DNA barcoding and the changing ontological commitments of taxonomy

DNA barcoding and the changing ontological commitments of taxonomy

James W. E. Lowe* & David S. Ingram
Science, Technology and Innovation Studies, University of Edinburgh, Edinburgh, UK
*Corresponding author: jwelowe1985@gmail.com

Abstract

This paper assesses the effect of DNA barcoding – the use of informative genetic markers to identify and discriminate between species – on taxonomy. Throughout, we interpret this in terms of *varipraxis*, a concept we introduce to make sense of the treatment of biological variation by scientists and other practitioners. From its inception, DNA barcoding was criticised for being reductive, in attempting to replace multiple forms of taxonomic evidence with just one: DNA sequence variation in one or a few indicative genes. We show, though, how DNA barcoding has not narrowed or reduced taxonomy in the way it was projected to. We examine the development and implementation of DNA barcoding across three kingdoms of life: animals, plants and protists. Through this, we demonstrate that for DNA barcoding to work, the range of acceptable intra-specific variation needs to be demarcated from variation deemed to be characteristic of inter-specific differences. Consequently, biological processes responsible for particular patterns of variation need to be investigated and understood. This encourages an integrative disposition towards understanding and explaining the evolutionary processes affecting the rate and nature of change at the nucleotide level. We detail how the impact of DNA barcoding has manifested differently across the three kingdoms we examine, assessing this in terms of the ontological commitments that are held and instantiated in practice. Based on this evaluation, we consider the problem of studying multi-kingdom communities, and assess the consequences of our analysis for understanding classification and taxonomy.

Keywords: taxonomy; DNA barcoding; variation; species discrimination; species classification; ontological commitments

Introduction

In this paper we evaluate the impact of DNA barcoding on taxonomy, assessing whether this 21st century approach has indeed had the revolutionary effects on the epistemology, methodology and ontology of taxonomy that was promised by its promotors – and feared by its detractors. We will address this general question by answering two specific questions:

- Has DNA barcoding reduced taxonomy to a single parameter indicative of species differences?
- Has it altered the ontological commitments of taxonomy?

DNA barcoding is a method by which a set of informative genetic markers associated with DNA sequence variation in a number of genes may be used to discriminate among species. To achieve this, the clustering of genomic variation within and across species is catalogued and evaluated. This paper explores the ontological consequences of such species-discrimination practices employed in biological communities studying various higher-level taxa. We have pursued this by an examination of the scientific and technical literature, augmented with semi-structured interviews of key participants identified through professional experience and inspection of that literature. Our perspectives were developed against the background of philosophical, historical, natural scientific

and social scientific enquiry into areas touching on taxonomy and classification. One of us approached this topic as a development of his historical and philosophical investigation into genomics and the products and capabilities it engenders; the other as a continuation of his ongoing conceptualisation of aspects of taxonomy and classification, a field he has been engaged with for decades as a practising plant scientist and a former head of a botanic garden. The proceeding work is a result of this productive interdisciplinary intersection.

Throughout the paper, we characterise classifications and classificatory practices in terms of ontological commitments (after Kendig 2020), which we define as established views of the kinds of entities that exist and the relationships between them in a given domain of theory and practice. These commitments owe something to the historical objectives of the scientific communities working on particular taxa, the materiality of the biological entities and processes that they engage with, and their specific *working world* problems (after Agar 2020). Working worlds are domains that pose and frame particular problems, for example the identification of measures to mitigate the spread, incidence and severity of Covid-19 infections, or improving crop plants and livestock.

At the outset of a particular endeavour, ontological commitments need not be particularly well-founded empirically, theoretically, or epistemically. Nor need they be explicit. However, ontological commitments may shift, and fruitful ontological commitments will evolve in a direction that enables the particular domain of inquiry they underpin to better serve the working world problems they are directed towards. For any endeavour, such as taxonomy, there may be diverging trajectories wherein different communities will adopt distinct sets of ontological commitments over time (e.g., where 'folk' taxonomies become either professional, standardised Linnaean codes or 'ethnobiologies', in which the advancement of the former becomes a working world in itself, while the latter may still continue to be tied to more particular and localised ecologies and cultural, social and economic aims). There may also be some internal differentiation, in which there is sufficient sharing of higher-level ontological commitments and/or overlapping of ontological commitments that they may still be said to be conducting the same general endeavour and operating with the same fundamental ontological commitments. In taxonomy, for instance, there are different codes and norms governing animal and plant classification.

In taxonomy, particular working worlds generate distinct theoretical articulations of biological variation – and practical means for its apprehension, analysis and interpretation – that affect the criteria by which markers of species-discrimination are selected and used.[1] To conceptualise the role of the apprehension, recording and analysis of variation in biological research, we coin a new term – *varipraxis* – to capture the entangled methodological, epistemic and ontological features of the elucidation of variation. We define varipraxis as the set of practices involved in the apprehension, measurement, investigation and analysis of different forms of variation. It both captures forms of variation that are identified and used, and the articulation of them in particular domains.

The apprehension, detection, measurement and comparison of variation involves the prior conception of a type, and thus involves two allied processes of abstraction; one in the definition of the type, the other in deciding which aspects of difference within the type constitute relevant variation to characterise, catalogue and investigate (DiTeresi 2010; Lowe 2015). Variation in nature is ordered in an indeterminate number of ways. It is up to the scientist's abstractive work to bring out a particular way in which the variation of biological objects of interest is ordered, and also to impose some order on it. These two aspects are not antagonistic or contradictory; the imposition of order will often mesh with the order that is established: lines will be sharpened, fuzzy groupings made less fuzzy and more concentrated, and boundaries drawn.

[1] For an articulation of the ways in which one key working world, nature conservation, affects the ontological commitments and practices of taxonomic research, see Conix (2019).

As a term, varipraxis is intended to capture particular manifestations of the delineation of types and variation in which the conceptual is entangled with the practical, as well as the ways in which this entanglement is manifested in biological work. The tasks of apprehending and exploring variation are central to biological investigation, across various different forms of enquiry that have been identified by other scholars, for example the "comparative" and "exemplary" modes characterised and related to each other by Strasser and de Chadarevian (2011) in their examination of the opening decades of molecular biology, and the "ways" and "styles" discerned by Pickstone (2000) and Hacking (1994), among others. Varipraxis emphasises, as these works have done to differing extents, that the more comparative and natural historical aspects of biological enquiry are inextricably linked to the more experimental and interventionist modes. It is necessary to define what parameters of variation and which type or types are being worked with in order to conduct, analyse, interpret and convey experiments, and draw more extended inferences from them, as well as to make forays into working worlds. In introducing varipraxis in the context of changes to taxonomy, a discipline that is more comparative and natural historical than most in the life sciences, we do not wish to imply that its applicability is limited to these kinds of pursuits. Rather, taxonomy and the impact of the advent of DNA barcoding allow us to draw the initial contours of the concept in a way that may not be possible – at this early stage in its inception – for fields in which the treatment of variation is more opaque or complex.

For our theoretical background, we also draw on Kendig (2020) and Minelli (2019), who highlight non-Linnaean "grey nomenclatures" manifested, for example, in DNA barcodes. A grey nomenclature deploys names or designators that depart from the nomenclature of the Linnaean international codes (Minelli 2019). Two simple examples of grey nomenclature would be using the vernacular 'daisy' for the common European plant '*Bellis perennis* L.', or 'thale cress' for the model organism '*Arabidopsis thaliana* (L.) Heynh.'. Grey nomenclature may also be based on molecular Operational Taxonomic Units (OTUs) or Barcode Identification Numbers (BINs), using molecular data to cluster samples into particular groupings that may or may not map onto existing Linnaean species designations.[2] Grey nomenclatures may therefore include 'folk' or ethnobiological taxonomic classifications, but also those deriving from and used in scientific research programmes. The latter differ from Linnaean codes in not constituting a universal standard; they will be more specifically associated with a particular varipraxis or related varipraxes. This does not mean that they are rejected or not widely accepted, but the scope of their deployment and significance may be circumscribed beyond the scientific community or communities particularly associated with them.

DNA barcoding, while providing a means to invigorate existing grey nomenclatures such as OTUs – or forge new ones like BINs – has also needed to connect to the established Linnaean codes. Has this conduit enabled its novel varipraxis and associated grey nomenclatures to transform the ontological commitments of existing taxonomy? To explore the significance and impact of DNA barcoding, we first identify the relationship of DNA barcoding to existing taxonomic practices and outcomes. We do not aim to present either as superior or preferable to the other, or to advocate for one form or another of their interaction, but merely to observe how the relationship between DNA barcoding and existing taxonomy has changed, and to note the irony inherent in the shifts we assess relative to the claims of proponents and critics of barcoding. As this has been the subject of substantial treatment elsewhere, we also elide the considerable socio-political complexity and contestation that was involved in the advent and development of DNA barcoding; for a detailed ethnographically-informed account of this, see Waterton et al. (2013).

