This is the accepted version of a paper accepted in *Phytochemistry Reviews* that appeared online on April 24, 2019 at https://doi.org/10.1007/s11101-019-09606-2.

The landscape of natural product diversity and their pharmacological relevance from a focus on the *Dictionary of Natural Products*®

François Chassagne, Guillaume Cabanac, Gilles Hubert, Bruno David, Guillaume Marti

F. Chassagne (*)

Center for the Study of Human Health, Emory University

615 Michael Street, Whitehead Building, Room 111

Atlanta, GA, 30322, USA.

email: francois.chassagne@emory.edu

F. Chassagne, G. Cabanac, G. Hubert, G. Marti

Informatics Research Institute of Toulouse (IRIT), Computer Science Department,

UMR 5505 CNRS, Université Toulouse 3 – Paul Sabatier,

118, Route de Narbonne, 31062 Toulouse, France

B. David

Green Mission Pierre Fabre

Institut de Recherche Pierre Fabre

3, Avenue Hubert Curien - BP 13562, 31562 Toulouse, France

G. Marti

PharmaDev, UMR 152 IRD, Université Toulouse 3 – Paul Sabatier,

35 Chemin des Maraîchers, 31400 Toulouse, France

Abstract

Nature is considered a prolific source of diverse biologically active chemotypes. While most reviews have focused on the characteristics of the chemical backbones of natural products (NPs), few have tried to provide an overview of their origins (the living organisms in which they are produced), chemical classes, and biological activities. This review discusses the current knowledge on NP diversity by focusing on the *Dictionary of Natural Products*® (DNP). We datamined the 300,000 NPs covered by the DNP to reveal relevant, albeit dormant, knowledge about NP diversity. This holistic picture of NPs allows us to discuss the most abundant biological sources of NPs investigated in relation to their chemical features and biological activities. In a nutshell, a large part of NPs originated from plants (67%), especially from the Compositae and Leguminosae families. Among all kingdoms, NPs isolated from *Streptomyces* spp. were largely represented, while terpenoids and alkaloids were the two most represented chemical classes. Out of all NPs documented, only 3,882 were reported to be bioactive (1,163 from plants and 1,006 from bacteria), with antibacterial, antibiotics, and antineoplastic agents being the most frequent therapeutic classes. In this paper, we also address the advantages and limitations of NP research from a pharmaceutical industry perspective. This work will provide useful insights and guidance to researchers involved in drug discovery from NPs.

Keywords

Biological activity, Drug Discovery, Genetic resources, Pharmaceutical industry, Plants

Abbreviations

Angiotensin Converting Enzyme	ACE
Chemical Abstracts Services	CAS
Dictionary of Natural Products®	DNP
HTS	High Throughput Screening
NP	Natural Product
PSK	Polysaccharide Krestin
PSP	Polysaccharopeptide

Introduction

Natural products (NPs) have been used in various branches of traditional medicine for millennia. The oldest written records of NPs as medicine date as far back as 4,600 years ago (Dias et al. 2012; Cragg and Newman 2013). Due to their chemical diversity, their structural complexity, and their biological selectivity, NPs are considered to be a great source of inspiration for the development of potential novel drugs (Clardy and Walsh 2004; Atanasov et al. 2015). From 1981 to 2014, about 51% of all new approved drugs were of (or derived from) natural sources, amounting to 65% of all antibacterial compounds and 73% of all anticancer compounds (Newman and Cragg 2016). It is expected that newly discovered natural compounds will continue to play an important role in drug discovery (Baker et al. 2007; Harvey et al. 2015).

Natural product databases have been developed to assist with *in silico* and *in vitro* screening in drug discovery. A number of NP libraries are already available and include general databases such as *SuperNatural 2* (≈ 326,000 NP) (Banerjee et al. 2015) and the *Universal Natural Products Database* (≈ 229,000 compounds) (Chen et al. 2017b). Specialized databases have also been designed, including those focusing on indigenous medicines such as *AfroDb* for African medicinal plants (Ntie-Kang et al. 2013), *NuBBE* for Brazilian biodiversity (Valli et al. 2013), *iSMART*, and *TCMID* for traditional Chinese medicine (Chang et al. 2011; Xue et al. 2013); and those focusing on specific types of NPs such as *HIT* for herbal ingredients and their targets (Ye et al. 2011), *SuperToxic* for toxic compounds (Schmidt et al. 2009), *NPACT* for anticancer NPs (Mangal et al. 2013), and *MarinLit* for marine NPs (Chen et al. 2017b).

The *Dictionary of Natural Products*® (CRC Press, v. 27.1) (DNP) is a compilation of all known compounds derived from natural sources and can be considered as one of the most comprehensive libraries of NPs available to date (Quinn et al. 2008; Gaudêncio and Pereira 2015; Chen et al. 2017b). The latest version of the DNP (v. 27.1, at the time of writing) provides data for nearly 300,000 natural compounds, and it contains information on the chemical, physical, and biological properties of compounds; along with their systematic and common names, literature references, molecular structures, and natural sources for purification (including family, genus, and species). Due to its rich content, the DNP can be considered as a body of knowledge for NPs and can be used to guide investigations in NP-based drug discovery.

Since its development, several studies have used and explored the content of the DNP. One of the first comprehensive reports was published by Henkel et al. (1999) who investigated the differences between the structural properties of NPs found in the DNP and those of synthetic substances; then Whittle et al. (2003) evaluated various similarity measures for screening molecular fingerprints from the DNP; later Koch et al. (2005) introduced a structural classification of NPs to chart biologically relevant chemical space. With the same objective, Rosén et al. (2009) compared the chemical space of bioactive medicinal chemistry compounds from the WOMBAT database to that of products from the DNP; in 2008, Quinn et al. (2008) used the DNP to develop a library of NPs that exhibit drug-like properties; also Kong et al. (2011) performed a historical analysis on the structural novelty of products from the DNP; more recently, Pascolutti et al. (2015) identified fragment-sized NPs from the DNP to capture the structural diversity of nature. These studies mainly focused on the characteristics of the chemical backbones of the NP

contained in this database, but none of them strove to provide an overview of their origins (the living organisms in which they are produced), chemical classes, and biological activities. To the best of our knowledge, only Bérdy (2005) summarized compounds isolated from natural sources, but his review was limited to bioactive compounds from microbial sources.

In this review, we unveil the current knowledge on NP diversity based on a descriptive study of the DNP (v. 27.1). This includes a characterization of all compounds (primary and secondary metabolites) from the DNP in relation to their natural sources (all types of organisms are considered), chemical classes, and biological activities (NPs without bioactivities are also reviewed). Natural products with biological activities annotated in the DNP are defined as "compounds with established activity and being used as drugs or under investigation for drug use" (Taylor & Francis group, pers. comm.). Finally, we also discuss the potential application of NPs in the pharmaceutical industry.

Defining the number of NPs already identified

A key issue in NP research is the estimation of the number of known NPs. First, Bérdy (2005) estimated the approximate number of known NPs to be close to one million. Later, the same author proposed that the number of published natural compounds was actually between 300,000 and 600,000 (Bérdy 2012). However, little information was provided on the methodology used to infer these figures. Other authors based their estimations on the number of compounds tabulated in the NP databases (Blunt et al. 2012). At present, in the general NP databases, the number of compounds described amounts to 326,000 NPs in *Super Natural 2*, 293,798 NPs in the DNP, 283,000 NPs in the *Chemical Abstract Services (CAS) registry*, and 220,000 NPs in *Reaxys* (Gaudêncio and Pereira 2015; Chen et al. 2017b). By comparing the number of unique NPs from a wide range of commercial and freely available virtual databases, Chen et al. (2017b) estimated a figure close to 250,000 NPs. Since the authors did not include *SuperNatural 2* and *Reaxys* in their analysis, the number of known NPs is likely to be higher. Furthermore, the number of NPs could even be larger if we consider the molecular fossils produced from natural products (Falk and Wolkenstein 2017).

In the DNP, out of the 194,977 NPs with information on organism classification, most of the NPs were from the Plantae kingdom (133,881 NPs, 67.3%), while Animalia ranked second (25,064 NPs, 12.6%), followed by Fungi (19,869 NPs, 9.99%) and Eubacteria (17,531, 8.81%) (Figure 1a). Regarding the chemistry of NPs, terpenoids and alkaloids groups were the two most represented chemical classes in these four kingdoms, and they represented more than half of all compounds isolated from the Plantae kingdom (Figure 1b). In contrast, some chemical classes did not have a cross distribution in the

different kingdoms of life. This is the case of steroid NPs which were seldomly represented in the Bacteria kingdom. Likewise, flavonoids and lignans were almost exclusively found in Plantae kingdom, while polyketides were mainly found in Bacteria, Fungi, and Protista kingdoms, which emphasize two distinct crossroads among the acetate pathway.

In the following sections, we provide a detailed analysis of the chemodiversity found in the most represented families and species of each kingdom with a special focus on their pharmacologically active compounds.

