A Return to Linnaeus's Focus on Diagnosis, Not Description: The Use of DNA Characters in the Formal Naming of Species

SUSANNE S. RENNER*

Systematic Botany and Mycology, University of Munich (LMU), Menzinger Street 67, 80638 Munich, Germany

*Correspondence to be sent to: Systematic Botany and Mycology, University of Munich (LMU), Menzinger Street 67, 80638 Munich, Germany.

E-mail: renner@lrz.uni-muenchen.de

Received 12 September 2015; reviews returned 2 April 2016; accepted 5 April 2016

Associate Editor: Tiffani Williams

Abstract—Descriptions and diagnoses are alternative choices in all Codes of Nomenclature because Linnaeus relied on diagnoses, not descriptions, to name ca. 13,400 animals, plants, and fungi. A diagnosis names characters in which a new taxon differs from the most similar known taxa; a description mixes taxonomically informative and uninformative features, usually without indicating which is which. The first formal diagnoses of new taxa that included DNA-based characters came out in 2001, and by November 2015, at least 98 names of species of acceles, lichens, angiosperms, annelids, alveolates, arachnids, centipedes, turtles, fishes, butterflies, mollusks, nematodes, and pathogenic fungi have been published based on diagnostic mitochondrial, plastid, or nuclear DNA substitutions, indels, or rarely genetic distances. Authors have found diverse ways to specify the diagnostic traits (all published studies are here tabulated). While descriptions try to “cover” within-species variation, a goal rarely accomplished because of (i) the stochastic nature of specimen availability (thousands of species are known from single collections) and (ii) the subjective circumscription of species, the purpose of diagnoses was and is speedy identification. Linnaeus tried to achieve this by citing images, geographic occurrence, and previous literature. The renewed attention to sharp diagnosis now coincides with worldwide barcoding efforts, may speed up formal naming, and matches the increasing reliance on DNA for both classification and identification. I argue for DNA-based diagnoses of new species becoming a recommendation in all Codes, not just the bacterial code. [Codes of Nomenclature; description; diagnosis; DNA-based diagnosis; naming new species; nomenclature.]

“Descriptions cannot be made full enough and accurate enough to satisfy later workers. Each generation of taxonomists must see the actual specimens used by earlier generations, and I think the tendency now is, or should be, to make descriptions short, but of course explicit and carefully calculated, and to make specimens widely available.”
P.J. Darlington, Jr., 1971, p. 146

The naming of organisms following standardized conventions is the basis for linking new information to existing knowledge. It is also the basis for biological classification, effective communication, and extrapolation of findings about organisms. The mere accession numbers of DNA sequences (or other strings of numbers lacking an agreed system of the numbers’ innate significance) do not permit extrapolation of information about morphological traits, biogeographic ranges, or sharing of published knowledge across disciplines, all of which is possible with a widely used naming convention. Most researchers are using the conventions of the Linnaean system, with the fixed starting points being Linnaeus’s treatments of plants and animals (1753, 1758; Persoon and Fries for certain fungi and the names of lichens).

Since about 2000, taxonomists have increasingly tried to combine morphological and molecular data for detecting and delimiting species (reviewed in Wheeler 2008; Begerow et al. 2010; Hibbett and Taylor 2013; Vences et al. 2013), and since 2003, DNA barcoding has become the method of choice for reliable identification, at least for insects, certain fungi, tropical trees, and many aquatic organisms (Hebert et al. 2003; Koljalg et al. 2013; Hausmann et al. 2013; Kress et al. 2015). Surprisingly, however, DNA characters have rarely been used in the formal description of species (Cook et al. 2010). Of 310 barcoding publications surveyed by Kress et al. (2015) that led to the discovery of new species, only one (Félix et al. 2014; Table 1) used DNA traits in species protologs.

To date, two papers have discussed DNA-based formal diagnoses (Cook et al. 2010; Tripp and Lendemer 2014). Both overlooked that the practice began 15 years ago (Westheide and Hass-Cordes 2001), and they either focused on a hypothetical example (Cook et al. 2010) or examples from 2012 and 2013 (Tripp and Lendemer 2014). Tripp and Lendemer (2012) also raised a potential problem with one type of DNA diagnosis, namely genetic distances, which I take up in the "Discussion" section. No previous paper has surveyed the conceptual and factual history of DNA-based formal naming, and the absence of a review of how taxonomists have
Table 1. All studies published so far that have included DNA-based diagnoses in species protologs, that is, for the formal diagnosis of type material

