useful to investigating fungal diversity (O'Brien et al., 2005; Unterseher et al., 2011). As high-throughput sequencing has been advanced on the 454 (Roche), MiSeq (Illumina) and Ion Torrent (Life Technologies) sequencing platforms, additional methods and protocols have been developed to multiplex samples for sequencing in high-throughput sequencing runs as well as many bioinformatics pipelines produced for analyzing the data. In theory, these techniques used for analyzing microbial diversity are applicable to other taxa as well. With the recent advances in the length of sequencing reads and decreasing prices, are we ready to start applying high-throughput sequencing to analyzing metazoan biodiversity?

The application of high-throughput sequencing to investigating animal species diversity depends on how we view biodiversity. Studies interested in quantifying changes in species richness across a landscape differ in their approach to those seeking to discover unknown species. Recognizing species will require utilizing genomic regions with higher rates of evolution that those investigating phylogenetic relationships between anciently diverged taxa. In the case of studying alpha-taxonomy, each DNA sequence needs to be directly linked to an individual specimen. Ever since Darwin, the core of taxonomy has been to examine variation to delineate species. Thus, our understanding of species diversity significantly increases when individuals are examined from many localities across their geographic range. Studying diversity at the species-level requires sequencing large numbers of individuals, and differs essentially from investigations at higher levels of taxonomic hierarchy which usually utilize much fewer individuals per taxon.

The development of new techniques to use high-throughput sequencing to sequence targeted regions of the genome instead of random portions has made it possible to focus on a few genetic markers which are homologous across taxa and known to be phylogenetically informative. With these approaches a modest number of loci may be utilized from each sample, but a larger number of samples are able to be analyzed. However, to investigate the relationships between closely related species, including the discovery of cryptic species, we need to be able to match individual sequences to individual specimens. Therefore, we cannot attach one MID tag to all samples found in a particular location and pool them for high-throughput sequencing. For higher-level taxonomic questions sequences might be easily matched to extremely divergent taxa without unique MID tags, but for closely related species in need of delineation, each individual needs its own MID tag. In metazoans, the cytochrome oxidase subunit I (*COI*) is a widely used marker for species delineation. Several studies have pioneered investigating metazoan diversity with high-throughput sequencing by applying MIDs to PCR products (targeted amplicon sequencing or parallel tagged sequencing) for several loci known to be useful for phylogenetic inference (Bybee et al., 2011; O'Neill et al., 2013; Clarke et al., 2014). These studies have proved that techniques developed to distinguish different environmental DNA samples in a high-throughput sequencing run can be adapted to identifying sequences from individual organisms for biodiversity studies.

The concept of using MIDs is an essential one for taxonomy. Using DNA sequences to delineate species requires being able to link each individual sequence back to an actual specimen. Attaching a unique MID to a sample containing a variety of species is quite economical, but attaching a unique MID to each individual specimen adds a significant cost. Compared to studies of microbial and fungal diversity that seek to look at species richness within a sample, many more MIDs are needed to pool PCR products from a large number of metazoans in a sequencing run in order to match sequences back to individuals. While the cost of sequencing has declined, the cost for producing primers to attach MIDs to PCR products has not declined greatly. One way to reduce the cost of indexes is to use a dual-indexing approach, which allows the MIDs to be used in different combinations to uniquely identify a larger number of samples (Fadrosh et al., 2014).

Protocols to add MIDs to the PCR products in a second step after amplification have also been developed (so that the MIDs can be reused for multiple genomic regions from the same individual), but they still do not reduce the cost significantly enough when a large number of individuals (hundreds) would be combined in a run. The costs for sequencing a high number of individuals in a high-throughput sequencing run is significantly higher than sequencing fewer individuals for a larger number of genes because of the need for a greater number of MID tags. In order to sequence 576 samples for several loci in a sequencing run would require a set of 24 forward and 24 reverse purified indexing primers, in addition to the original PCR primers with adapters to attach the indexes, which results in a high initial (financial) investment to adopt this methodology. Utilizing high-throughput sequencing without using MIDs to label each individual specimen would entail losing critical data. In fact, sequences would then be useless for taxonomic study. We would be limited to say how many species exist in a particular environmental sample, but we could not go back to the specimens

that have been collected and do the actual taxonomic work, especially describing new species. In comparison, the cost of investigating species diversity is still lower using uni-directional Sanger sequencing, following up with sequencing in the reverse direction from selected individuals and sequencing of additional loci as needed.