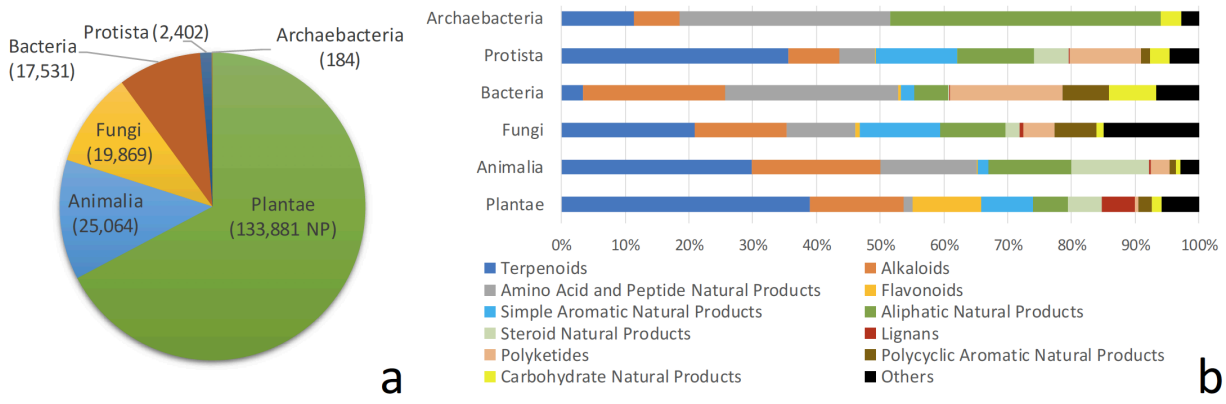

Figure 1. Overview of the Dictionary of Natural Products database (a) Distribution of natural products per kingdom of life; (b) Distribution of the main chemical classes of natural products in each kingdom of life.

NPs from the Plantae kingdom

A large percentage of the NPs (67.3%) recorded in the DNP are from the Plantae kingdom. This is consistent with Bérdy (2012) who estimated that approximately 70% of NP are of plant origin. From a historical perspective, this is not surprising since medicinal herbs were the first natural substances to be studied. It started with the isolation of morphine (alkaloid) in 1817 by a German pharmacist named Friedrich Sertuerner (Sertuerner, 1817). It gained momentum during the nineteenth century with the discovery of a wide range of natural compounds of plant origin (e.g., aspirin, atropine, caffeine, cocaine, colchicine, quinine, nicotine, and strychnine). Then, the discovery of penicillin in 1928 by Alexander Fleming and its purification by Howard Florey in 1938 led to an era of natural products discovery from microbial sources (David et al. 2015; Atanasov et al. 2015; Bernardini et al. 2017).

One of the distinctive features of the plant compounds compiled in the DNP is the high proportion of terpenoids, which are especially pervasive in the Gymnosperms group (57% of the total NPs) and in the Dicotyledons group (40%) (Figure 1b). Terpenoids are known as the largest and the most diverse class of NPs, and they play various roles in plant development and growth, as well as in communication and defence (Langenheim 1994; Gershenzon and Dudareva 2007). In the DNP, one third of terpenoid NPs have antineoplastic-related activities including, for instance (Huang et al. 2012):

- Limonene, a monoterpene from *Citrus limon* (L.) Osbeck.
- Tanshinone IIA, a diterpene from *Salvia miltiorrhiza* Bunge.
- Celastrol, a triterpene from *Tripterygium wilfordii* Hook. f.
- Lycopene, a tetraterpene from *Lycopersicon esculentum* Mill.

Most of the NPs isolated from the Plantae kingdom are from the dicotyledons group (Angiosperms clade) (83.7%), followed by the monocotyledons group (8.1%), and the Gymnosperms clade (3%). Liverworts, ferns, fern allies, and mosses represent only a small part of the NP recorded (3.2%), just as rhodophyta and chlorophytes groups (2%). Furthermore, three botanical families (Compositae, Leguminosae, and Labiatae) host about one quarter of the total compounds from the Plantae kingdom (Figure 2). With over 32,700 species recorded, Compositae (also known as Asteraceae) is the largest family of flowering plants worldwide, while Leguminosae (or Fabaceae) is the third-largest with more than 20,800 species (Roskov et al. 2018).

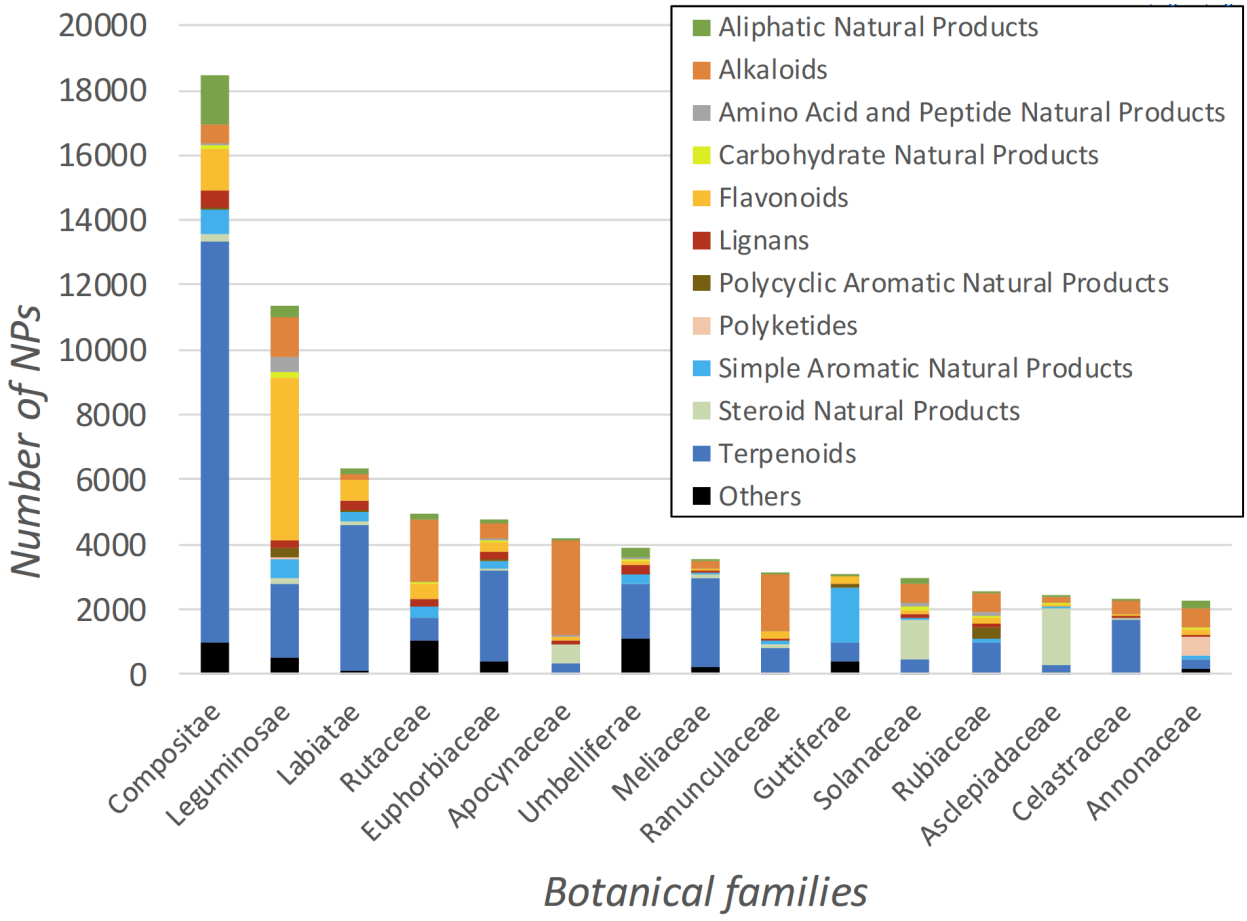

Figure 2. The top 15 botanical families containing natural products and the distribution of the different chemical classes.

In the Compositae family, the *Artemisia* genus, one of the largest genera in this family (≈ 500 species), was the most represented in the DNP (1,299 NP) (Suppl. Mat. S1A). The antimalarial compound, artemisinin, isolated from the traditional Chinese medicinal herb *Artemisia annua* L. in 1971 led to the development of various derivatives (e.g., artemether and sodium artesunate), and it has also attracted renewed interest into sesquiterpenes which are well represented in this genus (Kayser et al. 2003; Atanasov et

al. 2015). For example, arglabin isolated from *A. glabella* Kar. & Kir. is used in Kazakhstan for cancer chemotherapy, and santonin produced by *A. cina* Berg ex Poljakov was widely used in the past as an anthelminthic agent (Ivanescu et al. 2015).

According to the DNP, about half of the compounds found in the Leguminosae family are flavonoids. Indeed, a wide range of flavonoids are found in this botanical family including quercetin, kaempferol, and their derivatives (Wink 2013). Some of which are currently available or under investigation as drugs, such as genistein from *Glycine max* (L.) Merr. which is in clinical trials for use as an angiogenesis inhibitor (Veitch 2010; Russo et al. 2016). Other compounds isolated from the Leguminosae family that are of major importance in human health include the oleanane triterpenoids from *Glycyrrhiza glabra* L. (e.g., glycyrrhizin used as anti-inflammatory agent), polycyclic aromatic NP from *Cassia angustifolia* M.Vahl (i.e., sennosides used as laxatives), and cyclotryptamine alkaloids from *Physostigma venenosum* Balf. (i.e., physostigmine used to treat neurological diseases). Altogether, the Leguminosae family is known as the most drug-prolific botanical family with 44 drugs either approved or in clinical trials in 2011 (Zhu et al. 2011).