<table>
<thead>
<tr>
<th>Taxon</th>
<th>Number and geographic region of new species</th>
<th>Type of molecular diagnosis</th>
<th>Morphological description (yes/no); Morphological diagnosis (yes/no)</th>
<th>Deposition of type material (not always the specimen from which DNA was isolated)</th>
<th>Deposition of sequences or alignment(s)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acoelomorpha (Platyhelminthes) Nemertodermatida, Nemertodermatidae</td>
<td>9 species of Nemertinoides, worldwide in marine mesopsammon</td>
<td>Specified substitutions in LSU and SSU rRNA and histone 3</td>
<td>Yes Yes</td>
<td>Type material in a public collection</td>
<td>Sequences in GenBank</td>
<td>Meyer-Wachsmuth et al. 2014</td>
</tr>
<tr>
<td>Alveolata Dinophyceae</td>
<td>5 species of Alexandrium</td>
<td>&quot;The combined nucleotide sequences of the holotype strain .... The complete list of diagnostic D1-D2 LSU, ITS (Internal Transcribed Spacer)/5.8S and SSU Gen-Bank sequences which can be used as a genetic type for this species are reported in Supplementary Material.&quot; Similarly for the other species</td>
<td>Yes Yes</td>
<td>Holo- and isotypes on SEM stubs and in formalin in Senckenberg herbarium (FR)</td>
<td>Sequences in GenBank</td>
<td>John et al. 2014</td>
</tr>
<tr>
<td>Angiospermae Solanaceae</td>
<td>1 species of Brunfelsia from Brazil</td>
<td>Specified substitutions in the plastid ndhF gene and in nuclear ITS</td>
<td>Yes No</td>
<td>Type material in various herbaria</td>
<td>Sequences in GenBank</td>
<td>Filipowicz et al. 2012</td>
</tr>
<tr>
<td>Angiospermae Buxaceae</td>
<td>3 species of Buxus from Cuba</td>
<td>Specified substitutions in the plastid ndhF gene and in coding region of trnL spacer</td>
<td>Yes Yes</td>
<td>Type material in various herbaria</td>
<td>Sequences in GenBank</td>
<td>González Gutiérrez et al. 2013</td>
</tr>
<tr>
<td>Angiospermae Ehretiaceae</td>
<td>1 species of Rochefortia from Lesser Antilles</td>
<td>Specified substitutions in the nuclear ITS1 region</td>
<td>Yes Yes</td>
<td>Type material in various herbaria</td>
<td>Sequences in GenBank</td>
<td>Irimia and Gottschling 2016</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Taxonomic Group</th>
<th>Species</th>
<th>Specified Substitutions</th>
<th>Holotype</th>
<th>Type Material</th>
<th>Sequences In</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annelida, Syllidae (Bristle worms)</td>
<td>1 species of <em>Petitia</em></td>
<td>Specified substitutions in nuclear ITS and 8 diagnostic RFLP (Restriction Fragment Length Polymorphism) fragments with specified primers</td>
<td>Yes No</td>
<td>Holotype “used up,” three syntypes in Senckenberg Museum</td>
<td>Several nuclear ITS seqs in GenBank</td>
<td>Westheide and Hau-Cordes 2001</td>
</tr>
<tr>
<td>Annelida, Serpulidae (Tube worms)</td>
<td>1 species of <em>Galeolaria</em></td>
<td>Specified substitutions in mtDNA cytochrome b gene and nuclear ITS</td>
<td>Yes No (Description is called diagnosis)</td>
<td>Type material in public collection</td>
<td>Sequences in GenBank</td>
<td>Halt et al. 2009</td>
</tr>
<tr>
<td>Arachnida, Opiliones, Cyphophthalmi: Neogoveidae</td>
<td>2 species of <em>Metaxus</em></td>
<td>Specified substitutions in mtCOI gene</td>
<td>Yes No</td>
<td>Type material in a public collection</td>
<td>Sequences in GenBank</td>
<td>Clouse and Wheeler 2014</td>
</tr>
<tr>
<td>Arthropoda, Chilipoda, Craterostigmidae (Centipedes)</td>
<td>1 species of <em>Craterostigmus</em></td>
<td>Specified substitutions in nuclear 18S and 28S rRNA, mitochondrial (mt) 16S rRNA, and the protein-encoding cytochrome c oxidase subunit I (COI)</td>
<td>Yes Yes</td>
<td>Type material in a public collection (including DNA voucher from holotype)</td>
<td>Sequences in GenBank</td>
<td>Edgecombe and Garbet 2008</td>
</tr>
<tr>
<td>Ascomycota, Lecanoraceae (Lichens)</td>
<td>5 species of <em>Rhizoplaca</em></td>
<td>Unspecified substitutions in nuclear ITS</td>
<td>No No</td>
<td>Type material in public collection and Mycobank</td>
<td>Sequences in GenBank</td>
<td>Leavitt et al. 2013</td>
</tr>
<tr>
<td>Ascomycota, Lecanoraceae (Lichens)</td>
<td>1 species of <em>Lepraria</em></td>
<td>Specified substitutions in nuclear ITS (in Latin)</td>
<td>Yes No</td>
<td>Type material in public collections</td>
<td>Sequences in GenBank</td>
<td>Lendemer 2011</td>
</tr>
<tr>
<td>Ascomycota, Parmeliaceae (Lichens)</td>
<td>1 species of <em>Parmelia</em></td>
<td>Specified substitutions in a group 1 intron and nuclear ITS (in Latin)</td>
<td>No No</td>
<td>Type material in public collections</td>
<td>Sequences in GenBank</td>
<td>Molina et al. 2011</td>
</tr>
<tr>
<td>Taxonomy</td>
<td>Species</td>
<td>Specified substitutions in cyt b and RAG2 genes (listed in a table)</td>
