doi:10.5598/imafungus.2018.09.01.09 IMA FUNGUS · 9(1): 143–166 (2018)

ARTICLE

Formal description of sequence-based voucherless *Fungi*: promises and pitfalls, and how to resolve them

Robert Lücking[1] and David L. Hawksworth[2]

[1]Botanischer Garten und Botanisches Museum, Freie Universität Berlin, Königin-Luise-Strasse 6–8, 14195 Berlin, Germany; corresponding author e-mail: r.luecking@bgbm.org

[2]Department of Life Sciences, The Natural History Museum, Cromwell Road, London SW7 5BD, UK; and Comparative Plant and Fungal Biology, Royal Botanic Gardens, Kew, Surrey TW9 3DS, UK; Jilin Agricultural University, Changchun, Jilin Province,130118 China

Abstract: There is urgent need for a formal nomenclature of sequence-based, voucherless *Fungi*, given that environmental sequencing has accumulated more than one billion fungal ITS reads in the Sequence Read Archive, about 1,000 times as many as fungal ITS sequences in GenBank. These unnamed *Fungi* could help to bridge the gap between 115,000 to 140,000 currently accepted and 2.2 to 3.8 million predicted species, a gap that cannot realistically be filled using specimen or culture-based inventories. The *Code* never aimed at placing restrictions on the nature of characters chosen for taxonomy, and the requirement for physical types is now becoming a constraint on the advancement of science. We elaborate on the promises and pitfalls of sequence-based nomenclature and provide potential solutions to major concerns of the mycological community. Types of sequence-based taxa, which by default lack a physical specimen or culture, could be designated in four alternative ways: (1) the underlying sample ('bag' type), (2) the DNA extract, (3) fluorescent *in situ* hybridization (FISH), or (4) the type sequence itself. Only (4) would require changes to the *Code* and the latter would be the most straightforward approach, complying with three of the five principal functions of types better than physical specimens. A fifth way, representation of the sequence in an illustration, has been ruled as unacceptable in the *Code*.

Potential flaws in sequence data are analogous to flaws in physical types, and artifacts are manageable if a stringent analytical approach is applied. Conceptual errors such as homoplasy, intragenomic variation, gene duplication, hybridization, and horizontal gene transfer, apply to all molecular approaches and cannot be used as a specific argument against sequence-based nomenclature. The potential impact of these phenomena is manageable, as phylogenetic species delimitation has worked satisfactorily in *Fungi*. The most serious shortcoming of sequence-based nomenclature is the likelihood of parallel classifications, either by describing taxa that already have names based on physical types, or by using different markers to delimit species within the same lineage. The probability of inadvertently establishing sequence-based species that have names available is between 20.4 % and 1.5 % depending on the number of globally predicted fungal species. This compares favourably to a historical error rate of about 30 % based on physical types, and this rate could be reduced to practically zero by adding specific provisions to this approach in the *Code*. To avoid parallel classifications based on different markers, sequence-based nomenclature should be limited to a single marker, preferably the fungal ITS barcoding marker; this is possible since sequence-based nomenclature does not aim at accurate species delimitation but at naming lineages to generate a reference database, independent of whether these lineages represent species, closely related species complexes, or infraspecies. We argue that clustering methods are inappropriate for sequence-based nomenclature; this approach must instead use phylogenetic methods based on multiple alignments, combined with quantitative species recognition methods. We outline strategies to obtain higher-level phylogenies for ITS-based, voucherless species, including phylogenetic binning, 'hijacking' species delimitation methods, and temporal banding. We conclude that voucherless, sequence-based nomenclature is not a threat to specimen and culture-based fungal taxonomy, but a complementary approach capable of substantially closing the gap between known and predicted fungal diversity, an approach that requires careful work and high skill levels.

Key words:
biodiversity
ecologically cryptic *Fungi*
environmental sequencing
evolutionary placement algorithm
high throughput sequencing
IMC11
internal transcribed spacer
molecular barcoding
molecular sequence data
next generation sequencing
nomenclature
typification

.**Article info:** Submitted: 2 July 2017; Accepted: 15 May 2018; Published: 22 May 2018.

BACKGROUND

Fungal taxonomy and systematics based on DNA sequencing commenced about three decades ago (Kurtzman 1985, Michelmore & Hulbert 1987, Gouy & Li 1989, Hendriks *et al.* 1989, White *et al.* 1990, Bruns *et al.* 1991, Bowman *et al.* 1992). Large-scale analyses reshaped our understanding of fungal evolution and classification (e.g. Moncalvo *et al.* 2002, Lutzoni *et al.* 2004, Spatafora 2005, Blackwell *et al.* 2006, James *et al.* 2006, Hibbett *et al.* 2007, Schoch *et al.* 2009, Miądlikowska *et al.* 2014, Spatafora *et al.* 2016). Subsequently, focus shifted towards species delimitation (e.g. Taylor *et al*. 2000, Pringle *et al.* 2005, Geml *et al.* 2006, Bensch *et al.* 2010, Lombard *et al.* 2010, Lumbsch & Leavitt 2011, Leavitt *et al.* 2011, Nagy *et al.* 2012, Moncada *et al.* 2014, Quaedvlieg *et al.* 2014, O'Donnell *et al*. 2015, Del-Prado *et al.* 2016, Lücking *et al.* 2016a, Hawksworth & Lücking 2017). An important step was the community-wide adoption of the nuclear ITS as universal barcoding marker for *Fungi* (Pryce *et al.* 2003, Rossman 2007, Seifert 2008, 2009, Eberhardt 2010, Begerow *et al.* 2010, Vrålstad 2011, Schoch *et al.* 2012; Hibbett & Taylor 2013). Finally, next-generation, high throughput sequencing (NGS, HTS) opened a new dimension to molecular assessment of fungal diversity in environmental samples (e.g. Ronaghi & Elahi 2002, O'Brien *et al.* 2005, Sogin *et al.* 2006, Geml *et al.* 2008, Taylor *et al.* 2008, Buée *et al.* 2009, Amend *et al.* 2010, Lumini *et al.* 2010, Hibbett *et al.* 2011, Unterseher *et al.* 2011, McGuire *et al.* 2012, Hibbett & Taylor 2013, Tedersoo *et al.* 2014, 2017).

The structured query <*Fungi*[organism] AND (5.8S[title] OR ITS1[title] OR ITS2[title] OR ITS[title] OR "internal transcribed spacer"[title])> returned over 1 million (1 042 545) ITS sequences from GenBank (Benson *et al.* 2013, https://www.ncbi.nlm.nih.gov/genbank) on 19 Oct. 2017). The unstructured query <(*Fungi* or fungal) AND (5.8S OR ITS1 OR ITS2 OR ITS OR "internal transcribed spacer")> returned only a slightly higher number (1 065 267). This impressive number corresponds to approximately 30 years of sequencing work. Since 2009, the Sequence Read Archive (SRA; Leinonen *et al.* 2011, Kodama *et al*. 2012) stores data obtained from environmental sequencing studies. Using the unstructured query above (since the structured query does not work in the SRA), the SRA returned 246 studies, 2144 biosamples (= environmental samples), and 20 879 experiments (= NGS runs). Excluding 71 experiments with zero sequences and weighting the remaining 20 818 experiments as 1 (exclusively fungal), 0.5 (mixed fungal and bacterial), and 0 (likely low presence of fungal sequences); we estimated that these data contained over 1 billion fungal ITS reads (1 222 062 203), with an average length of 375 bases (SRA: https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=search_obj) on 19 Oct. 2017; see Suppl. File S1). Thus, at present there are 1000 times more NGS reads in the SRA than sequences in GenBank for the fungal barcoding marker (Fig. 1). Only three years ago, this ratio amounted to about 20:1 (Lücking 2014), which means it has grown by the factor 50 in this short time period and is expected to further increase exponentially, considering the growth rate of SRA data (https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi).

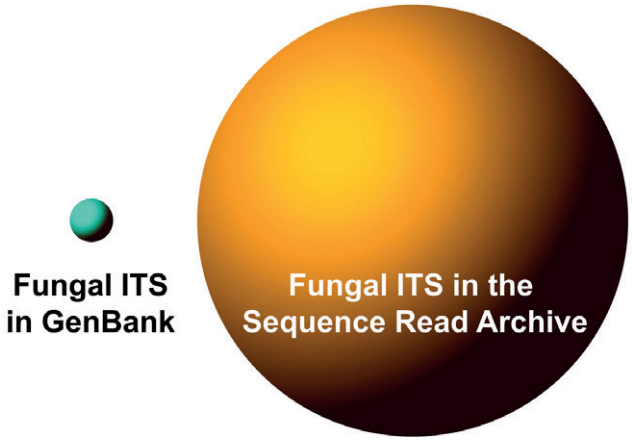

Fig. 1. As of 2017, the fungal ITS universe in GenBank and in the Sequence Read Archive roughly compare to the sizes of Earth *versus* Jupiter.

A substantial proportion of the approximately 1 million fungal ITS sequences in GenBank is unidentified or wrongly labeled or represents unrecognized contaminants (Harris 2003, Vilgalys 2003, Nilsson *et al.* 2005, 2006, 2012, 2014, Meier 2008, Bidartondo *et al.* 2008, Lücking *et al.* 2012, Kõljalg *et al.* 2013). Sixty percent of these correspond to 'uncultured' *Fungi* at best identified to genus level, but most often not identified at all. The number of taxa sequenced is only a portion of all currently accepted *Fungi*, about 35 000 out of 120 000 (C. Schoch, pers. comm., Hawksworth & Lücking 2017). Only properly identified and labeled sequences can be used as reference for accurate fungal identification using ITS barcoding, and clearly there is a need to quickly increase and eventually complete this reference library for all *Fungi* (Meier 2008, Begerow *et al.* 2010, Kõljalg *et al.* 2013, Tanabe & Toju 2013). While the situation is bad in GenBank, the over 1 billion fungal ITS reads in the SRA are not named at all. These data encompass thousands, perhaps tens or hundreds of thousands of novel taxonomic units, from species to class level, and hence provide a substantial source of potential reference sequences far beyond GenBank. To serve as such, they need to be named (Hibbett *et al.* 2011, 2016, Hawksworth *et al.* 2011, 2016, Hibbett & Taylor 2013, Lücking 2014, Minnis 2015, Hibbett 2016). An informal naming system that is not compatible with formal nomenclature, such as the 'species hypotheses' in UNITE (Kõljalg *et al*. 2013), is impractical as a reference library, as informal names or numbers remain obscure without a broadly accepted, formal naming framework. Another shortcoming of a curated database is the amount of data to be handled. UNITE has approximately 800 000 fungal ITS sequences, close to 75 % of what is deposited in GenBank, and corresponding to over 70 000 species hypotheses at 98.5 % similarity threshold (https://unite.ut.ee) on 19 Oct. 2017. To deal with SRA reads in a similar way, UNITE would have to add about 1000 times that number, with an exponential increase in the foreseeable future, an amount of data that is virtually 'incuratable' as so-called 'species hypotheses'. Also, clusters based on a fixed threshold do not necessarily correspond accurately to species (*see below*).

How many *Fungi* exist is unknown. The number of accepted species estimated oscillates between 115 000 and

140 000 (Roskov *et al.* 2016, Species Fungorum http://www.speciesfungorum.org), with a figure of 120 000 presumed reasonable (Hawksworth & Lücking 2017); these variations are attributable largely to allowances made for synonymy and separately named morphs fo the same species. Estimates for global fungal species richness range from 611 000 to 10 million (Hawksworth 1991, 2001, 2012, O'Brien *et al.* 2005, Schmit & Mueller 2007, Blackwell 2011, Bass & Richards 2011, Mora *et al.* 2011), with an often-cited number of 1.5 million and a recent estimate of 2.2–3.8 million (Hawksworth & Lücking 2017). Even with an estimate of 1.5 million, a complete inventory of all *Fungi* on Earth using traditional methods within a reasonable time frame is impossible, given that it took 250 years to discover and describe less than 10 % of that diversity. Furthermore, natural habitats harbouring unknown species are being destroyed at an accelerated rate before they can be inventoried, as a result of the Sixth Mass Extinction (Leakey & Lewin 1995, Wake & Vredenburg 2008, Barnosky *et al.* 2011).

While molecular approaches have revolutionized our understanding of fungal diversity, they have not substantially increased the speed of discovery and formal description of new species. In the two decades prior to the onset of molecular systematics, the average number of newly described species per year was about 1250, slightly increasing to about 1300 in the two decades between 1990 and 2010. With the growth of species delimitation approaches around 2010, this number stands now at about 1750 per year (Hawksworth & Lücking 2017). To classify most or all *Fungi* within a reasonable time, this rate would have to increase by an order of magnitude, an impossible prospect considering the already limited resources of the mycological community and the diminishing number of fungal taxonomists (Gams 1997, Korf 2005, Meier 2008, Hawksworth 2009, Gryzenhout *et al.* 2012, Rambold *et al.* 2013). NGS offers a new approach to fungal inventories, allowing fast detection of a broad range of taxa in a relatively short time and at low cost (Hibbett *et al.* 2011, Grantham *et al.* 2015, Hibbett 2016). Numerous novel *Fungi* have already been discovered from environmental samples, including at higher taxonomic ranks (Jones *et al.* 2011a, b, Rosling *et al.* 2011, Livermore & Mattes 2013, Glass *et al.* 2014, Tedersoo *et al.* 2014, 2017, Lazarus & James 2015, De Beer *et al.* 2016, Nilsson *et al.* 2016). The setback of this approach is that the only manifestation of these *Fungi* are sequence data, unless taxa are successfully cultured, resequenced, and matched to previously obtained sequences (Rosling *et al.* 2011, Menkis *et al.* 2014, De Beer *et al.* 2016).

Box 1. What is a reasonable time frame to classify most or all *Fungi*?
With a rate of 2000 species described per year, it would take 680, 1030, or 1830 years to complete cataloguing 1.5, 2.2, or 3.8 million species respectively. Increasing our current efforts by an order of magnitude, it would take between 68 and 183 years. A more realistic increase, e.g. 5,000 species per year, would result in 272 to 732 years. A "reasonable" time frame is arbitrary, but several factors can be considered. For one, conversion of natural ecosystems is accelerating, and the Sixth Mass Extinction is well under way. Second, global climate change will have substantial effects predicted within the next few decades. Third, there is the life-span of a scientific career, usually 50 years or less. Facing accelerated habitat destruction, global climate changes, and the fact that we would like to see substantial progress in terms of cataloguing fungal diversity within the life span of our own careers, including young mycologists now starting at the age of about 25, we can set a reasonable time frame at 50 years. At the current rate, we would then formally describe another 100,000 species if we stick to specimen- and culture-based inventories, for a total of 240 000, 16 % of 1.5 million and 6 % of 3.8 million. To get much closer to the goal, we would have to increase our efforts by an order of magnitude, achieving 1 million new species within 50 years or an impressive 76 % of 1.5 million, but still only 30 % of 3.8 million. A 10-fold increase of specimen- or culture-based inventories is, however, not going to happen; at best we could reach 5000 species per year, for a total of 250 000 in 50 years. With that, we approach 26 % of 1.5 million but only 10 % of 3.9 million, and we would barely reach the number of plant species. Therefore, if the mycological community cannot convincingly argue that we will reach a 10-fold increase of our specimen- and culture-based efforts almost instantly, there is no alternative to including voucherless, sequence-based formal nomenclature as an additional method of classifying fungal diversity if we want to see substantial progress during the next five decades.

An example illustrating the problem is the class *Archaeorhizomycetes*, originally established on a single genus and species, *Archaeorhizomyces finlayi*, with a second species described later, both based on permanently preserved cultures (Rosling *et al.* 2011, Menkis *et al.* 2014). Based on additional sequences from GenBank and a limited sample from the SRA, this class was estimated to contain close to 500 species (Menkis *et al.* 2014). SRA blast search on 9 Nov. 2014 using the ITS sequence of the type species retrieved 106 563 reads from environmental samples belonging to this class (Smith & Lücking, unpubl.; Suppl. File S2). With the overall increase in fungal ITS reads in the SRA by the factor 50 since 2014, this number could now potentially amount to about 5 million. Analysis of this data set using clustering through USEARCH (Edgar 2010) suggests the presence of between 28 435 species at 99 % threshold level and 2,658 species at 95 %; with the UNITE 'species hypothesis' threshold of 98.5 %, the estimated number of species would be 16 231. Preliminary phylogenetic analysis based on multiple sequence alignment suggests around 1000 taxa, apparently corresponding to separate genera, families, and perhaps orders within the class *Archaeorhizomycetes*. Irrespective of the accurate number, the magnitude of the problem is illustrated by the fact that, since the original

discovery of this class, only two valid names have been established (a rate of 0.3 per year). Therefore, adhering to the requirement of physical type specimens, including cultures, for the valid description of novel *Fungi* detected through environmental sequencing, is illusory.