The Labiatae family (or Lamiaceae) is the third largest botanical group, and according to the DNP a large number of the compounds (71%) isolated from this family are terpenoids. These results could be explained by the numerous reports on volatile oils (mainly monoterpenes and sesquiterpenes) found in genera of economic importance (e.g., *Lavandula*, *Mentha* and *Thymus*) (Wu et al. 2012). *Salvia* species account for almost 20% of the overall NPs isolated from Labiatae, with *Salvia miltiorrhiza* Bunge being the most represented. *S. miltiorrhiza* (Danshen in Chinese) is a very popular traditional Chinese herb used to treat several conditions including cardiovascular, cerebrovascular,

and hyperlipidaemia diseases (Wu et al. 2012). Lipophilic diterpenoids (i.e., tanshinones) and hydrophilic phenolic acids (i.e., salvianolic acids) are the main bioactive constituents of this species. These compounds have been shown to possess various pharmacological effects ranging from anticancer properties to effects on cardiovascular diseases (Su et al. 2015). More than thirty clinical trials of *S. miltiorrhiza* and its tanshinones have been undertaken to justify its use in stroke patients, angina, and other ischemic conditions (Adams et al. 2006; Tan et al. 2018). Moreover, Dantonic, a Chinese botanical drug containing *S. miltiorrhiza* extracts, has gone through Phase III clinical trials (completed in 2016) by the U.S. Food and Drug Administration (FDA) for the treatment of angina pectoris and cardiovascular diseases (Chao et al. 2017).

Of the most cited genera in the DNP, the *Euphorbia* genus ranks first in the Plantae kingdom and fourth amongst all kingdoms. With about 2,160 known species, *Euphorbia* is the largest genus in the Euphorbiaceae family and the third-largest amongst the flowering plants (Jassbi 2006; Ernst et al. 2015). In the DNP, approximately two thirds of the total NPs (1,008 compounds) isolated from this genus are diterpenoids. The diterpenoids are the most studied chemical class from *Euphorbia* species, especially the polycyclic diterpenoids (e.g., jatrophane, ingenane, tigliane, and lathyrane) (Shi et al. 2008). Indeed, these compounds are taxonomic markers of the Euphorbiaceae family (they occur only in Thymelaeaceae and Euphorbiaceae families), and are lead compounds in drug discovery from NPs (Vasas and Hohmann 2014). One diterpenoid compound (i.e., ingenol mebutate isolated from *E. peplus*) was approved by the FDA and the European Medicines Agency in 2012 for the treatment of actinic keratosis, and it is currently used in clinical practice (Cantisani et al. 2013).

Interestingly, *Tripterygium wilfordii* Hook. f. (Celastraceae) was the Plantae species that exhibited the highest number of NPs in the DNP (315) (Table 1). This liana has been used in traditional Chinese medicine for more than 2,000 years for the treatment of arthritis, muscle, and skeletal injury, as well as skin diseases (Chen 2001). In the 1960s, formulations of *T. wilfordii* began to be used in Chinese allopathic medicine to treat patients with inflammatory lesions caused by leprosy, and patients with rheumatoid arthritis (Tao and Lipsky 2000). Since then, numerous clinical trials have been conducted for the treatment of inflammatory diseases, and over 300 compounds have been identified from this plant, many of which displaying immunosuppressive and anti-inflammatory effects (Brinker et al. 2007). Its most abundant metabolites are diterpenoids including triptolide, tripdiolide, and triptonide (Goldbach-Mansky 2009). Triptolide analogues are currently in Phase II clinical trials for the treatment of rheumatoid arthritis and cancer (Chen et al. 2018).

Table 1: Top 15 natural sources (at the species level) of natural products.

Biological sources (genus)	Kingdom of life	Number and frequency of NP		Chemical class[a]		Biological activities[b]			
		N	%	Top 1	%	*N*	Top 1	%	Examples
Escherichia coli	Bacteria	487	2.2	AA	79.4	12	Antineop.	29.4	asparaginase
Streptomyces hygroscopicus	Bacteria	353	1.6	Polyk	58.1	44	Antibact.	70.5	nigericin, rapamycin
Tripterygium wilfordii	Plantae	315	1.4	Terp	71.1	7	Antineop.	36.4	celastrol, triptolide
Azadirachta indica	Plantae	278	1.3	Terp	84.8	6	Insectic.	25.0	azadirachtin
Helianthus annuus	Plantae	263	1.2	Terp	45.7	0	NA		
Bacillus subtilis	Bacteria	252	1.1	AA	61.7	14	Antibact.	35.0	bacitracin, rhizocticin
Nicotiana tabacum	Plantae	248	1.1	Terp	22.9	1	Gang. block.	50.0	nicotine
Aspergillus terreus	Fungi	244	1.1	Alk	20.8	9	Antibact.	33.3	patulin
Ganoderma lucidum	Fungi	239	1.1	Terp	86.9	1	Immuno.	100	ling zhi-8
Gynostemma pentaphyllum	Plantae	239	1.1	Terp	95.4	1	Antineop.	100	protopanaxadiol
Camellia sinensis	Plantae	236	1.1	Terp	37.1	2	Antifung.	50.0	octanoic acid
Panax ginseng	Plantae	221	1.0	Terp	71.6	4	Antineop.	100	ginsenosides
Saccharomyces cerevisiae	Fungi	220	1.0	AA	83.6	4	Antianaem.	20.0	sargramostim
Lyngbya majuscula	Bacteria	220	1.0	AA	38.2	2	Antifung.	33.3	laxaphycin B
Sorangium cellulosum	Bacteria	214	1.0	Polyk	48.7	2	Antineop.	100	epothilones

[a]Most represented chemical class in each genus with its percentage. AA=Amino Acid and Peptide; Alk=Alkaloids; Polyk=Polyketides; Terp=Terpenoids

[b]Number (N) of NP (for each species) with established activity and being used as drugs or under investigation for drug use, and most represented biological activities (with their percentage) in each genus. Antianaem.=Antianaemic agents; Antibact.=Antibacterial agents and Antibiotics; Antifung.=Antifungal agents; Antineop.=Antineoplastics agents; Gang. block.=Ganglion blocking agents; Immuno.=Immunomodulators; Insectic.=Insecticides

NA = Not Applicable

NPs from the Bacteria kingdom

According to the DNP, the number of NP from the Bacteria kingdom is less than that from the other kingdoms of life (only 9% of the total compounds from living organisms). However, the percentage of bioactive compounds from the Bacteria kingdom is three to five times higher than that from the other kingdoms of life (Suppl. Mat. S1B). Bérdy (2012) estimated that approximately 8% of NPs were from bacterial origin (actinobacteria and unicellular bacteria), but 47% of these compounds exhibited some kind of biological activity. Interestingly, the latter figure was lower in plants (7%) and in animal-derived compounds (3%).

Bioactive compounds from bacterial origin mainly display antibiotic and antibacterial properties (67% of bioactive NPs from the Bacteria kingdom in the DNP) (Figure 3). Regarding the number of FDA approved drugs from natural sources, Patridge et al. (2016) reported that 51% of antibacterial agents were from bacterial origin.

Many NPs from bacteria belong to the amino acid and peptide class (Figure 1). In the DNP, 27% of NPs from bacterial origin were amino acids and peptides, and cyclic-oligo- and polypeptides were the main represented chemical subclass with 1,147 NPs. Moreover, the amino acid and peptide group (236 NPs) and the polyketide group (248 NPs) combined amount to approximately half of the bioactive NPs from bacterial sources.

