

REVIEW

Open Access



Deep learning approaches for natural product discovery from plant endophytic microbiomes

Shiva Abdollahi Aghdam and Amanda May Vivian Brown* 

Abstract

Plant microbiomes are not only diverse, but also appear to host a vast pool of secondary metabolites holding great promise for bioactive natural products and drug discovery. Yet, most microbes within plants appear to be uncultivable, and for those that can be cultivated, their metabolic potential lies largely hidden through regulatory silencing of biosynthetic genes. The recent explosion of powerful interdisciplinary approaches, including multi-omics methods to address multi-trophic interactions and artificial intelligence-based computational approaches to infer distribution of function, together present a paradigm shift in high-throughput approaches to natural product discovery from plant-associated microbes. Arguably, the key to characterizing and harnessing this biochemical capacity depends on a novel, systematic approach to characterize the triggers that turn on secondary metabolite biosynthesis through molecular or genetic signals from the host plant, members of the rich 'in planta' community, or from the environment. This review explores breakthrough approaches for natural product discovery from plant microbiomes, emphasizing the promise of deep learning as a tool for endophyte bioprospecting, endophyte biochemical novelty prediction, and endophyte regulatory control. It concludes with a proposed pipeline to harness global databases (genomic, metabolomic, regulomic, and chemical) to uncover and unsilence desirable natural products.

Keywords: Endophytic fungi, Deep learning, Secondary metabolites, Natural product, Endohyphal bacteria, Mycovirus, miRNA, Multi-omics

Introduction

Microbiomes including communities of fungi and bacteria living asymptotically within plant tissues, are ubiquitous and important components of plants. Specialized microbes within plants harbor capacities to synthesize diverse and unique secondary metabolites (SMs), hence, they have been a major focus for anticancer, antibacterial, antifungal, and antiviral natural product (NP) discovery [1–6]. Even though most plant microbiome species are exceedingly challenging to work with, being difficult to grow and unlikely to express most SMs in culture, interest in

them as a source of medically important NPs has exploded, catapulted by the discovery of the breakthrough anticancer compound paclitaxel (Taxol) synthesized by the endophyte *Taxomyces andreanae* from Pacific yew trees (*Taxus brevifolia*) [7–10]. Research since the discovery of paclitaxel shows plant microbiomes, particularly the internal endophyte communities, offer a treasure trove of bioactive secondary metabolites with at least 60% of characterized species having medical and drug potential due to their novel and novel chemical structures [4, 11–13].

Familiar endophyte-derived medically important compounds include anti-cancer drugs paclitaxel, comptothecin, vinblastine, anti-viral drugs podophyllotoxin, isoindolone, talaromyolide, cytonic acid, and anti-bacterial

* Correspondence: Amanda.mv.brown@ttu.edu
Department of Biological Sciences, Texas Tech University, 2901 Main St, Lubbock, TX 79409, USA



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

drugs altersolanol, cryptocandin, and rutin [4, 14–18]. Indeed, microbes, rather than plants, dominate the pool of identified sources for drugs, representing about 75% of candidate drug sources, generating between 15 and 30 approved new drugs per year in the U.S. with indications for over 70 conditions or diseases [19]. It has been argued that plant microbiomes present a vast underexplored resource for discovery of chemically diverse NPs that may rival that from free-living microbes [20]. This phenomenal potential could be due to their ~400 million years of intimate service to plants [21, 22] in which endophytes evolved in a context of exceptional biochemical demands [23–27] leading to novel SM synthesis.

Whereas the majority of SMs exist in apparently silent gene clusters [28–31], if unsilenced, we estimate that global plant microbiomes may potentially yield 1.3 to 28.3×10^9 NPs that could lead to millions of drugs (see calculations in Tables 1 and 2). This biosynthesis needs only to be awakened – analogous to waking the sleeping giant – but so far, the path forward to harness this potential has been unclear. Significant barriers exist that

prevent progress in endophyte NP discovery. For example, genome sequencing and bioinformatics predict a vast pool of compounds missing in culture-based studies [47, 51, 55, 57–59] that fail to be expressed except *in planta*, or without providing substrates or precursors from plants or other microbes [28, 60]. Regulatory breakdowns that limit endophyte NP expression include spatially and temporally varying signals from the plant, other endophytic fungi, other endophytic bacteria, endohyphal bacteria [61–67], and perhaps even phage or mycoviruses [68–70]. There is also evidence for cooperative synthesis of compounds predicted in the hologenome [61, 71, 72].

This review will not present an exhaustive catalog of plant-associated microbes or NP chemical structures, which have been reviewed elsewhere [15, 73–77]. Nor will we cover detailed methodologies for extracting and analyzing endophyte secondary metabolites covered elsewhere [9, 78–80]. Instead, this review will present a novel analysis of the untapped potential of plant endophytic microbiomes for NP discovery, describing the

Table 1 Estimating plant microbiome diversity and NP potential on Earth

Here, we use simple, additive, non-combinatoric approaches to estimate endophyte species richness. In these calculations, only endophytic fungi and bacteria and considered with the simplifying assumption that each endophyte species acts alone (without the plant or other microbes) to synthesize its NPs

Historical estimates: The most widely cited estimate of endophyte species on Earth was proposed over 25 years ago, predicting 1.3 million endophytic fungi on Earth [11]. The simple calculation considered only culturable fungi from vascular plants and was based on researchers' experience suggesting each plant species hosts 2 to 5 unique host-specific endophytes (thus, for 270,000 plant species there would be up to $5 \times 270,000$ unique endophytes). While this study did not estimate global NPs, the following calculation attempts to do this. This study argued that each phylogenetic cluster of fungi produced largely similar set of known secondary metabolites which largely differed from that of other clusters. For the 8 well-studied groups of endophytic fungi in [11], comprising 8300 species there are an average of 1038 species per group, which would comprise 1253 groups (1.3 million/1083 species). Together these groups reportedly produced 5351 unique known secondary metabolites, or an average of 669 secondary metabolites per group, which for 1253 groups would produce an estimated 838,100 unique secondary metabolites on Earth. Among the shortcomings of these estimates are omission of bacterial endophytes and non-culturable endophytes, and omission of novel metabolites and predictions of silent or cryptic biosynthetic clusters.

Estimates based on next-generation sequencing: Based on amplicon sequencing of endophytes from plant tissues using 16S rRNA and ITS or 18S rRNA genes revealing large numbers of previously uncultured endophytes (i.e. OTU-based surveys, Fig. 1), a simple "back-of-the-napkin" estimate suggests there may be at least one non-culturable host-specific fungal endophyte for every one that is culturable [23], and perhaps 10 host-specific bacterial endophytes, such that for the estimated ~300,000 plant species on Earth, there may be 10 fungi + 10 bacteria (=20) \times 300,000 = 120 million endophyte species on Earth. Whereas this is two orders of magnitude greater than historical estimates, this would constitute only 0.012% of the estimated 1 trillion microbial species on Earth [32], suggesting it is not absurdly high. Based on estimates of known metabolites discussed above, this suggests endophytic fungi might produce 77,450,000 unique secondary metabolites ($110,800 \times 699$ per group) and estimating about half as many unique secondary metabolites per bacterial species, there would be perhaps 38,725,000 unique bacterial metabolites. However, considering studies that suggest ~90% of secondary metabolite biosynthetic capacity is silent or cryptic [33], the estimated endophyte-derived secondary metabolites on Earth might total 1.045×10^9 , or a billion potential endophyte secondary metabolites.

Model-fitting: Estimates of endophyte species richness and NP potential could incorporate OTU data (e.g. see Fig. 1) into models of species discovery or species accumulation curves. These can be based on number of leaves sampled for endophytes [34] or published new species or OTUs [35]. Alternatively, endophyte OTU data can be estimated using frequency counts, rank species abundance distributions, or Poisson lognormal (log-log) fitting approaches and scaling laws [32, 36–40]. In the latter case, it has been argued that microbes in microbiomes closely fit the pattern where S (number of species) scales with N (number of individuals) where commonness (resampling) is constrained by scaling N^z where $5 \sim N^z$ and $0.25 \leq z \leq 0.5$ (and for microbes $z = 0.38$ while for macroscopic organisms $z = 0.24$), and globally N_{max} (number of individuals of the most abundant species) = $0.38 * N^{0.93}$, $r^2 = 0.90$ [32]. Empirically, results scale at $S = 7.6 * N^{0.35}$, $r^2 = 0.38$. For endophytes, using estimated values of 10^4 to 10^8 endophytic bacterial cells per g of plant [41] plus ~10–100 fungal individuals per g, and an estimate of total Earth plant carbon (C) of 450 Gt [42] and assuming 0.43 g of C per 1 g plant matter [43], we estimate Earth's bacterial endophyte individuals, N at 1.044×10^{22} to 10^{26} , which with scaling laws results in an estimate of global endophytic bacteria species, S between 386 million and 9.7 billion and S for global endophytic fungal species between 34 and 77 million. These values produce not unreasonable estimates of numbers of microbial species per plant species, within the range of values summarized based on OTUs in Fig. 1 (i.e. for bacteria, 386 to 9700 million species divided by 300,000 plants = 1290 to 32,300 bacterial endophyte species per plant species – for example, similar OTU estimates in [44]; and for fungi 34 to 77 million species divided by 300,000 plant species = 113 to 257 fungal endophyte species per plant species). Extending the idea of endophyte secondary metabolite uniqueness per species-group as discussed above [11], there may be an estimated 124 million to 3.1 billion bacterial and 22 million to 50 million fungal secondary metabolites that could be expressed in cultures, and considering additional cryptic expression [33], up to 1.3 to 28.3×10^9 potential endophyte secondary metabolites to be discovered.

Table 2 Estimating global plant microbiome holometabolomes using combinatorics

Table 1 considered endophytes independently, ignoring plant-microbe and microbe-microbe cooperation in NP synthesis. Here, we estimate secondary metabolism that may be more than the sum of its parts through *in planta* multi-species co-synthesis, syntropy, and synergistic biosynthetic pathways. We first estimate the global number of endophyte communities, then these communities' additive and synergistic secondary holometabolomes

Estimating the number of distinct endophyte communities on Earth:

While the theoretical number of possible combinations of endophyte species within a plant would be 2^n for n endophytes, so that a plant that can host 500 species of endophytes would have 2^{500} possible endophyte communities, real life does not include all possible combinations. Instead, if we calculate the possible unique endophyte sets of size 500 from amongst a set of m unique microbes, we can use the binomial coefficient and take m choose k , with $k=500$, then solve for m based on the number of plant species or plant individuals on Earth. Given that m choose $k = m! / (k!(m-k)!)$, and assuming a limit of 300,000 combinations (one per plant species) we calculate m is between 502 and 503, meaning that the effective set of unique microbes per plant species would be just 2 or 3 under this endophyte set size of 500 per plant. Note: based on Table 1, we estimated from 113 to 257 fungi per plant species and from 1290 to 32,300 bacteria per plant species, which translates to effectively just over 3 unique fungi per plant species, and no unique bacteria per plant species. If plant individuals rather than plant species are better units for analysis, given the observed variance in endophytes across plant geographic ranges and predicted horizontal exchange of endophytes and we estimate an average of 50,000,000 plant individuals per species, there could be 15 trillion endophyte combinations on Earth, with as many as 6 to 7 unique endophytic fungi or over 3 unique endophytic bacteria per plant.

Estimating endophytic holometabolomes on Earth: Several studies have demonstrated the importance of plant-microbe and microbe-microbe shared metabolic pathways involving intermediate metabolite provision or positive regulatory cues [45]. Plant species diversity is positively associated with bacterial and fungal diversity and metabolism [25, 46], but how much of this is merely additive versus synergistic or cooperative? Results from OSMAC and co-culturing studies that show perhaps 90% of endophyte secondary metabolites depend on *in planta* conditions [28, 47, 48] pointing to *in planta* metabolic synergism. Combinatoric models have been helpful for exploring metabolism and biochemical space [49, 50], but these have seldom been applied to understanding synergies between microbes [51, 52]. Biosynthesis of secondary metabolites is particularly amenable to combinatoric synergy because it functions modularly through extending polymer backbones: -CH₂-(C=O)- units for polyketides, C₅ isoprene units for terpenoids, and non-ribosomal peptides, that later generate diverse chemicals with the assistance of tailoring enzymes. Recent combinatoric experiments on simple microbiomes also indicate that higher-order interactions in which each species impacts interactions among other species, at least for 2-way and 3-way interactions are widespread in a 5-species microbiome [52]. Here, we will apply a similarly small interaction network for endophytes and consider the holometabolome of a small localized section of plant tissue containing 5 interacting organisms: the plant, two fungal species and two bacterial species, that can interact and cooperate molecularly at close range. In this example, 2^n interactions could occur for n species. If each pair of these 5 species participates in one biosynthetic synergy leading to an additional metabolite, there would be n choose $2 = n! / 2!(n-2)! = 10$ unique metabolites arising synergistically for this interaction. Within a plant species, if a portion of its endophytes (e.g. 100) participated in these synergies with each of the 10 novel products synthesized once, $100 / 5 = 20$ such products would be generated, adding 6 million new secondary metabolites to the global tallies discussed in Table 1, or, at the upper limits, considering individual plants to have distinct communities and synergies, there could be 300 trillion unique *in planta* synergistic products on Earth. However, this value likely includes extensive redundancy, which is difficult to estimate without further empirical data, or models such as the deep learning models described in Table 3.

breakdowns in signaling that lead to endophyte secondary metabolite silencing and upcoming breakthrough methods including deep learning. We describe recent progress in identifying hidden endophyte NPs through heterologous expression experiments [81], methods of unsilencing genes in endophytes [82] especially including co-culturing and condition-modification [28, 83]. We then highlight breakthrough approaches and strategies needing more attention, including systems biology methods [84, 85] integrated with big data mining and deep learning [56] from an *in planta* perspective. Specifically, we illuminate recent breakthroughs in artificial intelligence-based methodologies; particularly deep learning applied to multiple phases of the discovery pipeline and multi-omics *in planta*. We will finish by outlining a new, integrated pipeline – a systematic, interdisciplinary approach using computational learning – that promises to “wake the sleeping giant” of endophyte NPs.

How much promise do endophytic microbiomes hold for natural product discovery?

Plant microbiomes may be one of the most promising and underdeveloped groups of organisms for natural product discovery, due to their long-evolved intimate interactions serving in chemical defense of plants [86–88]. For example, studies thus far on phyllosphere (i.e. above-ground microbiota) and root-associated microbiota have shown that endophytes provide bioactive secondary metabolites with unique structures such as Fusaric acid A & B, Pestalactams A & B, and polysaccharide DG2 [89–92]. But could they hold more promise for NPs than free-living microbes, as has been suggested [20]? This rhetorical question has practical importance: if endophytes do not hold exceptional promise as a source for novel NPs, it is pointless to invest exceptional effort to overcome the inherent challenges of their low culturability and high levels of silent gene clusters [93–95].

Answering this question requires consideration of how endophytic microbiota are distinct as a group. Once established in plant tissues, microbiome endophytes, in contrast to pathogens, can no longer increase their fitness by increasing biomass beyond the limited plant tissue growth, and instead can increase their fitness by switching their investment to benefits for the plant through increasing plant growth and synthesizing additional defense compounds [48, 84, 96, 97]. Plants and their microbiomes are distinctly limited in their options for escaping hostile interactions by means other than chemical innovation. Hence, endophytes show increased investments in defense roles, such as antiherbivory and antiviral activity, compared with free-living microbes [98, 99] ultimately showing enhanced directional or positive selection on defense compounds [87, 100],

whereas within the confines of plant tissues their biomass investment is downregulated by the plant [101]. Furthermore, endophytes that proliferate mainly (or solely) within hosts will have enhanced drift or bottleneck and accelerated evolution [102–104] enhanced by phases of high local or vertical transmission [2, 15, 105, 106]. In addition, long-term interactions within plants likely places evolutionary pressure specifically at the level of molecule-to-molecule interactions and pathway-to-pathway interactions, enhanced by the large and complex plant genome [104]. For example, some endophytic fungi produce plant hormones (gibberellins and indoleacetic acid) to promote host plant growth [97], and others synthesize plant-like defense compounds [101], famously including Taxol. For long-associated plant microbiome consortia, primary metabolism may decay, while secondary metabolism may be enhanced, sometimes on supernumerary chromosomes [107] or defense plasmids [108]. Thus, these distinct conditions in which endophytes have evolved should increase their secondary metabolite diversity. If so, why then do past surveys [109] suggest only ~5% of current medically relevant compounds are from endophytes? We explore answers to this question below, especially under-cataloging due to a focus on culture-based methods rather than analysis of the plant microbiome *in situ* or *in planta*.