<td>Yes/No</td>
<td>Type material in various collections</td>
<td>Sequences in GenBank</td>
<td>Reference</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>--------------------------</td>
<td>---------------------------------------------------------------------</td>
<td>----------</td>
<td>--------------------------------------</td>
<td>-----------------------------</td>
<td>------------------------------------</td>
</tr>
<tr>
<td>Chordata, Pisces, Ictaluridae</td>
<td>1 species of <em>Noturus</em></td>
<td>Specified substitutions and indels in nuclear 12S rDNA</td>
<td>Yes No</td>
<td>American-type culture collection</td>
<td>Yes</td>
<td>Fisher et al. 2002</td>
</tr>
<tr>
<td>Chordata, Reptilia, Pelomedusidae</td>
<td>6 species of <em>Pelomedusa</em></td>
<td>Specified substitutions in nuclear Chitinsynthase gene, two microsatellite differences (in Latin)</td>
<td>Yes No</td>
<td>Type material in various collections</td>
<td>12S rDNA sequences in GenBank</td>
<td>Egge and Simons 2006</td>
</tr>
<tr>
<td>Fungi Onygenales, Coccidioidomyces</td>
<td>1 species of <em>Coccidioides</em></td>
<td>Specified substitutions in nuclear rRNA ITS sequences with GenBank accessions GQ850318, GQ850355 &amp; GQ850368</td>
<td>No No</td>
<td>Holotype K(M)173535</td>
<td>Sequences in GenBank</td>
<td>Kirk 2012</td>
</tr>
<tr>
<td>Fungi Neocallimastigales</td>
<td>1 species of <em>Pirunyces</em></td>
<td>“The least inclusive clade containing organisms with nuclear rRNA ITS sequences with GenBank accession”</td>
<td>No No</td>
<td>Holotype IMI 39811</td>
<td>Sequences in GenBank</td>
<td>Bridge and Hughes 2012</td>
</tr>
<tr>
<td>Fungi Mortierellales</td>
<td>1 species of <em>Mortierella</em></td>
<td>“With an ITS sequence (GenBank KQ89360) that is distinct from other members of the gamsii/elongata clade, deviating in the ITS1 region from other species in the clade; with a 94–97% similarity. With a sister group relationship to a possibly polyphyletic clade containing Mortierella sclerotii (basal; GenBank HQ63038, ex type), <em>M. cogitans</em> (GenBank HQ63028, ex type), and <em>M. acrotona</em> (GenBank HQ63032, ex type)”</td>
<td>No No</td>
<td>Holotype IMI 39811</td>
<td>Sequences in GenBank</td>
<td>Bridge and Hughes 2012</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Lepidoptera: Hesperiidae: Endamine</th>
<th>10 species of <em>Astraptes</em></th>
<th>Specified substitutions in mitochondrial cytochrome oxidase subunit I gene fragment (COI)</th>
<th>No No</th>
<th>Type material in public collection</th>
<th>Sequences from GenBank (none new)</th>
<th>Brower 2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mollusca, Gastropoda, Abyssochrysidae Hydrothermal-vent snails</td>
<td>5 species of <em>Akthionana</em>, deep sea, hydrothermal vents, Western Pacific and Indian oceans</td>
<td>Specified substitutions in mt COI, 12S mt RNA, 16S rRNA, nuclear 28S, and 18S rRNA</td>
<td>Yes No</td>
<td>Type material in a public collection</td>
<td>Sequences in GenBank</td>
<td>Johnson et al. 2015</td>
</tr>
<tr>
<td>Mollusca, Gastropoda, Coralliophila Coralsnails</td>
<td>14 species of <em>Leptazoochusa</em>, parasitic snails living in corals in Indo-West Pacific of Egypt, the Maldives, Thailand, Palau, and Indonesia</td>
<td>Specified substitutions in mt COI and nuclear ITS</td>
<td>Yes No</td>
<td>685 snails collected from 327 hosts, incl. type material, deposited in various collections</td>
<td>Sequences in GenBank</td>
<td>Gittenberger and Gittenberger 2011</td>
</tr>
<tr>
<td>Mollusca, Gastropoda, Gasteropoda Nudibranchia</td>
<td>3 species of <em>Glaucus</em>, pelagic, Pacific Ocean</td>
<td>Specified substitutions in mt COI and 16S rRNA</td>
<td>Yes Yes</td>
<td>Type material in a public collection</td>
<td>Sequences in GenBank</td>
<td>Churchill et al. 2014</td>
</tr>
<tr>
<td>Mollusca, Gastropoda, Microhedylidae Micro slugs</td>
<td>9 species of <em>Protohedyia</em> from the mesopodamon in oceans worldwide</td>
<td>Specified substitutions in mt COI, 16S rRNA, nuclear 28S, and 18S rRNA</td>
<td>Yes Mostly no</td>
<td>Type material in various collections</td>
<td>Sequences in GenBank</td>
<td>Jörger and Schrödl 2013</td>
</tr>
<tr>
<td>Nematoda, Rhabditidae</td>
<td>15 species of <em>Caenorhabditis</em></td>
<td>This species differs by SSU, LSU, and ITS2 DNA sequences (JN636069) from all other species, listed in Tables 1 and 2. &quot;Note that these rhabdial DNA sequences may vary within the species.&quot;</td>
<td>No Yes (Diagnoses rely on specified fertile crosses with specified isolates, yielding fertile hybrid females and males)</td>
<td>The type culture specimens are deposited at the <em>Caenorhabditis</em> Genetics Center</td>
<td>Sequences in GenBank</td>
<td>Félix et al. 2014</td>
</tr>
</tbody>
</table>