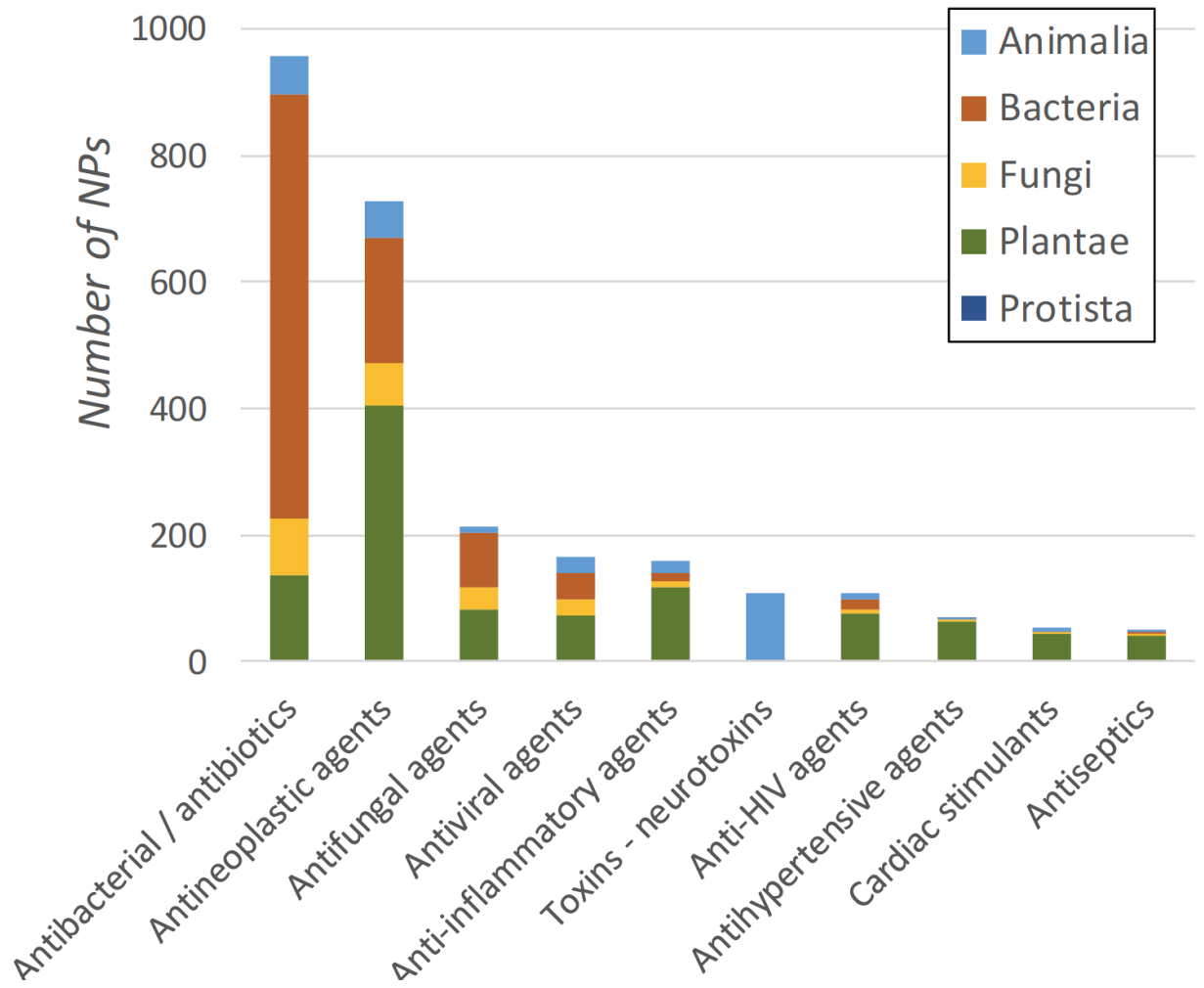

Figure 3. The top 10 activities of natural products and their distribution by kingdom of life.

Among all living organisms, the Streptomycetaceae family and *Streptomyces* species rank top as producers of NPs (7,951 NPs in total from *Streptomyces*) (Table 2). Bérdy (2005) estimated that 34% of metabolites from microbial sources were from this genus. In our study, we found that this number is even higher and reaches 45%. Indeed, with the discovery of streptothricin in 1942 and streptomycin in 1943 at the beginning of the antibiotic era, attention turned to *Streptomyces* species. Pharmaceutical companies

engaged in large efforts to discover NPs by developing research programs based on microbial fermentation, and by focusing mainly on antibacterial and antifungal targets (Clardy et al. 2006; Baker et al. 2007; Katz and Baltz 2016). In the 1950s and 1960s, about 70-80% of antibiotics were discovered from *Streptomyces* species and this percentage was still high at the beginning of the 1990s (≈ 40-50%) (Bérdy 2005). Several important compounds belonging to various antibiotic classes were discovered from *Streptomyces* species including β-lactams (cephamycin and carbapenems); aminoglycosides (neomycin and kanamycin); macrolides (tylosin and spiramycin); peptides (actinomycin); polyenes (candicidin, amphotericin B and nystatin); and tetracyclines (tetracycline, chlortetracycline and oxytetracycline) (Katz and Baltz 2016). Besides the discovery of antibiotics, a wide range of compounds related to various therapeutic classes have been isolated from this genus. Let us mention anthelminthic and antiparasitic drugs (e.g., ivermectins), anti-tumour products (e.g., bleomycin and doxorubicin), anti-obesity agents (e.g., lipstatin), immunosuppressive agents (e.g., rapamycin) and herbicides (e.g., bialaphos). Altogether, Watve et al. (2001) estimated that this genus is capable of producing approximately 100,000 antimicrobial compounds, of which only 1-3% have been discovered so far.

Table 2: Top 15 natural sources (at the genus level) of natural products.

Biological sources (genus)	Kingdom of life	Number and frequency of NP		Chemical class[a]		Biological activities[b]			
		N	%	Top 1	%	*N*	Top 1	%	Examples
Streptomyces	Bacteria	7951	3.4	Polyk	24.5	597	Antibact.	63.6	amphotericin B, avermectin, deferoxamine, doxorubicin, nystatin, tetracycline
Aspergillus	Fungi	2374	1.0	Alk	25.1	67	Antibact.	25.7	fumagillin, lovastatin
Penicillium	Fungi	2104	0.9	Alk	21.7	31	Antibact.	25.5	benzylpenicillin, griseofulvin
Euphorbia	Plantae	1434	0.6	Terp	76.5	8	Antivir.	28.5	ingenol
Artemisia	Plantae	1299	0.5	Terp	66.5	21	Antimal.	11.1	artemisinin
Salvia	Plantae	1166	0.5	Terp	79.4	9	Antineop.	21.4	salvinorin A, tanshinones
Pseudomonas	Bacteria	952	0.4	AA	37.5	40	Antibact.	45.1	aeruginoic acid, mupirocin
Aconitum	Plantae	948	0.4	Alk	92.8	7	Anaesth. g.	12.5	aconitine
Isodon	Plantae	895	0.4	Terp	96.2	4	Antineop.	57.1	oridonin, ponicidin
Garcinia	Plantae	892	0.4	Aro	68.9	7	Antivir.	27.2	gambogic acid
Senecio	Plantae	865	0.4	Terp	68.6	6	Antispasm.	28.5	platyphylline
Laurencia	Plantae	851	0.4	Terp	64.3	0	NA		
Bacillus	Bacteria	849	0.4	AA	53.4	39	Antibact.	47.5	bacitracin, spergualin
Piper	Plantae	847	0.4	Alk	30.8	17	Antipyr.	13.2	piperine
Taxus	Plantae	797	0.3	Terp	68.4	3	Antineop.	16.7	paclitaxel

[a]Most represented chemical class in each genus with its percentage. AA=Amino Acid and Peptide; Alk=Alkaloids; Aro=Simple Aromatic NP; Polyk=Polyketides; Terp=Terpenoids

[b]Number (N) of NP (for each genus) with established activity and being used as drugs or under investigation for drug use, and most represented biological activity (with its percentage) for each genus. Anaesth. g.= General anaesthetics ; Antibact.=Antibacterial agents and Antibiotics; Antimal.=Antimalarial agents; Antineop.=Antineoplastic agents; Antipyr.=Antipyretic agents; Antispasm.=Antispasmodic agents; Antivir.=Antiviral agents

NA = Not applicable

More generally, microbial organisms from the Actinomycetales order represent the largest group of bioactive bacterial compounds. According to the DNP 9,261 compounds have been isolated from these species, including the largest genus *Streptomyces*, but also rare actinomycetes such as *Micromonospora* (436 total NPs and 84 bioactive NPs), *Nocardia* (312 and 32) and *Actinomadura* (267 and 27) species. The discovery of rare actinomycetes increased significantly between the 1970s and the 1990s, and during this period about 30% of new antibiotics stemmed from these species (Bérdy 2005). Besides the fact that they are difficult to isolate and cultivate, one of the distinctive features of the rare actinomycetes is their production of diverse, unique, and complex compounds with excellent antibacterial potency and usually low toxicity (Lazzarini et al. 2000; Kurtböke 2012). Glycopeptide and orthosomycin antibiotics are produced almost exclusively by rare actinomycetales (Bérdy 2005). Some of the successful antibacterial agents currently on the market include gentamicin isolated from *Micromonospora purpurea*, erythromycin produced by *Saccharopolyspora erythraea*, rifamycins by *Amycolatopsis rifamycinica* and vancomycin by *Amycolatopsis orientalis* (Tiwari and Gupta 2012). Moreover, the rare actinomycetes are highly drug-prolific, with the Pseudonocardiaceae family (including *Amycolatopsis* and *Saccharopolyspora* species) being the most productive with 76 approved drugs in 2011 (Zhu et al. 2011). Nowadays, rare actinomycetes can be isolated from diverse environments (i.e., marine sources, soil samples, plant materials, and extreme environments) and intensively studied in antibiotic discovery programs (Jose and Jebakumar 2013; Dhakal et al. 2017).

Finally, other eubacteriales species including *Pseudomonas*, *Bacillus* and *Escherichia* species are among the highest producers of NPs. Bérdy (2005) already stressed the importance of the *Bacillus* and *Pseudomonas* species as the most prolific producers of bioactive metabolites from the unicellular bacteria group. Due to their ubiquity in the environment (i.e., soil, water, plants, and animals), *Pseudomonas* species are known to synthetize a wide range of metabolites (Gross and Loper 2009). In particular, *Pseudomonas* is one of the most common genera associated with plants as it contains epiphytic, endophytic, and pathogenic species (Strobel et al. 2004). They have been found to produce phytotoxic compounds as well as antimicrobial agents involved in the biocontrol of plant pathogens (Haas and Défago 2005). For example, *P. viridiflava*, associated with the leaves of grass species, produces ecomycins — a family of lipopeptides active against *Cryptococcus neoformans* and *Candida albicans* (Miller et al. 1998). *Bacillus* species produce mainly peptide antibiotics such as lantibiotics which contain the uncommon thioether amino acids lanthionine and ß-methyllanthionine (Stein 2005). Various lantibiotics have been isolated from *Bacillus* species including subtilin from *B. subtilis*, haloduracin from *B. halodurans*, lichenicidin from *B. licheniformis*, and cerecidins from *B. cereus* (Lawton et al. 2007; Dischinger et al. 2009; Wang et al. 2014). Since the 1970s, *Escherichia coli* have been widely used as a model for the development of DNA recombinant technology (Katz and Baltz 2016). Nowadays, it is the most employed host for the expression of drug candidates and small molecules (Atanasov et al. 2015). Thus, metabolites from *E. coli* have been widely studied and this may explain the high number of NPs from this species in the DNP (Reed and Palsson 2003).