It is inconceivable that a sizable proportion of *Fungi* from environmental sequencing reads will ever be documented through specimen-based fungal inventories or culturing. Culturing only detects a portion of the fungal diversity present in a sample (Arnold *et al.* 2007, De Beer *et al.* 2016), and the examples of *Archaeorhizomyces*, *Hawksworthiomyces*, and *Cryptomycota* (Livermore & Mattes 2013, Letcher *et al.* 2013, Lazarus & James 2015, De Beer *et al.* 2016), show that culturing hardly makes a dent into the huge number of *Fungi* to be formally described from environmental samples, simply because there are no capacities for a global approach to catalogue millions of species that way. The *Westerdijk Fungal Biodiversity Institute* (CBS; formerly the *CBS-KNAW Fungal Biodiversity Institute* and *Centraalbureau voor Schimmelcultures*) and the *ARS Culture Collection* (NRRL) are the largest public service fungal culture collections in the world, with about 50,000 and 68,000 strains, respectively (for CBS see below, for NRRL data provided by T. Adkins, pers. comm. 2017). Both have contributed substantially to fungal ITS sequences in GenBank (https://www.ncbi.nlm.nih.gov/biocollections/?term=cbs; https://www.ncbi.nlm.nih.gov/biocollections/3689). The search string <*Fungi* AND CBS AND (5.8S OR ITS1 OR ITS2 OR ITS OR "internal transcribed spacer")> returned 37 680 fungal entries, including almost 10 % of sequences from type material on 19 Oct. 2017; just <*Fungi* AND CBS> returns 863 723 fungal entries. For NRRL, there are 5340 ITS sequences and 209 624 fungal entries overall. Over 90 % of the CBS entries are identified to species, corresponding to nearly 9000 taxa. While this level of resolution is impressive, the identified taxa constitute just 7 % of the currently accepted *Fungi*, a proportion that decreases to far less than 1 % if we assume up to 3.8 million predicted species.

Even if CBS and other large fungal culture collections could increase their efforts by an order of magnitude, culture-based fungal inventories would still be incapable of dealing with even the most conservative species-richness predictions in a reasonable time frame. CBS had 51 908 fungal strains corresponding to 15 526 species on 5 Dec. 2017 (www.westerdijkinstitute.nl/Collections/localfiles/CBSStrainsJuly21st2016.zip). A ten-fold increase to about 500 000 strains, apart from being logistically challenging, if not impossible, may increase the number of taxa to about 150 000. If we assume three large culture collections, with a taxonomic overlap of 50 %, we would stand at 300 000 taxa. Thus, an already impossible effort by the two cited large culture collections to augment their capacities by the factor ten, plus adding a third such collection, would increase the proportion of known species to just 20 % (if we assume 1.5 million) or even less than 10 % (if we assume 3.8 million). Clearly, the bulk of *Fungi* detected through environmental sequencing cannot be formally named if not also based on sequence data. Whatever reservations there may be against this approach, it seems impossible to conceive a practical alternative. Leaving this diversity unnamed and unclassified is not an option, as it would continue to be an enormous and increasing impediment to communication and research in the field.

In order to address this problem, a proposal had been put forward to modify the *Code* to allow sequences as types (Hawksworth *et al.* 2016). This proposal was not supported by the Nomenclature Committee for *Fungi* (Turland & Wiersema 2017) and was rejected by the Nomenclature Section of the International Botanical Congress in Shenzhen 2017. However, the Congress established a Special Committee to examine the matter for all groups of organisms which is due to report to the next Congress in 2023 (Hawksworth *et al*. 2017). Some authors have nevertheless already described new species based on an environmental sample type, such as *Piromyces cryptodigmaticus* (Fliegerová *et al.* in Kirk 2012), or with a sequence type, such as *Hawksworthiomyces sequentia* (De Beer *et al.* 2016), a currently invalid name established in anticipation of changes to the *Code*. A potential loophole for the formal description of voucherless, ecologically cryptic microfungi based on sequence data was posited by invoking the 'illustration clause' in Art. 40.5 (De Beer *et al.* 2016, Lücking & Moncada 2017, Turland & Wiersema 2017). However, this led to a suggestion during the Nomenclature Section meeting to redefine what constitutes an illustration allowed as type; an example was inserted into the *Code* to close this potential loop-hole (Turland *et al.* 2018: Art.40.5 Ex. 5), making clear that a representation of a sequence was not to be interpreted as an illustration for the purposes of typification; this option cannot now be the subject of a proposal until the next International Botanical Congress in 2023.

In order to stimulate discussion of this issue prior to the XIth International Mycological Congress (IMC11) in Puerto Rico in July 2018, to which a similar proposal to allow sequences to serve as types of fungal names has been submitted (Hawksworth *et al*. 2018), we elaborate here on the promises and pitfalls of formal, sequence-based, voucherless nomenclature. We offer solutions to problems at hand that could lead to specific provisions being made in the *Code* at IMC11 to allow formal, sequence-based nomenclature for voucherless fungi.

THE TYPE CONCEPT IN SEQUENCE-BASED NOMENCLATURE

The purpose of a type is to fix the application of a name. Presently, this has to be a physical specimen (including a permanently preserved, metabolically inactive culture in the case of *Fungi*), or if none can be preserved, in certain circumstances an illustration (Lücking & Moncada 2017, Turland *et al*. 2018). Apart from linking a name to a specimen, a name-bearing type has the following functions:

- **Depiction of the phenotype, including morphological, anatomical, chemical, and physiological characters.**
- **Long-term (ideally perpetual) conservation of the original material.**

Table 1. Possible alternative types for sequence-based nomenclature and their advantages and disadvantages. Desirability refers to a purely scientific viewpoint, without considering the actual necessity; risk refers to the possibility of undesired outcomes, such as artifactual taxa or newly described synonyms; feasibility refers to the efforts required to designate and store a type; and effectivity refers to the proportion of effort versus gain in closing the gap of undescribed fungal diversity.

	Specimen	Culture	DNA extract	Environmental sample	FISH	Illustration	Sequence
Phenotype	yes	yes	no	no	no	no	no
Comparability	yes	yes	no	no	no	limited	limited
New characters	yes	yes	limited	limited	no	no	no
Quality control	yes	yes	limited	limited	yes	no	no
Artifactual taxa	rare	impossible	possible	possible	impossible	possible	possible
Technical requirements	low	high	medium	medium	high	low	low
Storage requirements	medium	high	medium	high	medium	low	low
Long-term preservation	limited	limited	limited	limited	limited	yes	yes
Capture of diversity	low	low	medium	medium	medium	high	high
Code compliancy	yes	yes	limited	limited	yes	limited	no
Changes to the *Code*	none	none	"mixtotype"?	"mixtotype"?	none	clarify	"sequence type"
Overall desirability	high	very high	medium	low	high	very low	very low
Overall risk	low	very low	medium	high	very low	high	high
Overall feasibility	very low	low	high	medium	medium	very high	very high
Overall effectivity	very low	low	medium	high	medium	very high	very high

- **Reassessment of characters whenever necessary.**
- **Comparison with other specimens to establish their identity.**
- **Assessment of additional and new characters, including through new technology.**

Seven possible types could be conceived to formally describe new *Fungi* from environmental sequencing data (Table 1). These can be divided into four groups: (1) physical type specimens (dried specimen, metabolically inactive culture); (2) undefined mixed samples (DNA extract, environmental sample); (3) a novel physical type derived *via* FISH technology (Spribille *et al*. 2016); and (4) the sequence data itself. Physical type specimens fulfil all five principal functions of a type, are *Code* compliant, and score high in terms of quality control and assessment of phenotype and novel characters (Table 1). However, to obtain physical types from taxa detected through environmental sequencing is not feasible at a large scale and this approach would defy the concept, since ultimately the type sequence would be obtained by resequencing the specimen and not from the original environmental sequence data. Thus, by default, sequence-based nomenclature cannot operate with traditional physical types, which in effect leaves only the five options in categories two to four above.

Some workers proposed using the environmental sample from which sequences were obtained as type material, in what can be referred to as a 'bag' type (Kirk 2012, Hawksworth *et al.* 2011, Hibbett & Taylor 2013, Minnis 2015, De Beer *et al.* 2016). This complies with the *Code* in having a physical type and hence fulfils a formal requisite for valid description: according to Art. 40.2, "*... indication of the type as required by Art. 40.1 can be achieved by reference to an entire gathering, or a part thereof, even if it consists of two or more specimens as defined in Art. 8 ...*" (Turland *et al*. 2018). Although valid, for practical purposes this is not feasible, for three reasons: (1) the precise specimen to which a sequence belongs cannot be located within the sample, except for techniques such as fluorescent *in situ* hybridization (FISH; e.g. Spribille *et al.* 2016); (2) it is uncertain whether a fungus detected in the portion of the sample used up in the study is actually present in the remaining sample (De Beer *et al.* 2016); and (3) samples would have to be stored in long-term preservation in a frozen state to allow for further access of DNA material, to render the type material actually useful. In any case, such a type would be ambiguous, which would require subsequent lectotypification, generating the very problems outlined above, in that a precise lectotype cannot be designated.

Another option is to designate the DNA extract from which a type sequence originated as type. While permanent storage of a DNA extract is more feasible than the corresponding environmental sample, and the DNA that produced the sequence is likely to be contained in the remaining extract, a DNA extract type has the same problems as a 'bag ' type, in that the precise piece of DNA corresponding to a particular taxon cannot be located within the extract. In addition, it might be argued whether a DNA extract type is still in compliance with the *Code* since, contrary to a 'bag' type, it does not contain an actual fungus. A type based on fluorescent *in situ* hybridization (FISH), a technique for instance performed in *Cryptomycota* and *Cyphobasidium* (Jones *et al.* 2011a, Spribille *et al.* 2016), would appear to be an ideal compromise between the extremes of a physical type and a sequence type. This technique would use the type sequence of a clade recognized as new taxon to precisely locate and visualize the corresponding physical structures (cells or hyphae) in the underlying sample, which could then be photographically documented and stored as a

permanent slide (similar to a metabolically inactive culture). The immediate advantage of this approach would be the implicit cross-check of the sequence data, since only real sequences would lead to a positive result. On the other hand, it would be difficult to validate this approach *a posteriori* unless the fluorescent effect is permanent in the type slide. In case of a 'bag' type, a DNA extract type, or a FISH type, valid description of new, sequence-based taxa would only be possible with simultaneous access to the original material and its subsequent deposition in an institutional collection where it would be permanently accessible to researchers.

A voucherless sequence type fails on depicting a phenotype and assessment of additional and new characters (by default, all of its characters have been assessed through initial analysis), but fulfils the other three criteria better than a physical type. Whereas a physical type degrades over time, a sequence type can be stored as a digital file in perpetuity without quality loss. Digital data may be subject to technical failure and cyberattacks, but this applies to any electronic data and is not specific to the issue at hand. There is an equal likelihood of damage to physical type specimens, e.g. through pests, mould, humidity, water damage, and fire (Metsger 1999), as evidenced by the loss of most of the Berlin (B) collections in World War II (Hiepko 1987). Since a digital type can be stored in multiple identical copies, the risk of a complete loss of the information is much lower than for any physical type material. Effectively, a digital sequence type is an 'exsiccate' with an unlimited number of copies. Type sequences are universally accessible and their restudy is not destructive. In contrast, physical types need to be located, borrowed, or require a visit to study them. In addition, restudy is destructive, and so reduces their value as a reference point over time, for example as sporing structures are removed and samples taken for thin-layer-chromatography. Study of physical types is also dependent on methodology, can be open to interpretation and can lead to ambiguous results. In contrast, a sequence corresponds to a defined set of features (four possible states per character expressed by the universal IUPAC letter code), and the characters and their features underlie specific rules for their assessment, for example by checking against an original trace file. Two workers assessing ascospore dimensions in the same type may obtain different results; in sequence data, at a given position, an 'A' is an 'A'.

Box 2. Environmental samples are real

There is a misconception that environmental sequences are not real but consensus sequences. This is not correct. Environmental reads deposited in e.g. the SRA, with an individual accession number, are as real as Sanger contigs deposited in GenBank. OTUs in analyses of environmental studies that are derived as consensus sequences from clustering are not real sequences and these are not acceptable as type sequences. A type sequence is always a real sequence and ideally it is the best sequence representing the lineage (or cluster). This is comparable to physical types: when a new species is described based on more than one collection, the description is a consensus of all specimens, but the type cannot be a consensus and instead must be a single collection or specimen, usually the one that is best developed. Another argument is that sequences do not correspond to real organisms but are trace manifestations. While this is true, there is another example of formal nomenclature of *Fungi* in which the types are trace manifestations. With some exceptions, fossils are traces of organisms obtained through impression or crystallization, and their interpretation is often speculative. Yet, their nomenclature is fully integrated into the *Code*, with some specific provisions.

An important argument regarding nomenclatural types is repeated, unlimited and free access. For physical types this criterion does not apply, although the problem is in part remedied by the availability of digital type images, e.g. through the *Global Plants Initiative* on JSTOR (http://gpi.myspecies.info; https://plants.jstor.org). In contrast, type sequences can be accessed and compared to other sequence data in unlimited ways and in reproducible fashion, using quantitative methods such as automated alignment, assessment of alignment ambiguity, phylogenetic analysis, and species recognition methods. As a consequence, while the ideal situation is to have physical types plus sequence data in order to apply a consolidated species concept (Quaedvlieg *et al.* 2014), type sequences, while not displaying phenotype features or harboring potential new characters, are superior to physical types in three of the five criteria listed above.

POTENTIAL PITFALLS OF DNA SEQUENCE TYPES

Sequence errors

Physical types may have flaws. A type does not usually encompass the phenotypic variation of a species. It need not be "typical", since a species usually becomes much better known after its original description. It might not exhibit all characters that define the species, especially if the taxon occurs in various sexual, asexual and vegetative morphs, or it might be a mix of more than one taxon or else aberrant or an monstrosity. Sequence data may have errors analogous to those of physical types.

One of the most serious problems is chimeras, occurring both in Sanger and NGS techniques, as well as base flow (homopolymer) errors, most typical of Roche 454 and Ion Torrent platforms, and tag switching (Carlsen *et al*. 2012, Luo *et al.* 2012, Yergeau *et al.* 2012, Salipante *et al.* 2014, Goodwin *et al.* 2016). Chimeras arise from template DNA representing more than one taxon during PCR (Haas *et al.* 2011). A mixed template in a Sanger PCR will cause double

Box 3. Physical *vs.* Sequence types
Opponents of sequence-based nomenclature argue that, in contrast to characters from physical types, environmental sequences cannot be reproduced. This is not correct. Environmental sequences come from samples and DNA extracts and if properly stored, sequences are theoretically reproducible. At the same time, physical types can have diagnostic characters that are not apparent in dried fungarium collections, such as basidiome cap colour, ascospore halos, conidial appendages which also cannot be resurrected. This is also so for metabolically inactive cultures designated as types from cultured *Fungi* where it can be impossible to recover some of the diagnostic features of a living culture from these types. The issue is therefore comparable to names based on physical types and cannot be upheld as an argument against sequence-based nomenclature.

or multiple peaks at a given position. The trace files are easily recognized and dismissed. In a rare constellation, a primer pair may have differential affinity to one template or another, generating clean trace files of different taxa for each primer. Sequence assembly will result in numerous ambiguous base calls except for conserved regions, making such chimeras again easily detectable. Very unlikely, but not impossible, is the above case but with reduced read length due to particular cycle conditions. For instance, in the case of ITS, the forward primer would sequence the ITS1 region and the reverse primer the ITS2. Through the conserved 5.8S region, such reads would be assembled into chimeras without immediate detection, but they can be identified in a phylogenetic context: since they have unique, artifactual sequence patterns and are pulled towards two separate lineages simultaneously, they will appear on long branches with low basal node support. Dividing a data set into ITS1 and ITS2 and analyzing these separately (e.g. Blaalid *et al*. 2013), with a subsequent test for topological conflict, is thus a straightforward strategy to detect chimeras.

NGS chimeras arise from mixed DNA templates in amplicon PCR, when the amplicon from one template finishes prematurely and in the next cycle another template attaches (Haas *et al.* 2011). Since this is a stochastic process, the PCR product results in a mix of templates with regions of close correspondence and regions of disparate base calls. It is unlikely that such a mixed PCR product produces a sequence read passing through quality filters, since many positions will have subpar signal due to the presence of mixed bases in a given position between individual DNA fragments. In the unlikely event that the PCR combines two different templates at the same position, a true chimera corresponding to a high quality read similar to Sanger chimeras would result, with the difference that the joining point was not obvious; the only way to spot such chimeras would be to divide the read into variable portions starting from the centre and simultaneously blast both. The proportion of chimeras in NGS amplicon sequencing typically ranges between 8 % and 17 % for raw reads, and there are various tools for chimera detection and removal that reduce the proportion of chimeric reads to about 1 % (Huber *et al.* 2004, Ashelford *et al.* 2005, Edgar *et al.* 2011, Quince *et al.* 2011, Schloss *et al.* 2011, Porazinska *et al.* 2012, Kim *et al.* 2013, Mysara *et al.* 2015, Edgar 2016).