Note: Full references are in the main text's reference list. Several studies rely entirely on DNA and provide no morphological diagnoses or descriptions.
incorporated molecular characters into protologs has led to uncertainty and reinventing of the wheel (Cook et al. 2010; Jörger and Schrödl 2013).

Species are always delimited against already known species (Linnaeus 1753, 1758; Mayr 1992; Naciri and Linder 2015). This holds true regardless of whether they are conceived as created (Linnaeus 1753, 1758) or as the result of evolution (Mayr 1992). Huge numbers of “cryptic” species—a term only meaningful relative to the particular technology used for studying organisms—can be distinguished with genomic data, and taxonomists are facing the challenge of naming at least some of this organismal diversity as it may be relevant for their research interests. It is useful then to consider how earlier taxonomists facing large numbers of new species mastered the task.

Foremost among taxonomists naming species is Linnaeus, who named ca. 6000 species of plants and 4400 species of animals (Müller-Wille 2006; Jarvis 2007). To do so, Linnaeus focused on diagnostic features in which a species differs from closest relatives. He was rightly proud of this idea and devoted much thought to the drafting of his diagnostic phrases, which were for him the true names of species. “Linnaeus held that these diagnosis should not exceed 12 words in length, and he and Jacquin even managed on occasion to reduce them to one word” (Stearn 1992, p. 144). In addition, he cited previous literature, available illustrations, and species’ ranges whenever known (Jarvis 2007). Longer descriptions based on multiple specimens and indicating the range of trait variation became widespread with the Prodromus project of the two De Candolles (J. Fris, personal communication, May 2015). Alphonse de Candolle was president of the International Botanical Congress in 1866 in London and wielded an immense influence (Nicolson 1991), and, obviously, a focus on within-species variation fit with Darwinian views on descent with modification.