NPs from the Fungi kingdom

In the DNP, approximately 20,000 compounds were reported to belong to the Fungi kingdom, and two divisions (Ascomycota and Basidiomycota) account for almost 90% of the total fungal products. The Fungi kingdom is thought to contain the largest number of microbial metabolites.

Compared to the other kingdoms, fungal products are structurally less complex and have smaller molecular weights (Bérdy 2005). A wide range of chemical classes are represented among fungi, including terpenoids (e.g., paclitaxel from *Taxomyces*), alkaloids (e.g., ergot alkaloids from *Claviceps*), simple aromatic NPs (e.g., griseofulvin from *Penicillium*), amino acids and peptides (e.g., echinocandins from *Aspergillus* and β-lactams from *Penicillium*), benzofuranoids (e.g., mycophenolic acid from *Penicillium*), carbohydrate NPs (e.g., ß-glucans from basidiomycetes), polycyclic aromatic NPs (e.g., parietin from lichens), and polyketides (e.g., aflatoxins and lovastatin from *Aspergillus*) (Figure 4) (Goyal et al. 2016). This plethora of diverse fungal metabolites is associated with various biological activities ranging from pharmacological agents to mycotoxins (Peláez 2005; Goyal et al. 2016).

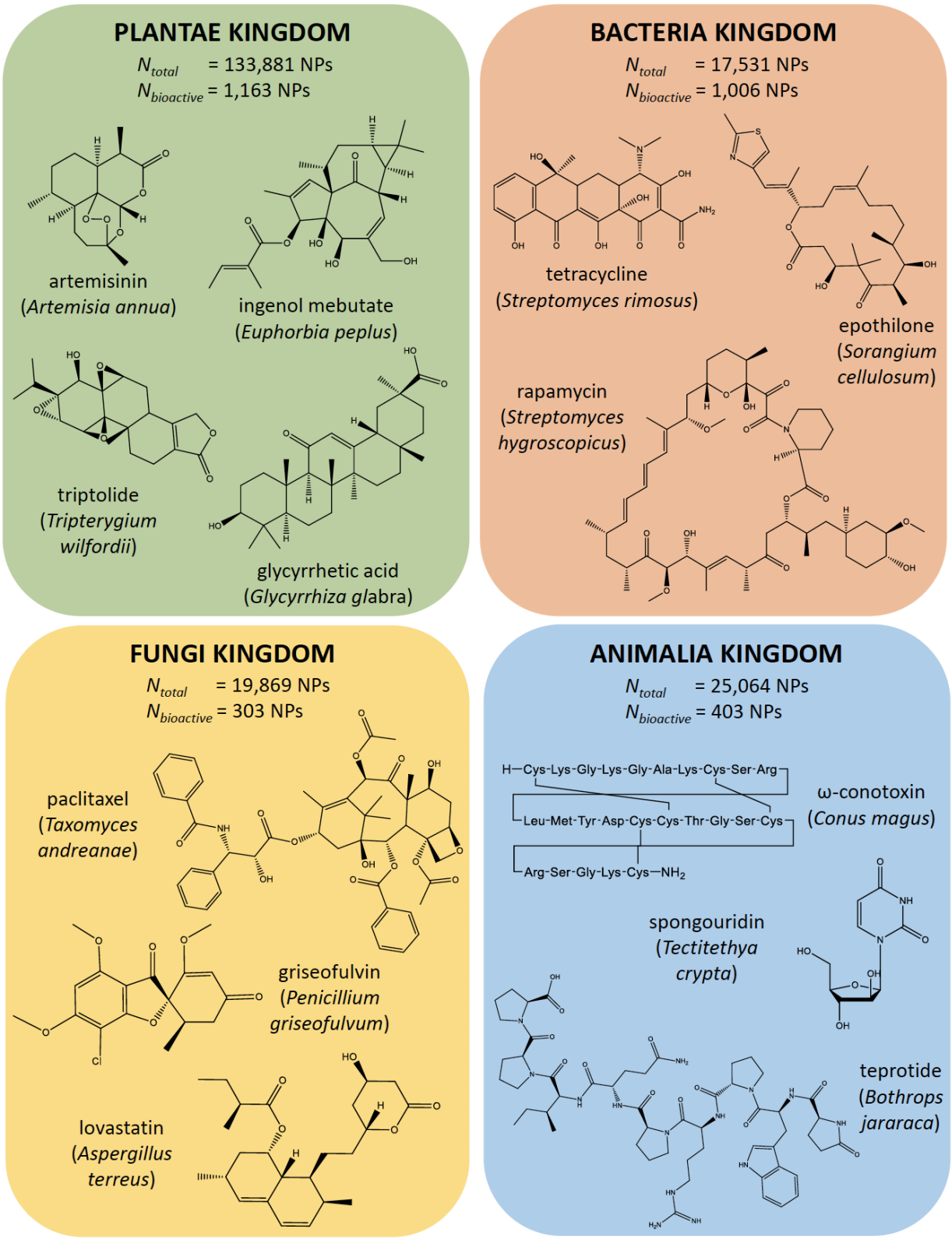

Figure 4. Structures of biologically active compounds of plant, bacterial, fungal and animal origin.

With more than 13,000 compounds, the Ascomycota division is the largest producer of NPs in the Fungi kingdom (Suppl. Mat. S1C). This group is also the largest phylum of Fungi (64,000 known species) and one of the most ubiquitous phyla of eukaryotes (Schoch et al. 2009). It comprises filamentous fungi (e.g., *Aspergillus* and *Penicillium*) but also yeasts (e.g., *Saccharomyces cerevisiae*), endophytic species (e.g., *Alternaria*, *Phoma*, *Stachybotrys* and *Trichoderma*), and lichen-forming fungi representing 40% of all ascomycetes (e.g., *Cladonia*, *Lecanora* and *Parmelia*) (Bérdy 2005; Blackwell 2011).

One of the milestones in the history of NPs from ascomycetes is the discovery of penicillin from the fungus *Penicillium chrysogenum* in 1928 by Alexander Fleming. This discovery led to the production and commercialization of synthetic penicillins in the early 1940s which saved a huge number of lives during WWII and ushered in the "Golden Age of Antibiotics" from the 1940s to the 1970s (Cragg et al. 2014; Bernardini et al. 2017). Besides the important discovery of β-lactams from *Penicillium* species, other NPs with pharmaceutical activities have also been isolated from this genus including mycophenolic acid, an immunosuppressant agent produced by *P. brevicompactum*; griseofulvin, an antifungal agent isolated from *P. griseofulvum*, and compactins known for their cholesterol-lowering activities produced by *P. brevicompactum* and *P. citrinum* (Frisvad et al. 2004; Chakravarti and Sahai 2004).

While the *Penicillium* genus is the second-largest producer of NPs in the Ascomycota division, *Aspergillus* is the most represented genus with *A. terreus* being the most prolific species from this genus in the DNP. Indeed, *Aspergillus* is renowned for its

medical, pathogenic, and industrial importance (Sanchez et al. 2012). As an example, *A. terreus* is a significant cause of aspergillosis, but it is also the main source of lovastatin, the first commercially marketed statin which was approved by the FDA in 1987 for the treatment of hypercholesterolemia (Demain and Sanchez 2009). Nowadays, synthetic derivatives of lovastatin are among the world's most sold drugs (Katz and Baltz 2016). Another example is that of *A. nidulans*, a model organism widely used to study genetics and cell biology, but it is also the main producer of echinocandins, a family of lipopeptides used in the treatment of candidiasis (Denning 2003). The first licensed semisynthetic echinocandin derivatives were caspofungin, micafungin and anidulafungin since 2001 (Butler 2008). Of all *Aspergillus* species, *A. fumigatus* is the primary causative agent of human infection, and gliotoxin is the main mycotoxin involved in mycosis (Dagenais and Keller 2009). *A. fumigatus* also produces fumagillin, first isolated in 1949 and used in the treatment of microsporidiosis (Mishra and Tiwari 2011). Finally, *Aspergillus* species (especially *A. flavus* and *A. parasiticus*) are also a unique source of aflatoxins, a major class of mycotoxins that have been described as human carcinogens and implicated in hepatocellular carcinoma (Henry et al. 2002; Mishra and Das 2003).