Carry-forward-incomplete-extension (CaFIE) errors are mostly generated in Roche 454 pyrosequencing and on Ion Torrent platforms, while apparently not occuring on Illumina platforms, although the latter has other sequencing errors (Minoche *et al.* 2011, Loman *et al.* 2012, Luo *et al.* 2012). During sequencing, the extension in a given well follows a Poisson distribution, with most fragments fully, but a small portion only partially extended. Depending on the proportion of extended fragments, this leads to a suboptimal light signal which will be interpreted as either base not present or as a homopolymer of shorter length (Margulies *et al.* 2005, Huse *et al.* 2007, Gomez-Alvares *et al.* 2009, Kunin *et al.* 2010, Niu *et al.* 2010, Tedersoo *et al.* 2010, Balzer *et al.* 2011, Lücking *et al.* 2014). Incompletely extended homopolymer fragments become desynchronized and are completed during the next cycle of the corresponding flow base, causing a misplaced signal several bases after the homopolymer, not detectable as error but mimicking a genuine substitution. The only way to detect such errors is through alignment of reads relative to a broad reference alignment which will then place misplaced base calls in largely gapped columns (*see below*). CAFIE errors depends on the location of homopolymers and their length, with the consequence that phased indels can appear at the same position in independent reads; as a consequence, erroneous sequences are not necessarily singletons and they do not exhibit random patterns, which makes their automated detection close to impossible. It has been shown that such erroneous sequences can inflate taxonomic diversity computed through clustering techniques by several orders of magnitude, whereas multiple alignment-based methods are not susceptible to this problem (Lücking *et al.* 2014; *see below*)

There are various approaches available to detect, filter, and manage artifactual sequences, so that the problem of inadvertently including artifactual data in an analysis leading to recognition of artificial taxa can be reduced to a manageable proportion of less than 5 %. The most commonly used approaches exclude singletons and rare sequence reads or reads that cannot be mapped to reference taxa (Tedersoo *et al.* 2010, Nilsson *et al.* 2011, Caporaso *et al.* 2012, Edgar 2013). However, this will also exclude genuinely rare taxa (Lim *et al*. 2012), as was the case in an unnamed Alaskan soil fungus (Glass *et al.* 2014). Other tools to be tested to potentially detect aberrant and artifactual sequences include assessing the secondary structure of ITS reads (Goertzen *et al.* 2003, Morrison 2009, Glass *et al.* 2013, Koetschan *et al.* 2014, Coleman 2015, Giudicelli *et al.* 2017). The likelihood of artifactual sequences obtained from different studies being so similar that they form a well-defined and well-supported species-level clade is remote. Therefore, instead of simply excluding rare sequences or singletons from submitted biosample runs, the best approach to filter potential artifacts is to only allow formal description of novel species if the

sequences defining a species-level clade have been detected in a number of independent samples, with that number small enough to consider rare species and high enough to provide effective quality control. This number could be determined by a simple formula relating to the probability of sequence reads coming from N separate, independent studies, forming a supported clade, and at the same time being artifactual; it can be shown that N ≥ 5 fulfils this requirement. This had been proposed as a recommendation (Hawksworth *et al.* 2016, 2018), but could be made a mandatory requirement for taxa based on NGS reads.

The following guidelines would help or substantially reduce the probability of describing artifactual taxa based on faulty sequence data:

- **Dismiss sequences with a high proportion of ambiguous base calls (e.g. over 5 %).**
- **Use filtering techniques for raw NGS sequence data to automatically detect and remove chimeras** (e.g. UCHIME, UCHIME2; Edgar *et al.* 2011; Edgar 2016).
- **Divide ITS data set into two portions, analyse them separately and test for topological conflict and rogue OTUs** (e.g. Blaalid *et al*. 2013).
- **DO NOT USE CLUSTERING TECHNIQUES! Instead, apply multiple alignment techniques aligning reads to a reference alignment and check for gap-rich indel columns** (e.g. PaPaRa, MAFFT '--add'; Berger & Stamatakis 2011, 2012, Katoh *et al*. 2017).

Conceptual errors in sequence-based species delimitation

Besides faulty sequences, sequence-based nomenclature is prone to conceptual errors that may lead to inaccurate recognition of taxa. The most commonly cited problems are homoplasy, intragenomic variation and gene duplication, and lack of resolution at species level (O'Donnell& Cigelnik 1997, O'Donnell *et al.* 1998, Hassanin *et al.* 1998, Kälersjö *et al.* 1999, Inderbitzin *et al.* 2009, Druzhinina *et al.* 2010, Gazis *et al.* 2011, Kovács *et al.* 2011, Coissac *et al.* 2012, Kiss 2012). These problems are not unique to sequence-based taxa but apply to phylogenetic species recognition in general; therefore, the possibility of conceptual errors cannot be held exclusively against the idea of sequence-based nomenclature. The difference is that sequence-based taxa do not allow for independent, specimen-based check of sequence data; also, multi-locus approaches to detect problems with individual markers are not (yet) possible in sequence-based nomenclature (*see below*). However, the probability for such conceptual errors to occur is not higher in sequence-based than in specimen-based taxa. For the latter, a pleothora of studies supports the value of molecular phylogeny for taxonomy, systematics and classification, in spite of the occasional shortcomings (Rossman 2007, Seifert 2008, 2009, Begerow *et al.* 2010, Schoch *et al.* 2012).

Some workers have argued that DNA homoplasy is more frequent than phenotype homoplasy (Baker *et al.* 1998; Wiens 2004). This is based on the misguided concept of total evidence, in which morphological homoplasy is potentially masked as incorrect phylogenetic signal. When mapping phenotype characters on trees derived from sequence data only, it can be shown that phenotype characters are often ecologically overformed and obscure true evolutionary relationships (Wake 1991, Hall 2003). An outstanding example is the molecular phylogeny of the *Fungi*, which has dramatically altered our understanding of fungal evolution and classification, from kingdom to species level. Yet, misunderstandings about sequence data persist. While the *Nomenclature Committee for Fungi* did not support the proposal on allowing sequences as types (Hawksworth *et al.* 2016), the Rapporteurs expressed their concern about a presumed "*... lack of control as to the type sequence being an informative sequence. Many taxa could have the same sequence*." (Turland & Wiersema 2017: 225). Homoplasy is of course present in DNA sequence data; for instance, in the third-codon position of protein-codon genes (Hassanin *et al.* 1998, Kälersjö *et al.* 1999), there is a 25 % probability that the same base call arose by chance in two unrelated sequences. The same applies to ITS sequences, in which saturation effects, even if they cannot be directly measured due to ambiguous alignment positions, may occur in the highly variable ITS1 and ITS2 regions. However, contrary to phenotype data, in which characters are subjectively weighted, DNA-based phylogenies are based on simultaneous, unweighted assessment of all characters. For instance, in a 1000-bases long protein marker, there are about 300 third codon positions that may evolve freely and develop homoplasy due to saturation effects. For each individual position, the probability of homoplasy is 25 %, but the probability of two entire, unrelated sequences to evolve a similar pattern over 300 positions by chance is effectively zero. Thus, whatever the notion on DNA homoplasy may be: it is virtually impossible to imagine that two highly similar sequences of a given length evolved independently by chance. Sequence similarity is therefore always to be interpreted as indicating common descent, albeit possibly obscured by mechanisms such as hybridization or horizontal gene transfer.

There is, however, the problem of a lack of phylogenetic resolution due to homoplasy in recently and actively evolving lineages with incomplete lineage sorting (Will *et al.* 2005, Inderbitzin *et al.* 2009, Druzhinina *et al.* 2010, Gazis *et al.* 2011, Dupuis *et al.* 2012, O'Donnell *et al.* 2015). However, sequence-based nomenclature does not aim at resolving species complexes, it aims at naming novel lineages. It is therefore of no practical consequence if, in some cases, clades in voucherless taxonomy are erroneously defined as species when in reality they represent a species complex. As long as there are no associated physical specimens, there is no way of knowing, and this would only lead to an under-recognition of novel taxa. Long-branch attraction due to presumed DNA homoplasy (Bergsten 2005, Kück *et al.* 2012, Susko 2014) is a different issue that does not apply here. Taxa falsely clustering on long branches have both a long shared stem branch and long terminal branches, a pattern

Box 4. Species delimitation
Voucherless sequence-based nomenclature is not about accurately delimiting species in *Fungi*. The sole purpose of this approach is to convert countless unnamed into named sequences so that they can serve as proper references. This approach only works if done with a single marker, with the obvious choice being the fungal barcoding marker ITS. It does not matter whether ITS may not fully resolve species in particular groups or may over-resolve species as infraspecific entities in others. What matters is that a particular ITS sequence, when compared to a reference database, returns names, instead of designations such as "uncultured fungus", "uncultured *Ascomycota*", or "ectomycorrhizal fungal clone" etc. Whether such names are actual species, species complexes, or infraspecies, is irrelevant to the current debate; they will always point to the correct lineage at the given level of resolution and this will greatly facilitate correct classification of environmental or otherwise unknown sequences. Potential conflict arises when specimens or cultures become available for an initially voucherless lineage and multilocus studies may show that ITS in such lineages does not properly resolve species. However, such conflict can be handled: if ITS is too variable, with additional markers resulting in less species, formal synonymization takes place. If ITS lacks resolution and a different marker is a better barcode, then the initial name is epitypified with one of the specimens for which the additional marker is available and the name is attached to that particular clade. This process is the same as in nomenclature based on physical types.

different from species-level clades, in which the terminal branches leading to individual sequences are short; hence, long-branch attraction cannot lead to artifactual species-level entities.

Intragenomic variation and gene duplication (paralogs, pseudogenes), as well as horizontal gene transfer, may be serious issues resulting in artifactual topologies. Horizontal gene transfer has been demonstrated in *Fungi* (Schmitt & Lumbsch 2009, Soanes & Richards 2014), but is not expected to pose problems in species delimitation studies, particularly not with a multiple-copy marker such as ITS. Gene duplication has been found in various protein coding genes, such as β-tubulin, TEF1 (EF1α), and PKS genes (Schmitt *et al.* 2005, James *et al.* 2006, Aguileta *et al.* 2008, Hubka & Kolarik 2012), and using such markers may result in duplicate clades mimicking separate taxa. In contrast to protein-coding genes, rDNA occurs in multiple copies in large arrays in the genome and is presumed to maintain a consistent sequence pattern due to concerted evolution (Hurst & Smith 1998, Liao 1999, 2008, Ganley & Kobayashi 2007). Evidence for potential intragenomic ITS variation in *Fungi* is inconclusive; studies have demonstrated both presence and absence of such variation, using techniques such as RFLP, cloning, NGS amplicon sequencing, whole genome sequencing including HTS, and specifically designed primers (O'Donnell & Cigelnik 1997, O'Donnell *et al.* 1998, Ganley & Kobayashi 2007, Simon & Weiß 2008, James *et al.* 2009, Lindner & Banik 2011, Kovács *et al.* 2011, Kiss 2012, Lindner *et al.* 2013). Some methods are subject to the observer effect, in that variation is generated by methodical errors rather than being intrinsic (Keirle *et al.* 2011, Lücking *et al.* 2014, Mark *et al.* 2016). Lücking *et al.* (2014) reported that up to 99.3 % of indel variation in 454 pyrosequencing ITS reads from a single target taxon were due to sequencing errors, particularly homopolymer (CAFIE) errors, with genuine variation almost entirely ascribed to substitutions (Fig. 2). In the cloning approach by Simon & Weiß (2008), the proportion of variant base calls indicated in the supplemental figure [mbe-08-0468-File005_msn188.pdf] is 0.11 %, at reported TAG polymerase error levels (Chen *et al*. 1991), and the authors admit that *in vitro* TAG misreadings might cause such variation (Simon & Weiß 2008: 2251). Even if the rDNA cistron underlies strict concerted evolution, it can be expected that each generation of ITS copies after replication will have a natural level of variation corresponding to DNA polymerase misreadings *in situ*. In any case, such point mutations or single nucleotide polymorphisms (SNPs), whether real or methodological, would not result in artifactual taxa when analysed in the context of multiple alignments, whereas clustering methods are highly sensitive to such variation (see below). Sanger sequencing usually gives a consistent signal corresponding to the dominant haplotype, which is supported by the numerous studies in which the ITS barcoding approach appears to work well (Rossman 2007, Begerow *et al.* 2010, Gazis *et al.* 2011, Schoch *et al.* 2012).

Pyrosequencing analyses demonstrated generally low intragenomic ITS variation in a broad set of fungal taxa (Lindner *et al.* 2013), and potential gene duplication involving the ITS has been reported from a few lineages only (O'Donnell& Cigelnik 1997, O'Donnell *et al.* 1998, Hughes & Petersen 2001, Ko & Jung 2002, Gomes *et al.* 2002, Smith *et al.* 2007a, Li *et al.* 2013). In most cases, this phenomenon is explained by past hybridization, and it appears to be highly constrained in the fungal genome (Wapinski *et al.* 2007) and hence would have minor impact on species delimitation approaches. In the study by Lindner & Banik (2011), considerable intragenomic ITS variation was reported for *Laetiporus cincinnatus*. Reanalysis of the original data (not shown) recovered these results which, however, suggest

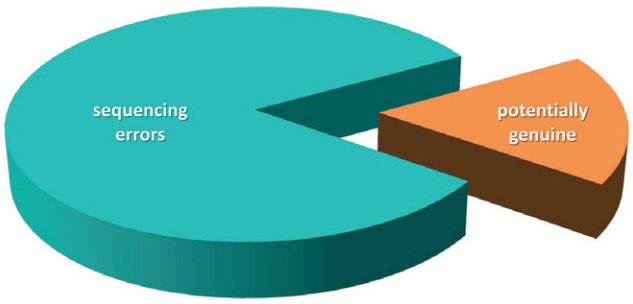

Fig. 2. Proportion of presumably genuine intragenomic variation versus sequencing errors in 18 933 indels and substituions detected among 16,665 pyrosequencing reads of the ITS in the basidiolichen fungus *Cora inversa* (after Lücking *et al.* 2014). Almost all genuine variation is ascribed to substitutions.

hybridization as the cause: all "rogue" haplotypes cluster with strong support with other *Laetiporus* species and so cannot be the result of intragenomic evolution of new ITS variants. If intragenomic ITS variants are causes by hybridization, detection of such variants in voucherless sequence data would not lead to artifactual taxa, since individual ITS clones would always belong to an existing species, even in a hybrid genome.

Many species-delimitation studies attempt to obtain fungal ITS barcoding and other markers from physical type specimens, indicating a community consensus that sequence data can properly place types within a phylogenetic framework, and hence allow for a proper application of the names attached to them. It is therefore not logical to argue that sequences as types would not work or would be inferior to physical types. DNA sequence data have already been used as sole diagnostic characters (Fliegerová *et al.* in Kirk 2012, Tripp & Lendemer 2012, 2014, Renner 2016, Lücking *et al.* 2016b). The argument that recovery and validation of a sequence from the material cannot be guaranteed is not relevant, as the same problem may exist with ephemeral phenotype characters of physical types, an issue not confined to fungi and seen, for example, in the highly diagnostic oil bodies in *Hepaticae* (von Konrat *et al.* 2012, He *et al.* 2013). There is also the 'reverse epitype' concept: currently, in a molecular framework, epitypes are designated based on specimens from which sequence data were obtained, to complement original physical type material. In analogy, when a fungus originally described from voucherless type sequences is eventually discovered as physical specimen, that material can be designated as an epitype to depict the phenotypic features of the fungus.

PARALLEL CLASSIFICATIONS

Apart from the possibility of formally establishing artifactual species based on erroneous sequence data or unrecognized conceptual pitfalls such as gene duplication, another major pitfall of sequence-based nomenclature is the establishment of parallel species-level classifications, either by describing new species that potentially have a name among the numerous unsequenced *Fungi* or by separately using different markers that cannot be traced back to a single taxon.