Hyperdiversity and its effects on holobiont metabolism *in planta*

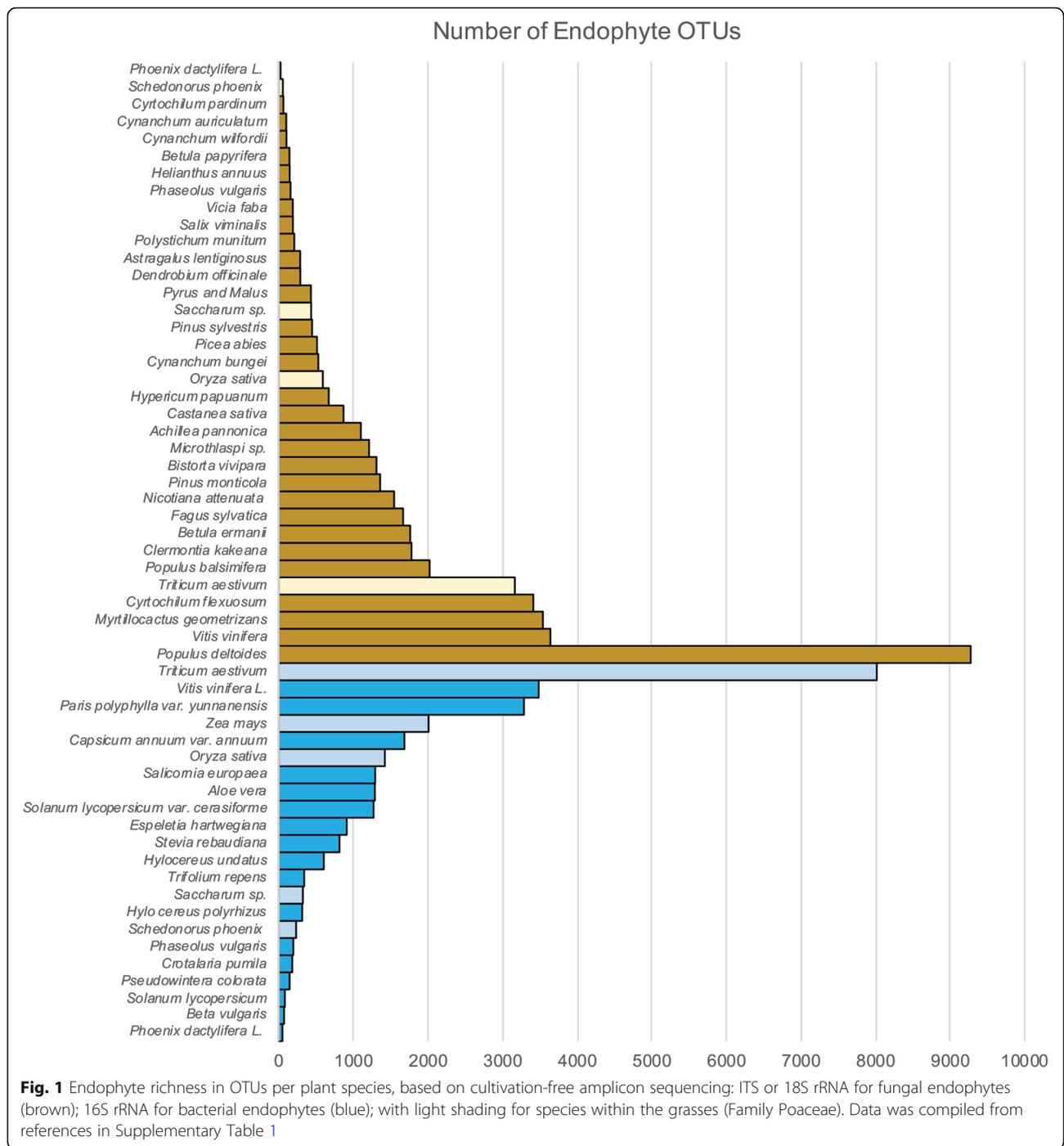
Estimating the taxonomic and functional diversity of plant microbiomes is critical because species and strain diversity are believed to predict secondary metabolite diversity [110, 111]. To date, we lack a systematic census of global plant microbiome secondary metabolite diversity. A recent meta-analysis suggests complex evolutionary and ecological forces may influence the endophyte assemblages [112] and another recent study suggests adaptive matching drives diversification of plants and endophytes [104]. Therefore, in this section we illuminate key empirical studies showing the hyperdiversity of fungal, bacterial, and viral inhabitants of plants (Fig. 1) and present a new estimate of global endophyte diversity (also see Table 1).

Endophytic fungi are ubiquitous and hyperdiverse

Fungi appear to be the dominant microbial inhabitants, in terms of culturable biomass, in plants [113], and hence, likely the most prolific sources of endophyte NPs. Evidence of fungi in fossilized tissues of plants from ~460 million years ago may explain why fungi have diversified to all plants in all habitats studied to date [21]. Reports describing endophytic fungi in the tropics as “hyperdiverse” [25] have raised much interest in drug discovery. For example a seminal culture-based survey

showed 418 endophyte morphospecies (~347 genetically distinct taxa) isolated from 83 healthy leaves of just two plants, *Heisteria concinna* and *Ouratea lucens*, in a tropical forest [25]. Despite these and other surveys [112], most of the world’s fungal endophyte taxonomic diversity – and therefore NP diversity – is uncharted. Clearly, fungal diversity estimates are wide-ranging and depend on census approach: culture-based studies suggest there may be ~5 to ~350 fungal endophyte species per plant, while culture-free amplicon-based deep sequencing based approaches, focused on 18S or ITS rRNA genes, suggest there may be ~40 to 1200 fungal endophyte species per plant (see references in Fig. 1).

Species counts alone do not estimate functional or metabolic diversity; specific fungal endophyte clades differ in roles, and therefore biosynthetic capacity. For example, fungal associations can be foliar, systemic, or root-limited and will differ in roles accordingly. Taxonomically, most endophytes fall into the non-balansiaceous group (non-grass endophytes), which include diverse hyphae-forming Ascomycota (the dominant phylum of fungal endophytes), Basidiomycota, and Glomeromycota [114]. Many of the common genera, such as *Acremonium*, *Alternaria*, *Cladosporium*, *Coniothyrium*, *Epicoccum*, *Fusarium*, *Geniculosporium*, *Phoma*, and *Pleospora* are ubiquitous [115] with some groups dominating in the tropics (Xylariaceae, *Colletotrichum*, *Phyllosticta*, and *Pestalotiopsis*) and others common to both tropical and temperate climates (e.g. *Fusarium*, *Phomopsis*, and *Phoma*) [115, 116]. Biosynthetic capacity relevant to natural product discovery appears to be distributed broadly across these fungi. For example, a study of endophytic fungi with antitumor activity showed dominance of Ascomycota (96%), but broad taxonomic distribution within this group, and others such as Basidiomycota (3%) and Glomeromycota (1%) [117]. The genera identified as antitumor compound-producing are broad (e.g. including *Pestalotiopsis*, *Aspergillus*, *Chaetomium*, *Fusarium*, *Penicillium*, *Alternaria*, *Phomopsis*, *Acremonium*, *Ceriporia*, *Colletotrichum*, *Cytospora*, *Emericella*, *Eurotium*, *Eutypella*, *Guignardia*, *Hypocrea*, *Periconia*, *Stemphylium*, *Talaromyces*, *Thielavia* and *Xylaria*) [117]. In contrast, Balansiaceous endophytes (or grass endophytes) are narrower taxonomically and include clavicipitaceous genera *Epichloë* and *Balansia*, with their anamorphs *Neotyphodium* and *Ephelis* predominating. Balansiaceous endophytes are notable for their vertical transmission with seeds and production of anti-insect alkaloids peramine and lolines, and the anti-vertebrate alkaloids lolitrem B and ergovaline [118]. In preparing this review, we found no comparative analysis of the classes of secondary metabolites or natural products grouped with endophyte tissue- or taxon-class, but presumably such patterns do exist.



There has not been a comprehensive model to estimate the diversity or richness of endophytic fungi, but an often cited calculation suggested there are 2–4 unique endophytic fungi per plant, which would suggest there are ~1 million species of endophytic fungi on Earth, based on an estimated 270,000 plant species [11]. However, these estimates predate next generation sequencing studies [119–122], and likely suffer from bias against non-culturable taxa. Thus, we have attempted to synthesize some of the recent

sequence-based data on endophytic fungal diversity within plants at a taxonomic level most relevant for NP discovery (i.e. strain-level), integrating established models (e.g. Poisson lognormal) in Table 1. These provisional calculations suggest far more diversity than previous estimates, with possibly 34 to 77 million endophytic fungal species and 10 to 20-fold more strains on Earth with capacity to synthesize 22 to 50 million biosynthetic gene clusters (BGCs) based on pangenome-level BGC analysis.

Endophytic bacteria are also ubiquitous and hyperdiverse

Bacteria are the other dominant and diverse microbes associated with plants, providing additional metabolic and biosynthetic capacity. Recent reviews have presented endophytic actinobacterial secondary metabolites in depth and described key interactions and metabolites in this group [6, 123]. Taxonomic profiling studies have tended to focus on crops, fruits and vegetables [124–126], or forest tree foliar endophytes [127] and cold adapted plants [122]. Nevertheless, endophytic bacteria are poorly known, despite the fact that bacteria are the most speciose and metabolically diverse domain of life, with perhaps 1 trillion species [32]. Bacterial endophyte diversity may be far more under-cataloged than endophytic fungal diversity due to the small size, low biomass, less clear ecological roles. However, some studies suggest bacteria are ubiquitous, colonizing all parts of plants as inter- and intra-cellular endophytic bacteria living in roots, stems, shoots/leaves, and vascular tissues [41, 128–131], as well as foliar epiphytes on leaf surfaces [132–134], rhizosphere associates on root surfaces and the more well-studied nodule-forming root endophytes (e.g. rhizobia in legumes) [135–137]. While endophytic bacterial diversity can be extremely high (e.g. 31,952 OTUs at 97% similarity) [44], typically, the number of distinct bacteria per plant ranges from 10 to 200 for culture-based studies and from 20 to 600 from amplicon sequencing-based studies (see references in Fig. 1). While no current models exist to estimate bacterial endophyte diversity, based on extant 16S rRNA surveys of bacterial endophytes and the framework used above for fungi, we estimate there may be perhaps 386 to 9700 million bacterial endophyte species on Earth, with perhaps 124 to 3.1 billion biosynthetic gene clusters (Table 1).

Endohyphal bacteria may enrich endophytic fungal diversity and metabolite synthesis

Endohyphal bacteria (EHB) live within free-living and endophytic fungi, adding to their biosynthetic capacity, function and regulatory complexity [62, 63, 67, 138]. Far from being rare, EHB appear to be widespread [64], potentially protecting the plant and endophytic fungi from pathogens [65] and interacting with plant hormones [66]. EHB have been described as the prokaryotic modulators of host fungal biology in hyphae of endophytes in many plant tissues and across many plant lineages [139, 140]. This endosymbiotic association was first detected inside the mycelium of mycorrhizal fungi wherein mycorrhiza helper bacteria were associated with the fungal nutrition transport [62]. A remarkable example is the ectomycorrhizal fungus, *Amanita muscaria*, and a mycorrhiza helper bacterium, *Streptomyces* strain AcH 505. Strain AcH 505 produces both fungal growth-stimulating compounds (e.g. auxofuran) and compounds

that suppress plant-pathogenic fungi, and alters gene expression in *A. muscaria* [63]. In some cases, EHBs may enhance stress tolerance of plant and fungus, production of phytotoxins and regulation of host reproductive machinery [61], influence the ecology of plant endophytes [64], or confer other types of protection to the host fungus or plant [65]. Although these bacteria play important roles in modulating the secondary metabolism of their host fungi, this is still poorly understood.

Viruses of plants and endophytes impact the holobiont metabolism

Viruses are widespread and diverse pathogens of plants, fungi, and bacteria and can impact their host populations and alter host SM biosynthesis [141–144]. Hypovirulent viruses and phage are of special interest for potentially serving to regularly unsilence NP clusters [145–147]. We consider three important types of viruses: (1) mycoviruses, i.e. viruses that infect fungi and show low virulence; (2) bacteriophage of endophytic bacteria and endohyphal bacteria; and (3) latent plant viruses. Mycoviruses are diverse and classified into seven families of double-strand RNA (dsRNA), single-strand RNA (ssRNA) and single-strand DNA (ssDNA) [70, 141, 148]. These hypovirulent mycoviruses have been diagnosed from all classes of endophytic fungi [142]. However, mycovirus diversity and host-specificity is still poorly understood, and the role of mycoviruses is poorly understood. For example, mycoviruses in the endophytes of *Ambrosia psilostachya* and its parasite *Cuscuta cuspidata* were shared between different fungi [149] suggesting they might not be specific to a single fungal taxon. In contrast, endohyphal viruses of related endophytes of Pine, *Diplodia scrobiculata* and *D. pinea* and appear not to be related [150]. Nevertheless, mycovirus species richness appears to be vast, with viruses identified in over 30–80% of fungal species [70]. Specialized mycoviruses that may impact fungus-plant interactions. A notable example is the fungal endophyte *Curvularia protuberante* of the tropical panic grass *Dichanthelium lanuginosum* in which its mycovirus allows the plant to grow at high soil temperature [68].

Bacterial viruses, or bacteriophage (phage), are hyperdiverse with perhaps 10 or more estimated unique phage per species of bacteria [151–153]. However, little is known about of phage that specialize on endophytic bacteria. Nevertheless, they almost certainly affect endophytic and endohyphal bacterial fitness, population dynamics, and aspects of secondary metabolite production that involve these bacteria.

Plant viruses, especially latent or persistent plant viruses that remain asymptomatic for extended periods of time, including Endornaviridae, Partitiviridae, and Luteoviridae, are diverse and ubiquitous [154–157].

Numerous studies suggest that together, plant viruses may impact plant resistance to infectious and beneficial bacteria and fungi, and may impact plant interactions with and colonization by endophytes [154–157]. Detailed studies of the impacts of plant viruses on plant secondary metabolism [158, 159] suggest ways in which the plant holobiont (including its resident endophytes) may shift gene expression, proteome, and metabolome, resulting in altered holobiont NP profile [155].

Are plant microbiome communities greater than the sum of their parts?

Much of secondary metabolism in cells contributes to the “holometabolome” (i.e. the net metabolome of the holobiont) additively. However, many studies suggest that *in planta* endophyte community interactions and regulatory cross-talk (see recent review [140]) that may influence secondary metabolite synthesis [45, 160–162]. Some of these major interactions within plants, such as plant-endophyte, fungi-fungi, fungi-bacteria, fungi-EHB, fungi-mycovirus, bacteria-phage, and miRNA and small-molecule signals, are shown in Fig. 2. Several studies suggest a portion of the holometabolome may arise through provisioning of substrates, such that secondary metabolism is not merely additive, but instead is greater than the sum of its parts. For example, endophytes may metabolize secondary compounds from the host, or the host and endophyte may share parts of a specific pathway – although this is not well-known [161]. One example of this is the putative combined synthesis of cardiotoxin by endophytic *Burkholderia* spp. and plants [123, 163, 164]. Generally, most evidence for cooperative exchange comes from laboratory co-cultivation studies, suggesting fungi-fungi and bacteria-fungi interactions

may impact SM production [165, 166]. Indeed, it is the rule, rather than the exception in microbial communities that multiple species may exchange a plethora metabolites – hence, classical models of inter-species metabolite exchange [167]. There has been speculation about the role of horizontal gene transfer as a key factor in the apparent convergence of endophyte and plant metabolites [168], but to date, this question has not been thoroughly examined. Co-regulation of independently evolved BGC homologs in plants and their microbes has also been described [169], but remains poorly understood. Secondly, endophytes may prime the host plant’s defense via ethylene-jasmonic acid transduction, mediators of biotic and abiotic stresses and ROS, modulating plant receptors for chitin and flagellin [61, 140], although this is better known for plant-pathogens than endophytes and similar studies for mutualistic endophytes are lacking.