Endophytic fungi are also well represented in the DNP with *Fusarium* (i.e., *F. oxysporum* and *F. solani*), *Trichoderma* (i.e., *T. harzianum* and *T. viride*), *Alternaria* (i.e., *A. alternata*), and *Chaetomium* genera (i.e., *C. globosum*) ranking 3rd, 4th, 5th and 6th respectively in terms of NPs per genus. Endophytic fungi are defined by their occurrence within internal tissues of plants without causing any immediate overtly negative effects (Stone et al. 2000). They represent a polyphyletic group of ascomycetous fungi and are

found in various plants including liverworts, mosses, ferns, and seed plants (Aly et al. 2010). Interest in endophytic strains began with the detection of paclitaxel (an anti-tumour agent originally isolated from the plant species *Taxus brevifolia*) from the endophytic fungus *Taxomyces andreanae* (Stierle et al. 1993). This discovery highlighted the potential of endophytic fungi as alternative sources of plant secondary metabolites, and led to a shift in the discovery of new compounds from fungal sources in the early 1990s (Bérdy 2005). In the last decade, about half of the newly discovered fungal metabolites (approximately 5,000 compounds) were from endophytic fungi (Aly et al. 2010; Bérdy 2012). This includes various compounds belonging to different chemical classes such as aldehydes (e.g., chaetopyranin from *Chaetomium globosum*), alkaloids (e.g., camptothecin from *Fusarium solani*), lignans (e.g., podophyllotoxin from *Fusarium oxysporum*), peptides (e.g., beauvericin from *Fusarium oxysporum*), polyketides (e.g., alternariol from *Alternaria* species), steroids (e.g., wortmannins from *Talaromyces wortmannii*), and terpenoids (e.g., pestalotiopsins from *Pestalotiopsis* spp.) (Tan and Zou 2001; Kharwar et al. 2011; Gao et al. 2018). Moreover, in a literature review of 135 metabolites isolated from endophytic fungi, it was shown that the proportion of novel chemical structures produced by endophytes is significantly higher (51%) than that produced by fungi isolated from soil (38%), thus confirming that endophytes are a good source of bioactive metabolites (Schulz et al. 2002).

In addition to ascomycetes, basidiomycetes are also of great importance since amongst the fungi kingdom they are the second-largest producer of NPs in the DNP (4,820 NP). This group is also the second-largest phylum of fungi (22,000 known

species), and it includes most of the macroscopic fungi (Bills et al. 2005). One of the main features of basidiomycetes is the presence of compounds with high molecular weight such as polysaccharides, proteins, and lipids, as well as low molecular weight metabolites including alkaloids, peptides, polyketides, steroids, and terpenoids (De Silva et al. 2013). Polysaccharides have been widely studied, and in particular the study of ß-glucans has led to the development of anticancer drugs used in Asia such as lentinan from *Lentinula edodes*, schizophyllan from *Schizophyllum commune* and two specific proteoglycans (polysaccharide krestin [PSK] and polysaccharopeptide [PSP]) from *Trametes versicolor* (Lindequist et al. 2005). Secondary metabolites produced by basidiomycetes have also been investigated and have proved to be a source of bioactive compounds with, e.g., strobilurins isolated from *Strobilurus tenacellus* used as fungicides; pleuromutilins produced by *Pleurotus mutilus* and its semisynthetic derivative (retapamulin) used in the topical treatment of impetigo; and illudins isolated from *Omphalotus* spp. and their analogue (irofulven) currently under investigation as anticancer drugs (Stadler and Hoffmeister 2015).

Ganoderma species account for almost 10% of the total compounds isolated from basidiomycetes, and in the DNP half of these compounds are from *G. lucidum*. Also known as “Lingzhi”, this species has been widely used in traditional Chinese medicine to promote good health and to treat many diseases including allergy, arthritis, cancer, hypertension, and inflammation (Paterson 2006). Polysaccharides (e.g., ß-glucans) and triterpenoids (e.g., ganoderic acids) are the two major groups of compounds that exhibit pharmacological effects, especially anticancer and immunomodulatory activities (Yuen

and Gohel 2005). Extracts of *G. lucidum*, and the medicinal peptide, Ling Zhi-8, are currently under investigation for use as chemopreventive and adjuvant agents in the treatment of cancer (Cheng and Sliva 2015; Chen et al. 2017a).

NPs from the Animalia kingdom

Of all the kingdoms Animalia ranks as the second-largest producer of NPs in the DNP. More than 25,000 NPs have been identified from animal sources, and approximately half of these compounds belong to the terpenoid and alkaloid groups.

In the DNP, one quarter of the bioactive NPs from animal sources were isolated from their venoms and toxins, among which amino acid and peptide NPs are highly represented (97%) (Figure 5). The best-known example of a successful venom-based drug is that of the angiotensin-converting enzyme (ACE) inhibitors which act as antihypertensive agents. A venom peptide (teprotide) was isolated from the snake *Bothrops jararaca* in the 1960s and led to the development of ACE inhibitors (i.e., captopril, enalapril and lisinopril) which are currently among the top best-selling drugs in the world (Lewis and Garcia 2003; Antunes et al. 2016). Other venomous animals have also been investigated for their pharmaceutical potential such as the Gila monster *Heloderma suspectum* which afforded exendin-4, an antidiabetic agent approved by the FDA in 2005; the medicinal leech *Hirudo medicinalis* which led to the development of bivalirudin, a synthetic analog of hirudin with anticoagulant activity approved by the FDA in 2000; and the cone snail *Conus magus* which produces ω-conotoxins, the synthetic form ziconotide was approved in 2004 for the treatment of refractory neuropathic pain (King 2011; Zambelli et al. 2016). Interestingly, among all of the kingdoms *Conus* is one of the genera that produces the largest number of bioactive compounds, since more than 100 bioactive NPs (accounting for more than 70% of the total NPs from this genus) have been recorded in the DNP. This high number of diverse pharmacologically active

compounds results from the high rate of hypermutations as well as the remarkable number of post-translational modifications in this genus which allow little overlap in conopeptides between *Conus* species (Buczek et al. 2005). To date, it is estimated that the number of conopeptides per species might exceed 1,000, and that a total of 35,000 compounds could be available from the 700 *Conus* species identified (Davis et al. 2009).

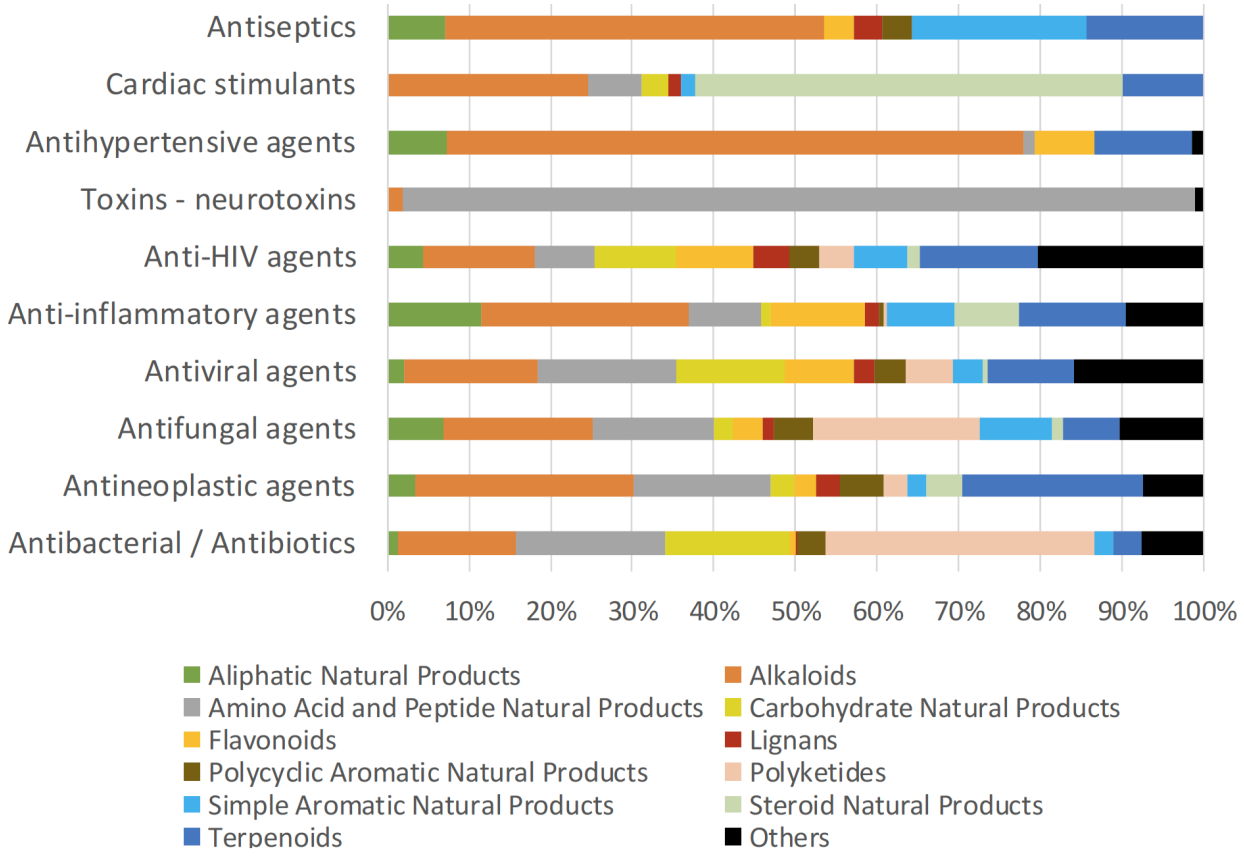

Figure 5. Distribution of the main chemical classes of natural products in the most important biological activities.

An interesting feature of animal-derived products is the high proportion of compounds originating from marine sources, with more than half of the animal-derived NPs in the DNP belonging to sponges (i.e., Ceractinomorpha, Tetractinomorpha, and Homoscleromorpha) and cnidarians (i.e., Octocorallia and Hexacorallia) (Suppl. Mat.