Accidental *de novo* descriptions of the same species

The number of fungal species has been estimated conservatively at 1.5 million, with other estimates ranging between 611 000 and up to 10 million (Hawksworth 1991, 2001, 2012, O'Brien *et al.* 2005, Schmit & Mueller 2007, Blackwell 2011, Mora *et al.* 2011, Hawksworth & Lücking 2017). With 120 000 species currently accepted, this means that in the best case scenario, over 500 000 species are still to be discovered; using the recently proposed range of 2.2 to 3.8 million, at least over 2 million await formal recognition. About 240 000 species-level names have been described in *Fungi*, apart from the 120 000 accepted species, another 120 000 considered synonyms or orphans[1] (Hawksworth & Lücking 2017). Presently, about 35 000 species have sequence data available. Thus, if we assume a scenario of 120,000 accepted species, of which 35 000 have been sequenced, with a total of 1 million species existing and half of the presumably synonymous names or orphans not being conspecific with any of the 120 000 accepted species, a random set of environmental sequencing data would resolve as follows if a random representation of fungal diversity was assumed: 3.5 % of the sequences would cluster with accepted species and 96.5 % appearing novel; of the latter, 8.5 % would correspond to accepted, yet unsequenced species, 6 % to species with names potentially available but not currently in use, and 82 % to genuinely novel taxa, resulting in a probability of 14.5 % of newly describing species that already have names. This probability decreases assuming a higher total of fungal species (Table 2). If among the 240 000 existing names, in addition to the 120 000 accepted species, there are further 60 000 hidden in synonyms and orphans, the overall error rate for taxonomy based on physical types over the past 250 years is 33 %. Therefore appears that a projected, statistical rate of between 14.5 % and 1.5 % newly generated synonyms for sequence-based nomenclature would be a considerable improvement over specimen-based nomenclature.

There are ways to deal with this problem. Apart from unknown fungal lineages, environmental sequencing techniques frequently yield *Fungi* in well-known taxa, such as ectomycorrhizal species of, e.g. the genus *Russula* in soil samples or species of *Xylaria* in endophyte studies (Arnold *et al.* 2003, Davis *et al.* 2003, O'Brien *et al.* 2005, Geml *et al.* 2010). Since only a fraction of described species in such genera has been sequenced, sequence-based nomenclature would allow establishment of new, voucherless taxa that may already have a name. This is not the objective of sequence-based nomenclature, which should aim at formally classifying genuinely novel taxa and not interfering with other, integrative approaches to classify *Fungi*. For instance, the genera *Russula* and *Xylaria* contain around 750 and 300 accepted species, respectively (Kirk *et al.* 2008), with 2 673 and 791 species-level names described in each (*Index Fungorum*). Of these, 135 (*Russula*) and 17 (*Xylaria*) have been sequenced (GenBank), i.e. in both cases there are numerous described species that have not been sequenced, plus hundreds of synonyms that may correspond to yet unrecognized species. As a consequence, until all these names have been sorted out in a phylogenetic or taxonomic context (e.g. as synonyms in other genera), establishing new species based on sequence data only should be avoided. In contrast, *Archaeorhizomyces*, *Hawksworthiomyces* and *Lawreymyces* are novel genera based on environmental sequencing or similar approaches and thus had no existing species names available prior to their description (Rosling *et al.* 2011, Menkis *et al.* 2014, De Beer *et al.* 2016, Lücking & Moncada 2017). The same applies to species of *Cyphobasidium* detected by Spribille

[1]*Orphan*: in taxonomy, a species name described in a genus to which it does not belong, and the placement of which has not been re-assessed. Examples include species in genera such as *Agaricus, Lichen, Sphaeria,* and *Verrucaria* not congeneric with the type species of those generic names.

Table 2. Probability of inadvertently describing sequence-based, voucherless taxa that already have names available, depending on predicted global species richness of *Fungi*, based on a random sample of environmental sequences. The calculations assumes 120 000 accepted species, 240 000 total names described (i.e. 120 000 synonyms and orphaned names), 60 000 potentially good species hidden among synonyms and orphaned names, and 35 000 species already sequenced.

	A	B	C	D	B + C
Predicted global species richness	**Proportion of accepted, sequenced species**	**Proportion of accepted, not sequenced species**	**Proportion of species hidden among synonyms**	**Proportion of genuinely novel taxa**	**Estimated error rate**
712 000	4.9 %	11.9 %	8.4 %	74.7 %	20.4 %
1 000 000	3.5 %	8.5 %	6.0 %	82.0 %	14.5 %
1 500 000	2.3 %	5.7 %	4.0 %	88.0 %	9.7 %
2 000 000	1.8 %	4.3 %	3.0 %	91.0 %	7.3 %
2 200 000	1.6 %	3.9 %	2.7 %	91.8 %	6.6 %
3 000 000	1.2 %	2.8 %	2.0 %	94.0 %	4.8 %
3 800 000	0.9 %	2.2 %	1.6 %	95.3 %	3.8 %
5 000 000	0.7 %	1.7 %	1.2 %	96.4 %	2.9 %
10 000 000	0.4 %	0.9 %	0.6 %	98.2 %	1.5 %

et al. (2016), as only two species based on physical type specimens have been described in this genus (Millanes *et al.* 2016). Two complementary or alternative provisions could take care of this concern among the mycological community.

Parallel classification based on different markers

One of the central issues of sequence-based nomenclature is a community-wide agreement which markers to use. Current NGS technologies do not yet allow sequencing different markers from the same template or entire genomes, and maximum read lengths on Illumina MiSeq and HiSeq and Ion-Torrent PGM platforms do not exceed 300–600 bases (100–200 bases on HTS platforms), compared to the phased-out Roche 454 Titanium platform with up to 700 bases or 1500 bases and more on PacBio RS (Loman *et al.* 2012, Luo *et al.* 2012, Quail *et al.* 2012, Yergeau *et al.* 2012, Salipante *et al.* 2014, Goodwin *et al.* 2016). Sequence data corresponding to different markers, or fragments thereof, obtained from the same environmental sample cannot be concatenated to produce multilocus phylogenies since they cannot be traced back to particular individuals.

The ITS has been selected as the universal fungal barcoding marker (Schoch *et al.* 2012), inspite of some shortcomings, such as potential infragenomic variation and lack of resolution in evolving species complexes (*see above*). For instance, ITS data suggest the mushroom *Schizophyllum commune* represents a single species, whereas IGS indicates several, geographically separated lineages (James *et al.* 2001). Intron-rich protein-coding markers such as TEF1 have been shown to be superior to ITS in delimiting species in *Fusarium* (O'Donnell *et al.* 2015). Notably, while arguments against ITS include potential intragenomic variation, TEF1 has been shown to contain paralogs (James *et al.* 2006, Aguileta *et al.* 2008).

Some workers argue not to limit sequence-based nomenclature to a single marker and instead select the best possible marker in each instance (De Beer *et al.* 2016, Hibbett *et al.* 2016). Hawksworth *et al.* (2016) proposed as recommendation 8C.3: "*DNA sequence data used for typification should be drawn from the molecular regions that are appropriate for delimiting species, based on prevailing best practices as determined by the relevant taxonomic communities*." This suggests the ITS barcoding locus as principal marker for the mycological community, but leaves the ultimate choice open to the specialists of a given taxonomic group. One could envision a scenario where ITS would be the default marker and more variable markers would be used in specific lineages. However, this could potentially lead to

Box 5. Introducing voucherless names in a known genus

For a new species based on a sequence type, without a physical voucher specimen, to be validly established in a genus where the type species are generally specimens: (A) All available names in the containing genus should ideally have been linked to a phylogenetically defined and named clade different from that with the new species (to be shown in the phylogenetic analysis accompanying the description), or must have been established as valid species or synonyms in other genera. (B) Names based on sequence types are not given priority over names based on physical types (specimen or culture), unless later epitypified with a matching specimen or culture.

There is also the option of names based on sequence types alone carrying a specific prefix (e.g. "DNA", "SEQ", "ENAS", *Candidatus*) denoting absence of priority and that prefix would be removed upon epitypification by a physical type. This procedure would, however, be confusing to non-systematists familiar with binomial nomenclature, especially ecologists likely to want to to refer to such taxa, and so we do not commend such an approach here.

irreconcilable, parallel classifications if, for instance, one study described new, broadly defined species of *Fusarium* using ITS, whereas another study found more narrowly defined species based on TEF1. In such a case, there would be no way of knowing which of the TEF1-based clades correspond to which of the ITS-based species, although this could be resolved by epitypification. Therefore, unless there is community-wide agreement that in particular taxa, another marker could be consistently used instead of, not in addition to, ITS, an approach using markers of choice is not feasible.

While resolution and accuracy of a barcoding marker is crucial to resolve species, this issue is less important in sequence-based nomenclature of voucherless *Fungi*. First, there are no phenotype characters that could result in conflict with phylogenetically defined species. Second, resolving difficult species complexes is not the objective of this endeavour (see box above). With further advancements of NGS technologies (Koren *et al.* 2013), it might eventually be possible to generate more than one marker or entire genomes from a single template and the limitation to a single marker could be removed.

Even when using ITS as a single marker, the problem of parallel classifications goes further. The approximately 1 billion fungal ITS reads in the SRA have an average length of 353 bases, which mostly corresponds to either the ITS1 or the ITS2 region. As a consequence, reads that correspond only to the ITS1 or ITS2 region cannot be used in parallel to establish species-level clades. Instead, besides using complete ITS sequences from Sanger sequencing and newer NGS technologies, there would have to be an agreement with regard to short reads whether to use either ITS1 or ITS2 only (Bazzicalupo *et al.* 2013). Conceptually, this does not impose a limitation on resolution; ITS1 and ITS2 separately are mostly congruent with full-length ITS data (Blaalid *et al.* 2013), although an eukaryote-wide study suggests that ITS1 is generally superior to ITS2 as barcode marker, particularly in the *Ascomycota* (Wang *et al.* 2015). Again, as outlined above, this issue is not relevant to the purpose of sequence-based nomenclature.

Simultaneous description of new species

In a traditional context, the description of new taxa depends on access to material, including types for comparative studies, and taxonomic expertise. It is therefore uncommon that the same species is described simultaneously with the corresponding authors being unaware of each other's work. In the case of describing new species based on environmental sequence data, such a situation is much more likely because there is universal, unrestricted and simultaneous worldwide access to data including type sequences (whereas a physical type can only be studied at one place at a given time) and the required expertise of phylogenetic analysis including species recognition methods is more widely dispersed and not limited to taxonomic experts of a group. Therefore, there is a greater possibility of different workers simultaneously studying the same data and describing the same taxon under different names. The principle of priority would take care of this as it does for names based on physical types, but it would be unfortunate to unnecessarily duplicate work.

There are several mechanisms that could be introduced to prevent this from happening or reduce the possibility:

- **A network in which ongoing studies are announced and defined.**
- **Immediate release of type sequences of taxa described as new so that similar or identical sequences can be immediately detected.**
- **Free accessibility to registered new taxa in manuscript stage prior to publication.**
- **Peer review by experts that have an overview of the field.**

This would, however, require changes in the procedures in the *Code* (Turland *et al.* 2018: Art. F.5) for the current mandatory system for the registration of names of new taxa in the approved repositories (Fungal Names. Index Fungorum, or MycoBank). It is not recommended that new taxa are registered prior to a paper being accepted for publication (Rec. F.5A.1), as changes are often made during the peer review process and there are many names in the repositories that have never been validly published. Further. names are not released by the repositories until they have been effectively published.

SPECIES DELIMITATION

Environmental sequencing studies yield tens to hundreds of thousands of reads each. With 20,879 experiments (NGS runs) containing 1 222 062 203 fungal ITS reads currently in the SRA (see above), the average number of reads per sequencing run is 58,531. Analysing sequences from the SRA representing a particular clade of interest could potentially retrieve millions of reads. Such huge amounts of data can only be classified by fast methods such as blasting and clustering (Li & Godzik 2006, Schloss *et al.* 2009, Edgar 2010, 2013, Caporaso *et al.* 2010, Huang *et al.* 2010, Huse *et al.* 2010, Kumar *et al.* 2011, Nilsson *et al.* 2011). Unfortunately, clustering is inferior to alignment-based phylogenetic methods, resulting in overestimations of taxonomic diversity (Quince *et al.* 2009, Engelbrektson *et al.* 2010, Kunin *et al.* 2010, Porter & Golding 2011, Powell *et al.* 2011, Unterseher *et al.* 2011, Zhou *et al.* 2011). Estimates of global species richness based on such approaches may lead to exaggerated numbers. For instance, O'Brien *et al.* (2005) estimated the number of fungal species at 5.1 million (Blackwell 2011, Hawksworth 2012), and a recent study by Locey & Lennon (2016) predicted a trillion(!) species on Earth, many of these ecologically cryptic *Fungi* and other microorganisms detected through environmental sequencing.

While the problem of overestimating taxonomic diversity based on clustering is well-documented, clustering continues to be the method of choice for analysing large amounts of NGS data. Clustering works fast and capable of sorting large amounts of data based on pairwise alignment. In pairwise

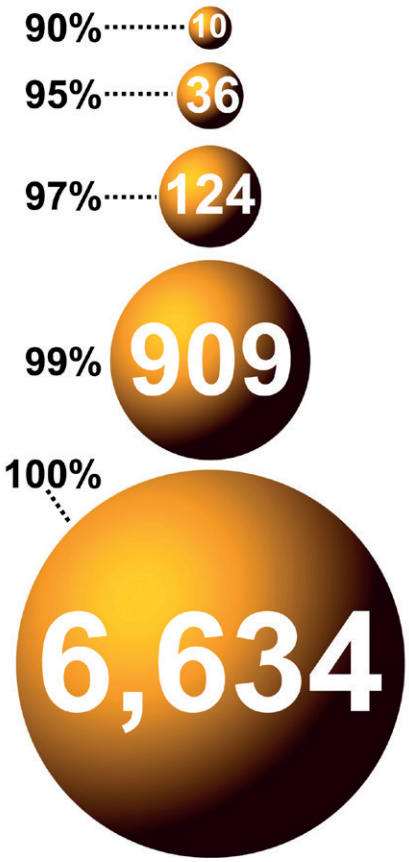

Fig. 3. Number of species-level clusters computed from pyrosequencing ITS reads belonging to a "single" species, the basidiolichen former *Cora inversa*; up to 99 % of the observed variation is due to sequencing errors (after Lücking *et al.* 2014). The same data cluster as a single species-level clade in multiple-alignment-based phylogenetic analysis (see Fig. 4).

alignment, sequencing errors such as CAFIE are interpreted as substitutions (Gazis *et al.* 2011, Lücking *et al.* 2014), unless the gap penalty is substantially lowered which, however, may lead to false interpretation of true substitutions as indels. Therefore, sequences of the same species containing errors are parsed out into different clusters, inflating taxonomic diversity (Fig. 3). This problem does not occur in multiple alignment-based phylogeny, since multiple alignments of closely related sequences place erroneous indels in gapped columns, where they have practically no effect on the resulting topology (Fig. 4, Lücking *et al.* 2014).

Another problem of clustering is the requirement of a fixed threshold value which, depending on the study, is usually set between 95 % and 99 % (O'Brien *et al.* 2005, Smith *et al.* 2007b, Morris *et al.* 2008, Ryberg *et al.* 2008, Walker *et al.* 2008). Such fixed thresholds do not exist in nature (Bruns *et al.* 2007, Nilsson *et al.* 2008, Hughes *et al.* 2009), since intraspecific and interspecific sequence divergence is a function of time, population size and geographic distribution. Fixed thresholds have also taken into consideration potential sequencing errors, at rates between 0.2 % and 1.5 %, in additive fashion, whereas in reality, effects of sequencing errors are augmented by their random positions relative to genuine substitutions and, due to the nature of pairwise alignment in clustering methods, can be multiplicative rather than additive (Lücking *et al.* 2014). Therefore, a fixed threshold cannot prevent sequencing errors to affect the outcome in a clustering approach.