Another issue for current high-throughput sequencing technologies is rarely mentioned. In order to use targeted amplicon sequencing of PCR products, the size of the PCR products must be smaller than the read length in Sanger sequencing. Clarke's study using the Ion Torrent only used 133 bp of the *COI* region compared to the 710 bp routinely sequenced using Folmer primers (Clarke et al., 2014). Using other sequencing technology, Bybee and collaborators sequenced the 28S and 18S regions frequently used in metazoan phylogenetics, but had to divide the genes into PCRs of 4 overlapping regions of 400–600 bp to assemble them in the 454 run (Bybee et al., 2011). The Illumina MiSeq has advanced to producing read lengths up to 600 bp (300 bp paired end reads) but, in practice, the length of a locus that can be sequenced with this technology must still be less than 600 bp in order to have overlap between the reads so that there would not be a gap in the middle of the sequence with low coverage. Sequencing technology will undoubtedly advance to the point that sequencing reads will be long enough to accommodate longer PCR fragments, but sequencing much smaller regions of a gene with high-throughput sequencing would compromise the number of informative sites that can be used for phylogenetic analysis compared to Sanger sequencing.

Many studies have championed the cost-savings that can be achieved using high-throughput sequencing. Certainly the amount of sequences produced for the cost has drastically declined on some sequencing platforms. However, in the case of sequencing for taxonomy, it is necessary to utilize the longest sequencing reads available so that the entire sequence with its' MID can be obtained and linked back to a given individual specimen. If these MIDs were to be re-used for many sequencing runs, the overall cost could eventually be cheaper than bi-directionally sequencing thousands of specimens with Sanger sequencing. Yet, the initial investment for the MIDs is quite high. It is therefore still currently much cheaper to sequence large numbers of specimens with uni-directional Sanger sequencing than to pool high numbers of MID-labeled PCR products for a few loci into a high-throughput sequencing run.

High throughput sequencing has revolutionized diverse fields of biology by allowing us to gain a great deal of genomic data from diverse taxa to understand deep evolutionary relationships and many aspects of organismal biology. There is great promise that this technology will one day revolutionize alpha-taxonomy by allowing us to sequence many genomic regions from a large number specimens and still be able to match sequences to individual specimens at a low cost.

5. Conclusion

In 2010, the United Nations declared the International Year of Biodiversity to highlight the multitude of threats species are facing and to encourage international collaboration to address challenges in conservation. Yet, even with public awareness of threats to species diversity and efforts to protect it, undescribed species are going extinct faster than they can be recognized (Vitousek et al., 1997; Hooper et al., 2012; Costello et al., 2013). The biodiversity crisis described thirty years ago by Wilson (1985) is still ongoing. In fact, its pace does not cease to increase. So, massive and effective efforts are needed

to explore the Earth's species diversity. Any constructive initiative is welcome!

Regardless of (or maybe thanks to) the heated exchanges that took place for a few years following the first publications on DNA barcoding and DNA taxonomy, single-marker DNA sequences have been embraced as an important step in evaluating species diversity, despite the imperfect nature of the data. In particular, DNA sequence data can greatly enhance our taxonomic knowledge when morphological differences are rare, confusing, rare, or simply absent. New applications of DNA barcode data continue to be developed, including metabarcoding of environmental samples to estimate community composition, monitor biodiversity over time or to detect invasive species, as well as match different life history stages and sexual dimorphism to a species (Taberlet et al., 2012; Ji et al., 2013; Bohmann et al., 2014; Glowska et al., 2014; Puncher et al., 2015).

Taxonomy and taxonomists are too often unfairly and negatively criticized. It goes without saying that our field could still be improved. However, taxonomists are not responsible for the two most important issues with respect to biodiversity: the fact that species diversity has been under attack for decades due to all kinds of human activities, and the lack of funds and jobs in taxonomy. Also, the fact that DNA sequences have been adopted so enthusiastically by taxonomists in the past decade shows that taxonomy can — and in fact does — change, even radically. Routinely using DNA sequences for evaluating species diversity is a great thing for taxonomy. Clearly, taxonomy rejuvenated in the past ten years.