While descriptions that mix taxonomically informative and uninformative traits became customary after about 1850, none of the Codes of Nomenclature stipulates that a new taxon must be described because such a requirement would have made Linnaeus’s names unavailable (under the zoological code) or invalid (under the botanical and mycological code). Instead, all Codes leave a choice between either a description or a diagnosis. The Code of Nomenclature for algae, fungi, and plants (McNeill et al. 2012, article 32.2) defines a diagnosis as “a statement of that which in the opinion of its author distinguishes the taxon from others.” The International Code of Zoological Nomenclature (ICZN 1999, article 13.1.3) defines it thus “When describing a new taxon, an author should make clear his or her purpose to differentiate the taxon by including with it a diagnosis, that is to say, a summary of the characters that differentiate the new nominal taxon from related or similar taxa,” and the Bacteriological Code (Lapage et al. 1992) states that any name proposal “must contain a brief diagnosis, i.e., a statement or list of those features that led the author to conclude that the proposed taxon is sufficiently different from other recognized taxa…” Besides a diagnosis or description, a type specimen must be clearly indicated, and it is the type material that provides the objective standard of reference for the application of the name it bears. The combination of the type method (i.e., name-bearing specimens deposited in one or more collections) and the discrete nature of nucleotide characters (substitutions or insertions/deletions of codons) begs reconsideration of Linnaeus’s focus of diagnosing species by features that distinguish them from their known closest relatives, instead of describing mixed sets of traits that vary at different hierarchical levels. Here, I consider the ways in which taxonomists have incorporated DNA characters directly into the publication of new species names, and I also review the history of DNA-based formal naming. I conclude with recommendations about best practice DNA-based diagnosis.

MATERIALS AND METHODS

Literature Search and Data Documentation

I compiled published molecular diagnoses through internet searches, surveying relevant journals and corresponding with colleagues. A molecular diagnosis involves the formal naming of a taxon by listing the DNA or protein characters in which it differs from its closest relative(s) in the protolog, thus associating it with a binomial Latinized name and the type material with its place of deposition. I checked that the molecular data indicated in the diagnosis were accessible in the cited database, usually the National Center for Biotechnology Information (NCBI: http://www.ncbi.nlm.nih.gov/, accessed 28 April 2016).

My survey focused on species names. An example of a molecular diagnosis of a higher taxon is that of the family Ambuchananiaceae Seppelt & H.A. Crum ex A.J. Shaw, “fam. nov. Planta heterogeneae in morphology, synapomorpha molecularibus in DNA nuclei mitochondri et plasti unitae. Type: Ambuchanania” (Shaw et al. 2010, p. 1523). Of course, this was before botanists abolished the Latin requirement on 1 January 2012. The baselines for bacterial names are Approved Lists, with a starting point of 1980, and new bacterial names are reviewed by a nomenclature committee and published in the IJSEM (Lapage et al. 1992). As mentioned in the “Introduction,” the proposal of a new bacterial name must contain a type designation and a brief diagnosis, that is, a statement or list of those features that led the author to conclude that the proposed taxon is sufficiently different from other recognized taxa (see also Stackebrandt and Goebel 1994). For the present review, I focus on eukaryotes.

RESULTS

The Use of DNA Characters in Species Diagnoses since 2001

The first to discuss how DNA characters might be used in species diagnosis were Don Reynolds and
John Taylor (1991) who clarified that the existing rules of the International Code of Botanical Nomenclature (as it was then still called) allowed DNA-based species naming and that DNA itself could serve as the type element. They provide two hypothetical examples of new fungal species names, one with a mix of DNA and morphological type materials and the other with DNA type material only, and call on herbaria to prepare for storing DNA material as types. Almost a quarter of a century has passed since this prescient article, but taxonomists are still feeling the need to defend the use of DNA characters in protologs (Cook et al. 2010; Jörger and Schrödl 2013; Tripp and Lendemer 2014), and the approach is only slowly becoming more common (Fig. 1).

Perhaps, surprisingly, Reynolds and Taylor (1991) devoted more discussion to the idea of using genomic material as type material than to the utility of nucleic acid characters as diagnostic tools because they thought it “unavoidable that DNA will serve as character source for contemporary taxonomic descriptions” (Reynolds and Taylor 1991, p. 311). Their hypothetical diagnosis, for a species collected on “a health food candy bar”, takes the shape, “5′-3′, ATGCCATAA ACTACCTAGC, AACT GATACTAATACC, Nucleotide positions 116-136, 1200-1216, Small Nuclear rDNA (1800 BP); 5′-3′, TATAGCCGCTAATCG CTAGATAA, Nucleotide positions 100-123, Mitochondrial Small rDNA (1648 BP).” (I.c., p. 314). The first molecular diagnosis of a real, not hypothetical, taxon is of a polychaete annelid from the Seychelles, the protolog of which differentiates it from morphologically similar individuals from Rhodos and Tenerife (Westheide and Hass-Cordes 2001). Twenty specimens were available for microscopy, and 13 others were used for RAPD fingerprinting or sequencing of the nuclear internal transcribed species region of ribosomal DNA. The diagnostic DNA characters consist of eight RAPD bands obtained with specified primers and of characteristic substitutions in the ITS2.