Empirical and theoretical analysis of endophyte taxonomic and functional diversity should inform bioprospecting strategies and be particularly helpful for identifying novel *in planta* communities that might produce novel natural products. However, few studies have examined this. One study estimated at least one unique endophyte community per plant species [2]. We re-estimate this in Table 2 using a combinatoric approach and suggest there may a range of 1 community per plant species to 1 community per plant individual or 300,000 to 15 trillion combinations on Earth. To evaluate global *holometabolome diversity*, we considered both the sum of endophyte metabololic potential alone and estimated possibly 1.3 to 28.3×10^9 metabolites (Table 1) and then we additional synergistic metabolism by considering only subcommunities within plants, and estimated these could add between 6 million to 300 trillion unique *in*

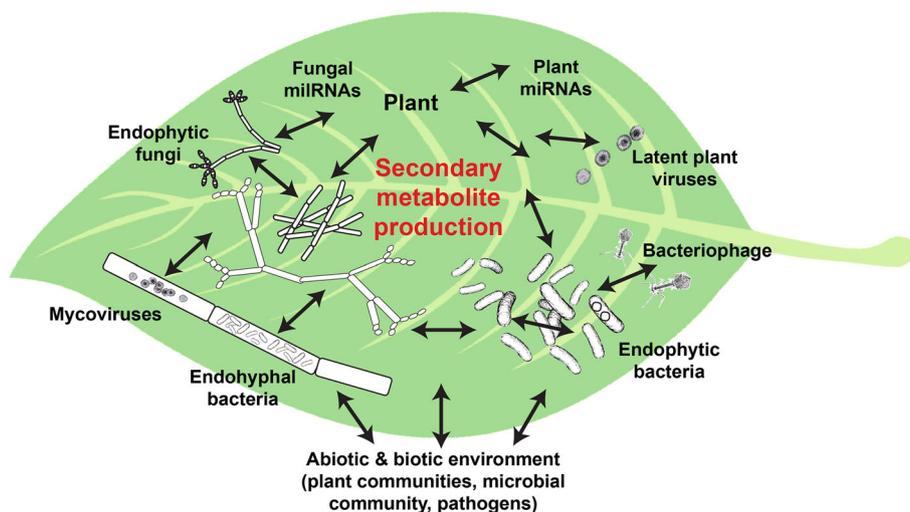


Fig. 2 Schematic of the plant microbiome showing *in planta* interactions leading to multipartite biosynthesis and regulation of endophyte-plant (holobiont) secondary metabolites

planta synergistic products on Earth (see Table 2). Co-regulation and downregulation will arguably reduce the biosynthesis overserved at any time, so these estimates would reflect long-term capacity under a variety of environmental conditions and triggers.

Chemical diversity in the plant microbiome: a universe of natural products

Compounds from endophytic consortia likely traverse the sphere of possible natural products. Chemical diversity, or chemical space (all molecules that might exist) has been estimated theoretically at $>10^{60}$ small compounds <500 Da. Natural products occupy a part of this theoretical space, mostly falling into four categories of secondary metabolites (alkaloids, terpenoids, phenylpropanoids, and polyketides). Current curated natural compound databases such as the Dictionary of Natural Products and Super Natural II [170], which include over $>325,000$ natural compounds with only perhaps about 5 to 10% of known bioactive products come from microbes [13, 171] with perhaps half from Actinomycete bacteria (particularly *Streptomyces*), and a growing proportion from fungi, but only a few chemical compounds recognized from endophytes. From 2014 to 2017, a total of 224 novel compounds were recognized from endophytic fungi [73]. Estimates of all possible undiscovered natural compounds on Earth could range from near the current asymptote of discovery (i.e. with only 25,000 more to be discovered) [172] up to one per undiscovered microbe [173], which, with 99.999% of Earth's microbes undiscovered [32], might yield 5000 to 2 million novel NP-derived drug candidates. But drug chemical space is much smaller than natural product space due to the limitations of oral administration and pharmacokinetics – following Lipinski's rule of five. Conversely, despite known natural products being a tiny portion of all theoretical compounds, they contribute more than half of FDA approved drugs likely because evolutionary forces promote natural compounds with specific bioactivities.

However, the curve of natural product discovery appears to be leveling off [172]. Arguably, one reason for the leveling is that we have reached the limits in methodology and screening approaches that focus mostly on the small proportion of microbes that can be easily cultured under laboratory conditions. For example, analyses of secondary metabolite libraries suggest that while we have reached some limits in examining planar compounds (2-dimensional or sp^2 -hybridized double bond-rich) that are effective in interacting with similar targets (e.g. kinases), we have under-examined the richer drug potential of diverse 3-dimensional compounds (e.g. those with fewer aromatic rings and more sp^3 -hybridized single bond carbons with higher stereochemical center diversity) that will in theory have vastly greater target

richness (e.g. protein-protein or transcription factor) [173]. Some of these may be expressed only under special conditions. Indeed, genome analysis has uncovered universal microbial processes to down-regulate or silence biosynthetic gene clusters [174]. In fact, genome mining studies suggest 92–96% of fungal secondary metabolite biosynthesis is routinely turned off [175, 176] through epigenetic regulators and absence of triggers from other organisms [177], presumably to reduce energetic costs during times when the products do not add to fitness. Furthermore, as argued in Table 2, chemical complexity may depend on community interactions that transform compounds [3], sometimes through enzymes or shunt metabolites (e.g. acetyl-CoA, shikimic acid, mevalonic acid, 1-deoxyxylulose-5-phosphate, in alkylation, decarboxylation, aldol, or Schiff base formation) [178], via natural biotransformation or bioconversion. Even Taxol biosynthesis seems to depend on microbe-microbe, microbe-plant, and abiotic factors [179, 180]. Cooperative biosynthesis has been described extensively in microbe and microbe-host systems [71, 181, 182]. Several studies suggest endophytes can in some cases can directly synthesize plant-like metabolites [183].

Studies of bioactive compounds from fungal endophytes of leaves and roots [184–187] show that while only a few strains have been extensively studied, typically each has several novel compounds (e.g. Li et al. 2018 reviewed 224 compounds from 109 endophyte strains). The taxonomic distribution of fungal endophyte derived chemical compound synthesis is dominated by Ascomycota (~97%) (with classes Sordariomycetes ~40%, Dothideomycetes ~31%, Eurotiomycetes ~24%, include notable pathogens as well as endophytes), with some Pezizomycetes and Agaricomycetes, and also Basidiomycota (~2%), and Mucoromycota (~1%) with the most richly represented compound-producing strains belonging to *Aspergillus*, *Penicillium*, *Pestalotiopsis*, followed by *Fusarium*, *Phomopsis*, and *Alternaria* [73, 117]. Notably, 5 of 14 strains of *Pestalotiopsis* produce the cancer drug Taxol. Similarly, recent studies of anti-cancer compounds isolated from endophytic fungi showed novel alkaloids and nitrogen-containing heterocycles (>27 new compounds including penicisulfuranols, penochalasin, aspergillines, etc.), polyketides (>25 new compounds including phomones, rhytidchromones, allahabadolactones, etc.), terpenoids and steroids (>18 new compounds including rhizovarins integracides, etc.), quinones, phenylpropanoids, and esters (>20 new compounds including versicoumarins, versicolols, pestalotrioprolides, etc.), and other classes of compound (>35 new compounds including muroxanthenones, etc.) [73]. Another review showed compounds from endophytic fungi of similar taxonomic breadth having potentially activity against neglected tropical diseases

(including compounds Citrinin, palmarumycins, Cochlioquinone, Grandisin, Altenusin, Pullularins, Pestalactams, Viridiol, Phomoarcherins, etc.) [188]. Further reviews have highlighted the wide array of therapeutics isolated from endophytes that mimic therapeutic plant-derived secondary metabolites, e.g. antioxidants (Lapachol, Cajanin stillbene acid, Resveratrol, Rutin, Phillyrin), antihypercholesteremics (Rosuvastatin, Piperin, Chartaractams, Phenylspirodrimanones, Lovastatin), antidiabetics (2,6-di-tert-butyl-p-cresol, Berberine, Cajanol, Aspergillus A, Rohitukine, Helvolic acid), and further compounds identical to plant-derived anticancer compounds (Taxol, Hypericin, Vincristine, Vinblastine, Camptothecin, Podophyllotoxin, Kaempferol, Azadirachtin, Rohitukine) [189–191] possibly as an ecological survival strategy [168]. In a few cases, research shows endophytic compounds to be exceedingly rare, yet especially useful medically, such as the unique mellein compounds of *Aspergillus flocculus* (Tawfike et al., 2019). From 2010 to 2017, 65 metabolites from endophytic fungi were identified as antimicrobial and anticancer agents with unique compounds such as Solamargine (alkaloid), Piperine (alkaloid), Cajanol (flavonoide), Vinblastin and vincristine (alkaloids), Forskolin (alkaloid), Homoharringtonine (alkaloid), Chrysin (flavonoid), and have antimicrobial and anticancer activities [84, 191–193].

Amongst bacterial endophytes, Actinomycete bacteria have been studied extensively, especially *Streptomyces*, *Micromonospora*, *Polymorphospora*, *Jishengella*, and *Actinoallomurus* which produce many remarkable bioactive compounds including highly modified alkaloids (diketopiperazines, lansai, spoxazomicins, dihydrooxazole alkaloids, spoxazomicins, pyrazine), peptides (such as cyclotetrapeptides), a wide array of polyketides (such as glycosylated and prenylated antibiotic coumarins, butyrolactone antibiotics, cedarmycins, pteridic acids, clethramycin, efomycin M, salaceyins, lorneic acid, stipitatic acid, secocycloheximides, maklamicin, linfuranones, germicidin, actinoalloses, alnumycin, lupinacidins), terpenoids (such as kandenols), and mixed synthesis metabolites (such as indolosesquiterpenes, xiamycin B, indosespene, sespenine, celastramycin, and trehangelins) [171].

Together, these studies show an increasing universe of natural products with novel bioactivities compounds from fungal and bacterial endophytes, even in the absence of *in planta* inputs such as precursors and regulatory molecules, or environmental cues. It remains unclear if this universe will continue to expand, or if the predictions in Table 2 will ever be realized, but we argue the primary challenge will be harnessing new potential from the vast unculturable majority of microbes.

Isolation is the problem

Isolating and culturing plant microbiome species to uncover their biosynthetic capacity is a poor strategy for

two reasons; first, most endophytes cannot be grown in culture, and second, most endophytes will not express many secondary metabolites outside the host plant tissue or environmental niche. The apparent failure of culturing for most microbiota within plants makes sense given the long association of these organisms and the widespread tendency of symbionts to lose the capacity for traits needed to live outside the host, due to relaxed purifying selection on those traits. Studies on the fungal endophytes that can be easily cultivated suggest taxa and their secondary compounds are tissue- and organ-specific, and seasonally, and geographically variable [15]. This pattern is likely mirrored by the even more host-adapted non-cultivable endophytic fungi and bacteria, and likely translates to further hidden biosynthetic diversity. For example, one study showed high NP diversity from non-cultured 3409 endophytic bacteria, but only 1.6% of the identified BGC clusters matched any known BGC [194]. The new era of advanced sequencing and computation discussed in this review should result in a sharp rise in discoveries for these difficult-to-culture microbes. However, traditionally, culturing has been required to confirm and analyze natural compounds. This problem is one of the major breakdowns in the NP discovery pipeline: breakdown of microbe-host molecular exchanges makes plant microbiomes difficult to study.

Endophyte NP diversity is under-cataloged, even for culturable species, presumably because culturing methods fail to adequately supply *in planta* molecular signals required to unsilence BGCs [14, 195–201]. This observation derives from sequencing studies and metabologenomic analyses showing evidence of BGCs for products that are not detected in cultures. As a primary example, polyketide synthases (PKSs) and nonribosomal peptide synthetases (NRPSs), which are multifunctional enzyme systems that assemble many of the secondary metabolites from simple building blocks including carboxylic acids and amino acids [202, 203], show limited expression under laboratory conditions [204]. Extensive efforts have been made to unsilence such clusters [205, 206]. Most genetic manipulation methods attempting to control PKSs and NRPSs as multifunctional enzymes to regulate expression of BGCs rely on multi-target approaches not specific to a single secondary metabolite and display complex interactions.

In fungi, control is often regulated by chromatin-based mechanisms and histone acetyltransferases, deacetylases, methyltransferases, and proteins involved in heterochromatin formation [207, 208], thus, modifying the chromatin landscape through chemical modifiers can regulate secondary metabolite synthesis [111]. Specifically, many putative silent BGCs are located in the distal regions of the chromosomes in the heterochromatin which is controlled by epigenetic regulation [209]. However, these modifications can lead to unpredictable changes in

expression of other genes [111]. This is true for the fungal blight pathogen, *Fusarium graminearum*, where increasing the expression of the heterochromatin protein homolog (HEP1) which plays an important role in the production of secondary metabolites. HEP1 influences expression of genes of aurofusarin with antibacterial/toxicological effects [210]. Other attempts at changing chromatin do not always unsilence cryptic fungal BGCs, since most secondary metabolite gene clusters remain silent by these approaches [211]. Many methods that include pleiotropic and pathway-specific approaches have had similarly limited effectiveness. For example, small-molecule elicitors released from plant hosts may affect endophyte SM transcription, many studies of endophytes grown outside plant tissues have used epigenetic modulators to attempt to activate the silent BGCs [212], with inconsistent results. Small molecule epigenetic regulators and in different expression-type strains of different PKS reduction states stimulated a variety of alternative VOCs [213], while heterologous expression experiments [81] and other unsilencing approaches [82, 214] have had mixed success.

In planta studies of the plant microbiome in situ, in contrast to studies of cultured endophytes, have revealed that broad gene expression derives from integrated, dynamic components of the plant-endophyte holobiont [215]. This integration of gene expression regulation may be ~ 460 million years old [21, 22], enough time for the evolution of cooperative synthesis of compounds and precursor supply (or regulation of degradation of precursors for secondary metabolism) [72], with the help of neighbors, such as the plant, other endophytic fungi and bacteria [61, 142]. Thus, breakdowns between endophyte and host metabolism, precursor supply, and signaling may drive biosynthetic gene clusters to be silenced as they are studied in culture. For example, studies show that endohyphal bacteria such as members of the Enterobacteriaceae, which may impact fungal gene expression [61–67], may diminish or change during culturing [216]. Clearly, expression of BGCs can be context-dependent. Even simple variations in the growth medium such as pH, temperature, aeration, and light can change the level of transcription of BGCs [217]. This point is evident from co-cultivation experiments that provide interspecies signals for SM synthesis [218], and in vitro multi-endophyte array experiments [191]. In many studies, co-cultivation of endophytic fungi with their plant hosts led to the activation of formerly silent gene clusters [219]. Another missing signal in cultured endophytes may be small RNAs. These have been observed to transmit bidirectionally [220] as a mode of trans-kingdom cross-talk [221, 222] and may transcriptionally activate silent clusters or regulate translation in response to infection [223]. Indeed, fungi encode

microRNA-like small RNAs (miRNAs) that may interact with other regulatory elements and affect transcription and post-transcriptional changes [224, 225]. Furthermore, miRNAs triggered by pathogens could unsilence endophyte fungi or unsilence plant signals directed at endophytes, that turn on genes for SMs. Some remarkable small RNAs in bacteria may impact hosts, and miRNAs from hosts may pass into endophytic bacterial cells and regulate their expression [223].

But why should endophyte BGCs be silenced during growth in culture? And why should plants down-regulate endophyte SM production except under specific conditions? The proximal cause of silencing in culture may be simple lack of signals or precursors, however, the ultimate evolutionary cause may be the need to redirect energy to growth [204]. Long-evolved intimate partners often chemically stabilize and control their interactions with neighboring organisms to coordinate or regulate growth [200] conserve energy and maintain the novel benefits of symbiosis.

