BACKGROUND: DEVELOPMENT OF T-BAS

T-BAS was developed as part of a Dimensions of Biodiversity project (an interdisciplinary study of hyperdiverse endophytic fungi and their function in boreal forests) funded by the National Science Foundation (www.enDoBiodiversity.org). The project focuses on foliar endophytes (i.e., fungi that live within asymptomatic photosynthetic tissues of plants) and endolichenic fungi (non-lichen fungi that occur within healthy lichen thalli), which comprise an immense phylogenetic richness around the circumboreal belt (Arnold and Lutzoni, 2007; Arnold, et al., 2009; Higgins, et al., 2007; U'Ren, et al., 2012). Data for this project and related studies have been obtained through both culture-dependent and -independent methods (see Chen, et al., 2015; Huang, et al., 2016; Massimo, et al., 2015; Oono, et al., 2014; U'Ren, et al., 2012), including next generation sequencing (NGS) of amplicon pools (e.g., U'Ren, et al., 2014).

The vast majority of foliar endophytic and endolichenic fungi, such as those described in the EnDoBiodiversity project above, are members of the largest fungal subphylum – the Pezizomycotina (filamentous ascomycetes, Ascomycota). The locus of choice for characterizing unknown Fungi is the nuclear ribosomal RNA internal transcribed spacers and nr5.8S gene (nrITS; Schoch, et al., 2012). However, because the two spacer regions (nrITS1 and nrITS2) cannot be aligned reliably across highly divergent taxa, only the conserved region – the short, RNA-coding nr5.8S gene – can be used for inferring large-scale phylogenetic trees. However, the nr5.8S is both too short and insufficiently informative to reconstruct known relationships among major fungal groups, leaving a large gap between studies using nrITS regions for diversity and phylogenetic information.

To obtain more reliable phylogenetic placements of taxa represented in the culture collection resulting from our study of endophytic and endolichenic fungi, we adopted an approach sequencing a single fragment including both the nrITS and an adjacent ca. 600 bp portion of the nrLSU using Sanger sequencing technology (see Arnold, et al., 2007). We refer to this region as the nrITS-LSU locus. This region has been sequenced systematically for all fungal strains in our studies (e.g., see U'Ren, et al., 2010; U'Ren, et al., 2012; U'Ren, et al., 2016; U'Ren, et al., 2014; see also Chen, et al., 2015; Huang, et al., 2016; Massimo, et al., 2015; Oono, et al., 2014; Sandberg, et al., 2014; Arnold et al., unpublished data), and aligned with published sequences from 396 representative Pezizomycotina species to infer phylogenetic placement (unpublished, but see Arnold, et al., 2009; Chen, et al., 2015; U'Ren, et al., 2016). However, inferring large-scale phylogenies using only nr5.8S and nrLSU sequence data results in low phylogenetic resolution and thus limits the scaling needed for ecological studies. To accomplish robust phylogenetic placement of unknown samples of fungi within the Pezizomycotina, a two-step approach can be used. First, a multi-locus, well-resolved tree with high phylogenetic confidence can be generated using representative taxa from each main lineages of this subphylum. Second, a well-supported backbone constraint tree (see Arnold, et al., 2007; Chen, et al., 2015; Martin, et al., 2015) can be extracted from the resulting multi-locus phylogeny and used to enhance the phylogenetic searches based solely on nr5.8S and nrLSU. This two-step process relies in part on the use of pure cultures to obtain reliable sequences from multiple loci, or increasingly from entire genomes, to eventually infer phylogenetic affinities of unknown taxa for which only Sanger sequences from the nrITS-

LSU locus are available, or short sequence reads are obtained directly from environmental samples.

Overview of Future Development for T-BAS

T-BAS version 1.0 includes a multi-locus tree for the Pezizomycotina and associated alignments, which provide custom alignments, taxonomic information, and GenBank accession numbers for all included loci (Fig. 1). Currently, online users can download curated alignments for specific reference species and loci combinations of their choice using the T-BAS tree visualization tool directly, or after placing on that reference tree unknown sequences using BLAST and/or a phylogeny-based placement methods. In the latter case, downloaded alignments can include unknown sequences if selected by the user. Therefore, T-BAS can be used to align sequences by integrating them in existing T-BAS master alignments.

The next version of T-BAS will also include a single-locus (nrITS-LSU) phylogeny (Fig. 1). Contrary to the multi-locus tree, which includes only reference taxa for which sequences are available for multiple loci, the nrITS-LSU tree will include reference taxa but also published unknown sequences for which ecological data are available. The nrITS-LSU tree, inferred with the help of a backbone constraint tree from the multi-locus topology, serves a complementary aim to phylogeny-based placements by including sequences from published but undescribed fungi in addition to reference taxa (Fig. 1). This tree can be used to help systematists describe new species because reference species for which only nrITS data are available can be included.