By November 2015, 98 molecular diagnoses of species of Acoelomorpha, Alveolata, Angiospermae, Annelida, Arachnida, Arthropoda, Ascomycota, Chordata (Reptilia and Pisces), Fungi, Lepidoptera, Mollusca, and Nematoda have been published (Table 1). Relatively few protologs refrain from also providing a morphological description (Brower 2010; Molina et al. 2011; Leavitt et al. 2013).

Names of Species of Fungi with DNA-Based Diagnoses

Mycologists were at the forefront of DNA-based formal species naming, probably because their organisms pose particular challenges, as pointed out by Reynolds and Taylor (1991, p. 315), “recognition of DNA as at least part of the type element is certain to diminish the reliance on sexual characters for classification and undermine the maintenance of a separate form-classification for fungi lacking sex.” Most mycologists see no problem in diagnosing species by specific DNA substitutions (Taylor 2011; Schoch et al. 2012; Koljalg et al. 2013), and mycologists have also developed some of the most creative DNA-based diagnoses (Table 1). Thus, Hibbett et al. (2011, p. 45) proposed this form, “The least inclusive group containing organisms with nuclear rRNA ITS sequences with GenBank accessions AB244041 and DQ054545.” This exact form was used by Kirk (2012) in the protolog of Piromyces cryptodigmaticus Fliegerová, K. Voigt &
P.M. Kirk, diagnosed as “The least inclusive clade containing organisms with nuclear rRNA ITS sequences with GenBank accessions GQ850318, GQ850335 & GQ850368; with a sister group relationship to the clade containing the proposed epitype of Pironjaceae communis with a nuclear rRNA ITS sequence with GenBank accession AY429665; the closest named common ancestor, Cylindraceae aberensis, with a nuclear rRNA ITS sequence with GenBank accession FJ483845. Holotype K(M) 173535.” A similar form has been used for five species of lichens (Leavitt et al. 2013, p. 11), “Rhizoplaca polymorpha consists of specimens recovered within ‘clade IVc’ in Leavitt et al. (2011), which is supported as a lineage distinct from all other populations according to coalescent-based genetic analysis of multiple genetic loci.”

This form of clade-based diagnosis (“The least inclusive clade containing...”) has been challenged by Tripp and Lendemer (2012), who have requested the Committee on the application of the Code of Nomenclature for algae, fungi, and plants to decide on the validity of this form, which in their view goes against the requirement in Article 32.2(d) that a diagnosis cannot describe properties such as purely aesthetic features, economic, medicinal or culinary usage, cultural significance, cultivation techniques, geographical origin, or geological age. This matter is currently unsolved, and I have not found examples from outside fungi and lichens of this form of diagnosis (Table 1).

Names of Animal Species with DNA-Based Diagnoses

While nuclear 18S and 28S rRNA, mitochondrial 16S rRNA, and protein-encoding cytochrome c oxidase subunit I (COI or cox1) sequences have all been used in the diagnoses of new animal species names, the barcoding region, cox1, which pinpoints the correct species in many groups of insects (Hausmann et al. 2013), has been used especially often (Table 1). In most studies, DNA diagnostic features serve to corroborate morphological differences. For example, diagnostic COI substitutions that agree with shell characters clearly diagnose species of parasitic snails, but “impoverished anatomical details [alone] do not allow identification” (Gittenberger and Gittenberger 2011; p. 21). Several large-bodied species, such as turtles, have also been diagnosed with molecular *cum* morphological traits (Petzold et al. 2014). DNA-derived traits, mixed with morphology, have also being used in a *key* to 205 described braconid Hymenoptera *Apanetes* from Mesoamerica (Fernández-Triana et al. 2014), but Fernández-Triana and colleagues decided not to use DNA barcoding traits as species diagnoses, instead using the form “sequences in BOLD: 2; barcode compliant sequences: 2.”