One of the major negative criticisms of taxonomy has often been that it is too slow. The promise that getting DNA (*COI*) sequences is going to speed up taxonomy to the point where unknown species diversity would rapidly become known is an illusion. Knowing that a species exists thanks to a few *COI* sequences merely is a first step towards knowing that species. Is it new to science or has it been already named? If already named, what have naturalists said about it in past years? What is its general biology? How can it be recognized in the field? In what microhabitat is it found? Is it common or rare? How does it differ from closely-related species? How does it interact with other species? What are its adaptations (physiological, behavioral, morphological) to its habitat? All those questions, among others, cannot be answered by *COI* sequences alone but they are the actual steps that are needed to know a species. The taxonomists who can answer those questions have skills and expertise that cannot be acquired overnight. And good taxonomic research takes time, like all research. In that regard, we wholeheartedly agree with Tan and collaborators when they claim that "DNA sequences will not speed-up taxonomic research, but will lead to the estimation of more accurate species boundaries based on a more satisfactory amount of data" (Tan et al., 2010, p. 59).

High-throughput next-generation sequencing has not yet reached the point where it can be used as quickly or cost-effectively as standard Sanger sequencing techniques for taxonomic studies of biodiversity. However, it clearly is going to become real in the near future. Sooner or later, it will be possible to sequence longer fragments of genes from well-tagged individual genomes that will be directly and unambiguously connected to specific individual specimens. Looking into the future, however, it is essential that DNA sequences obtained for biodiversity research, whether through Sanger sequencing or next-generation sequencing, can be matched to individual specimens so that species can continue to be properly described by trained experts.

There is a difference between knowing that a species exists and knowing that species. Integrative Taxonomy provides a general framework to help people to know species better. Let us hope we can get to know as many species as possible as their decline inexorably continues.

Acknowledgements

We thank Igor Pavlinov for the opportunity to reflect on Integrative Taxonomy.

The Dayrat lab is supported by the National Science Foundation (DEB 1419394) and the Eberly College of Science at the Pennsylvania State University.

References

Abellán P., Millán A., & Ribera I. 2009. Parallel habitat-driven differences in the phylogeographical structure of two independent lineages of Mediterranean saline water beetles. — Molecular Ecology, 18 (18): 3885–3902.

Avise J.C. 2000. Phylogeography: the history and formation of species. Cambridge: Harvard University Press. 447 p.

Avise J.C., Arnold J., Ball R.M. et al. 1987. Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics. — Annual Review of Ecology and Systematics, 18: 489–522.

Avise J.C., Ball R.M. 1990. Principles of genealogical concordance in species concepts and biological taxonomy. — Oxford surveys in evolutionary biology, 7: 45–67.

Baker C.S., Dalebout M.L., Lavery S., Ross H.A. 2003. www.DNA-surveillance: applied molecular taxonomy for species conservation and discovery. — Trends in Ecology and Evolution, 18 (6): 271–272.

Bergsten J., Bilton D.T., Fujisawa T. et al. 2012. The effect of geographical scale of sampling on DNA barcoding. — Systematic Biology, 61 (5): 851–869.

Bhadury P., Austen M.C., Bilton D.T. et al. 2006. Development and evaluation of a DNA-barcoding approach for the rapid identification of nematodes. — Marine Ecology Progress Series, 320: 1–9.

Bickford D., Lohman D.J., Sodhi N.S. et al. 2007. Cryptic species as a window on diversity and conservation. — Trends in Ecology and Evolution, 22 (3): 148–155.

Blaxter M., Mann J., Chapman T. et al. 2005. Defining operational taxonomic units using DNA barcode data. — Philosophical Transactions of the Royal Society, Ser. B: Biological Sciences, 360 (1462): 1935–1943.

Bohmann K., Evans A., Gilbert M. T. P. et al. 2014. Environmental DNA for wildlife biology and biodiversity monitoring. — Trends in Ecology & Evolution, 29 (6): 358–367.

Boyer S.L., Baker J.M., Giribet G. 2007. Deep genetic divergences in *Aoraki denticulata* (Arachnida, Opiliones, Cyphophthalmi): a widespread "mite harvestman" defies DNA taxonomy. — Molecular Ecology, 16 (23): 4999–5016.