As a consequence, description of new fungal species based on voucherless sequences must be based on approaches that employ rigorous, multiple alignment-based phylogenetic analysis and, in addition, should use quantitative species-delimitation methods such as GMYC or

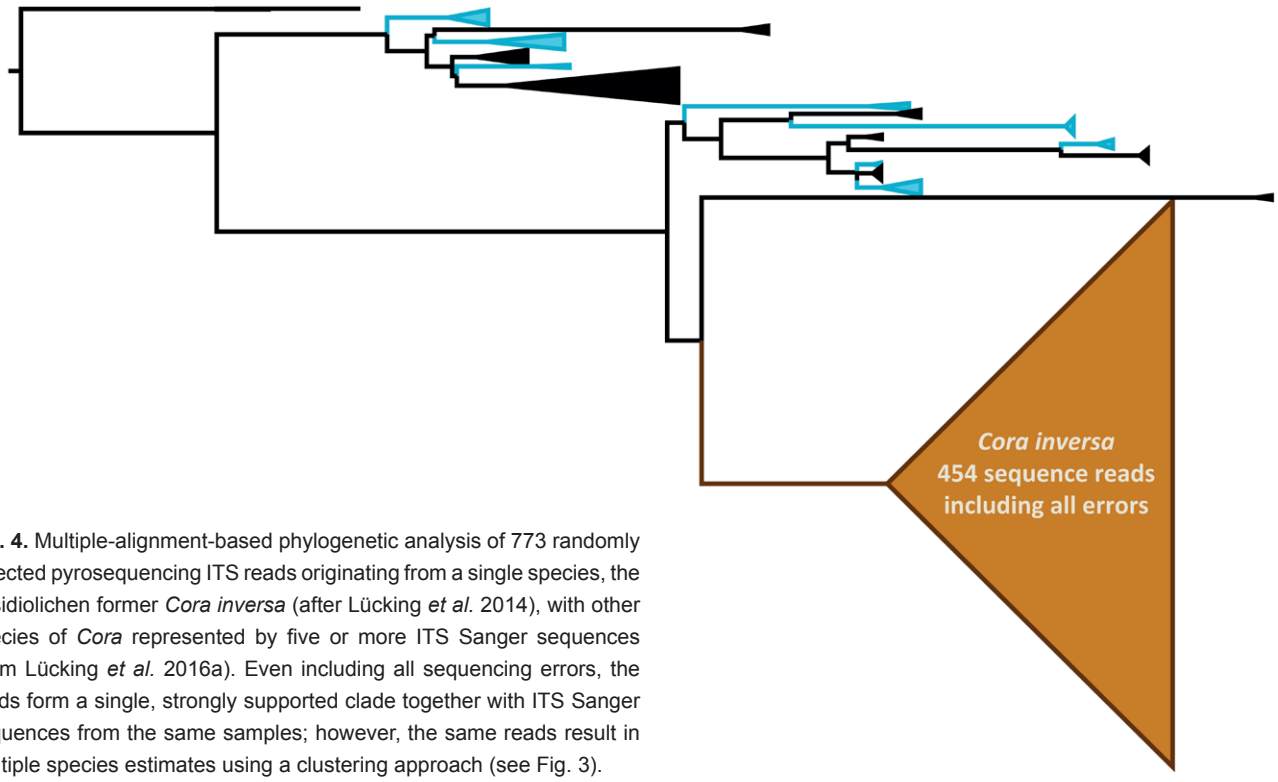

Fig. 4. Multiple-alignment-based phylogenetic analysis of 773 randomly selected pyrosequencing ITS reads originating from a single species, the basidiolichen former *Cora inversa* (after Lücking *et al.* 2014), with other species of *Cora* represented by five or more ITS Sanger sequences (from Lücking *et al.* 2016a). Even including all sequencing errors, the reads form a single, strongly supported clade together with ITS Sanger sequences from the same samples; however, the same reads result in multiple species estimates using a clustering approach (see Fig. 3).

Box 6. A protocol for describing new voucherless species of fungi

1. Assemble an initial data set of sequences of interest, e.g. through mining the SRA using a template sequence and retrieving all similar sequences setting a blast threshold; identify the sequences by study sample and read number.
2. Add related sequences from GenBank.
3. Digest the data set through clustering into smaller portions that can be handled by multiple alignment methods such as MAFFT, which currently allows alignments of up to 30,000 sequences but works best with 10,000 sequences or less.
4. Generate multiple alignments for data subsets using, e.g. Guidance and MAFFT.
5. Run phylogenetic analysis, e.g. maximum likelihood using RAxML.
6. Inspect resulting trees for potential anomalies, e.g. long branches with low support indicating potential chimeras; also, run ITS1 and ITS2 partitions separately and test for conflict.
7. Annotate and remove potentially problematic sequences.
8. Run final phylogenetic analysis, ideally both maximum likelihood and Bayesian.
9. Run various quantitative species delimitation methods, e.g. GMYC and PTP.
10. Identify clades consistently resolved as species that fulfil the requirement of the included sequences resulting from at least *N* independent study samples.
11. Combine all identified species-level clades into a single alignment and rerun (5) and (9).

PTP (Fujisawa & Barraclough 2013, Zhang *et al*. 2013). An idealized protocol is outlined in Box 6.

BACKBONE PHYLOGENY AND HIGHER CLASSIFICATION

ITS is generally not fully alignable across a broader taxon set above species level. Therefore, employing the fungal barcoding marker as principal locus to delimit and formally describe new species of voucherless *Fungi*, without the possibility of using concatenated data sets with more conserved loci, may generate problems when attempting to establish higher-level phylogenies for these new taxa, particularly if they represent novel lineages at the genus, family, order or class level (Hibbett *et al.* 2016, Nilsson *et al.* 2016, Tedersoo *et al.* 2017). In addition, voucherless fungal classification makes it impossible to rank hierarchically structured clades based on phenotype features. However, there are options to deal with these shortcomings. For instance, Wang *et al*. (2011) successfully employed a simultaneous alignment and tree building approach to delimit genera and species in *Geoglossomycetes* based on (largely environmental) ITS data only.

ITS sequence reads can be placed within a broad, multilocus phylogenetic framework generated from known *Fungi* using the evolutionary placement algorithm (EPA) implemented in RAxML (Stamatakis *et al.* 2010, Berger *et al.* 2011, Zhang *et al.* 2013, Stamatakis 2014), an ideal tool for environmental sequencing studies (e.g. Sunagawa *et al.* 2013). While a stand-alone, full alignment of ITS sequences across a broad taxonomic range is challenging, one alternative is adding new ITS reads to a fixed, multi-locus alignment of reference taxa, as implemented in tools such as PPlacer (Matsen *et al*. 2010), ML TreeMap (Stark *et al*. 2010), PaPaRa (Berger & Stamatakis 2011), MAFFT (Katoh & Frith 2012), or T-BAS (Carbone *et al*. 2017). An initial fixed ITS alignment could be elaborated from reference taxa by means of a combined alignment and tree building method, such as BAli-Phy or SATe (Suchard & Redelings 2006, Liu *et al*. 2009, 2012, Wang *et al*. 2011). Alternatively, ambiguously aligned regions can be recoded using PICS-Ord in a *de-novo* alignment, which as been shown to work effectively across broad taxon sets and large alignments of hundreds or thousands of sequences (Lücking *et al.* 2011).

A more reliable approach is *de-novo* alignment of the ITS across reference and query taxa using Guidance HoT scores for alignment confidence, which only retain columns aligned with high confidence (Penn *et al.* 2010a). Arguably, if used across an entire class or phylum, this approach would largely retain the conserved 5.8S region only, which is presumed to not contain sufficient resolution for a backbone phylogeny, but has been shown to work remarkably well in plants and *Fungi* (Hershkovitz & Lewis 1996). We tested this by analysing 210 ITS sequences of the genera *Tremella* (*Tremellales*), *Auricularia* (*Auriculariales*), *Albatrellus*, *Peniophora*, *Russula* (*Russulales*), *Athelia* (*Atheliales*), and *Boletus*, *Coniophora*, and *Suillus* (*Boletales*). The complete alignment for these taxa using MAFFT results in a length of 1354 columns, many of which are ambiguously aligned across the entire set. Running the sequences through the Guidance web server (Penn *et al.* 2010b) returns 407 columns aligned with a confidence of 95 % and higher, of which 174 columns represent a compact block present cross all taxa (Suppl. File S3). Analysing this alignment using RAxML (Stamatakis 2014), the resulting topology (Fig. 5) resolved the underlying

phylogeny remarkably well (except for the position of *Auricularia*), with the two orders *Boletales* and *Russulales* and most genera monophyletic except the collective genus *Athelia* (Rosenthal *et al.* 2017) and the genus *Coniophora* (resolved as paraphyletic grade), with moderate bootstrap support across genera (76 ± 17). Allowing a higher number of alignment columns by reducing the confidence limit to 70 % yields the same topology but strongly increases support across genera (95 ± 5). Thus, a much reduced ITS retaining only columns aligned with good to high confidence is not only capable of reconstructing the backbone phylogeny to a large extent but underlines the usefulness for the application of the EPA, with the added advantage that the entire process can be automated using a Guidance-MAFFT-RAxML pipeline.

There are several objective methods to hierarchically rank ITS backbone phylogenies in a consistent way. One approach is to "hijack" species delimitation methods such as GMYC, haplowebs, and PTP (Fujisawa & Barraclough 2013, Zhang *et al.* 2013, Dellicour & Flot 2015). Once a given set of sequences has been phylogenetically analysed and species-level clades have been identified, one sequence per species is retained. Applying the species delimitation method again will then denote higher level clades. Another approach is to run the Guidance HoT score analysis over a data set. The more closely related the included sequences, the lower the rank they represent as a whole, and the higher the number of columns that can be retained with confidence. Data of known taxa at various hierarchical levels can be used to establish correlations and thresholds. In the above example, aligning across *Agaricomycotina* (subphylum level) resulted in 30 % (407 of 1354) of all alignment columns retained at 95 % confidence, whereas for the genus *Russula* alone, 52 % (420 of 812) of the columns were retained. If these thresholds are consistent across taxa, an ITS data set of unidentified sequences retaining 30 % of columns with high confidence is likely to represent a class or subphylum, wheres 50 % point to a genus. Finally, temporal banding allows the definition of ranks based on divergence times obtained from an ultrametric or molecular clock tree, as recently suggested for *Ascomycota*, *Sordariomycetes*, *Lecanoromycetes*, and *Parmeliaceae* (Divakar *et al.* 2017, Hyde *et al*. 2017, Liu *et al*. 2017).

CONCLUSIONS

Voucherless, sequence-based nomenclature poses numerous challenges, but there appears to be no practicable alternative to formally naming the numerous novel fungal lineages now being detected in environmental sequencing studies. We showed that even if increased by an order of magnitude, specimen- and culture-based inventories will not be capable to formally classify a substantial portion of the predicted unknown fungal diversity within a reasonable time frame. The challenges of sequence-based nomenclature are manageable and there are numerous methods to classify voucherless *Fungi*, using a single marker such as the ITS, both at the species level and at higher taxonomic ranks. There have been arguments that voucherless, sequence-based nomenclature may threaten support to other branches of mycology, such as culture collections and their research, or on the contrary may favour large laboratories in North America or Europe and leave researchers in other countries behind. These arguments have no grounds, on

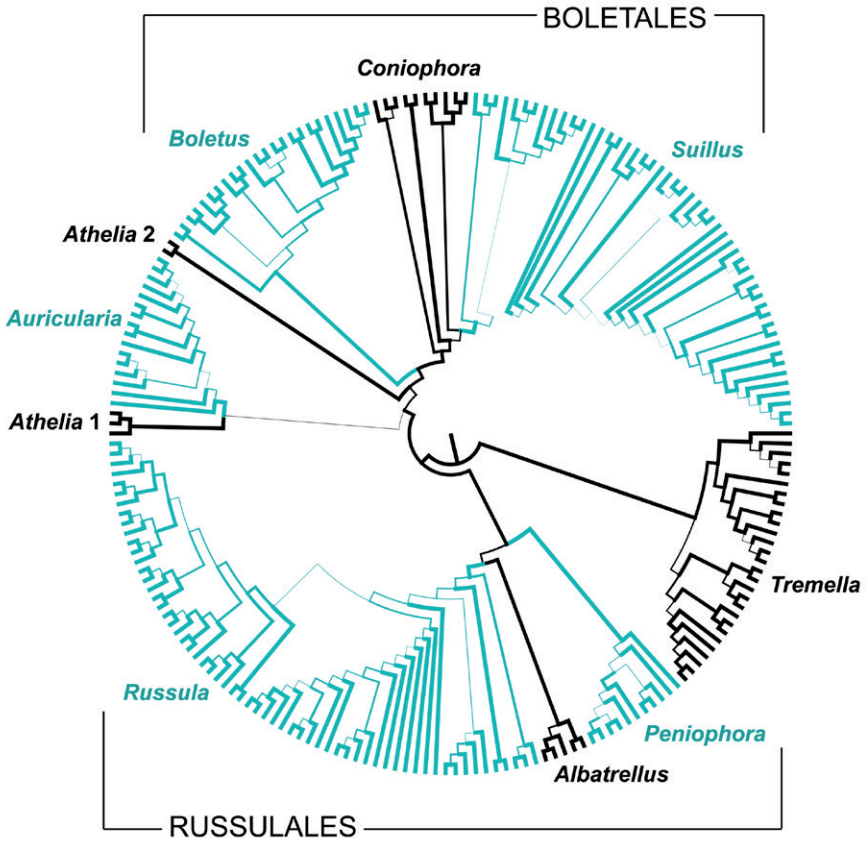

Fig. 5. Exemplar backbone phylogeny for selected genera of *Agaricomycotina* using only columns of the fungal ITS barcoding marker aligned with a Guidance HoT confidence level of 70 % and higher (Penn *et al.* 2010a, b).

the contrary. Funding for fungal research is mostly based on the importance of *Fungi* for ecosystem services and their potential applications. These can only be studied based on specimens and cultures, but not based on voucherless, sequence-based taxa. Therefore, sequence-based nomenclature will not diminish funding to other branches of mycology, but can be expected to generate additional funding in areas of computational biology related to sequence read placement, an area that is already now one of the hot spots of phylogenetic tools. Also, sequence-based nomenclature does not require any laboratory equipment but is entirely computational and hence accessible to virtually anybody, since both data and software are freely available and servers allow access to computational clusters to perform large scale analyses. Therefore, if anything, mycologists in any area of the world have equal access to this approach. As a whole, voucherless, sequence-based nomenclature is not a threat to specimen-based mycology, but rather a complement to substantially speed up cataloguing global fungal diversity in those lineages that are rarely detected using specimen-based methods. If considered desirable, simple and straightforward provisions in the *Code* or a Code of Practice developed by a body such as the International Commission on the Taxonomy of Fungi (ICTF) can help avoid the descry[tion of artifactual taxa or species for which names might already exist. Voucherless, sequence-based fungal taxonomy is universally accessible but is by no means "fast track" mycology, as this approach requires extremely careful work and high skill-levels comparable to those of specimen-based mycologists. However, control mechanisms and effective peer-review by the mycological community are crucial for a successful implementation of this approach, as in all other areas of research.

The time is right for the mycological community as a whole to consider and answer the following questions:

- **Do we recognize the potential of environmental sequences as a substantial source of fungal diversity information that cannot be addressed similarly by other means?**
- **If we recognize that potential, do we want to allow formal nomenclature to be based on types other than those currently allowed by the *Code* (i.e. dried specimens, microscopic preparations, illustrations, metabolically inactive cultures), to capture this diversity?**
- **If we agree to adjust formal nomenclature, what alternative types would be allowable (the underlying environmental sample or 'bag type'; the underlying DNA extract or 'DNA type'; a graphic illustration of the type sequence[2]; or the sequence itself or 'sequence type)?**
- **If we permit alternative types, what if any limitations on the formal establishment of sequence-based taxa do we want to hard-wire into the *Code* and what limitations do we want to trust to peer-review and scientific integrity?**

Most importantly, we should all recognize that established practices need to change to facilitate our science and should not be a hindrance to its progress. Mycologists have an enviable record amongst nomenclaturalists in showing willingness to adopt new ways of working, after due debate. Examples include the acceptability of metabolically inactive, permanently preserved cultures as name-bearing types, adoption of a single starting point date for the naming of fungi, the requirement to register new scientific names for them to be valid, the ability to propose lists of names for protection, and ending the separate naming of morphs of the same species. All these changes followed much debate at mycological meetings and exchanges in the literature, and in the end consensus was achieved and the rules that govern these changed. In some of these cases this process took many decades, and in the interim some authors chose to ignore the rules then in force leading to conflicting treatments. This is already starting to happen in the area of voucherless types, and we feel that the community needs to agree on an acceptable solution as a matter of urgency, as with advancing technology environmental sequencing is now accelerating exponentially.

ACKNOWLEDGEMENTS

We are indepted to Paul M. Kirk and Conrad L. Schoch for providing data from *Index Fungorum* and GenBank, respectively. Travis Adkins kindly supplied fungal culture numbers for the NRRL. Daniel Lindner is thanked for providing raw data from the study on intragenomic ITS variation in *Laetiporus sulphureus*. The members of the International Commission on the Taxonomy of Fungi (ICTF), including Catherine Aime, Takayuki Aoki, Lei Cai, Pedro Crous, Wilhelm DeBeer, David Geiser, Peter Johnston, Tom May, Andrew Miller, Conrad Schoch, Marco Thines, Keith Seifert, Marc Stadler, and Ning Zhang, engaged in fruitful discussions on the subject (referred to in the text or the boxes). Conrad Schoch and two anonymous reviewers provided valuable comments that helped to improve this manuscript.

REFERENCES

Aguileta G, Marthey S, Chiapello H, Lebrun MH, Rodolphe F, Fournier E, Gendrault-Jacquemard A, Giraud T (2008) Assessing the performance of single-copy genes for recovering robust phylogenies. *Systematic Biology* **57**: 613–627.

Amend A, Samson R, Seifert K, Bruns T (2010) Deep sequencing reveals diverse and geographically structured assemblages of *Fungi* in indoor dust. *Proceedings of the National Academy of Sciences, USA* **107**: 13748–13753.

Arnold AE, Henk DA, Eells RL, Lutzoni F, Vilgalys R (2007) Diversity and phylogenetic affinities of foliar fungal endophytes in loblolly pine inferred by culturing and environmental PCR. *Mycologia* **99**: 185–206.