[2] On the distinction between nomenclature and taxonomy, the former being the affixing of stable identifiers to specific reference objects (e.g., type specimens held in museums), see Sluys (2021).

After summarising DNA barcoding and its relationship to the wider taxonomic enterprise, we demonstrate how DNA barcoding has operated distinctly for different higher-level taxa, comparing plant DNA barcoding with that of animals and protists.[3] We focus on how DNA barcoding has related to the ontological commitments and working worlds associated with different groups of organisms represented by distinct barcoding efforts.

Although protists do not constitute a monophyletic clade, a working group coalesced around them to establish standard markers and barcoding procedures. We compare the means by which this was achieved, and the particular outcomes. These outcomes include criteria used to identify markers for use in barcodes, and data infrastructures established to enable the identification of species-specific barcodes.[4] We examine one kind of protist in particular, the photosynthetic and phytoplanktonic algae called diatoms, whose primary taxonomic interest lies in their role in environmental monitoring.

Finally, we relate how these barcoding efforts reflect back on the practices and ontological commitments that underpinned them, encouraging an interpretation of genomic variation within particular higher-level taxa that centres on a processual conceptualisation of taxonomic discrimination. This leads to 'thicker' explorations of the wider biology of the organisms involved, in a quest to understand the processes by which the nature of genomic variation is shaped between and within putative taxa. We conclude by adumbrating the consequences of what we have found: for taxonomy, species delimitation, and the investigation and treatment of variation in biology.

DNA barcoding: an introduction

DNA barcoding is the practice of selecting particular stretches of DNA – for example parts of genes – and using particular sequences of nucleotides contained in them as indicators of the species identity of the organism from which the DNA was derived. Samples derived from known species may therefore be used to assign 'barcodes' in data infrastructures pertaining to the sequences typical of members of this species at these selected locations. The language of barcoding reflects the use of barcode labels in retail, and similarly provides a link between a tangible, material object (a product in a shop; an organism or part thereof) and a digital object (the price and stock levels of an item; a species entry in a taxonomic database). DNA barcoding is a solution that relies on the advancements in sequencing technologies made throughout the 1990s and beyond, and on informatics capabilities and possibilities characterising the increasingly data-centric nature of biology in that era. What problems were identified for which DNA barcoding was posited as a solution? We explore this below, and then consider the objections made to it, and alternatives that were presented within taxonomy.

The advent of DNA barcoding, its critics and alternatives

In the early-2000s, multiple schemes were proposed to attempt to manage a perceived crisis in taxonomy. Species were – and still are – going extinct at a rapid rate (e.g., Antonelli et al. 2020), and it was felt that taxonomy was not doing enough to describe them before they ceased to exist, or to

[3] In the genesis of this paper, we began by considering DNA taxonomy in plants, and first developed the arguments in it for the Philosophy of Plant Biology workshop held by the University of Exeter in May 2021. In further developing the work, we have broadened out from plants to a wider perspective, while retaining plants as a key comparative focus.

[4] The key role of data infrastructures in DNA barcoding is not a unique feature within taxonomy, which has been characterised as an "information science" (Kendig and Witteveen 2020) and as a discipline that has faced challenges of managing and organising large amounts of data for hundreds of years (Charmantier and Müller-Wille 2014).

aid conservation efforts to prevent this. Proposed new schemes to invigorate taxonomy included unitary taxonomy and DNA taxonomy. DNA barcoding was another (Waterton et al. 2013).

DNA barcoding was pioneered by Paul Hebert and his team at the University of Guelph, Canada (see, for example: Hebert et al. 2003a). They believed that taxonomic expertise was at that time "collapsing" and had come to the firm conclusion that the only "prospect for a sustainable identification capability [lay] in the construction of systems that employ DNA sequences as taxon 'barcodes'." They determined that the mitochondrial gene cytochrome *c* oxidase I (*COI*) was capable of being used as the "core of a global bioidentification system for animals" and then showed that using *COI* sequences enabled the placing of newly studied taxa into what they termed "the appropriate phylum or order". Next, they established that species-level placings could be derived from examining *COI* sequences. They concluded that if a *COI* identification system were fully developed, it would provide a reliable and practical solution to what they perceived as a crisis in the identification of animal species. Moreover, they predicted that this would lead to new insights into the diversification of animals and the "rules" of molecular evolution. According to these advocates, therefore, establishing DNA barcoding as a new form of *varipraxis* to supplement existing taxonomy would enable the provision of more data of relevance to classification and the exploration of scientific questions relating to change in the DNA of organisms over time, by creating a more restricted, simplified and standardised domain of relevant comparable variation.

DNA barcoding has offered a number of advantages to the taxonomic community. It is easier and more practical for scientists to use in the field than genome sequencing methods. Further, it offers solutions to taxonomic problems that present when only a fragment of an organism is available for analysis, or if often salient morphological features such as size and colour are taxonomically irrelevant.

From its inception, though, some taxonomists criticised DNA barcoding, arguing that it was reductive, in that it was trying to reduce taxonomy to one parameter. In this respect, it was suggested that DNA barcoding represented a backward step. By promoting the use of a single gene as the basis of their barcodes, some said that "DNA barcoders are returning to an ancient, typological, single-character-system approach" (Will et al. 2005). In taxonomy, a character is "any observable difference between two groups of organisms" (Wagner 2001, p. 3), and character states are different forms of a given character. Many developments in taxonomy in the twentieth century transformed the range and use of characters in species delimitation, for example by discarding procedures based on using one or a few diagnostic characters deemed to be particularly biologically meaningful, and replacing them with approaches that used larger numbers of characters, sometimes without assigning greater priority or weight to any particular subset of them. For some critics of barcoding, the new varipraxis of DNA barcoding seemed to be based on a narrower domain of variation on which to base species attributions and classification. Consequently, they judged that rather than boosting endeavours to use a wider range of characters by creating abundant data for assessment of variation, DNA barcoding instead undermined them by ignoring the multi-dimensionality of variation.

Taxonomists are aware that "character variation between and within species" overlaps, meaning that for a given character state, two individual organisms of different species may be more similar than a pair of individuals of the same species. Such overlaps create problems when only a small number of characters are used to distinguish one species from another. In the terms of one polemical account, being "stuck with its single, simple character set […] DNA barcoding has no way to overcome this common phenomenon—unless of course it brings in other genes and morphological characters and becomes integrative taxonomy!" (Will et al. 2005, p. 846).[5]

[5] Or a handful of characters, if one takes the additional barcodes used for non-animals as separate characters.

Integrative taxonomy was articulated in response to DNA barcoding, and advocated the use of multiple characters of very different kinds to resolve the problem of overlapping characters. While some characters used might still overlap, the non-overlapping characters that are included in this approach could be used to help distinguish between different species. The ascertainment and analysis of a large range of characters would also be advantageous as "our level of confidence in species supported by different kinds of data is much higher than for species supported by only one kind" (Dayrat 2005, p. 409).

Integrative taxonomy is an explicitly holistic approach that uses a wide variety of forms of evidence and taxonomic methods. It has been defined by Dayrat (2005, p. 407) as "the science that aims to delimit the units of life's diversity from multiple and complementary perspectives (phylogeography, comparative morphology, population genetics, ecology, development, behaviour, etc.)". This was conceived as multi-disciplinary and pluralist, "because the complexity of species biology requires that species boundaries be studied from multiple, complementary perspectives". Exploring species boundaries would, therefore, involve an appreciation of pertinent patterns of biological diversity and variation, and the multi-disciplinary investigation of the processes that account for these patterns. This would necessitate a varipraxis in which not only is the establishment of the proper extent and diversity of forms of variation to be collected pertinent, but so is the wider biology that allows the patterns of variation so measured to be explained with reference to plural and interconnected mechanisms and processes.

Two early advocates for integrative taxonomy differed in their disposition towards DNA barcoding. Dayrat welcomed it as a potential part of integrative taxonomy, whereas Will and colleagues saw it as something antithetical to it. In this paper, we argue that DNA barcoding does not clash with an integrative approach to taxonomy. Indeed, we demonstrate that it in fact ushers practitioners *towards* an integrative approach, but not in a way that is merely defensive and corrective of the deficiencies of being an approach centred on one or a few characters. Instead, we argue that establishing the basis for being able to assign new barcode sequences to particular species and to cluster samples into separate groupings based on barcode sequences, requires an inspection of the different aspects of the biological processes that affect patterns of sequence variation. The varipraxis that DNA barcoding establishes therefore evolves and expands as practical issues connected to the potential uses of barcoding are encountered by researchers. Following this argument, we also qualify the criticism of DNA barcoding based on it using only one or a few characters, for even if DNA barcodes can be said to be characters, they are not of the kind for which this criticism is appropriate.