Bruno F., Marinella M., Santamaria M. 2015. e-DNA meta-barcoding: From NGS raw data to taxonomic profiling. — Picardi E. (ed.). RNA Bioinformatics. New York: Springer. P. 257–278.

Butcher B.A., Smith M.A., Sharkey M.J., Quicke D.L.J. 2012. A turbo-taxonomic study of Thai aleiodes (aleiodes) and aleiodes (arcaleiodes) (Hymenoptera: Braconidae: Rogadinae) based largely on *COI* barcoded specimens, with rapid descriptions of 179 new species. — Zootaxa, 3457: 1–232.

Bybee S.M., Bracken-Grissom H., Haynes B.D. et al. 2011. Targeted amplicon sequencing (TAS): A scalable next-gen approach to multilocus, multitaxa phylogenetics. — Genome Biology and Evolution, 3: 1312–1323.

Chakrabarty P., Warren M., Page L. M., Baldwin C. C. 2013. GenSeq: An updated nomenclature and ranking for genetic sequences from type and non-type sources. — ZooKeys, 346: 29–41.

Clarke L.J., Czechowski P., Soubrier J. et al. 2014. Modular tagging of amplicons using a single PCR for high-throughput sequencing. — Molecular Ecology Resources, 14 (1): 117–121.

Costello M.J., May R.M., Stork N.E. 2013. Can we name Earth's species before they go extinct? — Science, 339 (6118): 413–416.

Dayrat B. 2005. Towards integrative taxonomy. — Biological Journal of the Linnean Society, 85 (3): 407–415.

Dayrat B., Goulding T.C., White T.R. 2014. Diversity of Indo-West Pacific *Siphonaria* (Mollusca: Gastropoda: Euthyneura). — Zootaxa, 3779 (2): 246–276.

DeBoer T.S., Naguit M.R.A., Erdmann M. V. et al. 2014. Concordance between phylogeographic and biogeographic boundaries in the Coral Triangle: conservation implications based on comparative analyses of multiple giant clam species. — Bulletin of Marine Science, 90 (1): 277–300.

Degnan J.H., Rosenberg N.A. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. — Trends in Ecology and Evolution, 24 (6): 332–340.

Dellicour S., Flot J.-F. 2015. Delimiting species-poor data sets using single molecular markers: A study of barcode gaps, Haplowebs and GMYC. — Systematic Biology, 64 (6): 900–908.

DeWalt R.E. 2011. DNA barcoding: a taxonomic point of view. — Journal of the North American Benthological Society, 30 (1): 174–181.

Dincă V., Montagud S., Talavera G. et al. 2015. DNA barcode reference library for Iberian butterflies enables a continental-scale preview of potential cryptic diversity. — Scientific Reports, 5, 12395.

Ebach M.C., Holdrege C. 2005. More taxonomy, not DNA barcoding. — BioScience, 55 (10): 823.

Einfeldt A.L., Doucet J.R., Addison J.A. 2014. Phylogeography and cryptic introduction of the ragworm Hediste diversicolor (Annelida, Nereididae) in the Northwest Atlantic. — Invertebrate Biology, 133 (3): 232–241.

Fadrosh D.W., Ma B., Gajer P. et al. 2014. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. — Microbiome, 2 (1): 1-7.

Faurby S., Jørgensen A., Kristensen R.M., Funch P. 2011. Phylogeography of North Atlantic intertidal tardigrades: Refugia, cryptic speciation and the history of the Mid-Atlantic Islands. — Journal of Biogeography, 38 (8): 1613–1624.

De Filippo C., Cavalieri D., Di Paola M. et al. 2010. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. — Proceedings of the National Academy of Sciences of the United States of America, 107 (33): 14691–14696.

Finet C., Timme R.E., Delwiche C.F., Marlétaz F. 2010. Multigene phylogeny of the green lineage reveals the origin and diversification of land plants. — Current Biology, 20 (24): 2217–2222.

Frey M.A., Vermeij G.J. 2008. Molecular phylogenies and historical biogeography of a circumtropical group of gastropods (Genus: Nerita): implications for regional diversity patterns in the marine tropics. — Molecular Phylogenetics and Evolution, 48 (3): 1067–1086.

