Three reasons to re-evaluate fungal diversity ‘on Earth and in the ocean’

David BASS\textsuperscript{a,*}, Thomas A. RICHARDS\textsuperscript{b}

\textsuperscript{a}Department of Zoology, The Natural History Museum, Cromwell Road, London SW5 7BD, UK
\textsuperscript{b}Biological Sciences, University of Exeter, Geoffrey Pope Building, Exeter EX4 4QD, UK

\textbf{A R T I C L E   I N F O}

Article history:
Received 6 October 2011
Received in revised form 10 October 2011
Accepted 10 October 2011

Keywords:
Archeorhizomycetes
Cryptomycota
Fungal diversity
Number of species
454 Sequencing

\textbf{A B S T R A C T}

Attempts to assess fungal global species richness are confounded by several problems: uncertainty about the number of described species, incomplete fungal inventories even at a high taxonomic level, high diversity of unknown, often small and elusive taxa, high levels of morphological conservation, and incomplete knowledge of their ecological and biogeographical distributions. The two main bases for estimating total fungal diversity are (1) the number of described species and their taxonomic structure, and (2) extrapolating species-area relationships. We argue that knowledge of fungal taxonomy and environmental sampling of fungi are both too incomplete for either approach to be reliable. However, it is likely that the true number of fungal species on the planet is a seven-digit number, and may even be an order of magnitude higher.

© 2011 Published by Elsevier Ltd on behalf of The British Mycological Society.

\section{1. Introduction}

Recently, Mora \textit{et al.} (2011) estimated the total number of species on Earth using a predictive algorithm based on generalizing the ‘strong’ correlation between number of higher taxa and taxonomic rank across all of life. This approach relies on the fact that the number of higher taxa is much more complete and consistent than the total number of species. For each taxonomic level from phylum to genus they fitted asymptotic regression models to the temporal accumulation curves of higher taxa, then the predicted number of taxa at each taxonomic rank down to genus was regressed against the numerical rank, and the fitted models used to predict the number of species (Mora \textit{et al.}, 2011). This approach was shown to work best where the higher ranks were most completely known and stable, and their number was at or close to asymptote, and where species were clearly and reproducibly defined and their diversity relatively well investigated. Consequently, it works particularly poorly for microbes, especially prokaryotes, in which new evolutionary groups are steadily being described (Pace, 1997; Rappe and Giovannoni, 2003) and where ‘species’ are often not the unit of research focus and therefore under-developed as a concept, and also tend to be more genetically inclusive than in many eukaryote groups (e.g. Doolittle, 2008; Gevers \textit{et al.}, 2005; Koeppl \textit{et al.}, 2008; Konstantinidis and Tiedje, 2005; Staley, 2006; Vandamme \textit{et al.}, 1996).

Unlike in animals and plants, where the algorithms used by Mora \textit{et al.} (2011) can be argued to have a historical validity in that taxa at all levels have been described using similar...
approaches and in a relatively consistent manner over the
period of their analysis (1800–2000), fungi (along with protists
and prokaryotes) have not enjoyed such consistency. Here we
expand upon the caveats outlined in Mora et al. (2011)
regarding the estimation of total fungal diversity and identify
three main reasons why the total number of species is almost
certainly much higher than their predicted figure of 611,000 
\(\pm\) 297,000: (1) uncertainty about the number of
described fungal species; (2) the revelations provided by
molecular biology and environmental probing; and (3) the
expanding number of higher fungal taxa.

2. How many fungal species are actually
known?

Most effort in describing and cataloguing fungal species has
been directed towards the larger and more obvious forms.
The current (10th) edition of the Dictionary of the Fungi gives
a figure of 98,128 species (Kirk et al., 2008) excluding
all fungal analogues (Richards et al., in press). The majority of
this 98,128 are terrestrial ascomycete and basidiomycete
species (Kirk et al., 2008). This figure may be artificially inflated
by synonyms created by separately described anamorphs and
teleomorphs, and synanamorphs (as discussed by Blackwell
(2011) – see Fig. 11 of her paper). Applying the 65 % adjustment
to account for synonymy as described by Hawksworth (1992,
2004) would bring this figure down to \(~59,000\), but this is far
too drastic an adjustment for this well-curated source of data.
Therefore, the figure of 43,271 used by Mora et al. (2011) seems
too low as a starting point for their calculations, even if it was
both a conservative number with linked taxonomic data (ob-
tained from www.species2000.org) and therefore the only one
available which matched the requirements of their methods.