Names of Plant Species with DNA-Based Diagnoses

Between January 1935 and 2012, botanists (and mycologists) had to write any diagnosis in Latin (Table lists three such molecular diagnoses). Since 2012, however, a few plant species have been diagnosed with nucleotide substitutions described in English (Table 1), and one study even provides both molecular and morphological diagnoses and molecular and morphological descriptions (González et al. 2013).

**DISCUSSION**

*Advantages of the Sharper Diagnosis of Type Material*

A key advantage of molecular diagnoses is their utility for more precisely characterizing type material than is possible with morphological traits. The better a type collection (including syntypes and paratypes) is characterized, the more reliable the identification of future specimens. This does not mean that unidentified specimens in the future will need to be sequenced for identification. Instead, identification may continue to rely on morphological matching of preserved specimens or, increasingly, of images using machine learning. Having stringent diagnoses that specify DNA differences among closely related species (or subspecific taxa) can facilitate identification in those cases where the correct identification of a specimen is crucial, for example, for parasites of crops or of animals, especially us, but also for specimens that are incomplete, poorly preserved, or immature, so that diagnostic features are missing. Also, as pointed out by Cook et al. (2010), it is often quicker and cheaper to use diagnostic DNA features than to rely on the traditional expert-centered paradigm of identification.

The many studies that have clarified erroneous application of names or relationships among living and extinct species by sequencing DNA from type material attest to the importance of DNA diagnosis, now and in the future (Stuart and Fritz 2008; Hausmann et al. 2009; Sebastian et al. 2010; Stuckas and Fritz 2011; Stuckas et al. 2013; Fritz et al. 2014; Petzold et al. 2014; Heupink et al. 2014; Cappellini et al. 2014; Renner et al. 2014; Speidel et al. 2015; Erpenbeck et al. 2016).

**Easy Accessibility, Interpretability, and Utility in Automated Keys**

Several taxonomic journals have hypertext markup language that allows direct linkages between new species names and sequences in GenBank or other sequence databanks (Fenév et al. 2010). Sequences mentioned in diagnoses will serve as a standard for future reference, as pointed out by Reynolds and Taylor (1991) and Tautz et al. (2003), together with the type material deposited in one or, better, more museum collections (cf. the Darlington quote at the top of this paper). “DNA sequence information is digital and is not influenced by subjective assessments. It would be reproducible at any time and by any person, speaking any language. Hence, it would be a universal communication tool and resource for taxonomy, which
can be linked to any kind of biological or biodiversity information. Even if a query sequence does not produce an exact match, it will be possible to link an organism to closely related ones” (Tautz et al. 2003, p. 71).

These authors, therefore, proposed that an attempt be made to provide a DNA sequence alongside all future taxonomic samples and species descriptions. In my view, this should become a recommendation in all Codes. Taxonomists, however, have begun to go further by including DNA characters directly in the diagnosis of nominal new taxa. This makes the type material more valuable and is safer for the future than if sequences come from other specimens that may be less well-preserved than type material typically is (or should be). Most importantly, sharp diagnosis of the types of species names will help avoid the publication of unnecessary names (new synonyms).

Last, DNA sequence databases with automated matching can replace identification keys. The functionality of such species-naming pipelines has been demonstrated in fungi (Koljalg et al. 2013). For animals, the concept of a Barcode Index Number (BIN) has been proposed (Ratnasingham and Hebert 2013), namely a persistent, species-level taxonomic registry using patterns of nucleotide variation in the barcode region of the cytochrome c oxidase I (COI) gene. The system begins by examining the correspondence between groups of specimens identified to species through prior taxonomic work and those inferred from the analysis of COI sequence variation using several algorithms.

**Differences Between Barcoding and DNA-Based Diagnosis, and How the Two Approaches Will Increasingly Reinforce Each Other**

There are three differences between barcoding and using DNA features in the protologs of new species. First, barcoding relies on a few universally agreed markers; DNA-based diagnosis does not, but can instead use a mix of other DNA traits, even indels (cf. Table 1). Second, barcoding is about identifying unknown material by matching sequences to named sequences in a database. This is not the purpose of DNA-based diagnoses, which serve to better describe a new species’ type collection(s). For barcoding, one does not need to study type material or deposit a type in a designated public collection, as one does to name a species. Third, barcoding one’s material is not a requirement or recommendation in any of the Codes of Nomenclature, while diagnosis is a recommendation involving fungal names published without reference to specific characters distinguishing them from their closest relatives (see above, “Results” section for a specific example, *Rhizoplaca polymorpha*). Based on my reading of 98 molecular diagnoses, I agree with Tripp and Lendemer that discrete DNA features of type specimens are more useful than node-based diagnoses, which focus on phylogenetic context, not specimens. At least one study, however, has combined genetic distances and discrete trait states (Meyer-Wachsmuth et al. 2014), and the naming of bacteria has long relied on distances (see section “Materials and Methods”).