Arnold AE, Mejía LC, Kyllo D, Rojas EI, Maynard Z, Robbins N, Herre EA (2003) Fungal endophytes limit pathogen damage in a

[2] Currently explicitly excluded by the *Code* (see p. 146).

tropical tree. *Proceedings of the National Academy of Sciences* **100**: 15649–15654.

Ashelford KE, Chuzhanova NA, Fry JC, Jones AJ, Weightman AJ (2005) At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. *Applied and Environmental Microbiology* **71**: 7724–7736.

Baker RH, Yu X, DeSalle R (1998) Assessing the relative contribution of molecular and morphological characters in simultaneous analysis trees. *Molecular Phylogenetics and Evolution* **9**: 427–436.

Balzer S, Malde K, Jonassen I (2011) Systematic exploration of error sources in pyrosequencing flowgram data. *Bioinformatics* **27**: 304–309.

Barnosky AD, Matzke N, Tomiya S, Wogan GO, Swartz B, Quental TB, Marshall C, McGuire JL, Lindsey EL, Maguire KC, Mersey B (2011) Has the Earth's sixth mass extinction already arrived? *Nature* **471**: 51–57.

Bass D, Richards TA (2011) Three reasons to re-evaluate fungal diversity 'on Earth and in the ocean'. *Fungal Biology Reviews* **25**: 159–164.

Bazzicalupo AL, Bálint M, Schmitt I (2013) Comparison of ITS1 and ITS2 rDNA in 454 sequencing of hyperdiverse fungal communities. *Fungal Ecology* **6**: 102–109.

Begerow D, Nilsson RH, Unterseher M, Maier W (2010) Current state and perspectives of fungal DNA barcoding and rapid identification procedures. *Applied Microbiology and Biotechnology* **87**: 99–108.

Bensch K, Groenewald JZ, Dijksterhuis J, Starink-Willemse M, Andersen B, *et al.* (2010) Species and ecological diversity within the *Cladosporium cladosporioides* complex (*Davidiellaceae*, *Capnodiales*). *Studies in Mycology* **67**: 1–94.

Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, *et al.* (2013) GenBank. *Nucleic Acids Research* **41** (D1): D36–D42.

Berger SA, Stamatakis A (2011) Aligning short reads to reference alignments and trees. *Bioinformatics* **27**: 2068–2075.

Berger SA, Stamatakis A (2012) *PaPaRa 2.0: a vectorized algorithm for probabilistic phylogeny-aware alignment extension*. Heidelberg: Heidelberg Institute for Theoretical Studies, http://sco. h-its. org/exelixis/publications. html. Exelixis-RRDR-2012-2015.

Berger SA, Krompass D, Stamatakis A (2011) Performance, accuracy, and web server for evolutionary placement of short sequence reads under maximum likelihood. *Systematic Biology* **60**: 291–302.

Bergsten J (2005) A review of long-branch attraction. *Cladistics* **21**: 163–193.

Bidartondo M, Bruns TD, Blackwell M, Edwards I, Taylor AFS, *et al.* (2008) Preserving accuracy in GenBank. *Science* **319**: 1616.

Blaalid R, Kumar S, Nilsson RH, Abarenkov K, Kirk PM, Kauserud H (2013) ITS1 versus ITS2 as DNA metabarcodes for *Fungi*. *Molecular Ecology Resources* **13**: 218–224.

Blackwell M (2011) The *Fungi*: 1, 2, 3… 5.1 million species? *American Journal of Botany* **98**: 426–438.

Blackwell M, Hibbett DS, Taylor JW, Spatafora JW (2006) Research coordination networks: a phylogeny for kingdom *Fungi* (Deep Hypha). *Mycologia* **98**: 829–837.

Bowman BH, Taylor JW, Brownlee AG, Lee J, Lu SD, White TJ (1992) Molecular evolution of the *Fungi*: relationship of the basidiomycetes, ascomycetes, and chytridiomycetes. *Molecular Biology and Evolution* **9**: 285–296.

Bruns TD, White TJ, Taylor JW (1991) Fungal molecular systematics. *Annual Review of Ecology and Systematics* **22**: 525–564.

Bruns TD, Arnold AE, Hughes KW (2007) Fungal networks made of humans: UNITE, FESIN and frontiers in fungal ecology. *New Phytologist* **177**: 586–588.

Buée M, Reich M, Murat C, Morin E, Nilsson RH, *et al.* (2009) 454 pyrosequencing analyses of forest soils reveal an unexpectedly high fungal diversity. *New Phytologist* **184**: 449–456.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, *et al.* (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7**: 335–336.

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, *et al.* (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME Journal* **6**: 1621–1624.

Carbone I, White JB, Miadlikowska J, Arnold AE, Miller MA *et al.* (2017) T-BAS: Tree-Based Alignment Selector toolkit for phylogenetic-based placement, alignment downloads and metadata visualization: an example with the Pezizomycotina tree of life. *Bioinformatics* **33**: 1160–1168.

Carlsen T, Aas AB, Lindner D, Vrålstad T, Schumacher T, Kauserud H (2012) Don't make a mista (g) ke: is tag switching an overlooked source of error in amplicon pyrosequencing studies? *Fungal Ecology* **5**: 747–749.

Chen J, Sahota A, Stambrook PJ, Tischfield JA (1991) Polymerase chain reaction amplification and sequence analysis of human mutant adenine phosphoribosyltransferase genes: the nature and frequency of errors caused by Taq DNA polymerase. *Mutation Research* **31**: 169–176.

Coissac E, Riaz T, Puillandre N (2012) Bioinformatic challenges for DNA metabarcoding of plants and animals. *Molecular Ecology* **21**: 1834–1847.

Coleman AW (2015) Nuclear rRNA transcript processing versus internal transcribed spacer secondary structure. *Trends in Genetics* **31**: 157–163.

Davis EC, Franklin JB, Shaw AJ, Vilgalys R (2003) Endophytic *Xylaria* (*Xylariaceae*) among liverworts and angiosperms: phylogenetics, distribution, and symbiosis. *American Journal of Botany* **90**: 1661–1667.

De Beer ZW, Marincowitz S, Duong TA, Kim JJ, Rodrigues A, Wingfield MJ (2016) *Hawksworthiomyces* gen. nov. (*Ophiostomatales*), illustrates the urgency for a decision on how to name novel taxa known only from environmental nucleic acid sequences (ENAS). *Fungal Biology* **120**: 1323–1340.

Dellicour S, Flot JF (2015) Delimiting species-poor datasets using single molecular markers: a study of barcode gaps, haplowebs and GMYC. *Systematic Biology* **19**: syu130.

Del-Prado R, Divakar PK, Lumbsch HT, Crespo AM (2016) Hidden genetic diversity in an asexually reproducing lichen forming fungal group. *PloS One* **11** (8): e0161031.

Divakar PK, Crespo A, Kraichak E, Leavitt SD, Singh G, *et al.* (2017) Using a temporal phylogenetic method to harmonize family-and genus-level classification in the largest clade of lichen-forming *Fungi*. *Fungal Diversity* **84**: 101–117.

Druzhinina IS, Kubicek CP, Komoń-Zelazowska M, Mulaw TB, Bissett J (2010) The *Trichoderma harzianum* demon: complex speciation history resulting in coexistence of hypothetical biological species, recent agamospecies and numerous relict lineages. *BMC Evolutionary Biology* **10**: 94.

Dupuis JR, Roe AD, Sperling FA (2012) Multi-locus species

delimitation in closely related animals and *Fungi*: one marker is not enough. *Molecular Ecology* **21**: 4422–4436.

Eberhardt U (2010) A constructive step towards selecting a DNA barcode for *Fungi*. *New Phytologist* **187**: 265–268.

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460–2461.

Edgar RC (2013) UPARSE: Highly accurate OTU sequences from microbial amplicon reads. *Nature Methods* **10**: 996–998.

Edgar, RC (2016) UCHIME2: improved chimera prediction for amplicon sequencing. bioRxiv: 074252.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**: 2194–2200.

Engelbrektson A, Kunin V, Wrighton KC, Zvenigorodsky N, Chen F, *et al.* (2010) Experimental factors affecting PCR-based estimates of microbial species richness and evenness. *ISME Journal* **4**: 642–647.

Kirk PM (2012) Nomenclatural novelties. *Index of Fungi* **1**: 1.

Fujisawa T, Barraclough TG (2013) Delimiting species using single-locus data and the Generalized Mixed Yule Coalescent approach: a revised method and evaluation on simulated data sets. *Systematic Biology* **16**: syt033.

Gams W (1997) Fungal taxonomy in crisis. In: *Diagnosis and Identification of Plant Pathogens* (Dehne HW, Adam G, Diekmann M, Frahm J, Mauler-Machnik A, van Halteren P (eds): 17–21. Dordrecht: Springer.

Ganley ARD, Kobayashi T (2007) Highly efficient concerted evolution in the ribosomal DNA repeats: Total rDNA repeat variation revealed by whole-genome shotgun sequence data. *Genome Research* **17**: 184–191.

Gazis R, Rehner S, Chaverri P (2011) Species delimitation in fungal endophyte diversity studies and its implications in ecological and biogeographic inferences. *Molecular Ecology* **20**: 3001–3013.

Geiser DM, Jimenez-Gasco MD, Kang SC, Makalowska I, Veeraraghavan N, *et al.* (2004) FUSARIUM-ID v. 1.0: a DNA sequence database for identifying *Fusarium*. *European Journal of Plant Pathology* **110**: 473–479.

Geml J, Laursen GA, Taylor DL (2008) Molecular diversity assessment of arctic and boreal *Agaricus* taxa. *Mycologia* **100**: 577–589.

Geml J, Laursen GA, O'Neill K, Nusbaum HC, Taylor DL (2006) Beringian origins and cryptic speciation events in the fly agaric (*Amanita muscaria*). *Molecular Ecology* **15**: 225–239.

Geml J, Laursen GA, Herriott IC, McFarland JM, Booth MG, *et al.* (2010) Phylogenetic and ecological analyses of soil and sporocarp DNA sequences reveal high diversity and strong habitat partitioning in the boreal ectomycorrhizal genus *Russula* (*Russulales*; *Basidiomycota*). *New Phytologist* **187**: 494–507.

Giudicelli GC, Mäder G, Silva-Arias GA, Zamberlan PM, Bonatto SL, Freitas LB (2017) Secondary structure of nrDNA Internal Transcribed Spacers as a useful tool to align highly divergent species in phylogenetic studies. *Genetics and Molecular Biology* **40**: 191–199.

Glass DJ, Takebayashi N, Olson LE, Taylor DL (2013) Evaluation of the authenticity of a highly novel environmental sequence from boreal forest soil using ribosomal RNA secondary structure modeling. *Molecular Phylogenetics and Evolution* **67**: 234–245.

Glass, DJ, Taylor AD, Herriott IC, Ruess RW, Taylor DL (2014) Habitat preferences, distribution, and temporal persistence of a novel fungal taxon in Alaskan boreal forest soils. *Fungal Ecology* **12**: 70–77.

Goertzen LR, Cannone JJ, Gutell RR, Jansen RK (2003) ITS secondary structure derived from comparative analysis: implications for sequence alignment and phylogeny of the *Asteraceae*. *Molecular Phylogenetics and Evolution* **29**: 216–234.

Gomes EA, Kasuya MCM, Barros EGD, Borges AC, Araújo EF (2002) Polymorphism in the internal transcribed spacer (ITS) of the ribosomal DNA of 26 isolates of ectomycorrhizal *Fungi*. *Genetics and Molecular Biology* **25**: 477–483.

Gomez-Alvarez V, Teal TK, Schmidt TM (2009) Systematic artifacts in metagenomes from complex microbial communities. *ISME Journal* **3**: 1314–1317.

Goodwin S, McPherson JD, McCombie WR (2016) Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics* **17**: 333–351.

Gouy M, Li WH (1989) Molecular phylogeny of the kingdoms Animalia, Plantae, and *Fungi*. *Molecular Biology and Evolution* **6**: 109–122.

Grantham NS, Reich BJ, Pacifici K, Laber EB, Menninger HL, *et al.* (2015). *Fungi* identify the geographic origin of dust samples. *PloS One* **10** (4): e0122605.

Gryzenhout M, Jefwa JM, Yorou NS (2012) The status of mycology in Africa: A document to promote awareness. *IMA Fungus* **3**: 99–102.

Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward DV, *et al.* (2011) Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Research* **21**: 494–504.

Hall BK (2003) Descent with modification: the unity underlying homology and homoplasy as seen through an analysis of development and evolution. *Biological Reviews* **78**: 409–433.

Harris JD (2003) Can you bank on GenBank? *Trends in Ecology and Evolution* **18**: 317–319.

Hassanin A, Lecointre G, Tillier S (1998) The 'evolutionary signal' of homoplasy in protein coding gene sequences and its consequences for a priori weighting in phylogeny. *Comptes Rendus de l'Académie des Sciences, series III,Sciences de la Vie* **321**: 611–620.

Hawksworth DL (1991) The fungal dimension of biodiversity: magnitude, significance, and conservation. *Mycological Research* **95**: 641–655.

Hawksworth DL (2001) The magnitude of fungal diversity: the 1.5 million species estimate revisited. *Mycological Research* **105**: 1422–1432.

Hawksworth DL (2009) Mycology: a neglected megascience. In: *Applied Mycology* (Rai M, Pridge PD, eds): 1–16. Wallingford: CAB International.

Hawksworth DL (2012) Global species numbers of *Fungi*: are tropical studies and molecular approaches contributing to a more robust estimate? *Biodiversity and Conservation* **21**: 2425–2433.

Hawksworth DL (2017) DNA sequences as types: a potential loophole in the rules discovered. *IMA Fungus* **8**: (4).

Hawksworth DL, Lücking R (2017) Fungal diversity revisited: 2.2 to 3.8 million species. *Microbiology Spectrum* **5**: FUNK-0052-2016.

Hawksworth DL, Crous PW, Redhead SA, Reynolds DR, Samson RA, *et al.* (2011) The Amsterdam declaration on fungal nomenclature. *IMA Fungus* **2**: 105–112.

Hawksworth DL, Hibbett DS, Kirk PM, Lücking R (2016) (308–310) Proposals to permit DNA sequence data to serve as types of names of *Fungi*. *Taxon* **65**: 899–900.

Hawksworth DL, May TW, Redhead SA (2017) Fungal nomenclature evolving: changes adopted by the 19th International Botanical Congress in Shenzhen 2017, and procedures for the Fungal Nomenclature Session at the 11th International Mycological Congress in Puerto Rico 2018. *IMA Fungus* **8**: 211–218.

He X, Sun Y, Zhu RL (2013) The oil bodies of liverworts: unique and important organelles in land plants. *Critical Reviews in Plant Sciences* **32**: 293–302.

Hawksworth DL, Hibbett DS, Kirk PM, Lücking R. (2018) (F-005-006) Proposals to permit DNA sequence data to be used as types of names of fungi. *IMA Fungus, Myconames*: v–vi.

Hendriks L, Goris A, Neefs JM, Van De Peer Y, Hennebert G, De Wachter R (1989) The nucleotide sequence of the small ribosomal subunit RNA of the yeast *Candida albicans* and the evolutionary position of the *Fungi* among the eukaryotes. *Systematic and Applied Microbiology* **12**: 223–229.

Hershkovitz MA, Lewis LA (1996) Deep-level diagnostic value of the rDNA-ITS region. *Molecular Biology and Evolution* **13**: 1276–1295.

Hibbett DS (2016) The invisible dimension of fungal diversity. *Science* **351**: 1150–1151.

Hibbett DS, Taylor JW (2013) Fungal systematics: is a new age of enlightenment at hand? *Nature Reviews Microbiology* **11**: 129–133.

Hibbett DS, Bindera J, Bischoffb JF, Blackwellc M, Cannon PF, *et al.* (2007) A higher-level phylogenetic classification of the *Fungi*. *Mycological Research* **111**: 509–547.

Hibbett DS, Ohman A, Glotzer D, Nuhn M, Kirk P, Nilsson RH (2011) Progress in molecular and morphological taxon discovery in *Fungi* and options for formal classification of environmental sequences. *Fungal Biology Reviews* **25**: 38–47.

Hibbett D, Abarenkov K, Koljalg U, Öpik M, Chai B, *et al.* (2016) Sequence-based classification and identification of *Fungi*. *Mycologia* **108**: 1049–1068.

Hiepko P (1987) The collections of the Botanical Museum Berlin-Dahlem (B) and their history. *Englera* **7**: 219–252.

Huang Y, Niu B, Gao Y, Fu L, Li W (2010) CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* **26**: 680–682.

Huber T, Faulkner G, Hugenholtz P (2004) Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. *Bioinformatics* **20**: 2317–2319.