Frézal L., Leblois R. 2008. Four years of DNA barcoding: Current advances and prospects. — Infection, Genetics and Evolution, 8 (5): 727–736.

Fujisawa T., Barraclough T.G. 2013. Delimiting species using single-locus data and the generalized mixed Yule coalescent approach: A revised method and evaluation on simulated data sets. — Systematic Biology, 62 (5): 707–724.

Fujita M.K., Leaché A.D., Burbrink F.T. et al. 2012. Coalescent-based species delimitation in an integrative taxonomy. — Trends in Ecology and Evolution, 27 (9): 480–488.

Funk V. A., Hoch P. C., Prather L. A., Wagner W. L. 2005. The importance of vouchers. — Taxon, 54 (1): 127–129.

García-Morales A.E., Elías-Gutiérrez M. 2013. DNA barcoding of freshwater Rotifera in Mexico: Evidence of cryptic speciation in common rotifers. — Molecular Ecology Resources, 13 (6): 1097–1107.

Gill S.R., Pop M., Deboy R.T. et al. 2006. Metagenomic analysis of the human distal gut microbiome. — Science, 312 (5778): 1355–1359.

Glowska E., Dragun-Damian A., Broda L. et al. DNA barcodes reveal female dimorphism in syringophilid mites (Actinotrichida: Prostigmata: Cheyletoidea): *Stibarokris phoeni-*

conaias and *Ciconichenophilus phoeniconaias* are conspecific. — Folia Parasitologica, 61 (3): 272–276.

Hajibabaei M., Shokralla S., Zhou X. et al. 2011. Environmental barcoding: A next-generation sequencing approach for biomonitoring applications using river benthos. — PLoS ONE, 6 (4): e17497.

Hebert P.D.N., Cywinska A., Ball S.L., DeWaard J.R. 2003. Biological identifications through DNA barcodes. — Proceedings of the Royal Society, Ser. B: Biological Sciences, 270 (1512): 313–321.

Hebert P.D.N., Penton E.H., Burns J.M. et al. 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator. — Proceedings of the National Academy of Sciences of the United States of America, 101 (41): 14812–14817.

Hickerson M.J., Meyer C.P., Moritz C. 2006. DNA barcoding will often fail to discover new animal species over broad parameter space. — Systematic Biology, 55 (5): 729–739.

Hołyński R.B. 2010. Taxonomy and the mediocrity of DNA barcoding — Some remarks on PACKER et al. 2009: DNA barcoding and the mediocrity of morphology. — Arthropod Systematics and Phylogeny, 68 (1): 143–150.

Hooper D.U., Adair E.C., Cardinale B.J. et al. 2012. A global synthesis reveals biodiversity loss as a major driver of ecosystem change. — Nature, 486 (7401): 105–108.

Janzen D.H., Hallwachs W., Blandin P. et al. 2009. Integration of DNA barcoding into an ongoing inventory of complex tropical biodiversity. — Molecular Ecology Resources, 9 (s1): 1–26.

Ji Y., Ashton L., Pedley S. M. et al. 2013. Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. — Ecology Letters, 16 (10): 1245–1257.

Jinbo U., Kato T., Ito M. 2011. Current progress in DNA barcoding and future implications for entomology. — Entomological Science, 14 (2): 107–124.

Jörger K.M., Neusser T.P., Brezinger B., Schrödl M. 2014. Exploring the diversity of mesopsammic gastropods: How to collect, identify, and delimitate small and elusive sea slugs? — American Malacological Bulletin, 32 (2): 290–307.

Kerr K.C.R., Stoeckle M.Y., Dove C.J. et al. 2007. Comprehensive DNA barcode coverage of North American birds. — Molecular Ecology Notes, 7 (4): 535–543.

Kocot K.M., Cannon J.T., Todt C. et al. 2011. Phylogenomics reveals deep molluscan relationships. — Nature, 477 (7365): 452–456.

Kozich J.J., Westcott S.L., Baxter N.T. et al. 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. — Applied and Environmental Microbiology, 79 (17): 5112–5120.

Kress W.J., Erickson D.L. 2008. DNA barcodes: genes, genomics, and bioinformatics. — Proceedings of the National Academy of Sciences of the United States of America, 105 (8): 2761–2762.