**CONCLUSIONS**

DNA-based diagnoses along with (generously loaned) museum specimens and stably archived specimen images to my mind are more important today than attempts to “cover” morphological variation in populations (which obviously can and will continue). Such attempts are always limited by the availability of material, which causes taxonomists to wait, often for years, until that perfect second or third collection shows up. Even where several specimens are available, which and how much variation to document and describe remains subjective (Darlington 1971) and is a matter of idiosyncratic taxonomic practice. Reducing the time spent on long descriptions and instead focusing on sharp diagnoses might lead to faster naming of new species (also Riedel et al. 2013).

The process of species discovery (their delimitation from already known species) is a question of human interest and available technologies, and this implies that we will never know all species, even if “knowing” is defined as having lists of agreed upon names as done by bacteriologists and increasingly mycologists. My recommendation would be to include examples of DNA-based diagnoses in the Codes to help practitioners (see section “Differences Between Barcoding and DNA-Based Diagnosis, and How the Two Approaches Will Increasingly Reinforce Each Other”). Matching with existing names, or naming as new, the thousands of unnamed entities waiting in collections (each more or less incompletely represented) will become easier with both ongoing barcoding efforts and the inclusion of DNA traits in the diagnoses of types.

**Genetic Distance Less Suitable than Diagnostic Substitutions?**

Tripp and Lendemer (2012) have raised the question whether node-based diagnoses (Hibbett et al. 2011; Kirk 2012), rather than diagnostic substitutions, are valid and have submitted a request to the Nomenclature Commission (for plants and fungi) for clarification of two examples involving fungal names published without reference to specific characters distinguishing them from their closest relatives (see “Results” section for a specific example, *Rhizoplaca polymorpha*). Based on my reading of 98 molecular diagnoses, I agree with Tripp and Lendemer that discrete DNA features of type specimens are more useful than node-based diagnoses, which focus on phylogenetic context, not specimens. At least one study, however, has combined genetic distances and discrete trait states (Meyer-Wachsmuth et al. 2014), and the naming of bacteria has long relied on distances (see section “Materials and Methods”).
ACKNOWLEDGMENTS

I thank Frank Anderson for critical comments; K. Jörger, Zoological Natural History Collections, Munich, and D. Begerow, Systematic Mycology, University of Bielefeld, for pointing me to DNA diagnoses in their fields of expertise; Ib Friis and Henrik Balslev for the opportunity to discuss the topic with philosophers, historians, and sociologists at a meeting in Hannover in June 2015.

REFERENCES


Bridge P.D., Hughes K.A. 2012. Index Fungorum 7 (1).


Cook L.G., Edwards R.D., Crisp M.D., Hardy N.B. 2010. Need opportunity to discuss the topic with philosophers, historians, and sociologists at a meeting in Hannover in June 2015.


INDEX FUNGORUM

The opportunity to discuss the topic with philosophers, historians, and sociologists at a meeting in Hannover in June 2015.

Invertebr. Syst. 21:1–21.


INDEX FUNGORUM

The opportunity to discuss the topic with philosophers, historians, and sociologists at a meeting in Hannover in June 2015.


Lendemer J.C. 2011. A taxonomic revision of the North American species of Lepraria s.l. that produce divaricate acid, with notes on the type species of the genus L. incana. Mycologia 103:1226–1229.

Linnaeus C. 1758. Systema naturæ per regna tria naturæ, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis. Stockholm: Salvius.


Taylor J.W. 2011. One Fungus One Name: DNA and fungal nomenclature twenty years after PCR. IMA Fungus 2: 113–120.


Vences M., Guayasamin J.M., Micles B., De La Riva I. 2013. T o name or not to name: Criteria to promote economy of change in Linnaean classification schemes. Zootaxa 3636:201–244.


Taylor J.W. 2011. One Fungus = one Name: DNA and fungal nomenclature twenty years after PCR. IMA Fungus 2: 113–120.


Vences M., Guayasamin J.M., Micles B., De La Riva I. 2013. T o name or not to name: Criteria to promote economy of change in Linnaean classification schemes. Zootaxa 3636:201–244.