There have been several attempts to estimate the total
species richness of fungi on the basis of their association
with plants. The well-known estimate of 1.5 M (more accu-
trately 1.62 M) species by Hawksworth (1991), later revised to
2.27 M (Hawksworth, 2001), assumes certain ratios of fungi to
vascular plant species. Although the figure of \(~1.5 M\) species
has been widely accepted by fungal experts, and in many cases
exceeded (Table 1), this approach is open to criticism as being
too much of a generalisation from local data and not consis-
tently applicable at a global scale (May, 1991, 1994; Mueller
and Schmit, 2007; Schmit and Mueller, 2007). However, in
important respects – as recognized by Hawksworth himself –
this method underestimates: it doesn’t take into account fungi not
associated with plants, those in non-soil habitats, elusive

Fig. 1 – Schematic fungal tree of life, representing some recent major additions excluded from fungal counting exercises (Hirt
et al., 1999; Jones et al., 2011; Rosling et al., 2011; Schadt et al., 2003). Topology shown is derived from James et al. (2006b) but
includes many weakly supported branching relationships drawn in grey. Numbers shown in square brackets are taken from
Blackwell (2011), Kirk et al., (2008), and refer to described species only.
and cryptic species, and higher fungal diversity in the tropics (Arnold and Lutzoni, 2007; Blackwell, 2011; Hawksworth, 2001; Hawksworth and Rossman, 1997; Richards et al., in press). For example, analysis of the diversity of fungi associated with palm species has suggested much higher fungi to vascular plant species ratios ranging from 26:1 to 33:1 (Fröhlich and Hyde, 1999; Hyde et al., 1997). Similar studies of the Phyllocladusaceae have led to suggestions that fungi may possibly number 9.9 M species (Cannon, 1997). Because of the difficulty of extrapolating geographically and taxonomically limited plant-fungi associations to a global scale and the attendant risk of overestimate, Schmit and Mueller (2007) conservatively estimated a lower boundary for global fungal diversity at 712,000 extant species, which the majority of commentators (including themselves) believe to be too low. Therefore, the predicted total of Mora et al. (2011), 611,000 +/- SE = 297,000, is already lower than conservative estimates based on traditional taxonomic approaches (Table 1).

### Table 1 – Popular estimations of species diversity in the Fungi. Reproduced from Hawksworth (2001), including some additional estimates [*].

<table>
<thead>
<tr>
<th>Publication</th>
<th>Estimated species number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pascoe (1990)</td>
<td>2,700,000</td>
</tr>
<tr>
<td>Hawksworth (1991)</td>
<td>1,620,000</td>
</tr>
<tr>
<td>Hammond (1992)</td>
<td>1,000,000</td>
</tr>
<tr>
<td>Rossman (1994)</td>
<td>1,000,000</td>
</tr>
<tr>
<td>Hammond (1992)</td>
<td>1,500,000</td>
</tr>
<tr>
<td>Cannon (1997)</td>
<td>9,900,000</td>
</tr>
<tr>
<td>Fröhlich and Hyde (1999)</td>
<td>1,500,000</td>
</tr>
<tr>
<td>Hawksworth (2001)</td>
<td>2,270,000</td>
</tr>
<tr>
<td>May (2000)</td>
<td>500,000*</td>
</tr>
<tr>
<td>O’Brien et al. (2005)</td>
<td>3,500,000—5,100,000*</td>
</tr>
<tr>
<td>Schmit and Mueller (2007)</td>
<td>712,000*</td>
</tr>
<tr>
<td>Mora et al. (2011)</td>
<td>611,000 +/- SE = 297,000*</td>
</tr>
</tbody>
</table>

3. Molecular inventories of fungal diversity and the cryptic majority

Small and cryptic organisms are difficult to distinguish according to the techniques of classical taxonomy, but their diversity is potentially massive (Horton and Bruns, 2001; Rappe and Giovannoni, 2003; Sogin et al., 2006). One clear illustration of this is the effect of the molecular biology revolution on diversity estimates in microbial groups: environmental sequencing has revealed a far higher microbial species-level diversity than is suggested by their morphological diversity. The alpha-level diversity in these groups is often (a) very difficult to detect and enumerate without molecular techniques, and (b) hyper-diverse when measured by genetic distances that approximate to species-level differences in many multicellular groups.