DNA barcoding was initially created strictly for species identification purposes, which is conceptually distinct from species discrimination. As the taxonomist John Waugh (2007) has forthrightly stated: DNA barcoding "is not part of a DNA taxonomy nor is it a tool for phylogenetic reconstruction. It simply provides a means of linking sample specimens directly to existing voucher specimens and taxonomical information." Implementing this seemingly limited approach in practice, however, often becomes a task of both species-discrimination and individuation, even though these are in principle separate tasks. Furthermore, we show that the pragmatic species-discriminatory function of DNA barcoding cannot but encroach on issues of the distinction between intra-specific variation and inter-specific variation, and is therefore intimately entangled with wider DNA taxonomy and phylogenetic research.[6]

[6] For simplicity, we will only refer to the distinction between intra- and inter-specific variation. Taxonomists also recognise, however, infra-specific variation, which pertains to variation between sub-species types.

Implementing DNA barcoding – challenges and practices

One of the key requirements of DNA barcoding is the ability to establish and implement standard gene variants for particular higher-level taxa. This focuses attention on the distinction between patterns of **intra-specific** (within-species) and **inter-specific** (between-species) variation.[7] The challenge of establishing this difference in order to distinguish between sequences that pertain to one species and those that belong to another species constitutes the problem known as the 'barcoding gap'. In order to resolve this issue, taxonomists need to be able to provide valid and widely accepted answers to the following questions:

- What sequence differences between samples are sufficient to assign each to a different species?
- What would constitute valid intra-specific variation?
- What gap is meaningful, and what gaps are meaningful within particular taxa, since there are no uniform criteria that can be used across the diversity of life?

Species-discrimination practices differ according to the higher-level taxa of interest, existing ontological commitments, and the working worlds of biologists. These are not independent factors; the objects of investigation, intended uses of the resultant classifications, and theoretical articulations of the nature of species and evolutionary processes will be interrelated in any coherent scheme of species discrimination.[8] How can taxonomists ensure distinctions between sequence differences attributed to intra-specific or inter-specific variation are non-arbitrary and appropriate to a particular taxon? Some anchoring in a stable taxonomy is required in order to make assessments about the meaningfulness of distinctions between intra- and inter-specific variation and therefore identify meaningful gaps. How, therefore, in situations where the existing taxonomy is not established or stable, can standard DNA barcodes be established?

To begin to answer these questions, we need to appreciate the way that DNA barcoding has come to generate standardised workflows and protocols, appropriate to the production of data and metadata, that will enter and circulate in large-scale data infrastructures.

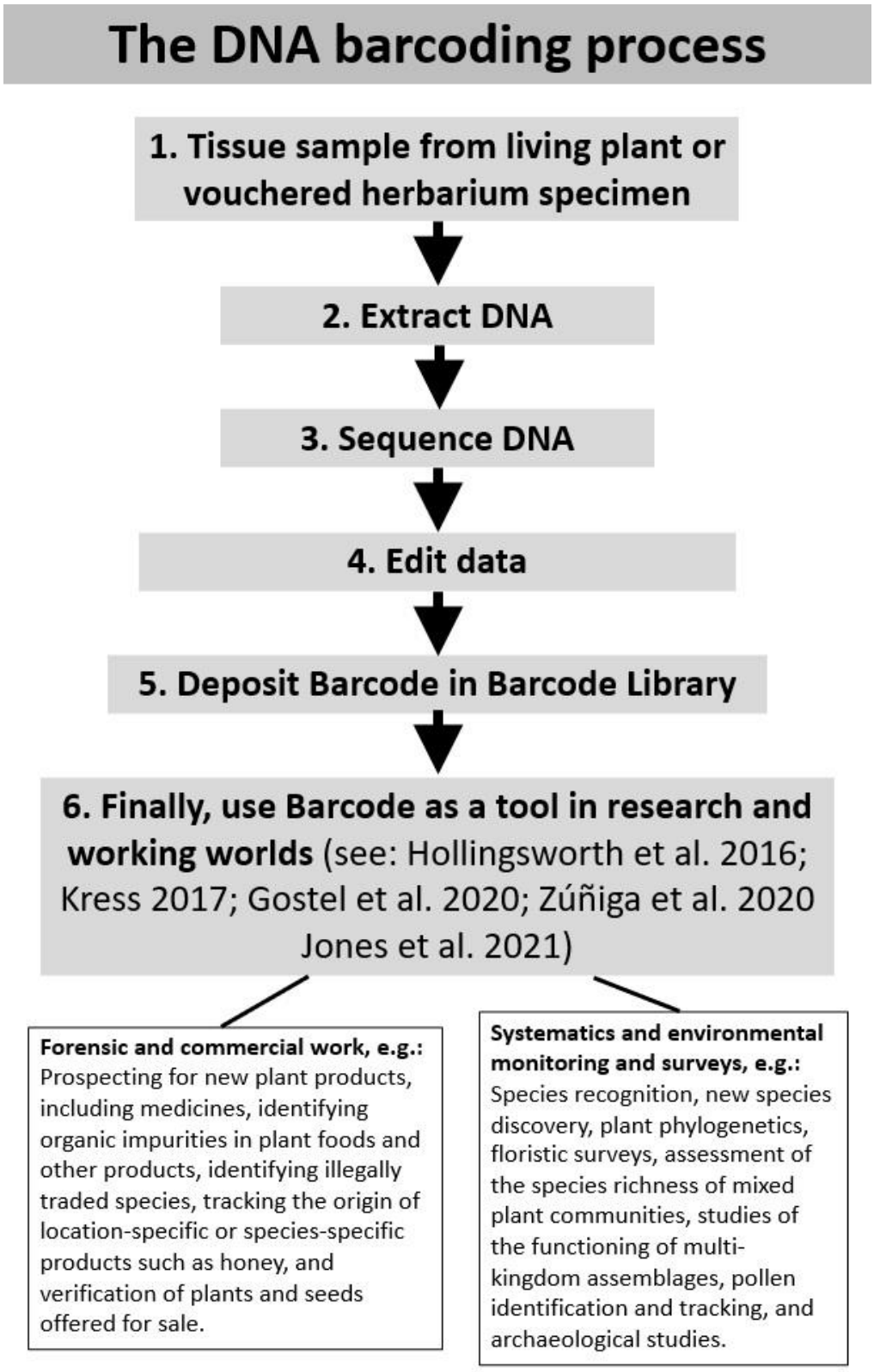
For plant DNA barcoding, the workflow starts with living plant tissue samples and vouchered herbarium specimens, and proceeds through generating DNA barcode sequences to build a barcode library for use in taxonomic identification, species discovery, and various applications (Figure 1). It is first necessary to take a tissue sample, either in the field or from an herbarium sheet, extract the DNA, sequence it, edit it, add metadata to it, and then add the resulting barcode to a barcode library in such a way that it can also be compared with all other barcodes recorded in that library. Thus, if

[7] The distinction between patterns characteristic of intra-specific variation and those of inter-specific variation does not mean that DNA barcoding aims to artificially introduce a discontinuity into known continuous change and variation between species, as alleged by some philosophers (Piotrowska 2009). Rather, it aims to discern alterations in the mode or pattern of continuous variation that ranges across putative species boundaries.

[8] Often these will be reflective of distinct "species concepts" either explicitly held or attributed to researchers and practitioners (for extensive analytical discussions of these and the history of species as a category, see Zachos 2016; Wilkins 2018). For example, the Biological Species Concept formulated by Ernst Mayr held reproductive compatibility to be the key criterion of species membership. He deemed geographical separation of lineages to be a main driver of speciation along these lines. This outlook was based on his primary research focus on sexually-reproducing animals and an emphasis on adaptation as a driver of species change. Distinct perspectives on the processes underpinning speciation have led to challenges to the Biological Species Concept, and the species discrimination approach implied by it, for instance: by microbiologists who study organisms that do not sexually reproduce but for which there is much 'horizontal' genetic transfer potentially between members of different species (Wilkins 2018, pp. 317-330); or by considering inheritance beyond the genome and dynamic relationships between organisms and environments (as in niche construction theory as outlined by Kevin Lala and others; Kendig 2014).

the species from which the sample was taken is unknown, it may then be compared with the records in the library as a whole in an attempt to assign it to a specific species (or alternative taxon). In the case of samples from herbarium sheets, links to the barcode record may then be attached to provide continuity with existing taxonomic practice in both practical and temporal terms.

Figure 1 – Depiction of the plant DNA barcoding process, developed from Kress (2017).



A key element of barcoding is a standard reference system, such as the Barcode of Life Data System (BOLD; http://www.boldsystems.org/), in which standard barcodes are attached to species names in the Linnaean hierarchy. Once it had dealt with the easy cases,[9] DNA barcoding moved beyond mere identification and fitting names into existing pigeonholes, as the lists of applications and references in Figure 1 above demonstrate.[10] DNA barcoding constitutes a platform technology (Fitzhugh 2006; Shokralla et al. 2014) that enables individual institutions that adopt it to diversify into a wide array of activities and possible collaborations.[11] In this sense, DNA barcodes function as polychrests – tools or materials with many potential uses (Werrett 2019) – leading to a diversity of applications and connections to working worlds.