Past and current solutions to discover NPs from plant microbiomes

Approaches focused on cultivatable endophytes

Standard pipelines for endophyte NP discovery are powerful, but usually low-throughput [29]. Historically, prior to next generation sequencing, methods for discovering endophyte-derived natural products would involve (1) field surveys to extract plant tissues, (2) endophyte (bacterial or fungal) culturing (e.g. for fungal endophyte culturing, see [188]), (3) extraction and separation of compounds for analysis, (4) chemical analysis and dereplication using any of many classical techniques such as UV spectroscopy, infrared spectroscopy, mass spectrometry (MS), and nuclear magnetic resonance spectroscopy (NMR) or more modern “on-line” hyphenated (i.e. coupled) approaches such as HPLC-NMR-MS (see [178]), (5) and finally bioactivity assays and testing on cells/animals. To speed up drug discovery, the search for natural product extracts was largely supplemented from the 1990s onward with synthetic combinatorial chemistry approaches which create large compound libraries that can be tested using automated high throughput screening (HTS). However, this approach has proven to have limitations [178].

Simultaneously, some of the limitations of natural product discovery have been overcome by increasingly sophisticated standard methods. Key methods in use are pleiotropic approaches such as “One Strain – Many Compounds” (OSMAC), chromatin remodeling, ribosome engineering, or targeting global regulatory genes or phosphotransferases, approaches that are specific to BGCs such as heterologous expression, promoter exchange, refactoring, and cluster-situated

regulators, and genome-wide targeting by reporter-guided mutant selection and elicitors [226]. The OSMAC approach, which centers on testing each isolated strain grown under a systematic array of culture conditions to increase the diversity of secondary metabolites produced has been one of the most effective NP discovery methods for culturable endophytes [28, 83]. In OSMAC, common modifications include high phosphate, modified media richness, pH value, temperature, salinity, metal ions, oxygen/aeration, or with addition of enzyme inhibitors [83, 227], or using UV mutagenesis, or with addition of plant or microbial extracts or cells or under co-cultivation, or affixed to various surfaces (i.e. as biofilms), or epigenetic modifiers (e.g. DNA methyltransferase inhibitor, histone deacetylase inhibitor, biosynthetic precursors). OSMAC's promise as a method ultimately derives from simulating not only abiotic but biotic plant niche-like triggers for endophyte gene expression.

Cocultivation approaches likely function in the same way, providing biological signals to modify gene expression [218]. In a remarkable recent example of co-culturing, Taxol gene expression was restored in *Aspergillus terreus* by culturing it in the presence of *Podocarpus gracilior* (African fern pine) leaves [228]. Similar triggers occur in heterologous expression experiments, for example, in *Aspergilli* [229]. Fungal-*E. coli* shuttle vectors (FACs) have been used to identify SMs and gene clusters combined with LC-MS (i.e. FAC-MS) that may force expression of silent clusters [230]. Using regulators and promoters can help researchers to control the level of gene expression. For example, in the rice fungus *Monascus pilosus* the monacolin K and terrequinone A gene clusters from *Aspergillus nidulans* were successfully overexpressed in *Aspergillus oryzae* using a constitutive active *pgk* promoter [231]. Genetic methods that have been used to unsilence BGCs include heterologous host ribosome engineering [229, 232], insertion of constitutive or inducible promoters [233], reporter-guided mutant selection [234], and interfering in the condensation state of the genomic DNA by inactivation of DNA-modifying enzymes [213]. Manipulation of genes involved in microorganism development is another promising unsilencing method [235]. Finally, for bacteria there are high-throughput methods not involving genetics, like high-throughput elicitor screening with imaging mass spectrometry (HiTES-IMS) that promise to induce the silent secondary metabolome in response to ~ 500 conditions [47]. Yet, most of these methods are either low throughput, or work only for culturable microbes.

Approaches using next generation sequencing, comparative genomics, genome-scale metabolic models, and metabolic network modeling

High-throughput sequencing and bioinformatics combined with other newer technologies over the past 15 years have been instrumental in identifying unculturable endophytes communities and opening new horizons for expression of silent BGCs. For example, through comparative genomics, we now know that much of the chemical diversity in microbes derives from enzyme clusters, or biosynthetic gene clusters (BGCs) that are conserved across many species, such as the tailoring enzymes consisting of non-ribosomal peptide synthetases (NRPS), polyketide synthases (PKS), and terpene synthases (TPS) and terpene cyclases (TCs), phenyltransferases (PTs) along with associated genes for regulation, uptake of substrates, and transport and secretion of products [236, 237]. Some are also synthesized, carried, or tailored by post-translationally modified peptides (RIPPs). There are other specialized or taxon-specific BGCs, but because these often remain silent or expressed at very low levels under laboratory conditions, it is often difficult to confirm that the genes are functional. Thus, many strategies to discover NPs from microbes begin with bioinformatic prediction of BGCs from genomic data, followed by experimental induction of predicted silent biosynthetic pathways through genetic engineering or an array of methods discussed above.

Continuing efforts at database and software development have been especially important in refining the search for plant microbiome-derived NPs. Various 'older' software include untargeted genome mining approaches using the ClustScan software and ClustScan Database (CSDB) [238], 'Database Of BioSynthesis clusters Curated and InTegrated' (doBISCUIT) [239] which identifies clusters involved in tailoring enzymes, and ClusterMine 360, which includes 200 PKS & NRPS [240]. Other older approaches include the software 'Secondary Metabolite Unknown Region Finder' (SMURF) [241] which is a web-based HMM tool to identify conserved domains in PKS, NRPS, hybrid-PKS/NRPS and terpenoid gene clusters in fungi and the updated Joint Genome Institute (JGI) 'Integrated Microbial Genomes - Atlas of Biosynthetic gene Clusters' (IMG-ABC) for identification of gene clusters [58]. An increasingly useful database is 'The Minimum Information on Biosynthetic Geneclusters' (MIBiG) [242, 243]. These approaches have been used for phylogeny-based BGC discovery [244], which has been shown to be effective in identifying inhibitors of multidrug resistant pathogens [245].

However, many of these tools have been superseded by or integrated with leading current comprehensive toolset and databases for genome-wide annotation and analysis of BGC, the 'antibiotics & Secondary Metabolite

Analysis Shell (antiSMASH), with current version 5.0 [55, 110]. antiSMASH works as a web-server or downloadable software, and primarily runs NCBI BLAST+, HMMer 3, Muscle 3, FastTree, PySVG and JQuery SVG, along with many other previously published secondary metabolite analysis tools. Genome-wide metabolic models (GEMs) can enhance these approaches, for example with the 'Reconstruction, Analysis and Visualization of Metabolic Networks' RAVEN 2.0 software [246, 247] and MetaFlux [248] which has been integrated into the comprehensive toolset Pathway Tools [54]. Of particular interest for community metagenomic holometabolism data from *in planta* studies and Pathway Tools v2.30's multi-pathway diagrams (pathway collages) and its new algorithm for generating mechanistic explanations of multi-omics data [54].

Network-algorithm-based software can improve the predictive power of these genome mining approaches by incorporating ecological interactions [216]. For example, secondary metabolite gene cluster similarity networks [249], and network simulation models have been useful in studying metabolic production during interaction [250]. These approaches can be combined with metabolic modeling approaches, such as flux-balance models [167] with predictive mechanistic frameworks that predict core metabolism. Metabolic interactions in microbial co-cultures are perhaps best modeled this way, with the Metabolic Support Index (MSI) used to predict the microbial interactions in a co-culture and understand which microbe receives maximum benefit from the interactions [251]. The MetQuest software explores possible benefits derived by microorganisms from interactions in a community [252], although such results require follow up using physiological experiments. Biokinetic models have also been developed for interspecific interactions among microorganisms sharing substrates in an ecosystem [253]. Single-cell analysis could augment our understanding of endophyte metabolism [192], particularly with the addition of context-specific transcriptomics. Remarkable insights have been made from transcriptomic studies. For example, fungal regulation appears to be conserved during SM production [72] and can be confirmed via *in planta* transcriptomics [254]. Further promising transcriptomic methods that can be integrated with *in planta* strategies include Iso-seq (long read transcript sequencing), illuminating alternative splicing in Taxol production [255], and miRNA target transcriptome-mining [256].

More powerful solutions

Deep learning for global plant microbiome NP bioprospecting

Despite our general predictions of potential plant endophyte diversity (Table 1) and endophyte community (i.e.

microbiome) diversity (Table 2), the true distribution of endophytes and their potential natural products remains largely unknown [112]. To focus future endophyte bioprospecting requires a new, rigorous framework to guide strategic field sampling. NP exploration strategies must also be sensitive to threatened species and habitats. Machine learning and deep learning approaches, which are defined and described in Table 3, offer an exciting option.

Ideally, machine learning or deep learning frameworks could begin to predict plant microbiome distribution patterns in the context of environmental niches, while also predicting endophyte-derived natural products, thus, replacing comprehensive, global-scale, molecular surveys of plant microbiomes, which are challenging for all but a few clades.

Initial training data sets could capitalize on existing the growing array of genomic, phylogenomic, and multi-omic surveys, particularly those with metabolomics from natural plant tissues, i.e. the holotranscriptome and holometabolome. To increase training data, complementary, strategic multi-omics studies could be performed based on identified hotspots. These data can be combined with network co-occurrence analysis, metabolic cooperation or complementarity analysis, and community biosynthetic pathway analysis [216, 249, 250, 252, 257].

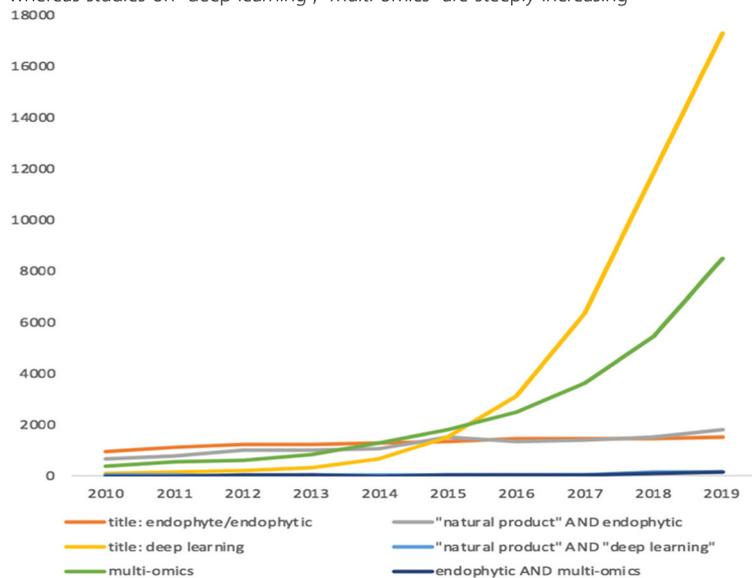
Several machine learning and deep learning software approaches are already in use for natural product discovery. For example, ClusterFinder [258] uses machine learning for known (curated) and unknown classes of BGCs, trained using a hidden Markov model-based probabilistic algorithm. DeepBGC [56] is a newer deep learning software tool that uses a Bidirectional Long Short-Term Memory (BiLSTM) neural network (RNN) and word2vec-like word embedding skip-gram neural network with three layers [56]. It uses an input layer of vectors of Pfam domains and genomic order, a layer of 128-dimensional hidden vectors, and the output layer of fully connected sigmoid functions, which is more sensitive (fewer false negatives) than ClusterFinder [56]. DeepBGC requires a large training data set for complex microbial communities.

In summary, the field of endophyte NP bioprospecting is ready for 'ecometabolomic' and 'phylo-metabolomic' deep learning, for example, using the H2O.ai deep learning framework [53]. Similar approaches are in use now in ecology [259] and there are increasingly more deep learning libraries for genomics, such as the recent python deep learning library, Janguu [260] which is compatible with other related python libraries; together, the goal will be to seamlessly integrate phylogenomic and hologenome predictions with interactome systems biology [261]. Arguably, the time to begin is now, given the rate of global plant habitat and biodiversity loss.

Table 3 Machine learning and deep learning approaches for plant microbiome-based natural product discovery

For predicting features of data that are too large to be completely sampled, one of the most promising approaches is computational learning or artificial intelligence, including machine learning and deep learning. These approaches deal with the problem of having an incomplete model to characterize unseen data, by evaluating diverse competing models on a set of training data. In other words, these approaches complete tasks without explicit instructions using patterns (models) learned from the training data. Specific machine learning approaches include Random forests, Hidden Markov Models, hierarchical cluster analyses, and support vector machines. Deep learning is a type of machine learning that handles additional complexity by using layers of data transformations. Specific deep learning approaches use convolutional neural networks where each layer learns from other, previous layers which are called hidden layers. One common framework for building such tools is the well-supported R Interface 'H2O' Scalable Machine Learning Platform (GitHub at h2oai/h2o-3) [53]. For **global endophyte NP bioprospecting**, we can integrate phylogenomic deep learning and genome-wide metabolic model deep learning frameworks. For example, using Pathway Tools v.23.0 [54] integrated with MetaFlux in antiSMASH [55] and DeepBGC [56]. For **predicting the chemical structural diversity of endophytes**, we can interface the approaches above into chemoinformatic and drug discovery deep learning frameworks. For **discovery of in planta unsilencing triggers – waking the sleeping giant**, we can integrate experimental system data, OSMAC, and multi-omics data (e.g. from data mining amplicon sequencing, shotgun sequencing, metatranscriptomic sequencing, and metabolomics)

Table 1 Inset: Recent trends in peer-reviewed studies with keywords/title "endophyte", "endophyte and natural product", showing limited increase, whereas studies on "deep learning", "multi-omics" are steeply increasing



Deep learning for predicting the chemical structural diversity of endophytes

Machine learning and deep learning approaches have been developed for chemoinformatics, anti-cancer and antibiotic drug discovery, and metabolomics [262–265]. In particular, these approaches have been useful for organic chemical exploration [264], bioactivity prediction based on chemical structure and mapping BGC combinations to chemical groups. We suggest the next critical frontier will be to develop chemoinformatics and bioactivity-focused informatics that integrate with and inform bioprospecting. Specifically, research could focus on systematic computational learning approaches for predicting chemical structural diversity from endophytes based on integrated comparative metabolomics and chemical compound analysis, combined with biotic interaction network analysis, building a model of correlations between *in planta* biochemistry and plant microbiome ecology. Furthermore, these frameworks can be tailored according to specific goals. For example, alternative deep learning frameworks could focus on chemical novelty and dereplication, or specific bioactivities (e.g.

antiviral vs. antifungal vs. anti-protozoan vs. antibacterial, or anticancer), or structures with the most complex synthesis such as (list chemical forms, bonds, or chirality groups).

Recent thinking on this topic is that it is critically important to avoid reductionism [266], because the power of these approaches is in their ability to address unknown interactions. Therefore, we suggest researchers should begin by training on encoded natural product chemical structural databases integrated with synthetic organic chemistry libraries and organismal metadata – particularly from habitat and metagenomic data. Because plants and plant-endophyte systems are targets for viral pathogens, they may hold promise for discovery of novel antiviral compounds, such as novel RNA-dependent RNA polymerase (RdRp) inhibitors, e.g. pyrazine family compounds related to pyrazinecarboxamides (e.g. favipiravir, currently in use as broad spectrum RdRp inhibitors against influenza and COVID-19). Similarly, plant-endophyte systems must defend against a wide range of fungal and bacterial pathogens and likely have evolved narrow-target antifungals and antibacterials. Animal-

specific cytotoxic compounds are likely diverse in these systems, to combat a range of possible herbivore pests.

But what about uncultivable endophytes, given that much research on endophyte NPs is motivated by the prospect that endophytes are easier to cultivate than plants [267, 268]? We argue that for uncultivable endophytes, computational learning-based chemical structure prediction will be especially helpful for overcoming the need for isolation and synthesis, but also such approaches can narrow the search for targets for downstream experimental (and computational) unsilencing, as described below.