Hubka V, Kolarik M (2012) β-tubulin paralogue tubC is frequently misidentified as the benA gene in *Aspergillus* section *Nigri* taxonomy: primer specificity testing and taxonomic consequences. *Persoonia* **29**: 1–10.

Hughes KW, Petersen RH (2001) Apparent recombination or gene conversion in the ribosomal ITS region of a *Flammulina* (*Fungi*, *Agaricales*) hybrid. *Molecular Biology and Evolution* **18**: 94–96

Hughes KW, Petersen RH, Lickey EB (2009) Using heterozygosity to estimate a percentage DNA sequence similarity for environmental species' delimitation across basidiomycete *Fungi*. *New Phytologist* **182**: 795–798.

Hurst LD, Smith NGC (1998) The evolution of concerted evolution. *Proceedings of the Royal Society of London, B, Biological Sciences* **265**: 121–127

Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM (2007) Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biology* **8**: R143.

Huse SM, Welch DM, Morrison HG, Sogin ML (2010) Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environmental Microbiology* **12**: 1889–1898.

Hyde KD, Maharachchikumbura SS, Hongsanan S, Samarakoon MC, Lücking R, *et al.* (2017) The ranking of *Fungi*: a tribute to David L. Hawksworth on his 70th birthday. *Fungal Diversity* **84**: 1–23.

Inderbitzin P, Mehta YR, Berbee ML (2009) *Pleospora* species with *Stemphylium* anamorphs: a four locus phylogeny resolves new lineages yet does not distinguish among species in the *Pleospora herbarum* clade. *Mycologia* **101**: 329–339.

James SA, O'Kelly MJ, Carter DM, Davey RP, van Oudenaarden A, Roberts IN, *et al.* (2009) Repetitive sequence variation and dynamics in the ribosomal DNA array of *Saccharomyces cerevisiae* as revealed by whole-genome resequencing. *Genome Research* **19**: 626–635.

James TY, Moncalvo JL, Li S, Vilgalys R 2001. Polymorphism at the ribosomal DNA spacers and its relation to breeding structure of the widespread mushroom *Schizophyllum commune*. *Genetics* **157**: 149–161.

James TY, Kauff F, Schoch CL, Matheny PB, Hofstetter V, *et al.* (2006) Reconstructing the early evolution of *Fungi* using a six-gene phylogeny. *Nature* **443**: 818–822.

Jones MD, Forn I, Gadelha C, Egan MJ, Bass D, *et al.* (2011a) Discovery of novel intermediate forms redefines the fungal tree of life. *Nature* **474**: 200–203.

Jones MD, Richards TA, Hawksworth DL, Bass D (2011b) Validation and justification of the phylum name Cryptomycota phyl. nov. *IMA Fungus* **2:** 173–175.

Kälersjö M, Albert VA, Farris JS (1999) Homoplasy increases phylogenetic structure. *Cladistics* **15**: 91–93.

Katoh K, Frith MC (2012) Adding unaligned sequences into an existing alignment using MAFFT and LAST. *Bioinformatics* **28**: 3144–3146.

Katoh K, Rozewicki R, Yamada KD (2017) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*: bbx108.

Keirle MR, Avis PG, Hemmes DE, Mueller GM (2011) Variability in the IGS1 region of *Rhodocollybia laulaha*: is it allelic, genomic or artificial? *Fungal Biology* **115**: 310–316.

Kim M, Lee KH, Yoon SW, Kim BS, Chun J, Yi H (2013) Analytical tools and databases for metagenomics in the next-generation sequencing era. *Genomics & Informatics* **11**: 102–113.

Kirk PM (2012) Nomenclatural novelties. *Index of Fungi* **1**: 1.

Kirk PM, Cannon PF, Minter DW, Stalpers JA (2008) *Ainsworth & Bisby's Dictionary of the Fungi*. 10th edn. Wallingford: CAB International.

Kiss L (2012) Limits of nuclear ribosomal DNA internal transcribed spacer (ITS) sequences as species barcodes for *Fungi*. *Proceedings of the National Academy of Sciences, USA* **109**: E1811–E1811.

Ko KS, Jung HS (2002) Three nonorthologous ITS1 types are present in a polypore fungus Trichaptum abietinum. *Molecular Phylogenetics and Evolution* **23**: 112–122

Kodama Y, Shumway M, Leinonen R (2012) The Sequence Read Archive: explosive growth of sequencing data. *Nucleic Acids Research* **40** (D1): D54–D56.

Koetschan C, Kittelmann S, Lu J, Al-Halbouni D, Jarvis GN, *et al.* (2014) Internal transcribed spacer 1 secondary structure analysis reveals a common core throughout the anaerobic *Fungi*

(*Neocallimastigomycota*). *PloS One* **9** (3): e91928.

Kõljalg U, Nilsson RH, Abarenkov K, Tedersoo L, Taylor AF, *et al.* (2013) Towards a unified paradigm for sequence-based identification of *Fungi*. *Molecular Ecology* **22**: 5271–5277.

Koren S, Harhay GP, Smith TP, Bono JL, Harhay DM, *et al.* (2013) Reducing assembly complexity of microbial genomes with single-molecule sequencing. *Genome Biology* **14**: R101.

Korf RP (2005) Reinventing taxonomy: a curmudgeon's view of 250 years of fungal taxonomy, the crisis in biodiversity, and the pitfalls of the phylogenetic age. *Mycotaxon* **93**: 407–416.

Kovács GM, Balázs TK, Calonge FD, Martín MP (2011) The diversity of *Terfezia* desert truffles: new species and a highly variable species complex with intrasporocarpic nrDNA ITS heterogeneity. *Mycologia* **103**: 841–853.

Kück P, Mayer C, Wägele JW, Misof B (2012) Long branch effects distort maximum likelihood phylogenies in simulations despite selection of the correct model. *PLoS One* **7(5)**: e36593.

Kumar S, Carlsen T, Mevik B, Enger P, Blaalid R, *et al.* (2011) CLOTU: an online pipeline for processing and clustering of 454 amplicon reads into OTUs followed by taxonomic annotation. *BMC Bioinformatics* **12**: 182.

Kunin V, Engelbrektson A, Ochman H, Hugenholtz P (2010) Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environmental Microbiology* **12**: 118–123.

Kurtzman CP (1985) Molecular taxonomy of the *Fungi*. In *Gene Manipulation in Fungi* (Bennett IW, Lasuse LL, eds): 35–63. Orlando: Academic Press.

Lazarus KL, James TY (2015) Surveying the biodiversity of the *Cryptomycota* using a targeted PCR approach. *Fungal Ecology* **14**: 62–70.

Leakey R, Lewin R (1995) *The Sixth Extinction: Patterns of Life and the Future of Humankind*. New York: Doubleday.

Leavitt SD, Johnson LA, Goward T, Clair LLS (2011) Species delimitation in taxonomically difficult lichen-forming *Fungi*: an example from morphologically and chemically diverse *Xanthoparmelia* (*Parmeliaceae*) in North America. *Molecular Phylogenetics and Evolution* **60**: 317–332.

Leinonen R, Sugawara H, Shumway M (2011) The sequence read archive. *Nucleic Acids Research* **39** (Database issue): D19–D21.

Letcher PM, Lopez S, Schmieder R, Lee PA, Behnke C, Powell MJ, McBride RC (2013) Characterization of *Amoeboaphelidium protococcarum*, an algal parasite new to the cryptomycota isolated from an outdoor algal pond used for the production of biofuel. *PLoS One* **8** (2): e56232.

Li W, Godzik A (2006) Cd-Hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**: 1658–1659.

Li Y, Jiao L, Yao YJ (2013) Non-concerted ITS evolution in *Fungi*, as revealed from the important medicinal fungus *Ophiocordyceps sinensis*. *Molecular Phylogenetics and Evolution* **68**: 373–379.

Liao D (1999) Concerted evolution: molecular mechanism and biological implications. *American Journal of Human Genetics* **64**: 24–30.

Liao D (2008) *Concerted Evolution*. Gainsville FL: Diaquing Lao.

Lim GS, Balke M, Meier R (2012) Determining species boundaries in a world full of rarity: singletons, species delimitation methods. *Systematic Biology* **61**: 165–169.

Lindner DL, Carlsen T, Nilsson HR, Davey M, Schumacher T, Kauserud H (2013) Employing 454 amplicon pyrosequencing to reveal intragenomic divergence in the internal transcribed spacer rDNA region in *Fungi*. *Ecology and Evolution* **3**: 1751–1764.

Lindner DL, Banik MT (2011) Intragenomic variation in the ITS rDNA region obscures phylogenetic relationships and inflates estimates of operational taxonomic units in the genus *Laetiporus*. *Mycologia* **103**: 731–740.

Liu JK, Hyde KD, Jeewon R, Phillips AJ, Maharachchikumbura SS, *et al.* (2017) Ranking higher taxa using divergence times: a case study in *Dothideomycetes*. *Fungal Diversity* **84**: 75–99.

Liu K, Raghavan S, Nelesen S, Linder CR, Warnow T (2009) Rapid and accurate large-scale coestimation of sequence alignments and phylogenetic trees. *Science* **19**: 1561–1564.

Liu K, Warnow TJ, Holder MT, Nelesen SM, Yu J, *et al.* (2012) SATe-II: very fast and accurate simultaneous estimation of multiple sequence alignments and phylogenetic trees. *Systematic Biology* **61**: 90–106.

Livermore JA, Mattes TE (2013) Phylogenetic detection of novel *Cryptomycota* in an Iowa (United States) aquifer and from previously collected marine and freshwater targeted high-throughput sequencing sets. *Environmental Microbiology* **15**: 2333–2341.

Locey KJ, Lennon JT (2016) Scaling laws predict global microbial diversity. *Proceedings of the National Academy of Sciences, USA* **113**: 5970–5975.

Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, *et al.* (2012) Performance comparison of benchtop high-throughput sequencing platforms. *Nature Biotechnology* **30**: 434–439.

Lombard L, Crous PW, Wingfield BD, Wingfield MJ (2010) Multigene phylogeny and mating tests reveal three cryptic species related to *Calonectria pauciramosa*. *Studies in Mycology* **66**: 15–30.

Lücking R, Moncada B (2017) Dismantling *Marchandiomphalina* into *Agonimia* (*Verrucariaceae*) and *Lawreymyces* gen. nov. (*Corticiaceae*): setting a precedent to the formal recognition of thousands of voucherless *Fungi* based on type sequences. *Fungal Diversity* **84**: 119–138.

Lücking R, Hodkinson BP, Stamatakis A, Cartwright RA (2011) PICS-Ord: unlimited coding of ambiguous regions by pairwise identity and cost scores ordination. *BMC Bioinformatics* **12** (1): 10.

Lücking R, Kalb K, Essene A (2012) The power of ITS: using megaphylogenies of barcoding genes to reveal inconsistencies in taxonomic identifications of genbank submissions. In: *7th IAL Symposium "Lichens: From Genome to Ecosystems in a Changing World", January 2012, Bangkok (Thailand), Book of Abstracts*: 3B-1-O2.

Lücking R (2014) A phylogenetic classification system for unvouchered environmental fungal sequences of unknown taxonomic affiliation. In: *10th International Mycological Congress, Bangkok, Thailand. IMC10 Book of Abstracts: O 8.6.1*: Abstract ID ABS0123; http://www.fabinet.up.ac.za/newsitem/112-IMC10 eBook of Abstracts.pdf.

Lücking R, Lawrey JD, Gillevet PM, Sikaroodi M, Dal Forno M, Berger SA (2014) Multiple ITS haplotypes in the genome of the lichenized basidiomycete *Cora inversa* (*Hygrophoraceae*): fact or artifact? *Journal of Molecular Evolution* **78:** 148–162.

Lücking R, Dal Forno M, Moncada B, Coca LF, Vargas-Mendoza LY, *et al.* (2016a) Turbo-taxonomy to assemble a megadiverse lichen genus: seventy new species of *Cora* (*Basidiomycota*: *Agaricales*: *Hygrophoraceae*), honouring David Leslie Hawksworth's seventieth birthday. *Fungal Diversity*: doi:10.1007/s13225-016-0374-9.

Lücking R, Nelsen MP, Aptroot A, Klee RB, Bawingan PA, *et al.* (2016b) A phylogenetic framework for reassessing generic concepts and species delimitation in the lichenized family Trypetheliaceae (Ascomycota: Dothideomycetes). *The Lichenologist* **48**: 739–762

Lumbsch HT, Leavitt SD (2011) Goodbye morphology? A paradigm shift in the delimitation of species in lichenized *Fungi*. *Fungal Diversity* **50**: 59.

Lumini E, Orgiazzi A, Borriello R, Bonfante P, Bianciotto V (2010) Disclosing arbuscular mycorrhizal fungal biodiversity in soil through a land-use gradient using a pyrosequencing approach. *Environmental Microbiology* **12**: 2165–2179.

Luo C, Tsementzi D, Kyrpides N, Read T, Konstantinidis KT (2012) Direct comparisons of Illumina *vs.* Roche 454 sequencing technologies on the same microbial community DNA sample. *PLoS One* **7**: e30087.

Lutzoni F, Kauff F, Cox CJ, McLaughlin D, Celio G, *et al.* (2004) Assembling the fungal tree of life: progress, classification, and evolution of subcellular traits. *American Journal of Botany* **91**: 1446–1480.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, *et al.* (2005) Genome sequencing inmicrofabricated high-density picolitre reactors. *Nature* **437**: 376–380.

Mark K, Cornejo C, Keller C, Flück D, Scheidegger C (2016) Barcoding lichen-forming *Fungi* using 454 pyrosequencing is challenged by artifactual and biological sequence variation. *Genome* **59**: 685–704.

Matsen FA, Kodner RB, Armbrust EV (2010) PPlacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics* **11**: 538.

McGuire KL, Fierer N, Bateman C, Treseder KK, Turner BL (2012) Fungal community composition in neotropical rain forests: the influence of tree diversity and precipitation. *Microbial Ecology* **63**: 804–812.

Meier R (2008) DNA sequences in taxonomy. In Wheeler QD (Ed.), *The New Taxonomy*, pp. 95–127. CRC Press, Boca Raton, FL.

Menkis A, Urbina H, James TY, Rosling A (2014) *Archaeorhizomyces borealis* sp. nov. and a sequence-based classification of related soil fungal species. *Fungal Biology* **118**: 943–955.

Metsger DA (1999) *Managing the Modern Herbarium: an inter-disciplinary approach*. Toronto: Royal Ontario Museum, Centre for Biodiversity and Conservation Biology, Ontario, Canada.

Miądlikowska J, Kauff F, Högnabba, Oliver JC, Molnár K, *et al.* (2014) A multigene phylogenetic synthesis for the class *Lecanoromycetes* (*Ascomycota*): 1307 *Fungi* representing 1139 infrageneric taxa, 317 genera and 66 families. *Molecular Phylogenetics and Evolution* **79**: 132–168.

Michelmore RW, Hulbert SH (1987) Molecular markers for genetic analysis of phytopathogenic *Fungi*. *Annual Review of Phytopathology* **25**: 383–404.

Millanes AM, Diederich P, Wedin M (2016) *Cyphobasidium* gen. nov., a new lichen-inhabiting lineage in the *Cystobasidiomycetes* (*Pucciniomycotina*, *Basidiomycota*, *Fungi*). *Fungal Biology* **120**: 1468–1477.

Minnis AM (2015) The shifting sands of fungal naming under the ICN and the One Name Era for *Fungi*. In: *The Mycota.* Vol VIIB. *Systematics and Evolution* (McLaughlin DJ, Spatafora JW, eds): 179–203. Berlin: Springer.

Minoche AE, Dohm JC, Himmelbauer H (2011) Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. *Genome Biology* **12** (11): R112.

Moncada B, Lücking R, Suárez A (2014) Molecular phylogeny of the genus *Sticta* (lichenized *Ascomycota*: *Lobariaceae*) in Colombia. *Fungal Diversity* **64**: 205–231.

Moncalvo JM, Vilgalys R, Redhead SA, Johnson JE, James TY, *et al.* (2002) One hundred and seventeen clades of euagarics. *Molecular Phylogenetics and Evolution* **23**: 357–400.

Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B (2011) How many species are there on earth and in the ocean. *PLoS Biology* **9** (8): e1001127.

Morris, HM, Smith ME, Rizzo DM, Rejmanek M, Bledsoe CS (2008) Contrasting ectomycorrhizal fungal communities on the roots of co-occurring oaks (*Quercus* spp.) in a California woodland. *New Phytologist* **178**: 167–176.

Morrison DA (2009) A framework for phylogenetic sequence alignment. *Plant Systematics and Evolution* **282**: 127–149.

Mysara M, Saeys Y, Leys N, Raes J, Monsieurs P (2015) CATCh, an ensemble classifier for chimera detection in 16S rRNA sequencing studies. *Applied and Environmental Microbiology* **81**: 1573–1584.