The revelations of molecular biology therefore constitute our second reason to re-consider fungal species richness. If we accept that internal transcribed spacer ribosomal DNA (ITS rDNA; which is situated between the small and large subunit rRNA genes) is a workable marker of species diversity in fungi, as has been reasoned by mycologists for over 20 y (see Horton and Bruns, 2001 for review) then environmental clone library sequencing, and more recently 454 sequencing, strongly suggest that the number of fungal species must be much greater than 611,000. The particular power of these molecular techniques is that they do not rely on visual detection or identification, and avoid all biases associated with isolation and cultivating in the laboratory.

A striking finding of many fungal-specific environmental sequencing studies is that they reveal high proportions of novel sequences, even when clustered quite conservatively. For example, Buée et al. (2009) used 454 sequencing to measure fungal diversity in forest soils. Their samples each comprised 4 g of forest soil from which they recovered a mean of 830 Operational Taxonomic Units (OTUs) (clustered at 97 % sequence similarity to control for sequencing artefact), for which the mean Chaol nonparametric OTU diversity estimate was 2240. 71.5 % of their sequences were not allocated to any particular fungal taxon when blasted against NCBI (http://www.ncbi.nlm.nih.gov/) or UNITE (http://unite.ut.ee/), and even when compared to a curated database of robustly identified sequences and excluding all ‘uncultured fungi’ 11 % remained unclassified and a further 20 % were grouped as unclassified Dikarya (Buée et al., 2009). Jumpponen and Jones (2009) used 454 sequencing to show that hundreds of fungal OTUs can be detected on a few square centimetres of Quercus macrocarpa leaf. Even after a conservative strategy of clustering at 95 % ITS rDNA similarity and excluding all singleton sequences, 11.3 % of their sequences were of uncertain affiliation. 48 % of their OTUs were singletons, which are likely to harbour a higher proportion of novel sequence types (Jumpponen and Jones, 2009). These studies provide a robust demonstration that the diversity of fungi in the vicinity of plants (although not necessarily involved in direct associations) is much higher than any of the ratios used for calculating fungal species numbers (Hawksworth, 2001). It should be noted that the samples analysed by molecular studies are usually very small and reveal relatively high levels of novelty within and between libraries. We currently have little idea how much sampling would be required to approach asymptote on a species accumulations curves even locally, and no appreciation of this on a global scale.

O’Brien et al. (2005), using clone library techniques, estimated 491 fungal OTUs in their pine forest soil samples and 616 in mixed hardwood plot samples. These results were from just a few grams of soil and are clearly underestimates as the ACE richness estimator continued to increase beyond the limits of their sampling. The authors then tentatively compare this data to the vascular plant richness in their plots assuming the validity of the 1.5 M fungal species estimate of Hawksworth (1991) and suggest that global fungal diversity may actually fall within the range of 3.5–5.1 M OTUs (O’Brien et al., 2005). As acknowledged by the authors, these figures are based on simple correlations between observations of fungal and plant diversity. But the point is made that ‘hidden’ fungal diversity in many (most?) environments may be an order of magnitude or more than suggested by estimates based on morphology and traditional taxonomy.

Even these molecular methods underestimate the diversity present. It is important to remember that primers used for most ‘fungal-specific’ molecular probing studies have been...
designed on the basis of known sequences, and are often biased towards Dikarya. More highly specific fungal group primers can reveal much higher levels of diversity within their group than general fungal primers do (e.g. Porter et al., 2008).

However, such highly specific approaches rarely detect more divergent fungi (e.g. novel phyla-level diversity), as can be revealed by eukaryote-wide studies (see Jones et al., 2011 and below). Therefore, to some extent poorly-known fungi are constrained to remain poorly known, emphasizing the importance of employing a range of approaches when assessing microbial diversity.

Molecular techniques also highlight the importance of distinguishing carefully between fungal microhabitats. O’Brien et al. (2005) detected a positive species–sample size relationship in their data at a relatively small scale; it would be interesting to apply this on the scale of a whole woodland, for example. Porras-Alfaro et al. (2011) used fungal clone libraries to describe communities in semi-arid grasslands. Concordantly with other molecular studies 40% of their OTUs were novel, i.e. less than 97% similar to other sequences in NCBI. They also showed that fungal communities in soil and root-derived samples were highly significantly different (Porras-Alfaro et al., 2011). But even different soil types can have highly different dominating fungal communities: in a 500-ITS rDNA sequence multi-library study of four separate woodlands and immediately adjacent grasslands, no ITS rDNA sequence was recovered from both grassland and woodland soils (Bass et al., 2011).