Deep learning for discovery of in planta unsilencing triggers – waking the sleeping giant

Hidden, or silenced biosynthetic capacities seem to be the rule, rather than the exception in plant microbiomes, as evidenced from bioinformatic identification of BGCs. This leads to a major research problem, that research has tried to overcome through co-cultivation, OSMAC experiments [28], heterologous expression experiments [232], high-throughput elicitor screening [47], transcription factor decoys [269], and *in planta* approaches [270]. Yet, to date, there has been little concerted effort to apply computational learning approaches to solve this problem. This would seem surprising, given that genome data mining methods exist to uncover a diversity of regulatory signaling processes, metabolic flux, metabolic pathway regulation, and holobiont metabolic interactions such as pathway complementation. Computational learning strategies could use training data that is already from high throughput elicitor or expression experiments, OSMAC arrays, combined with *in planta* or co-culture holometabolomic and holoregulomic data. One promising approach could be to incorporate trans-kingdom regulatory small RNA data, for example from miRNomics sequencing. Such approaches could be combined with unsilencing studies *in planta*, such as global effector studies on synthetic communities on gnotobiotic plants (SynCom), which have been used to analyze complex dynamics of effector secretion by pathogens and beneficials [270]. Finally, a major gap that could be addressed with deep learning is to investigate models of metabolic cooperation amongst endophytes and plants.

Thus, to increase the scope and throughput of BGC unsilencing experiments, we propose new *in silico* unsilencing pipelines that infuse comparative multi-omic analyses with deep learning. The result would be endophyte community-level ‘coregulomics’. With the blossoming world of software and bioinformatics approaches, this idea is arguably within reach.

Conclusions

To meet the demand of the world’s emergent and resistant diseases caused by viruses (e.g. COVID-19), bacteria

(e.g. tuberculosis), parasites (e.g. malaria), and other major illnesses and conditions, such as cancers, novel natural products will continue to be in demand. For plant microbiomes to fulfill their promise [20, 262] as a leading source of new antiviral, antibiotic, and anticancer drugs, higher throughput and computational approaches are needed. We have proposed integrating computational learning approaches (e.g. deep learning) into the pipeline for both predicting and validating novel endophyte metabolites. If implemented, such deep learning approaches could explore broader mysteries, for example, whether medicinal plant health benefits could derive from endophyte communities rather than plants, or whether cooperative biosynthetic pathways between host and microbe may be important in NP synthesis, for example, in Taxol. Endophyte-derived natural compounds may also be of value outside of medicine, for example, in buffering anthropogenic and climate effects or habitats and crops impacted by invasive pathogens [96, 271, 272]. All together, these points emphasize the need to conserve biodiversity with an enhanced focus on characterization and conservation of diverse endophyte-rich habitats.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40793-021-00375-0>.

Additional file 1.

Abbreviations

NP: Natural Product; SM: Secondary Metabolite; BGC: Biosynthetic Gene Cluster

Acknowledgements

We thank Carolin Frank for useful comments on the draft manuscript.

Conflict of interest

The authors declare no conflict of interests.

Authors’ contributions

SAA and AMVB conceived of and co-wrote and revised the manuscript. The authors read and approved the final manuscript.

Funding

Support for this work was through a Texas Tech University Graduate Student Research Support Award to SAA and startup funding support to AMVB.

Availability of data and materials

Not applicable.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 27 November 2020 Accepted: 21 February 2021

Published online: 18 March 2021

References

1. Staniek A, Woerdenbag HJ, Kayser O. Endophytes: exploiting biodiversity for the improvement of natural product-based drug discovery. *J Plant Interact.* 2008;3:75–93.
2. Aly AH, Debbab A, Proksch P. Fungal endophytes: unique plant inhabitants with great promises. *Appl Microbiol Biotechnol.* 2011;90:1829–45.
3. Pimentel MR, Molina G, Dionísio AP, Maróstica Junior MR, Pastore GM. The use of Endophytes to obtain bioactive compounds and their application in biotransformation process. *Biotechnol Res Int.* 2011;2011:1–11.
4. Strobel GA. Endophytes as sources of bioactive products. *Microbes Infect.* 2003;5:535–44.
5. Suryanarayanan TS, Thirunavukkarasu N, Govindarajulu MB, Sasse F, Jansen R, Murali TS. Fungal endophytes and bioprospecting. *Fungal Biol Rev.* 2009; 23:9–19. <https://doi.org/10.1016/j.fbr.2009.07.001>.
6. Qin S, Xing K, Jiang JH, Xu LH, Li WJ. Biodiversity, bioactive natural products and biotechnological potential of plant-associated endophytic actinobacteria. *Appl Microbiol Biotechnol.* 2011;89:457–73.
7. Stierle A, Strobel G, Stierle D. Taxol and Taxane production by *Taxomyces andreanae*, an Endophytic fungus of Pacific yew. *Science (80-)*. 1993;260: 214–6.
8. Kusari S, Singh S, Jayabaskaran C. Rethinking production of Taxol® (paclitaxel) using endophyte biotechnology. *Trends Biotechnol.* 2014;32: 304–11.
9. Zhou X, Zhu H, Liu L, Lin J, Tang K. A review: recent advances and future prospects of taxol-producing endophytic fungi. *Appl Microbiol Biotechnol.* 2010;86:1707–17.
10. Uzma F, Mohan CD, Hashem A, Konappa NM, Rangappa S, Kamath PV, et al. Endophytic fungi-alternative sources of cytotoxic compounds: a review. *Front Pharmacol.* 2018;9:1–37.
11. Dreyfuss MM, Chapela IH. Potential of fungi in the discovery of novel, low-molecular weight pharmaceuticals. In: *The discovery of natural products with therapeutic potential*; 1994. p. 49–80.
12. Schulz B, Boyle C, Draeger S, Römmert AK, Krohn K. Endophytic fungi: a source of novel biologically active secondary metabolites. *Mycol Res.* 2002; 106:996–1004.
13. Bérdy J. Bioactive microbial metabolites. *J Antibiot.* 2005;58:1–26 <https://www-nature-com.pugwash.lib.warwick.ac.uk/articles/ja20051.pdf>.
14. Prado S, Li Y, Nay B. Diversity and ecological significance of fungal endophyte natural products; 2012.
15. Gupta S, Chaturvedi P, Kulkarni MG, Van Staden J. A critical review on exploiting the pharmaceutical potential of plant endophytic fungi. *Biotechnol Adv.* 2020;39:107462. <https://doi.org/10.1016/j.biotechadv.2019.107462>.
16. Cain JW, Miller KI, Kalaitzis JA, Chau R, Neilan BA. Genome mining of a fungal endophyte of *Taxus yunnanensis* (Chinese yew) leads to the discovery of a novel azaphilone polyketide, lijiquinone. *J Microbiol Biotechnol.* 2020;13(5):1415–27.
17. He Q, Zeng Q, Shao Y, Zhou H, Li T, Song F, et al. Anti-cervical cancer activity of secondary metabolites of endophytic fungi from *Ginkgo biloba*. *Cancer Biomark.* 2020;Preprint:1–9.
18. Zhang G, Sun S, Zhu T, Lin Z, Gu J, Li D, et al. Antiviral isindolone derivatives from an endophytic fungus *Emericella* sp associated with *Aegiceris corniculatum*. *Phytochemistry.* 2011;72:1436–42.
19. Newman DJ, Cragg GM. Natural products as sources of new drugs from 1981 to 2014. *J Nat Prod.* 2016;79:629–61.
20. Nisa H, Kamili AN, Nawchoo IA, Shafi S, Shameem N, Bandh SA. Fungal endophytes as prolific source of phytochemicals and other bioactive natural products: a review. *Microb Pathog.* 2015;82:50–9.
21. Redecker D, Kodner R, Graham LE, Redecker D, Kodner R, Graham LE. Glomalean fungi from the Ordovician. *Science (80-)*. 2016;289:1920–1.
22. Krings M, Harper CJ, Taylor EL. Fungi and fungal interactions in the Rhyneic chert: a review of the evidence, with the description of *Perexiflasca tayloriana* gen. Et sp. nov. *Philos Trans R Soc B Biol Sci.* 2018;373:20160500.
23. Hawksworth DL, Lücking R. Fungal diversity revisited: 2.2 to 3.8 million species. *Microbiol Spectr.* 2017;5:1–17.
24. Blackwell M. The fungi: 1, 2, 3 ... 5.1 million species? *Am J Bot.* 2011;98: 426–38.
25. Arnold A, Maynard Z, Gilbert G, Coley P, Kursar T. Are tropical fungal endophytes hyperdiverse? *Ecol Lett.* 2000;3:267–74.
26. Higgins KL, Arnold AE, Miadlikowska J, Sarvate SD, Lutzoni F. Phylogenetic relationships, host affinity, and geographic structure of boreal and arctic endophytes from three major plant lineages. *Mol Phylogenet Evol.* 2007;42: 543–55.
27. Arnold AE, Zuleyka M, Gilbert GS. Fungal endophytes in dicotyledonous neotropical trees: patterns of abundance and diversity. *Mycol Res.* 2001;105: 1502–7.
28. Pan R, Bai X, Chen J, Zhang H, Wang H. Exploring structural diversity of microbe secondary metabolites using OSMAC strategy: a literature review. *Front Microbiol.* 2019;10:1–20.
29. Rashmi M, Venkateswara SV. Secondary metabolite production by Endophytic fungi: the gene clusters, nature, and expression. In: *Endophytes and secondary metabolites*; 2019. p. 475–90.
30. Carrión VJ, Perez-Jaramillo J, Cordovez V, Tracanna V, De Hollander M, Ruiz-Buck D, et al. Pathogen-induced activation of disease-suppressive functions in the endophytic root microbiome. *Science (80-)*. 2019;366:606–12.
31. Li YF, Tsai KJS, Harvey CJB, Li JJ, Ary BE, Berlew EE, et al. Comprehensive curation and analysis of fungal biosynthetic gene clusters of published natural products. *Fungal Genet Biol.* 2016;89:18–28.
32. Locey KJ, Lennon JT. Scaling laws predict global microbial diversity. *Proc Natl Acad Sci.* 2016;113:5970–5. <https://doi.org/10.1073/pnas.1521291113>.
33. Baltz RH. Gifted microbes for genome mining and natural product discovery. *J Ind Microbiol Biotechnol.* 2017;44:573–88.
34. Arnold AE, Maynard Z, Gilbert GS, Coley PD, Kursar TA. Are tropical fungal endophytes hyperdiverse? *Ecol Lett.* 2000;3:267–74.
35. Costello MJ, Wilson S, Houlding B. Predicting total global species richness using rates of species description and estimates of taxonomic effort. *Syst Biol.* 2012;61:871–83.
36. Willis A. Extrapolating abundance curves has no predictive power for estimating microbial biodiversity. *Proc Natl Acad Sci U S A.* 2016;113:E5096.
37. Locey KJ, Lennon JT. Powerful predictions of biodiversity from ecological models and scaling laws. *Proc Natl Acad Sci U S A.* 2016;113:E5097.
38. Shoemaker WR, Locey KJ, Lennon JT. A macroecological theory of microbial biodiversity. *Nat Ecol Evol.* 2017;1:1–6.
39. Wilson JB, Peet RK, Dengler J, Pärtel M. Plant species richness: the world records. *J Veg Sci.* 2012;23:796–802.
40. Preston FW. The commonness, and rarity, of species. *Ecology.* 1948;29:254–83.
41. Liu H, Carvalhais LC, Crawford M, Singh E, Dennis PG, Pieterse CMJ, et al. Inner plant values: Diversity, colonization and benefits from endophytic bacteria. *Front Microbiol.* 2017;8:1–17.
42. Bar-On YM, Phillips R, Milo R. The biomass distribution on earth. *Proc Natl Acad Sci U S A.* 2018;115:6506–11. <https://doi.org/10.1073/pnas.1711842115>.
43. Tang Z, Xu W, Zhou G, Bai Y, Li J, Tang X, et al. Patterns of plant carbon, nitrogen, and phosphorus concentration in relation to productivity in China's terrestrial ecosystems. *Proc Natl Acad Sci U S A.* 2018;115:4033–8.
44. Zhai Y, Wang W, Tan H, Cao L. A new approach to analyzing Endophytic Actinobacterial population in the roots of Banana plants (*Musa* sp., AAA). *J Biochem Mol Biol Res.* 2016;2:180–4.
45. Ludwig-Müller J. Interplay between Endophyte and host Plant in the Synthesis and Modification of metabolites. In: Schouten A, editor. *Endophyte. Biotechnology: Potential for Agriculture and Pharmacology.* CABI, Wallingford: CAB International; 2019;8:180.
46. He XY, Wang KL, Zhang W, Chen ZH, Zhu YG, Chen HS. Positive correlation between soil bacterial metabolic and plant species diversity and bacterial and fungal diversity in a vegetation succession on karst. *Plant and Soil.* 2008;307:123–34.
47. Xu F, Wu Y, Zhang C, Davis KM, Moon K, Bushin LB, et al. A genetics-free method for high-throughput discovery of cryptic microbial metabolites. *Nat Chem Biol.* 2019;15:161–8. <https://doi.org/10.1038/s41589-018-0193-2>.
48. Rodriguez RJ, White JF, Arnold A, Redman RS. Fungal endophytes: diversity and functional roles. *New Phytol.* 2009;182:314–30.
49. Bains W, Seager S. A combinatorial approach to biochemical space: description and application to the redox distribution of metabolism. *Astrobiology.* 2012;12:271–81.
50. Klamt S, Stelling J. Combinatorial complexity of pathway analysis in metabolic networks. *Mol Biol Rep.* 2002;29:233–6.
51. Skellam E. Strategies for engineering natural product biosynthesis in fungi. *Trends Biotechnol.* 2019;37:416–27. <https://doi.org/10.1016/j.tibtech.2018.09.003>.