The T-BAS nrITS-LSU tree will help mycologists determine if newly sampled fungi have not been found previously or belong to undescribed species for which no other loci have been sequenced. This tree can guide mycologists to determine which unknown and potentially new species need additional loci, or genomes, to be sequenced; and to evaluate when enough information has been gathered to move forward with new species delimitations and descriptions. Once new species are described and associated sequence data have been published, these new species and associated sequence- and metadata can be added to the T-BAS multi-locus tree, providing more comprehensive alignments and enhancing the backbone constraint trees for the T-BAS nrITS-LSU tree (Fig. 1).

Assembling all necessary reference sequences (see Schoch, et al., 2014) and generating new alignments represents one of the most time-consuming tasks of phylogenetic studies, especially for large-scale projects. Moreover, use of stand-alone phylogenetic tools and static databases impedes information flow between systematists and ecologists, and makes difficult tracking the discovery of unknown fungi globally, information critical to understanding biodiversity and conservation efforts. Such concerns motivate the long-term goals of T-BAS to provide large-scale publically distributed and dynamic online resources that can be shared among different working groups.

REFERENCES

Arnold, A.E.*, et al.* (2007) Diversity and phylogenetic affinities of foliar fungal endophytes in loblolly pine inferred by culturing and environmental PCR, *Mycologia*, **99**, 185-206.

Arnold, A.E. and Lutzoni, F. (2007) Diversity and host range of foliar fungal endophytes: are tropical leaves biodiversity hotspots?, *Ecology*, **88**, 541-549.

Arnold, A.E.*, et al.* (2009) A phylogenetic estimation of trophic transition networks for ascomycetous fungi: are lichens cradles of symbiotrophic fungal diversification?, *Syst Biol*, **58**, 283-297.

Chen, K.-H.*, et al.* (2015) Phylogenetic analyses of eurotiomycetous endophytes reveal their close affinities to Chaetothyriales, Eurotiales, and a new order - Phaeomoniellales, *Molecular Phylogenetics and Evolution*, **85**, 117-130.

Higgins, K.L.*, et al.* (2007) Phylogenetic relationships, host affinity, and geographic structure of boreal and arctic endophytes from three major plant lineages, *Molecular phylogenetics and evolution*, **42**, 543-555.

Huang, Y.L.*, et al.* (2016) Pervasive Effects of Wildfire on Foliar Endophyte Communities in Montane Forest Trees, *Microb Ecol*, **71**, 452-468.

Martin, R.*, et al.* (2015) Unexpected diversity of basidiomycetous endophytes in sapwood and leaves of Hevea, *Mycologia*, **107**, 284-297.

Massimo, N.C.*, et al.* (2015) Fungal endophytes in aboveground tissues of desert plants: infrequent in culture, but highly diverse and distinctive symbionts, *Microb Ecol*, **70**, 61-76.

Oono, R.*, et al.* (2014) Genetic variation in horizontally transmitted fungal endophytes of pine needles reveals population structure in cryptic species, *Am J Bot*, **101**, 1362-1374.

Sandberg, D.C., Battista, L.J. and Arnold, A.E. (2014) Fungal endophytes of aquatic macrophytes: diverse host-generalists characterized by tissue preferences and geographic structure, *Microb Ecol*, **67**, 735-747.

Schoch, C.L.*, et al.* (2014) Finding needles in haystacks: linking scientific names, reference specimens and molecular data for Fungi, *Database (Oxford)*, **2014**.

Schoch, C.L.*, et al.* (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi, *Proc Natl Acad Sci U S A*, **109**, 6241-6246.

U'Ren, J.M.*, et al.* (2010) Community analysis reveals close affinities between endophytic and endolichenic fungi in mosses and lichens, *Microb Ecol*, **60**, 340-353.

U'Ren, J.M.*, et al.* (2012) Host and geographic structure of endophytic and endolichenic fungi at a continental scale, *Am J Bot*, **99**, 898-914.

U'Ren, J.M.*, et al.* (2016) Contributions of North American endophytes to the phylogeny, ecology, and taxonomy of Xylariaceae (Sordariomycetes, Ascomycota), *Molecular phylogenetics and evolution*, **98**, 210-232.

U'Ren, J.M.*, et al.* (2014) Tissue storage and primer selection influence pyrosequencing-based inferences of diversity and community composition of endolichenic and endophytic fungi, *Mol Ecol Resour*, **14**, 1032-1048.