It is important to note at this point that institutions such as botanic gardens and natural history museums that have adopted DNA barcoding for plant taxonomic studies across multiple working worlds, are not only involved with the terrestrial environment, but with aquatic environments too.[12] Indeed, a key working world that we wish to consider here is aquatic environmental monitoring of rivers, lakes, and seas. This practice is laid down in legislation and regulation such as the Water Framework Directive of the European Union (https://eceuropa.eu), and involves the identification of species in particular areas as indicators of water quality, whether good or bad (Kelly et al. 2018). Both plant and diatom DNA barcoding have a role in environmental monitoring: plants for land-based ecosystems, from rainforests to meadows, and diatoms for aquatic ecosystems, from puddles to oceans. Plants have a well-established and relatively stable taxonomy resulting from extensive research based in such institutions as botanic gardens and natural history museums, as well as universities. Diatom taxonomy, by contrast, is less well resourced, and in consequence less well studied. Much of the support it does receive has come from its utility for environmental monitoring.

DNA barcoding across animals, plants and protists

There are different ways in which DNA barcoding has been developed and operated for different higher-level taxa (see Table 1). DNA barcoding began in zoological taxonomy (see Hebert et al. 2003a), where the barcode *COI* was chosen because it exhibits, consistently across the Kingdom Animalia, sufficient change in sequence variation over evolutionary time as to make it ideal for discriminating across all animals. It is also highly recoverable, meaning that its sequence can often be successfully retrieved from samples. In plants, however, the extent to which the *COI* gene varies across evolutionary time was found to be too inconsistent for it to be used as the universal barcode. Instead, the community identified two barcodes, *rbcL* (a chloroplast gene) and *matK* (a plastid gene), with two additional ones identified in case the original two did not perform as desired: the *tnrL–psbA* plastid and ITS (nuclear ribosomal internal transcribed spacer) (CBOL Plant Working Group 2009). These barcodes, like *COI*, met the requirement of recoverability, but also fulfilled the more direct criteria of being recoverable using a range of different primers for amplification of the initial sample, and ending up with high-quality sequences. The Consortium for the Barcode of Life (CBOL) Plant Working Group operated on the basis that the barcoding enterprise should ideally constitute "an open-access shared resource without constraints or patents limiting the use of regions and primer sets", though the usefulness of the *tnrL–psbA* plastid as a barcode meant that it became

[9] Easy cases would include barcoding samples in the field that match established barcodes for stable taxonomic designations, or barcoding samples of known species to add to a database.
[10] Comparison of the cited papers also shows that for plants, the ways in which the tools may be used to best effect are rapidly evolving.
[11] Though on the limitations and partialities of the resource and knowledge "commons" underpinning DNA barcoding's role as a platform, see Geary and Bubela (2019) and Geary et al. (2019).
[12] A significant proportion of the earth's photosynthesis occurs in aquatic environments (https://oceanservice.noaa.gov).

adopted as one in spite of the primers for part of it being patented, albeit made available for non-commercial use (Hollingsworth et al. 2011).

Protists include any eukaryotic organism not classified as animals, fungi or plants. As a result, it is very much a heterogenous category, including diatoms, and multifarious other organisms such as amoeba. Due to the diversity within this grouping – which does not form a monophyletic clade – the CBOL Protist Working Group that formed to develop and apply criteria to identify appropriate barcodes forged a hierarchical selective approach. Thus, using similar criteria employed in the search for standard plant barcodes, the Protist Working Group identified 18S (small subunit) ribosomal DNA (rDNA) as a 'pre-barcode' to identify the type of protist (e.g., amoebae or Stramenopiles, the latter being the clade that includes diatoms), with specific barcodes then being used depending on the type of protist so identified (Pawlowski et al. 2012). For diatoms, *rbcL* was in large part chosen because amid the relative paucity of taxonomic resources available for the group, there was sufficient existing data on this gene to make it viable to adopt. This decision reflects a general tendency in biology whereby, if an object has sufficient resources attached to it, further use, development and investment on and around it usually follows, as also occurs following the adoption and deployment of a model organism (Ankeny and Leonelli 2020).

Table 1 - Summary of the selection process, criteria and decisions for identifying and using DNA barcodes across the animal and plant kingdoms, and protists.

	Animals	**Plants**	**Protists (diatoms)**
Selection process	Work of Hebert et al. (e.g., 2003a & 2003b) establishing the mitochondrial gene Cytochrome c oxidase subunit I (*COI*).	Initial scoping work at 13 institutions. Plant Working Group of Consortium for the Barcode of Life (CBOL Plant Working Group 2009).	Application of population genetics 'model system', *Sellaphora pupula* complex. Application to other complexes. CBOL Protist Working Group (Hamsher et al. 2011; Pawlowski et al. 2012).
Selection criteria	Justification of *COI*: high recoverability; variability; discriminatory power across kingdom.	Universality (recoverability); discriminatory power; sequence quality; free and open access.	Universality (recoverability); discriminatory power; sequence quality; extent of existing data; already used in taxonomic practice.
Barcodes chosen	*COI*	*rbcL* and *matK* (CBOL PWG 2009) *tnrL–psbA* plastid and ITS (nuclear ribosomal internal transcribed spacer) added subsequently.	Variable V4 region of 18S rDNA 'pre-barcode' across protists. Then group-specific ones: *rbcL* for diatoms.
Deployment of barcodes	Single. Calls for additional barcodes due to limits of the mitochondrial gene.	2+2 – the two additional added for problematic tasks; now a group of four that can be selected from.	Hierarchical and selective. Narrow-down using 18S rDNA. Then use barcode for group (*rbcL*)

It is important to emphasise that the functions of the genes chosen for use in DNA barcoding are irrelevant; the key considerations instead being whether they allow discrimination between species or other designated groups, and that they are easy to work with.[13] Their relevance is primarily related to the pace and consistency of changes in their sequences over evolutionary time. Different selection criteria for identifying barcodes entail different modes of deployment. They also reflect the composition of the communities developing and implementing the barcoding, and the existing state of the taxonomy in given higher-level taxa.

Crucially, the elaboration and deployment of particular sets of selection criteria to identify appropriate barcodes undermines several linked criticisms of the use of DNA barcodes. Observing, as we have, the inseparability in practice of species identification and species discrimination, Piotrowska (2009) claims that this means that DNA barcoding uses implicit theoretical conceptions of the nature of species and speciation, and these are then used by barcoders with their limited number of characters – the barcodes – in allocating individuals to particular species. This critique echoes the case made against DNA barcoding by Will et al. (2005).

While some aspects of this argument will be addressed later in the paper, for now we note two counter-arguments. One is that many barcoders – especially those building barcode reference libraries – are taxonomists, and as we shall see, their barcoding activity and interpretations of its results are conducted as part of their wider taxonomic research and approach rather than being separate and distinct. To the extent that particular species concepts and theories of speciation are imported into distinctions made through barcoding, these are done so against the background of the taxonomist's appreciation of the nature of the biological kinds they work with and the pertinent evolutionary processes involved in shaping inter-specific and intra-specific variation and change at the sequence level. This background also enters the practices of DNA barcoding itself through the delineation of criteria for the selection and use of barcodes, and the choices subsequently made by barcoders. If particular conceptions of species enter implicitly into barcoding, then, they do so as part of the activities and ontological commitments of taxonomic research more broadly. As such, to the extent that this is a problem, it is not distinctive of barcoding.

This brings us to the second point, which is that given barcodes are selected to be informative about species distinctions for given higher-level taxa; they are intended to track the patterns of variation that result from multiple evolutionary processes that impinge on the DNA sequences in barcode regions and are imprinted in them. Far from being divorced from the multiple evolutionary processes that result in speciation and change across time, the non-arbitrary selection of barcodes is precisely intended to capture this. For this reason, it would not be accurate to claim that in specifying a particular DNA sequence or range of sequences as indicative of membership of a given species rather than another, that DNA barcodes constitute a "microstructural essence" (after Wilkins 2018, pp. 294-298) constitutive of that species. There is no meaningful causal, explanatory or definitional relationship that obtains from barcode to the species class and its membership. The arrow is, rather, operative in the reverse direction, with the DNA barcode identified precisely because it has the capacity to be indicative of the species class and its membership. The barcode or barcodes are pertinent precisely because they are a particularly informative token of the evolutionary processes that gave rise to species and reproduced them within particular higher-level taxa.

[13] The function(s) of the barcoding genes are irrelevant in and of themselves, but may be relevant more directly in terms of the processes of genomic evolution that shaped them. Although it uses gene sequences directly as the basis for its approach, as it is unconcerned with the causal connection between those sequences and the nature of some phenotypic trait, DNA barcoding does not therefore imply a commitment to any form of genetic determinism.