S1D). It is noteworthy that species from the Animalia kingdom are highly represented in aquatic environments, with 32 of the 33 animal phyla being marine (Cragg and Newman 2013). Historically, the first discovery of biologically active compounds from a marine environment was that of the nucleosides spongouridine and spongothymidine, isolated from the Caribbean sponge, *Tectitethya crypta* in the early 1950s (Dias et al. 2012). Two decades later, the systematic investigation of bioactive products from marine sources began thanks to the development of scuba diving techniques as well as the use of submersibles (Cragg and Newman 2013).

With 8,656 compounds in the DNP, the Porifera phylum (sponges) is by far the most represented amongst the Animalia kingdom. This phylum, comprising about 8,500 species, has also been reported to be the most prolific marine producer of natural compounds (approximately 30% of all NPs from marine sources) with an average of 200 compounds described each year (Van Soest et al. 2012; Blunt et al. 2015, 2017). In the DNP, *Dysidea* is the genus with the highest number of NPs from the Porifera phylum, while *Theonella swinhoei* (also from the Porifera phylum) is the largest producer of NPs from the Animalia kingdom. In a review of marine sponge-derived NPs, Mehbub et al. (2014) suggested that the high number of NPs found in key orders of the sponges including Dictyoceratida (including *Dysidea* spp.), Haplosclerida, Halichondrida, Poecilosclerida, and Astrophorida was the result of high species diversity in these orders. Another reason could be the diversity and host specificity of microbial symbionts found in these species, which are responsible for the synthesis of some of the metabolites isolated from these sponges (Thomas et al. 2010). This is the case for *Dysidea* spp. which host a

distinct cyanobacterial clade that may be responsible for the variable patterns in secondary metabolites found in this genus (Thacker and Starnes 2003). More generally, terpenoids, alkaloids, as well as peptides represent the main chemical classes of compounds found in the Porifera phylum, and more than half of the biological activities investigated were anticancer properties (Mehbub et al. 2014). Three compounds derived from marine sponges have been approved by the FDA including the anti-tumour compound eribulin mesylate, which is an analogue of halichondrin B first isolated from *Halichondria okadai* (Agrawal et al. 2016).

According to the DNP Cnidarians are the second-largest producer of NPs from animal sources (5,249 NPs). This group of relatively simple animals comprises over 11,000 species, and includes reef-forming corals, sea anemones, soft corals, jellyfishes, and marine hydroids (Daly et al. 2007; Appeltans et al. 2012). The Cnidarian phylum contains the *Sinularia* genus which, according to the DNP, is by far the biggest producer of NPs from the Animalia kingdom, while *Clavularia viridis* is the most prolific Cnidarian species. Both taxa belong to the Alcyonacea order (soft corals) which have been described as the main source of NPs from Cnidarians (Leal et al. 2012). With more than 160 known species, *Sinularia* are one of the richest genera from the Alcyonacea order (Roskov et al. 2018). They produce a wide range of secondary metabolites including sesquiterpenes, polyhydroxylated steroids, polyamine compounds, as well as cembranoid diterpenes which are the most frequently isolated NPs from *Sinularia* (Chen et al. 2012). Some of these compounds display various bioactivities such as antiulcer properties (e.g., sinulide), anti-inflammatory activities (e.g., gibberoketosterol) and anti-

tumour activities (e.g., flexilarin D) (Rocha et al. 2011). Altogether, most of the bioactive compounds found in Cnidarians belong to the terpenoids (i.e., monoterpenoids, diterpenoids, and sesquiterpenoids). Most of the interest in NPs isolated from Cnidarians surrounds their potential anti-tumour and anti-inflammatory activities (Rocha et al. 2015). To date, none of the compounds isolated from Cnidarians have been developed as drugs, however some compounds have been exhaustively investigated in preclinical studies, such as the diterpene glycoside pseudopterosin extracted from the sea whip *Pseudopterogorgia elisabethae*; eleutherobin derived from the soft coral *Eleutherobia* sp; and the diterpene sarcodictyn, found in some corals (Mariottini and Grice 2016). In addition, more than 70 compounds extracted from Cnidarians have been reported to possess promising bioactivities and thus might be of interest for future clinical studies (Rocha et al. 2011).

Targeting the most prolific biological organisms: popular taxa vs unexplored phyla

Targeting "popular" taxonomic groups seems to be a common way to uncover new natural compounds. In the DNP, a large number of NPs originate from species-rich biological groups. As an example, in the Plantae kingdom the highest number of natural compounds can be isolated from the Compositae and Leguminosae families, and they are also amongst the three largest botanical families in the world. In the Animalia kingdom, a high proportion of compounds originate from marine sources, but 32 of the 33 phyla in the Animalia kingdom include species from aquatic environments. Thus, the accessibility of biological sources (i.e., wide abundance and distribution) probably plays an important role in the selection of candidate species for further investigation. This is especially true for some marine taxa, for which the abundance, size, and colour have been reported to influence the selection process (Blunt et al. 2008). Indeed, "popular" taxa provide a wide range of NPs with high chemical diversity, thus suggesting the potential of these taxa as sources for drug discovery (Leal et al. 2012). The popularity of some organisms might also be related to their high-yield of bioactive compounds. For instance, *Streptomyces* spp., the most investigated genus in our analysis, exhibited the highest yield of bioactive compounds. Although it has been widely investigated, *Streptomyces* spp. continue to deliver novel scaffolds, such as the antibiotic platensimycin launched in 2006, as well as many other compounds currently under investigation in clinical trials (de Lima Procópio et al. 2012; Genilloud 2017). Furthermore, an increasing number of compounds thought to originate from plants or animals can now be reattributed to microbial organisms (e.g., endophytes, symbionts). Therefore, the high

number of compounds found in one particular taxonomic group could in fact be produced by microbial sources. As an example, 31 of the 32 polyketide metabolites isolated from the sponge *Theonella swinhoei*, the most represented species from the Animalia kingdom in our study, have been reattributed to an uncultured bacterium (Wilson et al. 2014).

While some researchers will prefer to focus on "popular" taxonomic groups exhibiting high chemical diversity, others will select unexplored phyla or geographical regions (Leal et al. 2012). Estimates for the total number of non-microbial species living on Earth range from 2 to 100 million (May 2010). However, some authors suggested that this number might be narrower and range between 5 and 10 million (Mora et al. 2011; Costello et al. 2013). If we compare the number of species catalogued with those predicted, the most interesting groups with a large number of species still to be discovered are likely to be fungi (≈ 100,000 species catalogued vs 0.8-5 million predicted) and animals (1-2 million vs 5-10 million) (Scheffers et al. 2012; Pimm et al. 2014). In contrast, more than two thirds of the predicted number of plant species (≈450,000) are already known (Pimm and Joppa 2015). Regarding bacteria, the total number of species is quite difficult to evaluate and estimates range between 10^3 to 10^{12} species (Pedrós-Alió 2006; Locey and Lennon 2016). In addition, some authors argued that marine environments and biodiversity hotspots could be a great source of new taxa (Mora et al. 2011; Scheffers et al. 2012). Species from specific taxonomic groups and geographical regions could be a source of NPs with potential bioactivity. In a recent paper, Pye et al. (2017) demonstrated that identifying new biological organisms leads to a burst in the discovery of structurally novel compounds, while investigating already studied classes of organisms

is likely to yield existing classes of compounds. Therefore, further studies should be performed to identify biological organisms with great potential. This could be done by mapping the distribution area of un-investigated (or less investigated) phyla, based on a datamining-guided search of an exhaustive as possible NP database.

NP research in the pharmaceutical industry

Historically, NPs have formed the basis of the therapeutic arsenal developed by pharmaceutical companies, and today they still represent a huge reservoir of inspirational bioactive chemodiversity. Out of the 293,798 NPs compiled in the DNP database, only 3,882 (≈1%) were reported to have biological activities, viz. compounds with established activity and being used as drugs or under investigation for drug use. This raises concerns about the future potential of NPs as drug candidates.

Historical background

Plants and NPs have historically been valuable sources of effective medicine. In the beginning of the nineteenth century, pure NPs started to be isolated from medicinal plants and administered as pure and unique active principles at precise dosages. This was the beginning of the single and "magic bullet" paradigm (Strebhardt and Ullrich 2008). In the meantime, herbalism with its mixture of multi-active NPs lost credit. Natural products then reached their moment of glory when hundreds of pharmaceutical companies were trying to associate clinical activity with the presence of a particular NP (Figure 6, phase 1800 to 1980). Inspired by the seminal huge screening campaigns initiated in 1938 by Jonathan Hartwell and the Cancer program at the National Cancer Institute in 1955, many companies ran large-scale drug discovery screening programs. The pharmaceutical industry turned towards the thousands of NPs found in genetic resources for their high throughput screening (HTS) drug discovery programs, made possible with the advent of miniaturization and robotics. During this phase (1980-2005),

drug discovery strategies made a 180-degree turn, starting with target molecules (e.g., enzymes, receptors, and ion channels) and moving up to cells, groups of cells, tissues, animals, and finally human patients. High throughput screening and molecular assays were performed in more and more dense microtiter plates (96, 384 or 1536 wells) at the farthest distance from the patient to conduct more and more experiments (up to 100,000 wells tested per day). Despite huge throughputs, and the advantage of screening very large numbers of compounds at a time, high attrition rates were observed when moving from hits active on a biological molecule towards the complexity of a human being. A HTS hit can easily access the target molecule in solution, but this is more difficult with cells, tissues, and complex biological organisms like laboratory animals (David and Ausseil 2014). Besides, by the end of the 1990s new techniques such as combinatorial chemistry, virtual screening, and even the deciphering of the human genome monopolized financial investments diverting funds away from NP screening programs. The Human Genome Project was supposed to generate thousands of new druggable proteins as pharmacological targets. But in fact, the number of targets rose from 300 to around 600 (Rask-Andersen et al. 2014).