Leavitt S.D., Esslinger T.L., Spribille T. et al. 2013. Multilocus phylogeny of the lichen-forming fungal genus *Melanohalea* (Parmeliaceae, Ascomycota): Insights on diversity, distributions, and a comparison of species tree and concatenated topologies. — Molecular Phylogenetics and Evolution, 66 (1): 138–152.

Lefébure T., Douady C.J., Gouy M., Gibert J. 2006. Relationship between morphological taxonomy and molecular divergence within Crustacea: Proposal of a molecular threshold to help species delimitation. — Molecular Phylogenetics and Evolution, 40 (2): 435–447.

Lemmon E.M., Lemmon A.R. 2013. High-throughput genomic data in systematics and phylogenetics. — Annual Review of Ecology, Evolution, and Systematics, 44 (1): 99–121.

Leray M., Boehm J.T., Mills S.C., Meyer C.P. 2012. Moorea BIOCODE barcode library as a tool for understanding predator–prey interactions: insights into the diet of common predatory coral reef fishes. — Coral Reefs, 31 (2): 383–388.

Liu H.-P., Hershler R., Lang B., Davies J. 2013. Molecular evidence for cryptic species in a

narrowly endemic western North American springsnail (*Pyrgulopsis gilae*). — Conservation Genetics, 14 (4): 917–923.

Martin N. A. 1990. Voucher specimens: A way to protect the value of your research. — Biology and Fertility of Soils, 9 (2): 93-94.

Meier R., Shiyang K., Vaidya G., Ng P.K.L. 2006. DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. — Systematic Biology, 55 (5): 715–728.

Meusemann K., Von Reumont B.M., Simon S. et al. 2010. A phylogenomic approach to resolve the arthropod tree of life. — Molecular Biology and Evolution, 27 (11): 2451–2464.

Meyer C.P., Paulay G. 2005. DNA barcoding: Error rates based on comprehensive sampling. — PLoS Biology, 3 (12): 2229–2238.

Miller S.E. 2007. DNA barcoding and the renaissance of taxonomy. — Proceedings of the National Academy of Sciences of the United States of America, 104 (12): 4775–4776.

Mitchell A. 2011. DNA barocoding is useful for taxonomy: a reply to Ebach. — Zootaxa, 2772: 67–68.

Moritz C. 1994. Defining "evolutionarily significant units" for conservation. — Trends in Ecology and Evolution, 9 (10): 373–375.

O'Brien H.E., Parrent J.L., Jackson J.A. et al. 2005. Fungal community analysis by large-scale sequencing of environmental samples. — Applied and Environmental Microbiology, 71 (9): 5544–5550.

O'Loughlin P.M., Paulay G., Davey N., Michonneau F. 2011. The Antarctic region as a marine biodiversity hotspot for echinoderms: Diversity and diversification of sea cucumbers. — Deep-Sea Research Part II: Topical Studies in Oceanography, 58 (1): 264–275.

O'Neill E.M., Schwartz R., Bullock C.T. et al. 2013. Parallel tagged amplicon sequencing reveals major lineages and phylogenetic structure in the North American tiger salamander (*Ambystoma tigrinum*) species complex. — Molecular Ecology, 22 (1): 111–129.

Padial J.M., Miralles A., De la Riva I., Vences M. 2010. The integrative future of taxonomy. — Frontiers in Zoology, 7: 1–14.

Piganeau G., Eyre-Walker A., Grimsley N., Moreau H. 2011. How and why DNA barcodes underestimate the diversity of microbial eukaryotes. — PLoS ONE, 6 (2): e16342.

Prasad A.B., Allard M.W., Green E.D. 2008. Confirming the phylogeny of mammals by use of large comparative sequence data sets. — Molecular Biology and Evolution, 25 (9): 1795–1808.

Puillandre N., Lambert A., Brouillet S., Achaz G. 2012. ABGD, Automatic Barcode Gap Discovery for primary species delimitation. — Molecular Ecology, 21 (8): 1864–1877.

Puncher G.N., Alemany F., Arrizabalaga H., Cariani A., Tinti F. 2015. Misidentification of bluefin tuna larvae: a call for caution and taxonomic reform. — Reviews in Fish Biology and Fisheries, 25 (3): 485–502.