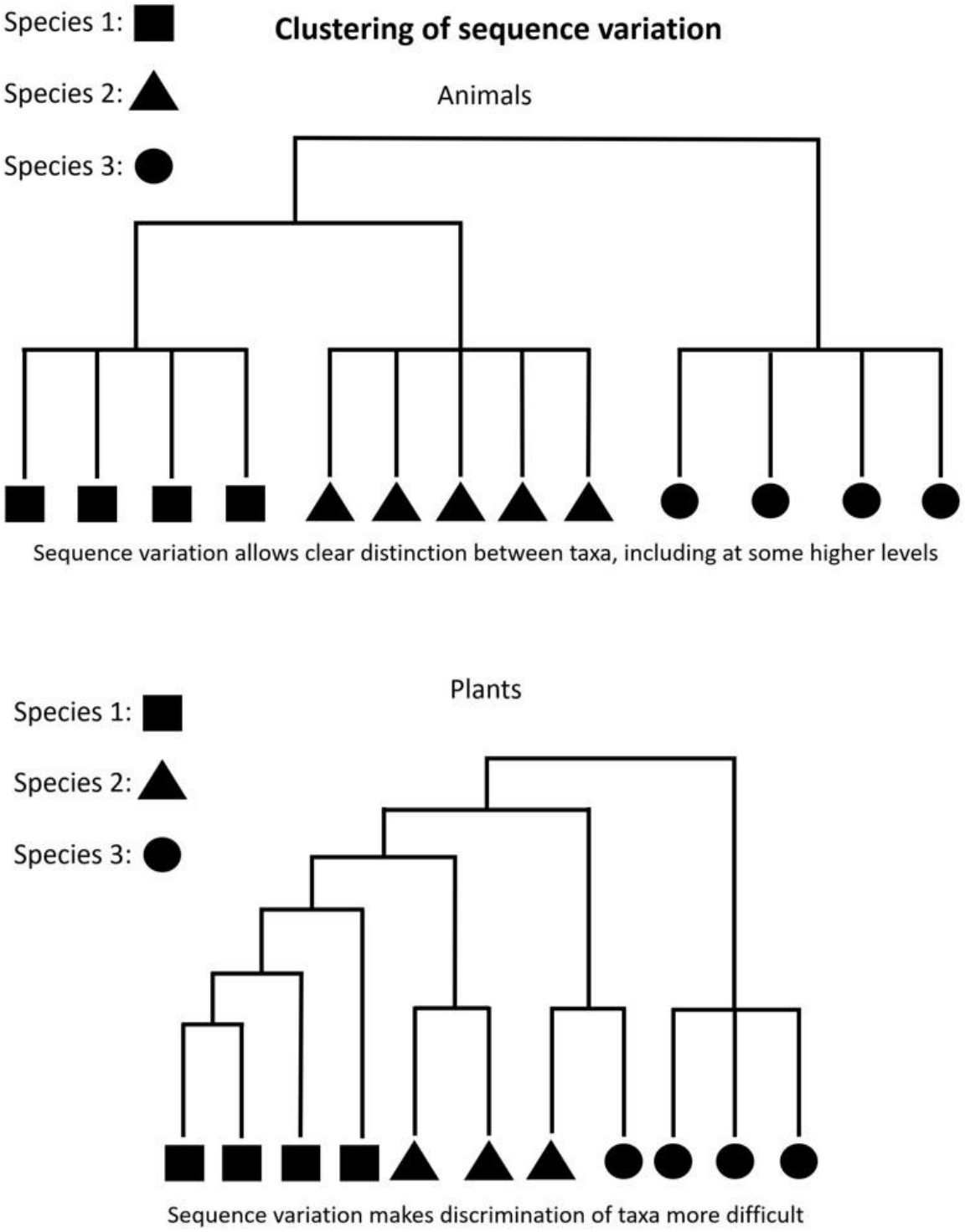
Furthermore, as the barcodes are carefully-chosen tools for tracking the results of biological processes, it is not the case that "DNA barcoders must be committed to the idea that a mutation at the *CO1* location that changes similarity to below 97.5% constitutes a speciation event" (Piotrowska 2009, p. 237; '*CO1*', not italicised in the original text cited, is an alternative identifier for the *COI* gene). The problem of identifying particular similarity thresholds is not trivial, and cases at the boundary do arise that make identification and discrimination difficult. Identifying particular thresholds – the problem of the barcoding gap – is, though, a spur to further research at the interface of barcoding practice and wider taxonomic research. It is not itself stipulative of species boundaries, but the result of research by taxonomists who interrogate such boundaries and the processes that define or blur them, and bring the results of these explorations into their shaping and implementation of barcoding practices.

Ontological commitments

To what extent has DNA barcoding affected the ontological commitments of taxonomy and taxonomists? In one respect, it has enabled the fresh delineation of grey nomenclatures, in the form of alternative categories to the Linnaean species, such as Barcode Identification Numbers (BINs) and Operational Taxonomic Units (OTUs). OTUs were introduced in the 1960s, but their identification and cataloguing gained impetus through the availability of the kinds of data generated by DNA barcoding, and the infrastructures put in place to implement it (Blaxter et al. 2005). OTUs originated as part of the development of numerical taxonomy, in which classifications were made on the basis of quantitative measures of similarity between organisms based on the assessment of multiple characters, none of which was weighted or attributed a significance more highly than any other. This commitment to using a variety of characters and character states, without *a priori* identification of particularly salient ones as prior methodological approaches in taxonomy had favoured (Sneath 1995; Vernon 1988; for a philosophical assessment of the commitments of numerical taxonomy, see Sterner 2014), was an important innovation of numerical taxonomy. As a result, numerical taxonomists aimed to cluster organisms into groups based on data relating to manifold character states, and were able to harness the computational resources that became available in the 1960s to achieve this (although some historians have suggested that advances in computing prompted the development of numerical taxonomy in the first place: Vernon 1988). The later advent of integrative taxonomy, in response to the development of approaches such as DNA barcoding, also emphasised the use of a wide range of characters reflecting the biology and ecology of organisms more holistically (Dayrat 2005; Will et al. 2005).

In using data to cluster organisms based on analyses of their similarity, however, numerical taxonomy appears to have been a forerunner of the ontological commitments and practices of DNA barcoding. When the DNA sequencing data concerning the barcode genes emerge, practitioners normally cluster the obtained barcoding data. Such clustering patterns help to explain some of the aspects and problems of species discrimination that were experienced before barcoding was developed. In animals, the clustering of sequence variation across species is very clear; it also enables the distinction between higher-level types. In plants, however, the clustering is less distinct and is therefore difficult to disentangle. Metaphors used to characterise clustering patterns are, for animals, a "garden rake" and for plants a "witch's broom" (e.g., Hollingsworth et al. 2016; see Figure 2).

Figure 2 – Depiction of the patterns produced by the clustering of sequence variation in animals and plants.



Information from clustering patterns may help to explain certain characteristics concerning species discrimination in particular high-level taxa, for example that among:

- **animals**, species are relatively easy to discriminate;
- **plants**, species-discrimination is more difficult than for animals due to greater prevalence of morphotypes, extensive hybridisation, alternation of generations and complexity of life histories;

- **diatoms**, the taxonomy is very unstable as a result of considerable morphological intra-specific variation, notably in size and form (due to the nature of the asexual reproduction phase), confusing distributions, and the presence of species complexes – groups of related organisms among which it is difficult, if not impossible, to demarcate discrete species – in particular environments.

Animals exhibit clearer differences between intra- and inter-specific variation than plants do. For diatoms, however, in addition to the reasons adduced above, certain features of the taxonomic research on it may also be responsible for the instability of its taxonomy. Some diatom taxonomists have contended that these features may be ameliorated by DNA barcoding (e.g., Evans et al. 2007), as we describe now.

Diatom taxonomy and barcoding

As noted above, diatom classification, like that of many other protists, presents multiple challenges for taxonomists. In addition to the problems of considerable morphological diversity and the existence of species complexes, taxonomist David Mann has suggested that their unstable taxonomy is due to limited sampling (Mann 1999). Specimens used may be idiosyncratically selected and this prevents a proper appreciation of the pattern of variation in and across putative types. In being able to contribute considerably more data than laborious morphological descriptions may provide, DNA barcoding therefore holds out the promise of at least removing the partiality of sampling as a factor inhibiting the further development and stabilising of diatom taxonomy.

In species discrimination, it should be possible to distinguish between intra- and inter-specific variation. The greater sampling made possible by DNA barcoding is therefore important in helping to appreciate the distinction between these, and therefore between species. Different patterns of variation might be expected when comparing individuals within species from when comparing individuals in different species. Species discrimination depends not only on the apprehension and identification of variation between species, but within them as well, and understanding the patterns of variation at and between different levels in the taxonomic hierarchy. This, in turn, implies the investigation of the processes by which the observed patterns of variation are manifested.

Even before the development of DNA barcoding, Mann (1999) argued that a holistic understanding of diatom biology – "diatomics" – was crucial to making taxonomic distinctions:

> "the only mechanism that integrates populations and leads to the formation of a boundary between hierarchical and non-hierarchical variation is gene flow, brought about principally by sexual reproduction, in conjunction with dispersal and migration. Thus, it can never be irrelevant for taxonomists to know about the mating system, phenology, and spatial dynamics of the organisms they study."