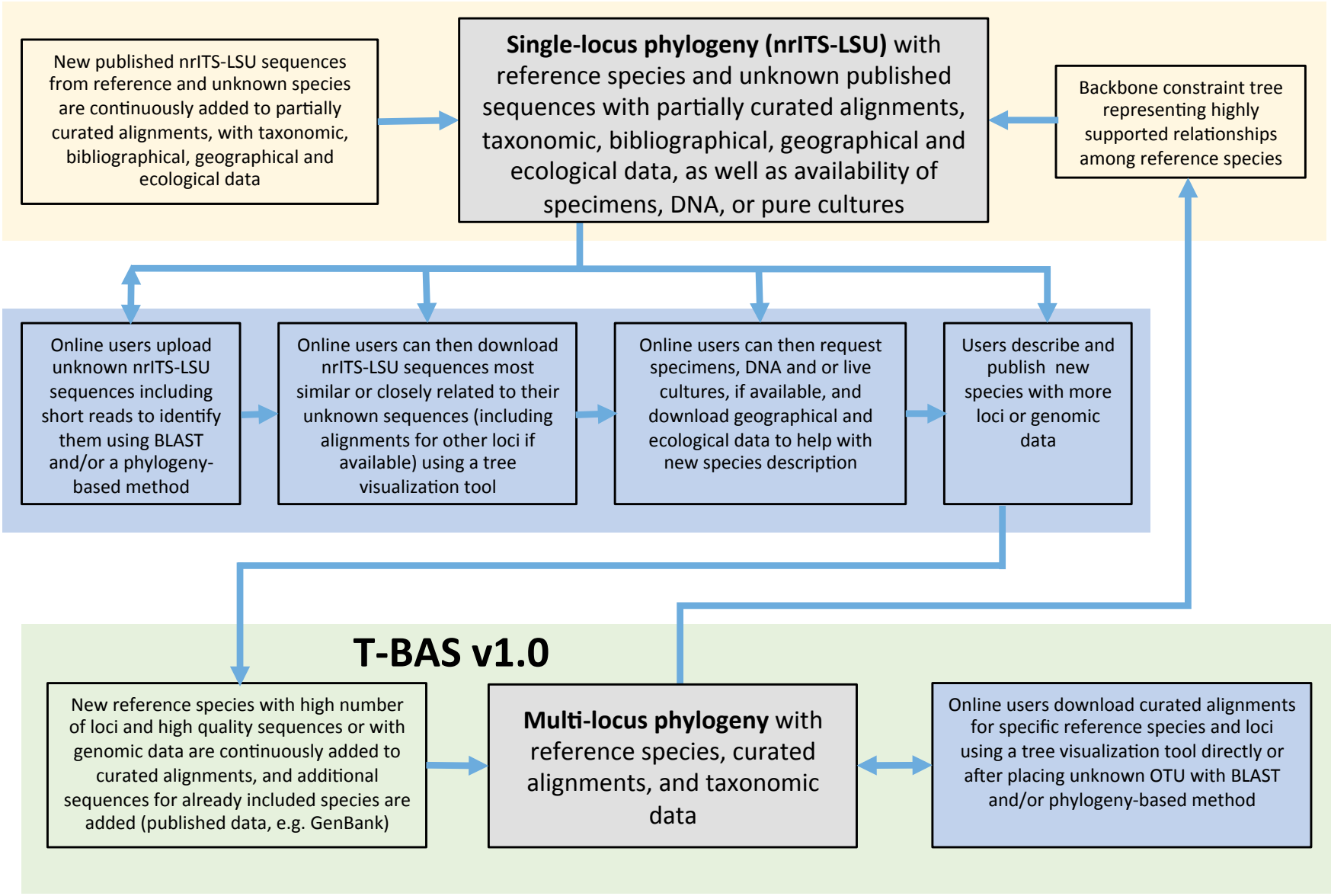
New published nrITS-LSU sequences from reference and unknown species are continuously added to partially curated alignments, with taxonomic, bibliographical, geographical and ecological data
Single-locus phylogeny (nrITS-LSU) with reference species and unknown published sequences with partially curated alignments, taxonomic, bibliographical, geographical and ecological data, as well as availability of specimens, DNA, or pure cultures
Backbone constraint tree representing highly supported relationships among reference species
Online users upload unknown nrITS-LSU sequences including short reads to identify them using BLAST and/or a phylogeny-based method
Online users can then download nrITS-LSU sequences most similar or closely related to their unknown sequences (including alignments for other loci if available) using a tree visualization tool
Online users can then request specimens, DNA and or live cultures, if available, and download geographical and ecological data to help with new species description
Users describe and publish new species with more loci or genomic data
T-BAS v1.0
New reference species with high number of loci and high quality sequences or with genomic data are continuously added to curated alignments, and additional sequences for already included species are added (published data, e.g. GenBank)
Multi-locus phylogeny with reference species, curated alignments, and taxonomic data
Online users download curated alignments for specific reference species and loci using a tree visualization tool directly or after placing unknown OTU with BLAST and/or phylogeny-based method

FIGURE 1. Schematic showing organization and future functions of T-BAS. The shaded portion in green represents services available in the first implementation of T-BAS (v1.0). The yellow shading indicates planned services in a future deployment of T-BAS. T-BAS functions for users (blue shading) are centered on two phylogenetic tree visualization tools (grey shading): (1) a multi-locus tree of reference taxa and (2) a single-locus (nrITS-LSU) tree of reference species and published sequences of unidentified fungi. T-BAS is currently developed for filamentous ascomycete fungi (Pezizomycotina, Ascomycota), but the framework is translatable to user-specified trees. Submission of such trees represents a focal plan for T-BAS v2.0.