It is also important to note that Hawksworth on several occasion states that the 1.5 million estimate does not take account of fungal diversity associated with insects (Hawksworth, 1991, 2001). Traditional approaches to sampling suggests between 40,000 and 100,000 beetle species are hosts to at least one ectoparasitic laboulbeniales species. Data on host specificity infer that there is a similar number of 10,000–50,000 beetle-associated laboulbeniales (Weir and Hammond, 1997). Furthermore, Suh and co workers demonstrated that beetle guts contain a diverse community of yeast species. Sequencing of a 600 bp fragment of the large subunit (LSU) ribosomal RNA (rRNA) gene revealed that 68% of the strains isolated had >5 bp differences compared to previously sampled fungi according to this relatively conserved molecular marker. This resulted in 200 previously undescribed phylotypes, arguably increasing the known census of yeast species by 30% (Suh et al., 2005).

Therefore molecular data, particularly that generated by next generation sequencing technologies, offer a powerful insight into fungal diversity (for review see Jones and Richards, 2011). However, the true power of this approach will only be revealed in time, when sampling coverage becomes sufficiently large. This will reduce the need to resort to unreliable extrapolations and assumptions about fungal diversity and its distribution. The main important factors to take into account to engage the full potential of environmental sequence data for global species diversity estimates are:

1. habitat heterogeneity,
2. species-area relationships,
3. primer specificity and inclusivity,
4. depth of sequencing (sampling saturation),
5. global biogeographical structuring,
6. relatively unexplored environments (e.g. Le Calvez et al., 2009; Suh et al., 2005),
7. comparability of molecular sampling methods.

Meta-analyses such as that for marine fungi clone library data in Richards et al. (in press) remove some of these biases, and have the potential to show the extent of overlap in diversity between different studies and environmental samples. This study identified 36 distinct and novel marine lineages (defined at 97% similarity level using the relatively conserved SSU rRNA molecular marker), the majority (24) and most divergent of which branch with the chytrids (Richards et al., in press). It is therefore likely that a large proportion of currently unknown fungal diversity resides within poorly studied groups, or even unrecognized ones, as discussed further in our third point below.

4. The implication of adding new groups at the highest taxonomic grades

The approach employed by Mora et al. (2011) relies heavily on a stable census of higher taxonomic groups for the species estimation methodology to be reliable. Insights from molecular techniques have implications at all taxonomic levels in fungi, not just around the species level. Blackwell (2011) states that ‘there cannot be any doubt that ascomycetes and basidiomycetes (Dikarya) comprise the vast majority of fungal diversity’. This is certainly true in terms of described species, but is not necessarily true when considering the entirety of the fungal kingdom. Only recently large higher-level taxa were described that were originally detected by environmental sequence studies: Cryptomycota (Jones et al., 2011; Lara et al., 2010), a chytrid group, and Archaeorhizomycetes and other soil ascomycete groups (Porter et al., 2008; Rosling et al., 2011; Schadt et al., 2003) (Fig. 1). More high level taxa are likely to be detected and described in the future, as suggested by the diverse and deep-branching marine chytrid groups shown from molecular studies by Le Calvez et al. (2009) and the meta-analysis of Richards et al. (in press). These new groups comprise physically small, cryptic, and elusive elements of fungal diversity, but that diversity – in terms of SSU rDNA variation – can be huge, as demonstrated for the Cryptomycota by Jones et al. (2011). The recognition of Cryptomycota alone could radically increase the size of the fungal kingdom. Furthermore, there is increasing evidence and consensus that another very diverse group of elusive organisms, the endoparasitic and protist-like Microsporidia belong within fungi (Hirt et al., 1999; Keeling, 2003), perhaps with a particular affinity with zygomycetes (Keeling, 2003; Lee et al., 2008) or Rozella (James et al., 2006a), but they are not counted as fungi in the main databases. Microsporidia are currently thought to comprise around 150 genera with 1200–1300 species (Lee et al., 2009), but the real figures are likely to be much higher as the diversity of their hosts is currently incompletely known and molecular diversity studies of this group are in their infancy (Krebes et al., 2010; McClymont et al., 2005).

5. Conclusion

In summary, we assert that there is at least one order of magnitude more fungal species than are currently known.
The predicted 611,000 of Mora et al. (2011) should be seen as a call-to-arms for a concerted and continued effort to improve our knowledge of fungal diversity by all means available. It is essential that modern taxonomic methods and databases are employed and maintained so that improvement in our knowledge of this all-pervasive and hugely influential group of organisms is embedded within a robust bioinformatics infrastructure. Such approaches will generate new ideas and hypotheses, rather than reflecting a historically confounded view of fungal diversity.

**References**