Being able to appreciate what the patterns of variation are in diatoms, both intra-specifically and inter-specifically, is therefore vital to being able to make taxonomic judgements. This, in turn, requires an understanding of the processes by which that variation is generated and maintained in that particular form; why and at what rate substitutions and other changes are made at the nucleotide level; whether there are species-specific substitutions in the DNA; and whether there are different frequency distributions of these across species.

Diatomics illustrates how the practicalities of species discrimination opens up the need to understand the processes by which genomic variation is structured within and across species. It is a

processual conceptualisation of taxonomic discrimination. This leads to deeper explorations of the wider biology of the organisms involved, in a quest to understand the processes by which the nature of genomic variation is shaped between and within putative taxa.

It is important to note that this approach was advocated before DNA barcoding was developed, because of the difficulties of classifying diatoms according to Linnaean nomenclature. The advent of DNA barcoding has encouraged this approach for other higher-level taxa too.

The impact of DNA barcoding

It is clear from the above that an interest in the biological processes underlying observed patterns of variation existed before DNA barcoding was developed. DNA barcoding provided added impetus to this holistic impulse. It did so negatively, in encouraging the formulation of integrative taxonomy, which closely resembles diatomics in encouraging the examination of the processes by which genomic variation between conspecifics and individuals of different species manifests, and attempting to relate these multiple lines of evidence to taxonomic concerns. However, DNA barcoding has also – inadvertently – encouraged a more holistic and contextual approach to taxonomy in a positive way, especially for animal and plant taxonomy. Compared to existing taxonomic practices, barcoding was founded as a new varipraxis with a more restricted basis of variation to work from, that was intended to enable more abundant collection, recording and integration into informatics infrastructures. The practicalities of implementing it, however, opened the door to this new varipraxis becoming more expansive and connected to other varipraxes.

How and why should this be so? In pursuing species discrimination using DNA barcoding, taxonomists become aware of the need to explain the patterns of variation they find, and in so doing pursue more fundamental biological research. Through that, they revise their ontological commitments.

For species identification, it is first necessary to identify the means to discriminate between species. To do that, the patterns and extent of intra-specific variation need to be identified and analysed. Also, there is a need to determine what constitutes the difference between intra- and inter-specific variation in any given taxa. This will differ between and among taxa, and in order to avoid being stipulative and arbitrary it needs to be biologically meaningful, which prompts a considerable body of biological research into what is underpinning those patterns of variation. This encourages not only a more integrative approach, but also feeds into long-standing approaches. For example, in nematode taxonomy, DNA barcoding can be used to narrow down the options for performing breeding experiments to identify which individuals belong to which species, using more traditional taxonomic criteria. Through this, DNA barcoding radically reduces the number of experiments required from millions to a much smaller and manageable figure (Stevens et al. 2019).

For *Lumbricus rubellus*, the red earthworm, barcoding analysis identified at least two distinct populations that were found to exhibit minimal interbreeding in the wild (King et al. 2008).[14] This led researchers to pose questions about why these genomically-distinct, but morphologically-indistinguishable types had arisen. Since this species is used to assess the toxicity of agrichemicals to soil organisms, the existence of cryptic species of *L. rubellus* prompted the concern that these different species may react in distinct ways to environmental challenges. In turn, this called the reproducibility of testing into question, as different regulatory organisations may unknowingly use different cryptic taxa in their assays. Indeed, different *L. rubellus* cryptic taxa exhibit differential

[14] Though some hybrids were found, and they could be induced to breed in the laboratory, albeit less readily than individuals within these populations; interview with Mark Blaxter, conducted over Zoom by James Lowe and David Ingram, 26th May 2020.

tolerance to heavy metals and other chemicals harmful to many lifeforms (Andre et al. 2010; Anderson et al. 2017). Beyond this practical, working world consequence of the findings of DNA barcoding, the exploration of the mechanisms of genomic differentiation within a previously unproblematically defined species can help identify particular processes and mechanisms that could inform taxonomic work on other organisms for which large-scale sampling of populations may not be possible.[15]

We therefore observe that not only has DNA barcoding given succour to a form of integrative taxonomy that was supposedly its antithesis, but it has also fed into more established ways of doing taxonomy. The identification, selection and analysis of the patterns of variation used in barcoding are not independent of: taxonomic practice and history, the working world concerns that drive taxonomic research more broadly, or species-discrimination and barcoding research more specifically.[16] This may not have been the case had *COI* been adopted as a truly universal barcode as planned, but for a variety of reasons – including evolutionary history – it was not. Indeed,

We further suggest that the nature of DNA barcodes means that, to the extent that they can be deemed to be characters, they are characters of a distinct and special type. They do not suffer from the usual limitations of using only a limited set of characters, such as the overlapping problem or the *a priori* designation of a particular structure or function as being especially salient for classification. The critique of Will et al. (2005) or the marginalisation of barcoding as merely being one component of an integrative taxonomy (Dayrat 2005) therefore cannot be sustained. We adduce this peculiar nature of barcodes as characters, as a potential explanation of why the exigencies of DNA barcoding has encouraged work in a more integrative direction.

What is special about DNA barcodes is that, even though particular barcodes have been chosen due to their being especially informative for species discrimination, there is nothing about the barcodes themselves that mean that they have to be applied only at this level. Indeed, in protists, the 'pre-barcode' is intended to discriminate at a much higher level than that of the species. DNA barcodes are, singly or in combination with a few other barcodes, applicable across a whole kingdom. In this way, we deem them to be considerably more indifferent to – or potentially independent of – taxonomic groups below the level of kingdom than other characters. This is aided by their lack of specificity compared to many diagnostic characters, which may depend on particular aspects of the morphology of related species that can be differentiated, e.g., numbers of bristles or pistil shape.[17]

Patterns of variation characteristic of the higher-level taxon and existing taxonomic resources affect changes to ontological commitments and practices. Overall, Linnaean nomenclature and international codes have continued to be central to the taxonomic enterprise. There are clear constraints on the revolutionary, eliminative or reductive impact of barcoding on taxonomy, and there has been continuity with the legacy practices and institutions of taxonomy. The taxonomic origin of DNA barcoding is important; additionally, other working world contexts are salient. Taxonomists formulated DNA barcoding and promoted it to other taxonomists. The continuity, precedence and stability emphasised in taxonomic practice tempers the transformative potential of

[15] Interview with Mark Blaxter, conducted over Zoom by James Lowe and David Ingram, 26th May 2020.
[16] Key here is the treatment of the *pattern of variation*, not merely the identification of intra-specific variation or differences between species. This encompasses the pattern of variation that goes beyond putative species boundaries.
[17] The selection and demarcation of characters cannot be theory-free or theory-independent (Harris and Mishler 2009). We do not intend to imply that barcodes depart from this; in line with our discussion of barcodes as characters above, however, we emphasise that the theory and values implicit in their delineation and use come from their embedding in and relation to taxonomy and – indirectly – other wider working worlds.

genomics, but makes it more immediately applicable to existing communities. Furthermore, in taxonomy new resources and designations must be made compatible with existing ones, and enable compatibility with potential future forms.

Table 2 – Table depicting the characteristic sequence clustering patterns for the animal, plant and protist kingdoms, alongside commentary concerning the effects of barcoding on the taxonomy of these kingdoms.

	Animals	Plants	Protists (diatoms)
Pattern of sequence clustering	**Garden rake** **Distinct**: clear discontinuities between different species.	**Witch's broom** **Less distinct**: variety of more to less clear separation between intra- and inter-specific variation across the kingdom.	**Reticulate – networked** **Coarse**: very difficult to distinguish between species on any consistent basis.
Existing stability and quality of taxonomy	**Very high**, with extensive resources.	**High**, with extensive resources.	**Low**, relatively sparse existing taxonomic resources.
Post-barcoding alterations to varipraxis	Increasing confidence of being able to define **grey nomenclatures** on the basis of barcoding data in addition to complementing existing taxonomic approaches.	Barcoding data **supplement existing classifications and taxonomic practices**. Impetus towards adopting **integrative taxonomy**.	In absence of stable pre-barcoding taxonomy, though barcoding data less determinative than for plants, it has revealed extensive variation and therefore **placed existing taxonomic debates & practices on a different footing. 'Diatomics' re-affirmed.**

What are the impacts of DNA barcoding on particular taxa (see Table 2)? We note a few particular points here. For plants, DNA sequence clustering is generally less discriminatory than for animals: grey nomenclatures based on this are therefore more uncertain. The impact of DNA barcoding is most felt when it constitutes a platform technology for entering into wider array of other areas. For diatoms, the pre-barcoding taxonomy was poorer and less stable. Barcoding has therefore had a greater impact on shaping the taxonomy. Additionally, barcoding has enriched an existing integrative disposition.