52. Gould AL, Zhang V, Lamberti L, Jones EW, Obadia B, Korasidis N, et al. Microbiome interactions shape host fitness. *Proc Natl Acad Sci U S A*. 2018; 115:E11951–60.
53. Cook D. Practical machine learning with H2O: powerful, scalable techniques for deep learning and AI. " O'Reilly Media, Inc."; 2016.
54. Karp PD, Midford PE, Billington R, Kothari A, Krummenacker M, Latendresse M, et al. Pathway Tools version 23.0 update: software for pathway/genome informatics and systems biology. *Brief Bioinform*. 2019. <https://doi.org/10.1093/bib/bbz104>.
55. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, et al. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res*. 2019;47:W81–7.
56. Hannigan GD, Prihoda D, Palicka A, Soukup J, Klempir O, Rampula L, et al. A deep learning genome-mining strategy for biosynthetic gene cluster prediction. *Nucleic Acids Res*. 2019;47:e110.
57. Inglis DO, Binkley J, Skrzypek MS, Arnaud MB, Cerqueira GC, Shah P, et al. Comprehensive annotation of secondary metabolite biosynthetic genes and gene clusters of *Aspergillus nidulans*, *A. fumigatus*, *A. niger* and *A. oryzae*. *BMC Microbiol*. 2013;13:1–23.
58. Hadjithomas M, Chen IMA, Chu K, Huang J, Ratner A, Palaniappan K, et al. IMG-ABC: new features for bacterial secondary metabolism analysis and targeted biosynthetic gene cluster discovery in thousands of microbial genomes. *Nucleic Acids Res*. 2017;45:D560–5.
59. Andersen MR, Nielsen JB, Klitgaard A, Petersen LM, Zachariassen M, Hansen TJ, et al. Accurate prediction of secondary metabolite gene clusters in filamentous fungi. *Proc Natl Acad Sci U S A*. 2013;110:E99–107.
60. Kusari P, Kusari S, Spitteller M, Kayser O. Implications of endophyte-plant crosstalk in light of quorum responses for plant biotechnology. *Appl Microbiol Biotechnol*. 2015;99:5383–90.
61. Arora P, Riyaz-Ul-Hassan S. Endohyphal bacteria; the prokaryotic modulators of host fungal biology. *Fungal Biol Rev*. 2019;33:72–81. <https://doi.org/10.1016/j.fbr.2018.08.003>.
62. MacDonald RM, Chandler MR, Mosse B. The occurrence of bacterium-like organelles in vesicular-arbuscular mycorrhizal fungi. *New Phytol*. 1982;90: 659–63.
63. Riedlinger J, Schrey SD, Tarkka MT, Hampp R, Kapur M, Fiedler HP. Auxofuran, a novel metabolite that stimulates the growth of fly agaric, is produced by the mycorrhizal helper bacterium *Streptomyces* strain Ach 505. *Appl Environ Microbiol*. 2006;72:3550–7.
64. Hoffman MT, Arnold AE. Diverse bacteria inhabit living hyphae of phylogenetically diverse fungal endophytes. *Appl Environ Microbiol*. 2010; 76:4063–75.
65. Pakvaz S, Soltani J. Endohyphal bacteria from fungal endophytes of the Mediterranean cypress (*Cupressus sempervirens*) exhibit in vitro bioactivity. *For Pathol*. 2016;46:569–81.
66. Hoffman MT, Gunatilaka MK, Wijeratne K, Gunatilaka L, Arnold AE. Endohyphal bacterium enhances production of Indole-3-acetic acid by a foliar fungal Endophyte. *PLoS One*. 2013;8:31–3.
67. Arendt KR, Hockett KL, Araldi-Brondolo SJ, Baltrus DA, Arnold AE. Isolation of Endohyphal bacteria from foliar Ascomycota and in vitro. *Appl Environ Microbiol*. 2016;82:2943–9.
68. Márquez LM, Redman RS, Rodriguez RJ, Roossinck MJ. A virus in a fungus in a plant: three-way symbiosis required for thermal tolerance. *Science* (80-). 2007;315:513–6.
69. Morsy MR, Oswald J, He J, Tang Y, Roossinck MJ. Teasing apart a three-way symbiosis: Transcriptome analyses of *Curvularia protuberata* in response to viral infection and heat stress. *Biochem Biophys Res Commun*. 2010;401: 225–30.
70. Ghabrial SA, Castón JR, Jiang D, Nibert ML, Suzuki N. 50-plus years of fungal viruses. *Virology*. 2015;479–480:356–68. <https://doi.org/10.1016/j.virol.2015.02.034>.
71. Venturi V, da Silva DP. Incoming pathogens team up with harmless "resident" bacteria. *Trends Microbiol*. 2012;20:160–4.
72. Nielsen JC, Prigent S, Grijsseels S, Workman M, Ji B, Nielsen J. Comparative Transcriptome analysis shows conserved metabolic regulation during production of secondary metabolites in filamentous fungi. *mSystems*. 2019; 4:1–14.
73. Li SJ, Zhang X, Wang XH, Zhao CQ. Novel natural compounds from endophytic fungi with anticancer activity. *Eur J Med Chem*. 2018;156:316–43. <https://doi.org/10.1016/j.ejmech.2018.07.015>.
74. Gao H, Li G, Lou HX. Structural diversity and biological activities of novel 1336 secondary metabolites from endophytes. *Molecules*. 2018;23(3):646.
75. Caruso G, Abdelhamid M, Kalisz A, Sekara A. Linking Endophytic fungi to medicinal plants therapeutic activity. A case study on Asteraceae. *Agriculture*. 2020;10:286.
76. Jia M, Chen L, Xin HL, Zheng CJ, Rahman K, Han T, et al. A friendly relationship between endophytic fungi and medicinal plants: a systematic review. *Front Microbiol*. 2016;7:1–14.
77. Gutierrez RMP, Gonzalez AMN, Ramirez AM. Compounds derived from Endophytes: a review of Phytochemistry and pharmacology. *Curr Med Chem*. 2012;19:2992–3030.
78. Kjer J, Debbab A, Aly AH, Proksch P. Methods for isolation of marine-derived endophytic fungi and their bioactive secondary products. *Nat Protoc*. 2010; 5:479–90.
79. Jouda JB, de Tamokou J, Mbazoa CD, Sarkar P, Bag PK, Wandji J. Anticancer and antibacterial secondary metabolites from the endophytic fungus *penicillium* sp. CAM64 against multi-drug resistant gram-negative bacteria. *Afr Health Sci*. 2016;16:734–43.
80. Prasher IB, Dhanda RK. GC-MS analysis of secondary metabolites of Endophytic *Nigrospora sphaerica* isolated from *Parthenium hysterophorus*. *Int J Pharm Sci Rev Res*. 2017;44:217–23.
81. Qiao YM, Yu RL, Zhu P. Advances in targeting and heterologous expression of genes involved in the synthesis of fungal secondary metabolites. *RSC Adv*. 2019;9:35124–34.
82. Mao D, Okada BK, Wu Y, Xu F, Seyedsayamdost MR. Recent advances in activating silent biosynthetic gene clusters in bacteria. *Curr Opin Microbiol*. 2018;45:156–63. <https://doi.org/10.1016/j.mib.2018.05.001>.
83. Bode HB, Bethe B, Höfs R, Zeeck A. Big effects from small changes: possible ways to explore nature's chemical diversity. *ChemBioChem*. 2002;3:619–27.
84. Porras-Alfaro A, Bayman P. Hidden fungi, emergent properties: Endophytes and microbiomes. *Annu Rev Phytopathol*. 2011;49:291–315.
85. Agler MT, Ruhe J, Kroll S, Morhenn C, Kim ST, Weigel D, et al. Microbial hub taxa link host and abiotic factors to plant microbiome variation. *PLoS Biol*. 2016;14:1–31.
86. White JF, Torres MS. Is plant endophyte-mediated defensive mutualism the result of oxidative stress protection? *Physiol Plant*. 2010;138:440–6.
87. Munir S, Li Y, He P, He P, Ahmed A, Wu Y, et al. Unraveling the metabolite signature of citrus showing defense response towards *Candidatus Liberibacter asiaticus* after application of endophyte *Bacillus subtilis* L1–21. *Microbiol Res*. 2020;234:126425. <https://doi.org/10.1016/j.micres.2020.126425>.
88. Hiruma K, Kobae Y, Toju H. Beneficial associations between Brassicaceae plants and fungal endophytes under nutrient-limiting conditions: evolutionary origins and host–symbiont molecular mechanisms. *Curr Opin Plant Biol*. 2018;44:145–54.
89. Guo B, Wang Y, Sun X, Tang K. Bioactive natural products from endophytes: a review. *Prikl Biokhim Mikrobiol*. 2008;44:153–8.
90. Zhu X, Zhong Y, Xie Z, Wu M, Hu Z, Ding W, et al. Fusarixins A and B: novel cyclic Hexadepsipeptides from the mangrove Endophytic fungus *Fusarium* sp. R5 with antifungal activities. *Planta Med*. 2018;84:1355–62.
91. Davis RA, Carroll AR, Andrews KT, Boyle GM, Tran TL, Healy PC, et al. Pestalactams A–C: novel caprolactams from the endophytic fungus *Pestalotiopsis* sp. *Org Biomol Chem*. 2010;8:1785–90.
92. Zeng Y-J, Yang H-R, Wang H-F, Zong M-H, Lou W-Y. Immune enhancement activity of a novel polysaccharide produced by *Dendrobium officinale* endophytic fungus *Fusarium solani* DO7. *J Funct Foods*. 2019;53:266–75.
93. Netzker T, Fischer J, Weber J, Mattern DJ, König CC, Valiante V, et al. Microbial communication leading to the activation of silent fungal secondary metabolite gene clusters. *Front Microbiol*. 2015;6:1–13.
94. Cristina Stroe M, Netzker T, Scherlach K, Krüger T, Hertweck C, Valiante V, et al. Targeted induction of a silent fungal gene cluster encoding the bacteria-specific germination inhibitor fumigermin. *Elife*. 2020;9:1–20.
95. Ren H, Wang B, Zhao H. Breaking the silence: new strategies for discovering novel natural products. *Curr Opin Biotechnol*. 2017;48:21–7. <https://doi.org/10.1016/j.copbio.2017.02.008>.
96. Muvea AM, Meyhöfer R, Subramanian S, Poehling HM, Ekesi S, Maniania NK. Colonization of onions by endophytic fungi and their impacts on the biology of thrips *tabaci*. *PLoS One*. 2014;9:e108242.
97. Waqas M, Khan AL, Kamran M, Hamayun M, Kang SM, Kim YH, et al. Endophytic fungi produce gibberellins and indoleacetic acid and promotes host-plant growth during stress. *Molecules*. 2012;17:10754–73.
98. Gange AC, Koricheva J, Currie AF, Jaber LR, Vidal S. Meta-analysis of the role of entomopathogenic and unspecialized fungal endophytes as plant bodyguards. *New Phytol*. 2019;223:2002–10.

99. Gundel PE, Sun P, Charlton ND, Young CA, Miller TEX, Rudgers JA. Simulated folivory increases vertical transmission of fungal endophytes that deter herbivores and alter tolerance to herbivory in *Poa autumnalis*. *Ann Bot*. 2020;125:981–91. <https://doi.org/10.1093/aob/mcaa021>.
100. Ehsan T, Reza RN, Das A, Ahmed O, Baten AKMA, Ferdous AS, et al. Genome and secretome analysis of jute endophyte *Grammothele lineata* strain SDLCO-2015-1: insights into its lignocellulolytic structure and secondary metabolite profile. *Genomics*. 2020;112(4):2794–803.
101. Schouten A. Saving resources: the exploitation of Endophytes by plants for the biosynthesis of multi-functional Defence compounds. In: Schouten A, editor. *Endophyte biotechnology: potential for agriculture and pharmacology*: CAB International; 2019. p. 122–44.
102. Higgins SA, Schadt CW, Matheny PB, Löffler FE. Phylogenomics reveal the dynamic evolution of fungal nitric oxide reductases and their relationship to secondary metabolism. *Genome Biol Evol*. 2018;10:2474–89.
103. Stajich JE. Fungal genomes and insights into the evolution of the Kingdom. *Fungal Kingd*. 2017;619–33. <https://doi.org/10.1128/9781555819583.ch29>.
104. Lajoie G, Maglione R, Kembel SW. Adaptive matching between phyllosphere bacteria and their tree hosts in a neotropical forest. *Microbiome*. 2020;8:1–10.
105. Kaul S, Gupta S, Ahmed M, Dhar MK. Endophytic fungi from medicinal plants: a treasure hunt for bioactive metabolites. *Phytochem Rev*. 2012;11:487–505.
106. Rout ME, Chrzanowski TH, Westlie TK, DeLuca TH, Callaway RM, Holben WE. Bacterial endophytes enhance competition by invasive plants. *Am J Bot*. 2013;100:1726–37.
107. Schouten A. Endophytic fungi: definitions, diversity, distribution and their significance in plant life. In: Schouten A, editor. *Endophyte biotechnology: potential for agriculture and pharmacology*: CAB International; 2019. p. 6–31.
108. Carlier AL, Eberl L. The eroded genome of a *Psychotria* leaf symbiont: hypotheses about lifestyle and interactions with its plant host. *Environ Microbiol*. 2012;14:2757–69.
109. Maheshwari DK, Maheshwari R. *Endophytes: biology and biotechnology*. Berlin: Springer; 2017.
110. Medema MH, Blin K, Cimermancic P, De Jager V, Zakrzewski P, Fischbach MA, et al. AntiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res*. 2011;39(SUPPL. 2):339–46.
111. Keller NP. Fungal secondary metabolism: regulation, function and drug discovery. *Nat Rev Microbiol*. 2019;17:167–80.
112. Harrison JG, Griffin EA. The diversity and distribution of endophytes across biomes, plant phylogeny and host tissues: how far have we come and where do we go from here? *Environ Microbiol*. 2020;22:2107–23.
113. Faville MJ, Briggs L, Cao M, Koulman A, Jahufer MZZ, Koolaard J, et al. A QTL analysis of host plant effects on fungal endophyte biomass and alkaloid expression in perennial ryegrass. *Mol Breed*. 2015;35:1–18.
114. Huang Y, Wang J, Li G, Zheng Z, Su W. Antitumor and antifungal activities in endophytic fungi isolated from pharmaceutical plants *Taxus mairei*, *Cephalataxus fortunei* and *Torreya grandis*. *FEMS Immunol Med Microbiol*. 2001;31:163–7.
115. Tenguria RK, Khan FN, Qureshi S. Endophytes- mines of pharmacological therapeutics. *World J Sci Technol*. 2011;1:127–49.
116. Liu AR, Chen SC, Lin XM, Wu SY, Xu T, M. CF, et al. Endophytic *Pestalotiopsis* species associated with plants of *Palmae*, *Rhizophoraceae*, *Planchonellae* and *Podocarpaceae* in Hainan, China. *Afr J Microbiol Res*. 2010;4:2661–9.
117. Chen L, Zhang QY, Jia M, Ming QL, Yue W, Rahman K, et al. Endophytic fungi with antitumor activities: their occurrence and anticancer compounds. *Crit Rev Microbiol*. 2016;42:454–73.
118. Panaccione DG, Johnson RD, Wang J, Young CA, Damrongkool P, Scott B, et al. Elimination of ergovaline from a grass-*Neotyphodium* endophyte symbiosis by genetic modification of the endophyte. *Proc Natl Acad Sci U S A*. 2001;98:12820–5.
119. Dissanayake AJ, Purahong W, Wubet T, Hyde KD, Zhang W, Xu H, et al. Direct comparison of culture-dependent and culture-independent molecular approaches reveal the diversity of fungal endophytic communities in stems of grapevine (*Vitis vinifera*). *Fungal Divers*. 2018;90:85–107. <https://doi.org/10.1007/s13225-018-0399-3>.
120. Glynnou K, Nam B, Thines M, Maciá-Vicente JG. Facultative root-colonizing fungi dominate endophytic assemblages in roots of nonmycorrhizal *Microthlaspi* species. *New Phytol*. 2018;217:1190–202.
121. Johansson VA, Bahram M, Tedersoo L, Kõljalg U, Eriksson O. Specificity of fungal associations of *Pyroleae* and *Monotropia hypopitys* during germination and seedling development. *Mol Ecol*. 2017;26:2591–604. <https://doi.org/10.1111/mec.14050>.
122. Nissinen RM, Männistö MK, van Elsas JD. Endophytic bacterial communities in three arctic plants from low arctic fell tundra are cold-adapted and host-plant specific. *FEMS Microbiol Ecol*. 2012;82:510–22.
123. Brader G, Compant S, Mitter B, Trognitz F, Sessitsch A. Metabolic potential of endophytic bacteria. *Curr Opin Biotechnol*. 2014;27:30–7.
124. Gorni C, Allemand D, Rossi D, Mariani P. Microbiome profiling in fresh-cut products. *Trends Food Sci Technol*. 2015;46:295–301. <https://doi.org/10.1016/j.tifs.2015.10.013>.
125. Thomas P, Soly TA. Endophytic bacteria associated with growing shoot tips of Banana (*Musa sp.*) cv. Grand Naine and the affinity of Endophytes to the host. *Microb Ecol*. 2009;58:952–64.
126. Woźniak M, Gałązka A, Grzażdziel J, Głodowska M. The identification and genetic diversity of endophytic bacteria isolated from selected crops. *J Agric Sci*. 2018;156:547–56.
127. Carper DL, Carrell AA, Kueppers LM, Frank AC. Bacterial endophyte communities in *Pinus flexilis* are structured by host age, tissue type, and environmental factors. *Plant and Soil*. 2018;428:335–52.
128. Magnani GS, Didonet CM, Cruz LM, Picheth CF, Pedrosa FO, Souza EM. Diversity of endophytic bacteria in Brazilian sugarcane. *Genet Mol Res*. 2010;9:250–8.
129. Sun L, Qiu F, Zhang X, Dai X, Dong X, Song W. Endophytic bacterial diversity in rice (*Oryza sativa* L.) roots estimated by 16S rDNA sequence analysis. *Microb Ecol*. 2008;55:415–24.
130. Izumi H. Diversity of Endophytic bacteria in Forest trees. In: Pirttilä AM, Frank AC, editors. *Endophytes of Forest trees: biology and applications*. Dordrecht: Springer Netherlands; 2011. p. 95–105. https://doi.org/10.1007/978-94-007-1599-8_6.
131. Shehzadi M, Fatima K, Imran A, Mirza MS, Khan QM, Afzal M. Ecology of bacterial endophytes associated with wetland plants growing in textile effluent for pollutant-degradation and plant growth-promotion potentials. *Plant Biosyst - An Int J Deal with all Asp Plant Biol*. 2016;150:1261–70. <https://doi.org/10.1080/11263504.2015.1022238>.
132. Helmann TC, Deutschbauer AM, Lindow SE. Genome-wide identification of *Pseudomonas syringae* genes required for fitness during colonization of the leaf surface and apoplast. *Proc Natl Acad Sci U S A*. 2019;116:18900–10.
133. Griffin EA, Traw MB, Morin PJ, Pruitt JN, Wright SJ, Carson WP. Foliar bacteria and soil fertility mediate seedling performance: a new and cryptic dimension of niche differentiation. *Ecology*. 2016;97:2998–3008.
134. Scheublin TR, Leveau JHJ. Isolation of *Arthrobacter* species from the phyllosphere and demonstration of their epiphytic fitness. *Microbiologyopen*. 2013;2:205–13.
135. Peix A, Ramírez-Bahena MH, Velázquez E, Bedmar EJ. Bacterial associations with legumes. *CRC Crit Rev Plant Sci*. 2015;34:17–42.
136. Wang D, Yang S, Tang F, Zhu H. Symbiosis specificity in the legume - rhizobial mutualism. *Cell Microbiol*. 2012;14:334–42.
137. Poole P, Ramachandran V, Terpolilli J. Rhizobia: from saprophytes to endosymbionts. *Nat Rev Microbiol*. 2018;16:291–303. <https://doi.org/10.1038/nrmicro.2017.171>.
138. Baltrus DA, Dougherty K, Arendt KR, Huntemann M, Clum A, Pillay M, et al. Absence of genome reduction in diverse, facultative endohyphal bacteria. *Microb Genomics*. 2017;3.
139. Bianciotto V, Lumini E, Bonfante P, Vandamme P. “*Candidatus Glomeribacter gigasporarum*” gen. Nov., sp. nov., an endosymbiont of arbuscular mycorrhizal fungi. *Int J Syst Evol Microbiol*. 2003;53:121–4.
140. Khare E, Mishra J, Arora NK. Multifaceted interactions between endophytes and plant: Developments and Prospects. *Front Microbiol*. 2018;9:1–12.
141. Son M, Yu J, Kim KH. Five questions about Mycoviruses. *PLoS Pathog*. 2015;11:5–11. <https://doi.org/10.1371/journal.ppat.1005172>.
142. Bao X, Roossinck MJ. Multiplexed interactions: viruses of Endophytic fungi. *Adv Virus Res*. 2013;86:37–58.
143. Segers GC, Zhang X, Deng F, Sun Q, Nuss DL. Evidence that RNA silencing functions as an antiviral defense mechanism in fungi. *Proc Natl Acad Sci U S A*. 2007;104:12902–6.
144. Zhang DX, Nuss DL. Engineering super mycovirus donor strains of chestnut blight fungus by systematic disruption of multilocus vic genes. *Proc Natl Acad Sci U S A*. 2016;113:2062–7.
145. Xie J, Jiang D. New insights into mycoviruses and exploration for the biological control of crop fungal diseases. *Annu Rev Phytopathol*. 2014;52:45–68.