Nagy LG, Házi J, Vágvölgyi C, Papp T (2012) Phylogeny and species delimitation in the genus *Coprinellus* with special emphasis on the haired species. *Mycologia* **104**: 254–275.

Nilsson RH, Kristiansson E, Ryberg M, Larsson KH (2005) Approaching the taxonomic affiliation of unidentified sequences in public databases–an example from the mycorrhizal *Fungi*. *BMC Bioinformatics* **6**: 178.

Nilsson RH, Ryberg M, Kristiansson E, Abarenkov K, Larsson K-H, Kõljalg U (2006) Taxonomic reliability of DNA sequences in public sequence databases: a fungal perspective. *PLoS One* **1**: e59.

Nilsson RH, Kristiansson E, Ryberg M, Hallenberg N, Larsson KH (2008) Intraspecific ITS variability in the kingdom *Fungi* as expressed in the international sequence databases and its implications for molecular species identification. *Evolutionary Bioinformatics* **4**: 193–201.

Nilsson RH, Tedersoo L, Lindahl BD, Kjøller R, Carlsen T, *et al.* (2011) Towards standardization of the description and publication of next-generation sequencing datasets of fungal communities. *New Phytologist* **191**: 314–318.

Nilsson RH, Tedersoo L, Abarenkov K, Ryberg M, Kristiansson E, Hartmann M, Schoch CL, Nylander JA, Bergsten J, Porter TM, Jumpponen A (2012) Five simple guidelines for establishing basic authenticity and reliability of newly generated fungal ITS sequences. *MycoKeys* **4**: 37–63.

Nilsson RH, Hyde KD, Pawłowska J, Ryberg M, Tedersoo L, *et al.* (2014) Improving ITS sequence data for identification of plant pathogenic *Fungi*. *Fungal Diversity* **67**: 11–19.

Nilsson RH, Wurzbacher C, Bahram M, Coimbra VR, Larsson E, *et al.* (2016) Top 50 most wanted *Fungi*. *MycoKeys* **12**: 29.

Niu B, Fu L, Sun S, Li W (2010) Artificial and natural duplicates in pyrosequencing reads of metagenomic data. *BMC Bioinformatics* **11**: 187.

O'Brien HE, Parrent JL, Jackson JA, Moncalvo JM, Vilgalys R (2005) Fungal community analysis by large-scale sequencing of environmental samples. *Applied and Environmental Microbiology* **71**: 5544–5550.

O'Donnell K, Cigelnik E (1997) Two divergent intragenomic rDNA ITS2 types within a monophyletic lineage of the fungus *Fusarium* are nonorthologous. *Molecular Phylogenetics and Evolution* **7**: 103–116.

O'Donnell K, Cigelnik E, Nirenberg HI (1998) Molecular systematics and phylogeography of the *Gibberella fujikuroi* species complex. *Mycologia* **90**: 465–493.

O'Donnell K, Ward TJ, Robert VARG, Crous PW, Geiser DM, Kang S (2015) DNA sequence-based identification of *Fusarium*: current status and future directions. *Phytoparasitica* **43**: 583–595.

Penn O, Privman E, Landan G, Graur D, Pupko T (2010a) An alignment confidence score capturing robustness to guide-tree uncertainty. *Molecular Biology and Evolution* **27**: 1759–1767.

Penn O, Privman E, Ashkenazy H, Landan G, Graur D, Pupko T (2010b) GUIDANCE: a web server for assessing alignment confidence scores. *Nucleic Acids Research* **38**: W23–W28.

Porazinska DL, Giblin-Davis RM, Sung W, Thomas WK (2012) The nature and frequency of chimeras in eukaryotic metagenetic samples. *Journal of Nematology* **44**: 18–25.

Porter TM, Golding GB (2011) Are similarity- or phylogeny-based methods more appropriate for classifying internal transcribed spacer (ITS) metagenomic amplicons? *New Phytologist* **192**: 775–782.

Powell JR, Monaghan MT, Opik M, Rillig MC (2011) Evolutionary criteria outperform operational approaches in producing ecologically relevant fungal species inventories. *Molecular Ecology* **20**: 655–666.

Pringle A, Baker DM, Platt JL, Wares JP, Latge JP, Taylor JW (2005) Cryptic speciation in the cosmopolitan and clonal human pathogenic fungus *Aspergillus fumigatus*. *Evolution* **59**: 1886–1899.

Pryce TM, Palladino S, Kay ID, Coombs GW (2003) Rapid identification of *Fungi* by sequencing the ITS1 and ITS2 regions using an automated capillary electrophoresis system. *Medical Mycology* **41**: 369–381.

Quaedvlieg W, Binder M, Groenwakd JZ, Summerell BA, Carnegie AJ, *et al.* (2014) Introducing the consolidated species concept to resolve species in the *Teratosphaericaeae*. *Persoonia* **33**: 1–40.

Quail MA, Smith M, Coupland P, Otto TD, Harris SR, *et al.* (2012) A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* **13**: 341.

Quince C, Lanzén A, Curtis TP, Davenport RJ, Hall N, *et al.* (2009) Accurate determination of microbial diversity from 454 pyrosequencing data. *Nature Methods* **6**: 639–641.

Quince C, Lanzen A, Davenport RJ, Turnbaugh PJ (2011) Removing noise from pyrosequenced amplicons. *BMC Bioinformatics* **12**: 38.

Rambold G, Stadler M, Begerow D (2013) Mycology should be recognized as a field in biology at eye level with other major disciplines – a memorandum. *Mycological Progress* **12**: 455–463.

Renner SS (2016) A return to Linnaeus's focus on diagnosis, not description: the use of DNA characters in the formal naming of species. *Systematic Biology* **65**: 1085–1095.

Ronaghi M, Elahi E (2002) Pyrosequencing for microbial typing. *Journal of Chromatography B* **782**: 67–72.

Rosenthal LM, Larsson KH, Branco S, Chung JA, Glassman SI, *et al.* (2017) Survey of corticioid *Fungi* in North American pinaceous forests reveals hyperdiversity, underpopulated sequence databases, and species that are potentially ectomycorrhizal. *Mycologia* **109**: 115–127.

Roskov Y, Abucay L, Orrell T, Nicolson D, Flann C, *et al.* (eds) (2016) *Species 2000 & ITIS Catalogue of Life, 2016 Annual Checklist*. Leiden: Naturalis; Digital resource at www.catalogueoflife.org/annual-checklist/2016.

Rosling A, Cox F, Cruz-Martinez K, Ihrmark K, Grelet GA, *et al.* (2011) *Archaeorhizomycetes*: unearthing an ancient class of ubiquitous soil *Fungi*. *Science* **333**: 876–879.

Rossman AY (2007) Report of the planning workshop for all *Fungi* DNA Barcoding. *Inoculum* **58**: 1–5.

Ryberg M, Nilsson RH, Kristiansson E, Topel M, Jacobsson S, Larsson E (2008) Mining metadata from unidentified ITS sequences in GenBank: a case study in *Inocybe* (Basidiomycota). *BMC Evolutionary Biology* **8**: 50–64.

Salipante SJ, Kawashima T, Rosenthal C, Hoogestraat DR, Cummings LA, *et al.* (2014) Performance comparison of Illumina and ion torrent next-generation sequencing platforms for 16S rRNA-based bacterial community profiling. *Applied and Environmental Microbiology* **80**: 7583–7591.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, *et al.* (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied Environmental Microbiology* **75**: 7537–7541.

Schloss PD, Gevers D, Westcott SL (2011) Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PloS One* **6** (12): e27310.

Schmit JP, Mueller GM (2007) An estimate of the lower limit of fungal diversity. *Biodiversity and Conservation* **16**: 99–111.

Schmitt I, Lumbsch HT (2009) Ancient horizontal gene transfer from bacteria enhances biosynthetic capabilities of *Fungi*. *PloS One* 4 (2): e4437.

Schmitt I, Martín MP, Kautz S, Lumbsch HT (2005) Diversity of non-reducing polyketide synthase genes in the *Pertusariales* (lichenized *Ascomycota*): a phylogenetic perspective. *Phytochemistry* **66**: 1241–1253.

Schoch CL, Sung GH, López-Giráldez F, Townsend JP, Miadlikowska J, *et al.* (2009) The *Ascomycota* tree of life: a phylum-wide phylogeny clarifies the origin and evolution of fundamental reproductive and ecological traits. *Systematic Biology* **58**: 224–239.

Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, *et al.* (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for *Fungi*. *Proceedings of the National Academy of Sciences, USA* **109**: 6241–6246.

Seifert KA (2008) The all-*Fungi* barcoding campaign (FunBOL). *Persoonia* **20**: 106.

Seifert KA (2009) Progress towards DNA barcoding of *Fungi*. *Molecular Ecology Resources* **9** (S1): 83–89.

Simon UK, Weiß M (2008) Intragenomic variation of fungal ribosomal genes is higher than previously thought. *Molecular Biology and Evolution* **25**: 2251–2254.

Smith ME, Douhan GW, Rizzo DM (2007a) Intra-specific and intra-sporocarp ITS variation of ectomycorrhizal *Fungi* as assessed by rDNA sequencing of sporocarps and pooled ectomycorrhizal roots from a *Quercus* woodland. *Mycorrhiza* **18**: 15–22.

Smith ME, Douhan GW, Rizzo DM (2007b) Ectomycorrhizal community structure in a xeric *Quercus* woodland based on rDNA sequence analysis of sporocarps and pooled roots. *New Phytologist* **174**: 847–863.

Soanes D, Richards TA (2014) Horizontal gene transfer in eukaryotic plant pathogens. *Annual Review of Phytopathology* **52**: 583–614.

Sogin ML, Morrison HG, Huber JA, Welch MD, Huse SM, *et al.* (2006) Microbial diversity in the deep sea and the underexplored 'rare

biosphere'. *Proceedings of the National Academy of Sciences, USA* **103**: 12115–12120.

Spatafora JW (2005) Assembling the fungal tree of life (AFTOL). *Mycological Research* **109**: 755–756.

Spatafora JW, Chang Y, Benny GL, Lazarus K, Smith ME, *et al.* (2016) A phylum-level phylogenetic classification of zygomycete *Fungi* based on genome-scale data. *Mycologia* **108**: 1028–1046.

Spribille T, Tuovinen V, Resl P, Vanderpool D, Wolinski H, *et al.* 2016) Basidiomycete yeasts in the cortex of ascomycete macrolichens. *Science*: 10.1126/science.aaf8287

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313.

Stamatakis A, Komornik Z, Berger SA (2010) Evolutionary placement of short sequence reads on multi-core architectures. In: *Computer Systems and Applications (AICCSA), 2010 IEEE/ACS International Conference*: 1–8. Hammamet, Tunisia: Institute of Electrical and Electronics Engineers.

Stark M, Berger SA, Stamatakis A, von Mering C (2010) MLTreeMap – accurate Maximum Likelihood placement of environmental DNA sequences into taxonomic and functional reference phylogenies. *BMC Genomics* **11**: 461.

Suchard MA, Redelings BD (2006) BAli-Phy: simultaneous Bayesian inference of alignment and phylogeny. *Bioinformatics* **22**: 2047–2048.

Sunagawa S, Mende DR, Zeller G, Izquierdo-Carrasco F, Berger SA, *et al.* (2013) Metagenomic species profiling using universal phylogenetic marker genes. *Nature Methods* **10**: 1196–1199.

Susko E (2014) Bayesian long branch attraction bias and corrections. *Systematic Biology* **64**: 243–255.

Tanabe AS, Toju H (2013) Two new computational methods for universal DNA barcoding: A benchmark using barcode sequences of bacteria, archaea, animals, *Fungi*, and land plants. *PloS One* **8** (10): e76910.

Taylor DL, Booth MG, McFarland JW, Herriott IC, Lennon NJ, *et al.* (2008) Increasing ecological infrence from high throughput sequencing of *Fungi* in the environment through a tagging approach. *Molecular Ecology Resources* **8**: 742–752.

Taylor JW, Jacobson DJ, Kroken S, Kasuga T, Geiser DM, (2000) Phylogenetic species recognition and species concepts in fungi. *Fungal Genetics and Biology* **31**: 21–32.

Tedersoo L, Nilsson RH, Abarenkov K, Jairus T, Sadam A, *et al.* (2010) 454 Pyrosequencing and Sanger sequencing of tropical mycorrhizal *Fungi* provide similar results but reveal substantial methodological biases. *New Phytologist* **188**: 291–301.

Tedersoo L, Bahram M, Põlme S, Kõljalg U, Yorou NS, *et al.* (2014) Global diversity and geography of soil *Fungi*. *Science* **346**: 1256688.

Tedersoo L, Bahram M, Puusepp R, Nilsson RH, James TY (2017) Novel soil-inhabiting clades fill gaps in the fungal tree of life. *Microbiome* **5**: 42.

Tripp EA, Lendemer JC (2012) (4–5) Request for binding decisions on the descriptive statements associated with *Mortierella sigyensis* (*Fungi*: *Mortierellaceae*) and *Piromyces cryptodigmaticus* (*Fungi*: *Neocallimastigaceae*). *Taxon* **61**: 886–888.

Tripp EA, Lendemer JC (2014) Sleepless nights: when you can't find anything to use but molecules to describe new taxa. *Taxon* **63**: 969–971.

Turland NJ, Wiersema JH (2017) Synopsis of Proposals on Nomenclature – Shenzhen 2017: A review of the proposals concerning the International Code of Nomenclature for algae, *Fungi*, and plants submitted to the XIX International Botanical Congress. *Taxon* **66**: 217–274.

Turland NJ, Wiersema JH, Barrie FR, Greuter W, Hawksworth DL, *et al.* (eds) (2018) *International Code of Nomenclature for algae, fungi, and plants (Shenzhen Code) adopted by the Nineteenth International Botanical Congress Shenzhen, China, July 2017.* [Regnum Vegetabile no. 159.] Glashütten: Koeltz Botanical Books.

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

Vilgalys R (2003) Taxonomic misidentification in public DNA databases. *New Phytologist* **160**: 4–5.

Von Konrat M, de Lange P, Greif M, Strozier L, *et al.* (2012) *Frullania knightbridgei*, a new liverwort (*Frullaniaceae*, *Marchantiophyta*) species from the deep south of Aotearoa-New Zealand based on an integrated evidence-based approach. *PhytoKeys* **2012** (8): 13–36.

Vrålstad T (2011) ITS, OTUs and beyond – fungal hyperdiversity calls for supplementary solutions. *Molecular Ecology* **20**: 2873–2875.

Wake DB (1991) Homoplasy: the result of natural selection, or evidence of design limitations? *The American Naturalist* **138**: 543–567.

Wake DB, Vredenburg VT (2008) Are we in the midst of the sixth mass extinction? A view from the world of amphibians. *Proceedings of the National Academy of Sciences, USA* **105**: 11466–11473.

Walker JF, Miller OK, Horton JL (2008) Seasonal dynamics of ectomycorrhizal fungus assemblages on oak seedlings in the southeastern Appalachian Mountains. *Mycorrhizia* **19**: 123–132.

Wang XC, Liu C, Huang L, Bengtsson-Palme J, Chen H, *et al.* (2015) ITS1: a DNA barcode better than ITS2 in eukaryotes? *Molecular Ecology Resources* **15**: 573–586.

Wang Z, Nilsson RH, Lopez-Giraldez F, Zhuang WY, Dai YC, *et al.* (2011) Tasting soil fungal diversity with earth tongues: phylogenetic test of SATe alignments for environmental ITS data. Plos One **6**: e19039.

Wapinski I, Pfeffer A, Friedman N, Regev A (2007) Natural history and evolutionary principles of gene duplication in *Fungi*. *Nature* **449**: 54–61.

White TJ, Bruns T, Lee SJWT, Taylor JW (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR Protocols: a guide to methods and applications* (Innis MA, Gelfand DH, Snisky JJ, White TJ, eds): 315–322. San Diego: Academic Press.

Wiens JJ (2004) The role of morphological data in phylogeny reconstruction. *Systematic Biology* **53**: 653–661.

Will KW, Mishler BD, Wheeler QD (2005) The perils of DNA barcoding and the need for integrative taxonomy. *Systematic Biology* **54**: 844–851.

Yergeau E, Lawrence JR, Sanschagrin S, Waiser MJ, Korber DR, Greer CW (2012) Next-generation sequencing of microbial communities in the Athabasca River and its tributaries in relation to oil sands mining activities. *Applied and Environmental Microbiology* **78**: 7626–7637.

Zhang J, Kapli P, Pavlidis P, Stamatakis A (2013) A general species delimitation method with applications to phylogenetic placements. *Bioinformatics* **29**: 2869–2876.

Zhou J, Wu L, Deng Y, Zhi X, Jiang YH, *et al.* (2011) Reproducibility and quantitation of amplicon sequencing-based detection. *ISME Journal* **1**: 11.

NOTE: the Supplementary files S1–S3 can be found on the IMA Fungus website, www.imafungus.org.