Why are grey nomenclatures not adopted instead of this biologicisation of taxonomic problems? We suspect that this may be due to pre-existing ontological commitments to Linnaean nomenclatures, as mentioned previously. These may result in part from the role of taxonomists themselves in developing the standards and protocols for DNA barcoding, and applying these to their work. It may also reflect the demands of the relevant working worlds, (Linnaean) species discrimination being required for practical – and often legal and regulatory – processes. Using OTUs or BINs simply based on sequence clustering without regard to the existing Linnaean hierarchy or codes would not be appropriate if doing work of a regulatory or legal nature that required the identification of particular Linnaean species.

It is therefore interesting that grey nomenclatures do seem to have been more enthusiastically adopted by animal barcoders. Why should this be so? One possibility is that they are more likely to be insulated from working world concerns (e.g., research on *C. elegans* species in the wild). More important, perhaps, is that zoologists have a more universal barcode to work with than botanists, and the patterns of change in that barcode are more consistent than for plants. Grey nomenclatures are therefore more reliable and trustworthy.

The delineation of sub-typic variation and types inspires and is informed by the creation of knowledge concerning the biological processes by which measurable differences are generated. In turn, the greater appreciation of the structures and patterns of variation at different hierarchical levels provides a firmer grounding for biological research concerning the processes by which this observed variation results. DNA barcoding therefore instantiates a varipraxis with a dynamic cycle operating between the ongoing establishment of the structure of variation and its causes, which requires understanding evolutionary and ecological dynamics as well as forming a richer basis of knowledge concerning organismal biology and molecular genetics. The expansion outwards towards the wider biology around taxonomy therefore cannot just operate within single taxa if patterns of variation must be apprehended and understood. Community dynamics are crucial to understanding the biology of plants and their interactants that underpin these patterns of variation, for instance. An example is the relationship between plants and pathogens. For example, one hypothesis that has long guided plant ecology research is that interspecific competition is the principal driving force in the organisation of unmanaged and partially managed plant communities (Tilman 1982). Various mechanisms have been suggested as being responsible for this, including resource partitioning, life history variation, and the effects of animal feeders. However, during recent decades, some ecologists have focused on the possibility that plant pathogens may also be involved (Ingram 2022). If this is so, it seems likely that the coevolution of plants and pathogens in such communities would need to be controlled by a trade-off between "the costs of virulence to the pathogens and of resistance to the plant hosts" (Ingram 2022, p. 103).

Evidence for this possibility includes that (Bever et al. 2015):

- Plant pathogens are major components of, and contribute significantly to, community complexity;
- Balanced co-existence of plants and pathogens is the norm in communities;
- Experimental evidence of feedback between plant and soil microbial communities in forests;
- Vigour of invasive alien plants, free of pathogen loads (enemy release hypothesis) compared with the relative co-existence of native plants & pathogens.

Metabarcoding of communities incorporating various higher-level taxa is therefore required both to understand the biology and structure of these communities in themselves, but also the biological processes explaining observed patterns of variation for taxonomic purposes. This will likely require that ontological commitments and varipraxes be aligned at multiple levels and dimensions of analysis and investigation.

Consequences

We conclude by adumbrating the consequences of what we have argued for taxonomy, species delimitation, and the role of the apprehension, recording and analysis of variation in biological research. To conceptualise the latter, we have introduced the concept of *varipraxis* to capture the entangled methodological, epistemic and ontological features of the elucidation of variation and its meaning for research problems in the biological sciences.

This paper has shown that DNA barcoding, far from instantiating and institutionalising a narrow and reductive vision of taxonomy, has opened up practitioners to a more integrative approach. This is not to correct for deficiencies inherent in the narrowness of the enterprise, but to establish an empirical and theoretical basis for delineating between intra-specific and inter-specific variation for particular taxa, and from this to define principles for the analysis of other taxa.

The exigencies of their work have led DNA barcoding practitioners to advance their endeavour in a more integrated way, including by establishing collaborations with biological researchers examining particular aspects of the biology of target organisms. An example of this, early in the days of DNA barcoding, was the advent of a research programme on the nature and causes of genomic differentiation among apparently morphologically indistinguishable populations of the earthworm *Lumbricus rubellus*. As well as inspiring such ecological and physiological research, barcoding methods have also been neatly integrated with existing approaches in taxonomy, including breeding experiments and field studies. It has done so without reducing these other facets of taxonomy to itself. It has not become a mere cog in a wider integrative taxonomy either, but a distinct endeavour. We suggest that one of the ways in which DNA barcoding has led to the adoption of integrative approaches – without liquidating itself to a mere method in a larger toolkit – has been because of the distinct object of taxonomic information and inference that the barcodes constitute. Far from representing a single point of information as a character, as the critiques from advocates of integrative taxonomy have stated, these barcodes can vary in sufficient ways across different samples and are taxon-indifferent even if they were selected to present sufficient differences between species. Through the process by which they are selected, the nature of the variation in them is greater and more continuous than a small number of alternative character states (including those for which a continuous series has been split into discrete states for the purposes of taxonomic analysis) would be. To the extent that they are characters, barcodes are not the kind of characters that one should be concerned to use a limited set of.

However, a more detailed examination of how DNA barcoding and other genomic approaches relate to the history, practices, traditions and institutions of taxonomy, including changes in the articulation and use of characters, is warranted. This will include an inspection of the extent to which DNA barcoding may transform a tension in taxonomy whereby "that every effort to make taxonomy more stable made it appear even less scientific, while every effort to make it more scientific seemed to make it less stable" (Endersby 2018, p. 448). Our examination of DNA barcoding, together with the integrative taxonomy that was partly founded in response to it, appears to affirm John S. Wilkins' contention that species are explicanda, "objects that have phenomenal salience" [italics in original removed] whose distinctness is a matter to be explained, rather than them being objects from which explanations can be derived (Wilkins 2018, p. 344). In constituting a practice that notices differences between objects that are salient phenomena – such as the existence of separate cryptic species, which has significance in various working worlds, such as environmental monitoring – DNA barcoding is therefore a practice that makes meaningful contributions to species delimitation.

Due to their vital working world applications, Linnaean species designations remain central to taxonomy. Although various grey nomenclatures – whether informal in the sense of vernacular descriptions, or more formal through the delineation of new categories such as OTUs and BINs – have been advocated, none have eclipsed the traditional classificatory framework. The ongoing development and use of grey nomenclatures reflects a desire to capture distinctions between in-group variation and out-group variation that does not necessarily coincide with the means of distinguishing different categories in the Linnaean framework. For example, the recent emphasis in conservation on conserving variation means that other means of categorising the diversity of nature may be appropriate. DNA barcoding, through using a grey nomenclature (BINs) and needing to connect the objects discerned in that to standard nomenclatures, may provide insights into the ways

in which translations between different nomenclatures – including those with distinct ontological commitments – may be effected.

One potential avenue for DNA barcoding is not merely species discrimination concerning samples that clearly belong to a particular kingdom, but in examining multi-kingdom communities. As we have shown, there are specificities in the implementation and impact of barcoding across kingdoms, and there is a need for an integrative – non-reductive – approach in order to make barcoding work. Is it therefore the case that in order to barcode communities, one needs to study and understand community dynamics? This may indeed be the case for barcoding within kingdoms as well, since community dynamics may shape the processes by which genomic variation is itself shaped within taxa, affecting the clustering of variation and therefore the selection and deployment of barcodes.

We have shown how establishing and implementing standard barcodes for higher-level taxa generates distinct articulations of how to distinguish intra-specific (within-species) and inter-specific (cross-species) variation**.** The varipraxes of DNA barcoding, while manifesting in particular ways in different higher-level taxa as a result, follow the same general principles and broad trajectories. In all cases, an initial varipraxis based on a narrowing of the domain of variation that is identified and recorded in taxonomic research did not remain so restricted. Indeed, because the task of DNA barcoding needed the identification of specific genes to serve as barcodes, the selection needed to be justified in ways that entailed the deeper exploration of multiple aspects of the biology of a taxon and its constituent populations, in order to explain particular patterns of variation and establish the means to demarcate these intra- and inter-specifically.

DNA barcoding and, we would hypothesise, any new technology that opens up the ability to record, store and analyse new forms of variation, opens up new regimes of varipraxis. We suggest that developing the concept of varipraxis could enable an extension of insights from promiscuous realism or pluralist realism concerning classification (e.g., Dupré 1993; Wilkins 2018) out towards wider practices in the biological sciences.

Acknowledgements

The authors would like to thank the anonymous reviewers, Peter Hollingsworth and Catherine Kendig for their helpful comments on a draft version, and the participants and organiser (Özlem Yılmaz) of the Philosophy of Plant Biology Workshop at the University of Exeter, at which we presented some of the ideas that we subsequently developed into this paper. We are additionally grateful to our interviewees in the world of DNA barcoding and taxonomy. Lowe was able to conduct his part of the research through his work on the "TRANSGENE: Medical Translation in the History of Modern Genomics" Starting Grant, funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, grant agreement No. 678757. We appreciate the opportunity provided by this grant and by the support of the Principal Investigator, Miguel García-Sancho.