146. Kanematsu S, Arakawa M, Oikawa Y, Onoue M, Osaki H, Nakamura H, et al. A reovirus causes hypovirulence of *Rosellinia necatrix*. *Phytopathology*. 2004; 94:561–8.
147. Xie J, Xiao X, Fu Y, Liu H, Cheng J, Ghabrial SA, et al. A novel mycovirus closely related to hypoviruses that infects the plant pathogenic fungus *Sclerotinia sclerotiorum*. *Virology*. 2011;418:49–56. <https://doi.org/10.1016/j.virol.2011.07.008>.
148. Siddique AB. Viruses of endophytic and pathogenic forest fungi. *Viruses*. 2020;56:407–16. <https://doi.org/10.1007/s11262-020-01763-3>.
149. Feldman TS, Morsy MR, Roossinck MJ. Are communities of microbial symbionts more diverse than communities of macrobial hosts? *Fungal Biol*. 2012;116:465–77. <https://doi.org/10.1016/j.funbio.2012.01.005>.
150. de Wet J, Bihon W, Preisig O, Wingfield BD, Wingfield MJ. Characterization of a novel dsRNA element in the pine endophytic fungus *Diplodia scrobiculata*. *Arch Virol*. 2011;156:199–208.
151. Paez-Espino D, Eloe-Fadrosh EA, Pavlopoulos GA, Thomas AD, Huntemann M, Mikhailova N, et al. Uncovering Earth's virome. *Nature*. 2016;536:425–30.
152. Ofir G, Sorek R. Contemporary phage biology: from classic models to new insights. *Cell*. 2018;172:1260–70. <https://doi.org/10.1016/j.cell.2017.10.045>.
153. Dion MB, Oechslin F, Moineau S. Phage diversity, genomics and phylogeny. *Nat Rev Microbiol*. 2020;18:125–38.
154. Takahashi H, Fukuhara T, Kitazawa H, Kormelink R. Virus latency and the impact on plants. *Front Microbiol*. 2019;10:2764.
155. Roossinck MJ. Plants, viruses and the environment: ecology and mutualism. *Virology*. 2015;479–480:271–7.
156. Roossinck MJ. Lifestyles of plant viruses. *Philos Trans R Soc B Biol Sci*. 2010; 365:1899–905.
157. Roossinck MJ. The good viruses: viral mutualistic symbioses. *Nat Rev Microbiol*. 2011;9:99–108.
158. Llave C. Dynamic cross-talk between host primary metabolism and viruses during infections in plants. *Curr Opin Virol*. 2016;19:50–5.
159. Montero R, Pérez-Bueno ML, Barón M, Florez-Sarasa I, Tohge T, Fernie AR, et al. Alterations in primary and secondary metabolism in *Vitis vinifera* 'Malvasía de Banyalbufar' upon infection with grapevine leafroll-associated virus 3. *Physiol Plant*. 2016;157:442–52.
160. Rosenblueth M, Martínez-Romero E. Bacterial Endophytes and their interactions with hosts. *Mol Plant Microbe Interact*. 2006;19:827–37. <https://doi.org/10.1094/MPMI-19-0827>.
161. Ludwig-Müller J. Plants and endophytes: equal partners in secondary metabolite production? *Biotechnol Lett*. 2015;37:1325–34.
162. Vandenkoornhuyse P, Quaiser A, Duhamel M, Le Van A, Dufresne A. The importance of the microbiome of the plant holobiont. *New Phytol*. 2015; 206:1196–206.
163. Verstraete B, van Elst D, Steyn H, van Wyk B, Lemaire B, Smets E, et al. Endophytic bacteria in toxic south african plants: identification, phylogeny and possible involvement in gossiekte. *PLoS One*. 2011;6:e19265.
164. Van Elst D, Nuyens S, van Wyk B, Verstraete B, Dessein S, Prinsen E. Distribution of the cardiotoxin pavettamine in the coffee family (Rubiaceae) and its significance for gossiekte, a fatal poisoning of ruminants. *Plant Physiol Biochem*. 2013;67:15–9. <https://doi.org/10.1016/j.plaphy.2013.02.022>.
165. Akone SH, Mándi A, Kurtán T, Hartmann R, Lin W, Daletos G, et al. Inducing secondary metabolite production by the endophytic fungus *Chaetomium* sp. through fungal–bacterial co-culture and epigenetic modification. *Tetrahedron*. 2016;72:6340–7.
166. do Nascimento JS, Silva FM, Magallanes-Noguera CA, Kurina-Sanz M, dos Santos EG, Caldas IS, et al. Natural trypanocidal product produced by endophytic fungi through co-culturing. *Folia Microbiol (Praha)*. 2020;65:323–8.
167. Stolyar S, Van Dien S, Hillesland KL, Pintel N, Lie TJ, Leigh JA, et al. Metabolic modeling of a mutualistic microbial community. *Mol Syst Biol*. 2007;3:1–14.
168. Naik S, Shaanker RU, Ravikanth G, Dayanandan S. How and why do endophytes produce plant secondary metabolites? *Symbiosis*. 2019;78:193–201.
169. Howitz KT, Sinclair DA. Xenohormesis: sensing the chemical cues of other species. *Cell*. 2008;133:387–91.
170. Banerjee P, Erehman J, Gohlke BO, Wilhelm T, Preissner R, Dunkel M. Super natural II—a database of natural products. *Nucleic Acids Res*. 2015;43:D935–9.
171. Bernardi DI, das Chagas FO, Monteiro AF, dos Santos GF, de Souza Berlinck RG. Isolation, synthesis, biosynthesis, and biological activities: secondary metabolites of endophytic actinomycetes; 2019.
172. Pye CR, Bertin MJ, Lokey RS, Gerwick WH, Lington RG. Retrospective analysis of natural products provides insights for future discovery trends. *Proc Natl Acad Sci U S A*. 2017;114:5601–6.
173. Boufridi A, Quinn RJ. Harnessing the properties of natural products. *Annu Rev Pharmacol Toxicol*. 2018;58:451–70.
174. Rutledge PJ, Challis GL. Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat Rev Microbiol*. 2015;13: 509–23.
175. Brakhage AA, Schroeckh V. Fungal secondary metabolites - strategies to activate silent gene clusters. *Fungal Genet Biol*. 2011;48:15–22. <https://doi.org/10.1016/j.fgb.2010.04.004>.
176. Pidroni A, Faber B, Brosch G, Bauer I, Graessle S. A class 1 histone deacetylase as major regulator of secondary metabolite production in *Aspergillus nidulans*. *Front Microbiol*. 2018;9:1–18.
177. Albarano L, Esposito R, Ruocco N, Costantini M. Genome mining as new challenge in natural products discovery. *Mar Drugs*. 2020;18:1–17.
178. Dias DA, Urban S, Roessner U. A historical overview of natural products in drug discovery. *Metabolites*. 2012;2:303–36.
179. Soliman SSM, Raizada MN. Interactions between co-habiting fungi elicit synthesis of Taxol from an endophytic fungus in host *Taxus* plants. *Front Microbiol*. 2013;4:1–14.
180. Ruiz-Sanchez J, Flores-Bustamante ZR, Dendooven L, Favela-Torres E, Soca-Chafre G, Galindez-Mayer J, et al. A comparative study of Taxol production in liquid and solid-state fermentation with *Nigrospora* sp. a fungus isolated from *Taxus globosa*. *J Appl Microbiol*. 2010;109:2144–50.
181. Hassani MA, Durán P, Hacquard S. Holobiont. *Encycl Syst Biol*; 2013. p. 902.
182. Caruana JC, Walper SA. Bacterial Membrane Vesicles as Mediators of Microbe – Microbe and Microbe – Host Community Interactions. *Front Microbiol*. 2020;11:1–24.
183. Aly AH, Debbab A, Proksch P. Fungal endophytes - secret producers of bioactive plant metabolites. *Pharmazie*. 2013;68:499–505.
184. Caicedo NH, Davalos AF, Puente PA, Rodríguez AY, Caicedo PA. Antioxidant activity of exo-metabolites produced by *Fusarium oxysporum*: An endophytic fungus isolated from leaves of *Otoba gracilipes*. *Microbiologyopen*. 2019;8:1–7.
185. Casella TM, Eparvier V, Mandavid H, Bendelac A, Odonne G, Dayan L, et al. Antimicrobial and cytotoxic secondary metabolites from tropical leaf endophytes: isolation of antibacterial agent pyrrocidine C from *Lewia infectoria* SNB-GTC2402. *Phytochemistry*. 2013;96:370–7. <https://doi.org/10.1016/j.phytochem.2013.10.004>.
186. Ma WJ, Schwander T. Patterns and mechanisms in instances of endosymbiont-induced parthenogenesis. *J Evol Biol*. 2017;30:868–88.
187. Sharma D, Pramanik A, Agrawal PK. Evaluation of bioactive secondary metabolites from endophytic fungus *Pestalotiopsis neglecta* BAB-5510 isolated from leaves of *Cupressus torulosa* D.Don. *3 Biotech*. 2016;6:1–14.
188. Rosa LH, Vieira MLA, Cota BB, Johann S, Alves TMA, Zani CL, et al. Endophytic fungi of tropical forests: a promising source of bioactive prototype molecules for the treatment of neglected diseases. In: *Drug Development - A Case Study Based Insight into Modern Strategies*; 2011. p. 469–86.
189. Mohammed SI, Patil MP, Patil RH, Maheshwari VL. Endophytes: Potential Source of Therapeutically Important Secondary Metabolites of Plant Origin. In: *Endophytes: Crop Productivity and Protection*; Springer; 2017. p. 95–110. <https://doi.org/10.1007/978-3-319-66544-3>.
190. Kusari S, Lamshöft M, Zühlke S, Spiteller M. An endophytic fungus from *Hypericum perforatum* that produces hypericin. *J Nat Prod*. 2008;71:159–62.
191. Kusari S, Pandey SP, Spiteller M. Untapped mutualistic paradigms linking host plant and endophytic fungal production of similar bioactive secondary metabolites. *Phytochemistry*. 2013;91:81–7. <https://doi.org/10.1016/j.phytochem.2012.07.021>.
192. Mori T, Cahn JKB, Wilson MC, Meoded RA, Wiebach V, Martinez AFC, et al. Single-bacterial genomics validates rich and varied specialized metabolism of uncultivated *Entotheonella* sponge symbionts. *Proc Natl Acad Sci U S A*. 2018;115:1718–23.
193. Stępień Ł, Lalak-Karćzugowska J, Witaszak N, Urbaniak M. So Close but So Far Away: *Fusarium* Secondary Metabolism Biosynthetic Pathways; 2020.
194. Blair PM, Land ML, Piatek MJ, Jacobson DA, Lu T-YS, Doktycz MJ, et al. Exploration of the Biosynthetic Potential of the *Populus* Microbiome. *mSystems*. 2018;3:1–17.
195. Lorenz N, Haarmann T, Pažoutová S, Jung M, Tudzynski P. The ergot alkaloid gene cluster: functional analyses and evolutionary aspects. *Phytochemistry*. 2009;70:1822–32.
196. Fleetwood DJ, Scott B, Lane GA, Tanaka A, Johnson RD. A complex ergovaline gene cluster in *Epichloë* endophytes of grasses. *Appl Environ Microbiol*. 2007;73:2571–9.