Rajendhran J., Gunasekaran P. 2011. Microbial phylogeny and diversity: small subunit ribosomal RNA sequence analysis and beyond. — Microbiological research, 166 (2): 99-110.

Riedel A., Sagata K., Suhardjono Y. R. et al. 2013. Integrative taxonomy on the fast track — towards more sustainability in biodiversity research. — Frontiers in Zoology, 10: 15.

Rodman J.E. 2007. Reflections on PEET, the partnerships for enhancing expertise in taxonomy. — Zootaxa, 46 (1668): 41–46.

Roesch L.F.W., Fulthorpe R.R., Riva A. et al. 2007. Pyrosequencing enumerates and contrasts soil microbial diversity. — The ISME Journal, 1 (4): 283–290.

Rubinoff D. 2006. Utility of mitochondrial DNA barcodes in species conservation. — Conservation Biology, 20 (4): 1026–1033.

Smith M.A., Fernandez-Triana J., Roughley R., Hebert P.D.N. 2009. DNA barcode accumulation curves for understudied taxa and areas. — Molecular Ecology Resources, 9 (s1): 208–216.

Stoeck T., Bass D., Nebel M. et al. 2010. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. — Molecular Ecology, 19 (s1): 21–31.

Taberlet P., Coissac E., Pompanon F. et al. 2012. Towards next-generation biodiversity assessment using DNA metabarcoding. — Molecular Ecology, 21 (8): 2045–2050.

Tamar K., Carranza S., Bosch H. et al. 2015. Hidden relationships and genetic diversity: Molecular phylogeny and phylogeography of the Levantine lizards of the genus *Phoenicolacerta* (Squamata: Lacertidae). — Molecular Phylogenetics and Evolution, 91: 86–97.

Tan D.S.H., Ang Y., Lim G.S. et al. 2010. From "cryptic species" to integrative taxonomy: An iterative process involving DNA sequences, morphology, and behaviour leads to the resurrection of *Sepsis pyrrhosoma* (Sepsidae: Diptera). — Zoologica Scripta, 39 (1): 51–61.

Tautz D., Arctander P., Minelli A. et al. 2003. A plea for DNA taxonomy. — Trends in Ecology and Evolution, 18 (2): 70–74.

Teletchea F. 2010. After 7 years and 1000 citations: Comparative assessment of the DNA barcoding and the DNA taxonomy proposals for taxonomists and non-taxonomists. — Mitochondrial DNA, 21 (6): 206–226.

Turnbaugh P.J., Ley R.E., Mahowald M.A. et al. 2006. An obesity-associated gut microbiome with increased capacity for energy harvest. — Nature, 444 (7122): 1027–1031.

Unterseher M., Jumpponen A., Öpik M. et al. 2011. Species abundance distributions and richness estimations in fungal metagenomics — lessons learned from community ecology. — Molecular Ecology, 20 (2): 275–285.

Valentini A., Pompanon F., Taberlet P. 2009. DNA barcoding for ecologists. — Trends in Ecology and Evolution, 24 (2): 110–117.

van Velzen R., Weitschek E., Felici G., Bakker F.T. 2012. DNA barcoding of recently diverged species: relative performance of matching methods. — PLoS ONE, 7 (1): e30490.

Vitousek P.M., Mooney H.A., Lubchenco J., Metillo J.M. 1997. Human domination of Earth's ecosystems. — Science, 277 (5325): 494–499.

Ward R.D., Hanner R., Hebert P.D.N. 2009. The campaign to DNA barcode all fishes, FISH-BOL. — Journal of Fish Biology, 74 (2): 329–356.

Waugh J. 2007. DNA barcoding in animal species: progress, potential and pitfalls. — BioEssays, 29 (2): 188–197.

Webster N.S., Taylor M.W., Behnam F. et al. 2010. Deep sequencing reveals exceptional diversity and modes of transmission for bacterial sponge symbionts. — Environmental Microbiology, 12 (8): 2070–2082.

Wilson E.O. 1985. The biological diversity crisis. — BioScience, 35 (11): 700–706.

Zimmermann J., Abarca N., Enk N. et al. 2014. Taxonomic reference libraries for environmental barcoding: A best practice example from diatom research. — PLoS ONE, 9 (9): e108793.