References

Agar J (2020) What is science for? The Lighthill report on artificial intelligence reinterpreted. The British Journal for the History of Science, 53(2): 289-310

Anderson C, Cunha L, Sechi P, Kille P, Spurgeon D (2017) Genetic variation in populations of the earthworm, *Lumbricus rubellus*, across contaminated mine sites. BMC Genetics, 18: 97

Andre J, King RA, Stürzenbaum SR, Kille P, Hodson ME, Morgan AJ (2010) Molecular genetic differentiation in earthworms inhabiting a heterogeneous Pb-polluted landscape. Environmental Pollution, 158(3): 883-890

Ankeny RA, Leonelli S (2020) Model Organisms. Cambridge University Press.

Antonelli A, Fry C, Smith RJ, Simmonds MSJ, Kersey PJ, Pritchard HW et al. (2020) State of the World's Plants and Fungi 2020. Royal Botanic Gardens, Kew.

Bever JD, Mangan SA, Alexander HM (2015) Maintenance of plant species diversity by pathogens. Annual Review of Ecology, Evolution & Systematics, 46: 305-325

Blaxter M, Mann J, Chapman T, Thomas F, Whitton C, Floyd R, Abebe E (2005) Defining operational taxonomic units using DNA barcode data. Philosophical Transactions of the Royal Society B: Biological Sciences 360(1462): 1935-1943

CBOL Plant Working Group (2009) A DNA barcode for land plants. Proceedings of the National Academy of Sciences, 106: 12794-12797

Charmantier I, Müller-Wille S (2014) Carl Linnaeus's botanical paper slips. Intellectual History Review, 24(2): 215-238

Conix S (2019) Taxonomy and conservation science: interdependent and value-laden. History and Philosophy of the Life Sciences, 41: 15

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

DiTeresi, C. (2010) Taming Variation: Typological Thinking and Scientific Practice in Developmental Biology. PhD thesis, University of Chicago.

Dupré J (1993) The Disorder of Things: Metaphysical Foundations of the Disunity of Science. Harvard University Press, Cambridge UK.

Endersby J (2018) Descriptive and prescriptive taxonomies. In: Curry HA, Jardine N, Secord JA, Spary EC (eds) Worlds of Natural History. Cambridge University Press, pp 447–459.

Evans KM, Alexandra H. Wortley AH, Mann DG (2007) An Assessment of Potential Diatom ''Barcode'' Genes (*cox1*, *rbc*L, 18S and ITS rDNA) and their Effectiveness in Determining Relationships in *Sellaphora* (Bacillariophyta). Protist, 158(3): 349-364

Fitzhugh K (2006) DNA Barcoding: An Instance of Technology-driven Science? BioScience, 56(6): 462-463

Geary J, Bubela T (2019) Governance of a global genetic resource commons for non-commercial research: A case-study of the DNA barcode commons. International Journal of the Commons, 13(1): 205-243

Geary J, Reay T, Bubela T (2019) The Impact of Heterogeneity in a Global Knowledge Commons: Implications for Governance of the DNA Barcode Commons. International Journal of the Commons, 13(2): 909-930

Gostel MR, Zúñiga JD, Kress WJ, Funk VA, Puente-Lelievre C (2020) Microfluid enrichment barcoding (MEBarcoding): a new method for high throughput plant DNA barcoding. Scientific Reports 10: 8701

Hacking I (1994) Styles of Scientific Thinking or Reasoning: A New Analytical Tool for Historians and Philosophers of the Sciences. In: Gavroglu K, Christianidis J, Nicolaidis E (eds) Trends in the Historiography of Science. Boston Studies in the Philosophy of Science, vol 151. Springer, pp 31–48.

Hamsher SE, Evans KM, Mann DG, Poulíčkovác A, Saunders GW (2011) Barcoding Diatoms: Exploring Alternatives to COI-5P. Protist, 162: 405-422

Harris ESJ, Mishler BD (2009) The Delimitation of Phylogenetic Characters. Biological Theory, 4(3): 230-234

Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003a) Biological identifications through DNA barcodes. Proceedings of the Royal Society of London B, 270: 313-321

Hebert, PDN, Ratnasingham S, de Waard JR (2003b) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proceedings of the Royal Society of London B, 270: S96-99

Hollingsworth PM, Graham SW, Little DP (2011) Choosing and Using a Plant DNA Barcode. PLoS ONE, 6(5): e19254

Hollingsworth PM, Li D-Z, van der Bank M, Twyford AD (2016) Telling plant species apart with DNA: from barcodes to genomes. Proceedings of the Royal Society of London B, 371: 20150338

Ingram DS (2022) A case for conserving plant pathogens. Plant Pathology, 71: 98-110

Jones L, Twyford AD, Ford CR, Rich TCG, Davies H, Forrest LL, Hart ML, McHaffie H, Brown MR, Hollingsworth PM, de Vere N (2021) Barcode UK: A complete DNA barcoding resource for the flowering plants and conifers of the United Kingdom, Molecular Ecological Resources, 21: 2050-2062

Kelly M, Boonham N, Juggins S, Kille P, Mann D, Pass D, Sapp M, Sato S, Glover R (2018) A DNA based diatom metabarcoding approach for Water Framework Directive classification of rivers. Environment Agency, UK.

Kendig C (2014) Towards a Multidimensional Metaconception of Species. Ratio, 27(2): 155-172

Kendig C (2020) Ontology and values anchor indigenous and grey nomenclatures: a case study in lichen naming practices among the Samí, Sherpa, Scots, and Okanagan. Studies in History and Philosophy of Biological and Biomedical Sciences, 84: 101340

Kendig C, Witteveen J (2020) The history and philosophy of taxonomy as an information science. History and Philosophy of the Life Sciences, 42: 40

King RA, Tibble AL, Symondson WOC (2008) Opening a can of worms: unprecedented sympatric cryptic diversity within British lumbricid earthworms. Molecular Ecology, 17: 4684-4698

Kress WJ (2017) Plant DNA barcodes: Applications today and in the future. Journal of Systematics and Evolution, 55(4): 291-307

Lowe, JWE (2015) Managing variation in the investigation of organismal development: problems and opportunities. History and Philosophy of the Life Sciences, 37: 449–473

Mann DG (1999) The species concept in diatoms. Phycologia, 38(6): 437-495

Minelli A (2019) The galaxy of the non-Linnaean nomenclature. History and Philosophy of the Life Sciences, 41: 31

Pawlowski J et al. (2012) CBOL Protist Working Group: Barcoding Eukaryotic Richness beyond the Animal, Plant, and Fungal Kingdoms. PLOS Biology, 10(11): e1001419

Pickstone JV (2000) Ways of Knowing: A New History of Science, Technology and Medicine. Manchester University Press.

Piotrowska M (2009) The Theoretical Costs of DNA Barcoding. Biological Theory, 4(3): 235-239

Shokralla S, Gibson JF, Nikbakht H, Janzen DH, Hallwachs W, Hajibabaei M (2014) Next-generation DNA barcoding: using next-generation sequencing to enhance and accelerate DNA barcode capture from single specimens. Molecular Ecology Resources 14(5): 892-901

Sluys R (2021) Attaching Names to Biological Species: The Use and Value of Type Specimens in Systematic Zoology and Natural History Collections. Biological Theory, 16: 49-61

Sneath PHA (1995) Thirty Years of Numerical Taxonomy. Systematic Biology, 44(3): 281-298

Sterner B (2014) Well-Structured Biology: Numerical Taxonomy's Epistemic Vision for Systematics In: Hamilton A (ed) The Evolution of Phylogenetic Systematics. University of California Press, pp. 213-244

Stevens L et al. (2019) Comparative genomics of 10 new *Caenorhabditis* species. Evolution Letters, 3: 217-236

Strasser BJ, de Chadarevian S (2011) The Comparative and the Exemplary: Revisiting the Early History of Molecular Biology. History of Science, 49(3): 317–336

Tilman GD (1982) Resource Competition and Community Structure. Princeton University Press.

Vernon K (1988) The Founding of Numerical Taxonomy. The British Journal for the History of Science, 21(2): 143-159

Wagner GP (2001) Characters, Units and Natural Kinds: an Introduction. In: Wagner GP (ed) The Character Concept in Evolutionary Biology. Academic Press, pp 1-10

Waterton C, Ellis R, Wynne B (2013) Barcoding Nature: Shifting cultures of taxonomy in an age of biodiversity loss. Routledge.

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

Werrett S (2019) Thrifty Science: Making the Most of Materials in the History of Experiment. The University of Chicago Press.

Wilkins JS (2018) Species: The Evolution of the Idea, Second Edition. CRC Press.

Will KP, Mishler BD, Wheeler QD (2005) The perils of DNA Barcoding and the need for integrative taxonomy. Systematic Biology, 54: 844-851

Zachos FE (2016) Species Concepts in Biology: Historical Development, Theoretical Foundations and Practical Relevance. Springer.