197. Young CA, Felitti S, Shields K, Spangenberg G, Johnson RD, Bryan GT, et al. A complex gene cluster for indole-diterpene biosynthesis in the grass endophyte *Neotyphodium lolii*. *Fungal Genet Biol*. 2006;43:679–93.
198. Staniek A, Woerdenbag HJ, Kayser O. *Taxomyces andreanae*: a presumed paclitaxel producer demystified? *Planta Med*. 2009;75:1561–6.
199. Kogel KH, Franken P, Hückelhoven R. Endophyte or parasite - what decides? *Curr Opin Plant Biol*. 2006;9:358–63.
200. Eaton CJ, Cox MP, Scott B. What triggers grass endophytes to switch from mutualism to pathogenism? *Plant Sci*. 2011;180:190–5. <https://doi.org/10.1016/j.plantsci.2010.10.002>.
201. Scherlach K, Hertweck C. Triggering cryptic natural product biosynthesis in microorganisms. *Org Biomol Chem*. 2009;7:1753–60.
202. Hwang S, Lee N, Cho S, Palsson B, Cho BK. Repurposing modular polyketide synthases and non-ribosomal peptide synthetases for novel chemical biosynthesis. *Front Mol Biosci*. 2020;7:1–27.
203. Nielsen ML, Isbrandt T, Petersen LM, Mortensen UH, Andersen MR, Hoof JB, et al. Linker flexibility facilitates module exchange in fungal hybrid PKS-NRPS engineering. *PLoS One*. 2016;11:1–18.
204. Gacek A, Strauss J. The chromatin code of fungal secondary metabolite gene clusters. *Appl Microbiol Biotechnol*. 2012;95:1389–404.
205. Dinesh R, Srinivasan V, Sheeja TE, Anandaraj M, Srmbikkal H. Endophytic actinobacteria: diversity, secondary metabolism and mechanisms to unsilence biosynthetic gene clusters. *Crit Rev Microbiol*. 2017;43:546–66.
206. Borah A, Thakur D. Phylogenetic and functional characterization of culturable endophytic actinobacteria associated with *Camellia* spp. for growth promotion in commercial tea cultivars. *Front Microbiol*. 2020;11:1–23.
207. Armeev GA, Gribkova AK, Pospelova I, Komarova GA, Shaytan AK. Linking chromatin composition and structural dynamics at the nucleosome level. *Curr Opin Struct Biol*. 2019;56:46–55.
208. Jeon J, Choi J, Lee GW, Park SY, Huh A, Dean RA, et al. Genome-wide profiling of DNA methylation provides insights into epigenetic regulation of fungal development in a plant pathogenic fungus. *Magnaporthe oryzae*. *Sci Rep*. 2015;5:1–11.
209. Shwab EK, Jin WB, Tribus M, Galehr J, Graessle S, Keller NP. Histone deacetylase activity regulates chemical diversity in *Aspergillus*. *Eukaryot Cell*. 2007;6:1656–64.
210. Reyes-Dominguez Y, Boedi S, Sulyok M, Wiesenberger G, Stoppacher N, Krska R, et al. Heterochromatin influences the secondary metabolite profile in the plant pathogen *Fusarium graminearum*. *Fungal Genet Biol*. 2012;49:39–47.
211. Collemare J, Seidl MF. Chromatin-dependent regulation of secondary metabolite biosynthesis in fungi: is the picture complete? *FEMS Microbiol Rev*. 2019;43(6):591–607.
212. Asai T, Chung YM, Sakurai H, Ozeki T, Chang FR, Wu YC, et al. Highly oxidized ergosterols and isariotin analogs from an entomopathogenic fungus, *Gibellula formosana*, cultivated in the presence of epigenetic modifying agents. *Tetrahedron*. 2012;68:5817–23.
213. Qadri M, Nalli Y, Jain SK, Chaubey A, Ali A, Strobel GA, et al. An insight into the secondary metabolism of *Muscodora yucatanensis*: small-molecule epigenetic modifiers induce expression of secondary metabolism-related genes and production of new metabolites in the Endophyte. *Microb Ecol*. 2017;73:954–65.
214. Venugopalan A, Srivastava S. Endophytes as in vitro production platforms of high value plant secondary metabolites. *Biotechnol Adv*. 2015;33:873–87. <https://doi.org/10.1016/j.biotechadv.2015.07.004>.
215. Carvalho TLG, Ballesteros HGF, Thiebaut F, Ferreira PCG, Hemery AS. Nice to meet you: genetic, epigenetic and metabolic controls of plant perception of beneficial associative and endophytic diazotrophic bacteria in non-leguminous plants. *Plant Mol Biol*. 2016;90:561–74.
216. Faust K, Raes J. Microbial interactions: from networks to models. *Nat Rev Microbiol*. 2012;10:538–50. <https://doi.org/10.1038/nrmicro2832>.
217. Manganiello G, Marra R, Staropoli A, Lombardi N, Vinale F, Nicoletti R. The shifting mycotoxin profiles of endophytic *Fusarium* strains: a case study. *Agric*. 2019;9:1–13.
218. Akone SH, Pham C-D, Chen H, Ola ARB, Ntie-Kang F, Proksch P. Epigenetic modification, co-culture and genomic methods for natural product discovery. *Phys Sci Rev*. 2018;4:1–13. <https://doi.org/10.1515/psr-2018-0118>.
219. Brakhage AA. Regulation of fungal secondary metabolism. *Nat Rev Microbiol*. 2013;11:21–32.
220. Hua C, Zhao JH, Guo HS. Trans-kingdom RNA silencing in plant–fungal pathogen interactions. *Mol Plant*. 2018;11:235–44. <https://doi.org/10.1016/j.molp.2017.12.001>.
221. Knip M, Constantin ME, Thordal-Christensen H. Trans-kingdom cross-talk: small RNAs on the move. *PLoS Genet*. 2014;10:e1004602.
222. Zeng G, Jiang Y, Gong Z. Cross-kingdom small RNAs among animals, plants and microbes. *Cells*. 2019;8:371.
223. Aguilar C, Mano M, Eulalio A. MicroRNAs at the host–bacteria interface: host defense or bacterial offense. *Trends Microbiol*. 2019;27:206–18. <https://doi.org/10.1016/j.tim.2018.10.011>.
224. Wang L, Xu X, Yang J, Chen L, Liu B, Liu T, et al. Integrated microRNA and mRNA analysis in the pathogenic filamentous fungus *Trichophyton rubrum*. *BMC Genomics*. 2018;19:1–14.
225. Jin Y, Zhao JH, Zhao P, Zhang T, Wang S, Guo HS. A fungal miRNA mediates epigenetic repression of a virulence gene in *Verticillium dahliae*. *Philos Trans R Soc B Biol Sci*. 2019;374:20180309.
226. Baral B, Akhgari A, Metsä-Ketelä M. Activation of microbial secondary metabolic pathways: avenues and challenges. *Synth Syst Biotechnol*. 2018;3:163–78. <https://doi.org/10.1016/j.synbio.2018.09.001>.
227. Frisvad JC. Fungal secondary metabolism. *Fungal Sec Metab Methods Protoc Methods Mol Biol*. 2012;944:47–58. <https://doi.org/10.1007/978-1-62703-122-6>.
228. El-Sayed ASA, Mohamed NZ, Safan S, Yassin MA, Shaban L, Shindia AA, et al. Restoring the taxol biosynthetic machinery of *Aspergillus terreus* by *Podocarpus gracilior* pilger microbiome, with retrieving the ribosome biogenesis proteins of WD40 superfamily. *Sci Rep*. 2019;9:1–12. <https://doi.org/10.1038/s41598-019-47816-y>.
229. Anyaogu DC, Mortensen UH. Heterologous production of fungal secondary metabolites in *Aspergilli*. *Front Microbiol*. 2015;6:1–6.
230. Clevenger KD, Bok JW, Ye R, Miley GP, Verdan MH, Velk T, et al. A scalable platform to identify fungal secondary metabolites and their gene clusters. *Nat Chem Biol*. 2017;17:139–48.
231. Chen YP, Tseng CP, Liaw LL, Wang CL, Chen IC, Wu WJ, et al. Cloning and characterization of monacolin K biosynthetic gene cluster from *Monascus pilosus*. *J Agric Food Chem*. 2008;56:5639–46.
232. Harvey CJB, Tang M, Schlecht U, Horecka J, Fischer CR, Lin HC, et al. HEx: a heterologous expression platform for the discovery of fungal natural products. *Sci Adv*. 2018;4:ear5459.
233. Corre C, Song L, O'Rourke S, Chater KF, Challis GL. 2-Alkyl-4-hydroxymethylfuran-3-carboxylic acids, antibiotic production inducers discovered by *Streptomyces coelicolor* genome mining. *Proc Natl Acad Sci U S A*. 2008;105:17510–5.
234. Guo F, Xiang S, Li L, Wang B, Rajasärkkä J, Gröndahl-Yli-Hannuksela K, et al. Targeted activation of silent natural product biosynthesis pathways by reporter-guided mutant selection. *Metab Eng*. 2015;28:134–42. <https://doi.org/10.1016/j.ymben.2014.12.006>.
235. Zheng Y, Ma K, Lyu H, Huang Y, Liu H, Liu L, et al. Genetic manipulation of the COP9 Signalosome subunit PFCsE leads to the discovery of pestaloficins in *Pestalotiopsis fici*. *Org Lett*. 2017;19:4700–3.
236. Hoffmeister D, Keller NP. Natural products of filamentous fungi: enzymes, genes, and their regulation. *Nat Prod Rep*. 2007;24:393–416.
237. Bills G, Li Y, Chen L, Yue Q, Niu XM, An Z. New insights into the echinocandins and other fungal non-ribosomal peptides and peptaibiotics. *Nat Prod Rep*. 2014;31:1348–75.
238. Starcevic A, Zucko J, Simunkovic J, Long PF, Cullum J, Hranueli D. ClustScan: An integrated program package for the semi-automatic annotation of modular biosynthetic gene clusters and in silico prediction of novel chemical structures. *Nucleic Acids Res*. 2008;36:6882–92.
239. Ichikawa N, Sasagawa M, Yamamoto M, Komaki H, Yoshida Y, Yamazaki S, et al. DoBISCUIT: a database of secondary metabolite biosynthetic gene clusters. *Nucleic Acids Res*. 2013;41:408–14.
240. Conway KR, Boddy CN. ClusterMine360: a database of microbial PKS/NRPS biosynthesis. *Nucleic Acids Res*. 2013;41:402–7.
241. Khaldi N, Seifuddin FT, Turner G, Haft D, Nierman WC, Wolfe KH, et al. SMURF: genomic mapping of fungal secondary metabolite clusters. *Fungal Genet Biol*. 2010;47:736–41.
242. Medema MH, Kottmann R, Yilmaz P, Cummings M, Biggins JB, Blin K, et al. Minimum information about a biosynthetic gene cluster. *Nat Chem Biol*. 2015;11:625–31. <https://doi.org/10.1038/nchembio.1890>.
243. Epstein SC, Charkoudian LK, Medema MH. A standardized workflow for submitting data to the minimum information about a biosynthetic gene

- cluster (MIBiG) repository: prospects for research-based educational experiences. *Stand Genomic Sci.* 2018;13:1–13.
244. Adamek M, Alanjary M, Ziemert N. Applied evolution: phylogeny-based approaches in natural products research. *Nat Prod Rep.* 2019;36:1295–312.
 245. Basalla J, Chatterjee P, Burgess E, Khan M, Verbrugge E, Wiegmann DD, et al. Loci encoding compounds potentially active against drug-resistant pathogens amidst a decreasing Pool of novel antibiotics. *Appl Environ Microbiol.* 2019;85:1–17.
 246. Agren R, Liu L, Shoaie S, Vongsangnak W, Nookaew I, Nielsen J. The RAVEN toolbox and its use for generating a genome-scale metabolic model for *Penicillium chrysogenum*. *PLoS Comput Biol.* 2013;9:e1002980.
 247. Wang H, Marcišauskas S, Sánchez BJ, Domenzain I, Hermansson D, Agren R, et al. RAVEN 2.0: a versatile toolbox for metabolic network reconstruction and a case study on *Streptomyces coelicolor*. *PLoS Comput Biol.* 2018;14:1–17.
 248. Latendresse M, Krummenacker M, Trupp M, Karp PD. Construction and completion of flux balance models from pathway databases. *Bioinformatics.* 2012;28:388–96.
 249. Theobald S, Vesth TC, Rendsvig JK, Nielsen KF, Riley R, de Abreu LM, et al. Uncovering secondary metabolite evolution and biosynthesis using gene cluster networks and genetic dereplication. *Sci Rep.* 2018;8:1–12.
 250. Opatovsky I, Santos-Garcia D, Ruan Z, Lahav T, Ofaim S, Mouton L, et al. Modeling trophic dependencies and exchanges among insects' bacterial symbionts in a host-simulated environment. *BMC Genomics.* 2018;19:1–14.
 251. Ravikrishnan A, Blank LM, Srivastava S, Raman K. Investigating metabolic interactions in a microbial co-culture through integrated modelling and experiments. *Comput Struct Biotechnol J.* 2020;18:1249–58.
 252. Ravikrishnan A, Nasre M, Raman K. Enumerating all possible biosynthetic pathways in metabolic networks. *Sci Rep.* 2018;8:1–11. <https://doi.org/10.1038/s41598-018-28007-7>.
 253. Koo T, Lee J, Hwang S. Development of an interspecies interaction model: An experiment on *Clostridium cadaveris* and *Clostridium sporogenes* under anaerobic condition. *J Environ Manage.* 2018;2019(237):247–54.
 254. Chapelle E, Alunni B, Malfatti P, Solier L, Pédrón J, Kraepiel Y, et al. A straightforward and reliable method for bacterial in planta transcriptomics: application to the *Dickeya dadantii*/Arabidopsis thaliana pathosystem. *Plant J.* 2015;82:352–62.
 255. Kuang X, Sun S, Wei J, Li Y, Sun C. Iso-Seq analysis of the *Taxus cuspidata* transcriptome reveals the complexity of Taxol biosynthesis. *BMC Plant Biol.* 2019;19:1–16.
 256. Petřijová L, Jurčáková Z, Čellárová E. Computational screening of miRNAs and their targets in leaves of *Hypericum* spp. by transcriptome-mining: a pilot study. *Planta.* 2020;251. <https://doi.org/10.1007/s00425-020-03342-0>.
 257. Toju H, Tanabe AS, Sato H. Network hubs in root-associated fungal metacommunities. *Microbiome.* 2018;6:1–16.
 258. Cimermancic P, Medema MH, Claesen J, Kurita K, Brown LCW, Mavrommatis K, et al. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell.* 2014;158:412–21. <https://doi.org/10.1016/j.cell.2014.06.034>.insights.
 259. Rammer W, Seidl R. Harnessing Deep Learning in Ecology: An Example Predicting Bark Beetle Outbreaks. *Front Plant Sci.* 2019;10:1–9.
 260. Kopp W, Monti R, Tamburrini A, Ohler U, Akalin A. Deep learning for genomics using Janggu. *Nat Commun.* 2020;11:3488. <https://doi.org/10.1038/s41467-020-17155-y>.
 261. Mishra B, Kumar N, Mukhtar MS. Systems biology and machine learning in plant–pathogen interactions. *Mol Plant Microbe Interact.* 2019;32:45–55.
 262. Rayan A, Raijn J, Falah M. Nature is the best source of anticancer drugs: indexing natural products for their anticancer bioactivity. *PLoS One.* 2017;12:1–12.
 263. Lusci A, Pollastri G, Baldi P. Deep architectures and deep learning in chemoinformatics: the prediction of aqueous solubility for drug-like molecules. *J Chem Inf Model.* 2013;53:1563–75.
 264. Mitchell JBO. Machine learning methods in chemoinformatics. *Wiley Interdiscip Rev Comput Mol Sci.* 2014;4:468–81.
 265. Stokes JM, Yang K, Swanson K, Jin W, Cubillos-Ruiz A, Donghia NM, et al. A deep learning approach to antibiotic discovery. *Cell.* 2020;180:688–702.e13. <https://doi.org/10.1016/j.cell.2020.01.021>.
 266. Martinez-Mayorga K, Madariaga-Mazon A, Medina-Franco JL, Maggiora G. The impact of chemoinformatics on drug discovery in the pharmaceutical industry. *Expert Opin Drug Discovery.* 2020;15:293–306.
 267. Gangadevi V, Muthumary J. A novel endophytic taxol-producing fungus *Chaetomella raphigera* isolated from a medicinal plant. *Terminalia arjuna*. *Appl Biochem Biotechnol.* 2009;158:675–84.
 268. Jasim B, Geethu PR, Mathew J, Radhakrishnan EK. Effect of endophytic bacillus sp. from selected medicinal plants on growth promotion and diosgenin production in *Trigonella foenum-graecum*. *Plant Cell Tiss Org Cult.* 2015;122:565–72.
 269. Wang B, Guo F, Dong SH, Zhao H. Activation of silent biosynthetic gene clusters using transcription factor decoys. *Nat Chem Biol.* 2019;15:111–4.
 270. Rodriguez PA, Rothballer M, Chowdhury SP, Nussbaumer T, Gutjahr C, Falter-Braun P. Systems biology of plant-microbiome interactions. *Mol Plant.* 2019;12:804–21. <https://doi.org/10.1016/j.molp.2019.05.006>.
 271. Sbaraini N, Andreis FC, Thompson CE, Guedes RLM, Junges A, Campos T, et al. Genome-wide analysis of secondary metabolite gene clusters in *Ophiostoma ulmi* and *Ophiostoma novo-ulmi* reveals a fujikurin-like gene cluster with a putative role in infection. *Front Microbiol.* 2017;8:1–12.
 272. Malhadas C, Malheiro R, Pereira JA, Guedes de Pinho P, Baptista P. Antimicrobial activity of endophytic fungi from olive tree leaves. *World J Microbiol Biotechnol.* 2017;33:46.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