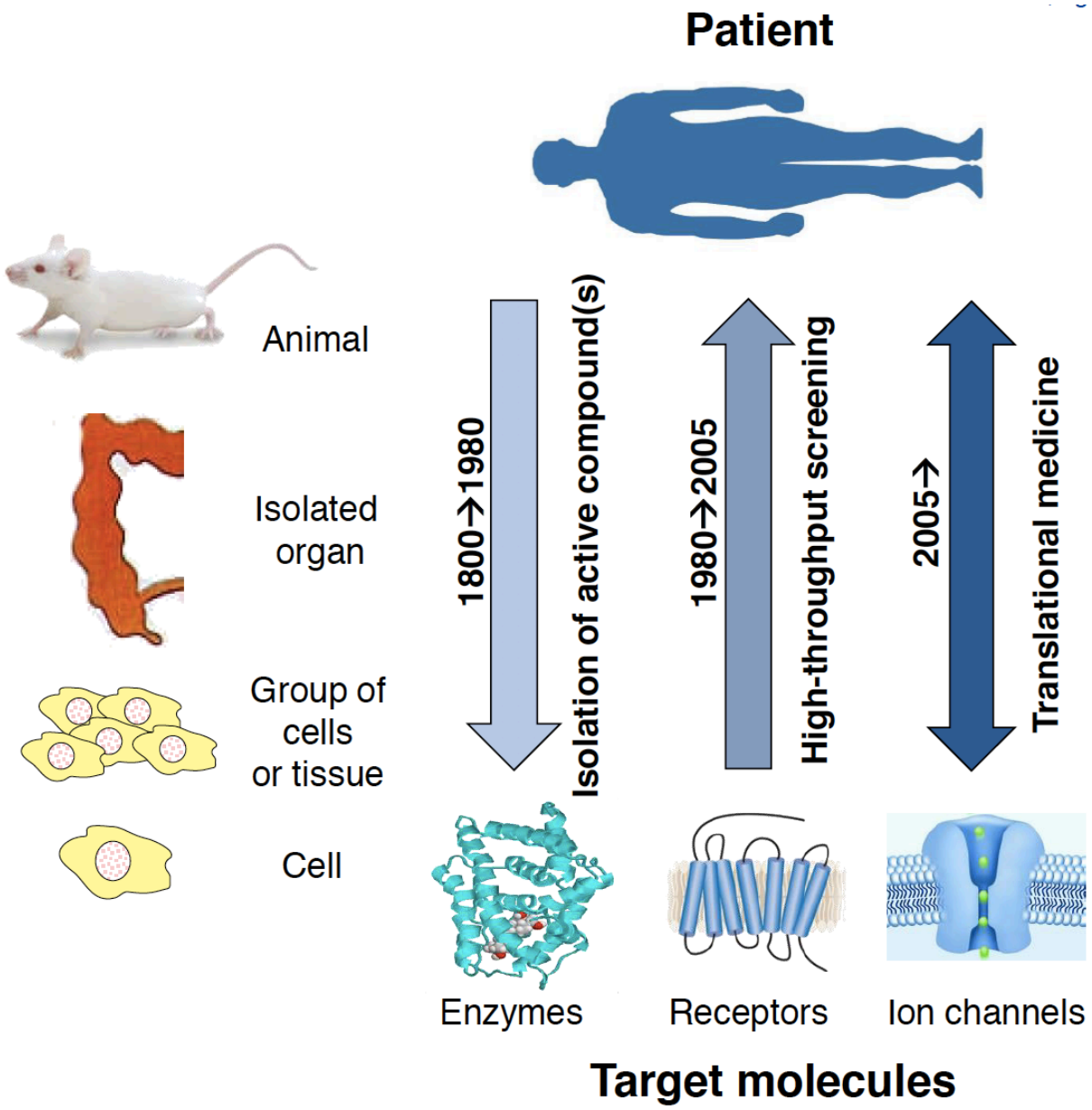

Figure 6. Evolution of the research strategies utilized by Pharma companies.

Current state of NP research in Pharma companies

Since the beginning of the twenty-first century, research into NPs has largely been downsized. The probability of a hit from a HTS screening program reaching the market is

around one in a million and the cost from hit to approval about $ 3 billion (Oprea 2000; DiMasi et al. 2016).

Paradigms need to change as a single approach is no longer possible. In drug discovery hits at a molecular level need to be validated up to the patient level. Drug discovery needs to connect the patients' scale to the scale of molecules by translational medicine. Moreover, computational and combinatorial chemistry techniques are now being combined and applied to drug discovery.

Nowadays studies on traditional knowledge are restricted to academic research mostly on tropical diseases. But in the meantime, the economic and social demand for "green" food supplements, botanical drugs, and NP-based medicine is expanding.

Main reasons for the decline in NP research

It is clear that for decades (or even centuries) traditional medicines have been raided by pharmaceutical companies and turned into active molecules. Therefore, most of the easiest "low-hanging fruits" have been picked (David and Ausseil 2014). An analysis of new structures described in the commercial database *AntiMarin* (measured by the Tanimoto index) reveals that newly reported NPs are becoming more and more similar to structures already described (Pye et al. 2017).

High throughput screening and even phenotypic high content screening on small animals have not been able to generate a return on investment with the introduction of new chemical entities on the market. In this strategy, researchers wait for an improbable "alignment of planets" between the druggability of NPs and the specific biological target

screened. The number of potential targets is not boundless, even after the deciphering of the human genome (around 600 druggable targets). Even huge phenotypic screening campaigns on small animals were not able to optimize this passive “lottery” approach. This strategy was based on the premise that a large library of NPs should contain the perfect gem dedicated to the pharmacological target being studied.

The complexity and inherent slowness of working with NPs was one of the reasons for the downsizing or shutdown of NP drug discovery programs. The supply, re-supply and authentication of genetic resources, as well as the isolation and identification of active molecules through bio-guided fractionation is not an easy game (Atanasov et al. 2015). The implementation of tedious access laws for genetic resources also had a negative effect on bioprospection and NP research, along with paradoxical effects on biodiversity conservation (David 2018).

Trends for the future

For the last two decades, research efforts into NPs have gone phases of decline and renaissance. Nevertheless, NPs still represent a huge source of potential drug leads, but research strategies need to evolve with the times and new paradigms need to be invented.

Since it was discontinued in the beginning of the 2000s, bioprospection will remain marginal in Big Pharma. Only Novartis and Pierre Fabre are still working on higher plants for drug discovery. The few companies involved in NP research are rather focusing on microorganisms or marine organisms which offer more “low-hanging fruits” in

underexplored biotopes. More than 90% of microorganisms are not cultivable, but heterologous models make it possible to produce NPs from these organisms should allow the production of novel or “cryptic” NPs (Palazzolo et al. 2017).

For scientific and administrative reasons, for two centuries pharmaceutical companies have focused their research efforts on the discovery of pure active compounds. This is the “magic bullet” paradigm. But plants act in more subtle ways. However, it should be admitted that the “single bullet” approach is not always the best approach, and indeed multitherapy is favoured in the treatment of AIDS and other infective diseases. Thus, in some chronic or metabolic diseases the single high affinity bullet should be replaced by a swarm of less active bullets as in phytotherapy and traditional African, Indian or Chinese medicines. The synergetic combination approach is gaining more and more interest. It is likely that paradigms will evolve away from single molecules to well-defined extracts simultaneously addressing multiple pharmacological targets, as is the proposed mechanism of action for botanical drugs.

Novel scientific methods for the discovery, validation, characterization, and standardization of these multicomponent botanical drugs will surely become more and more recognised by Health agencies. Products from vegetal origin such as Acheflan®, Angipars®, Epogam®, Fulyzaq®, Iberogast®, Picato®, Rosaderm® and Veregen® are paving the way towards the new paradigm.

Meanwhile, the economic and social need for food supplements, phytotherapy, and botanical drugs is growing. The cultivation of medicinal plants for the production of standardized enhanced traditional medicines is recommended by the World Health

Organization and was advocated in 2008 with the Beijing declaration (World Health Organization 2008). This makes sense in terms of cost, availability, and environmental impact. Pharmacological screening is generally conducted with a huge number of compounds on one specific biological target. But the opposite is possible with selectivity profiling, *id est* the screening of one NP on a lot of targets (Pouny et al. 2014). Re-purposing old NPs which benefit from decades of pharmacovigilance is also a fruitful approach (Cragg et al. 2014).

Fragment-based drug discovery looks very promising. This ground-breaking technique uses a restricted library of approximately 500 small natural molecules (150-250 Da called “fragments”). Optimized libraries of natural fragments cover a larger chemical space than synthetic fragments and offer straightforward growing possibilities to cover a huge chemodiversity space (Pascolutti et al. 2015).

Translational medicine, neural network analyses, and the generation of large data sets have started to allow molecular interactome studies, and links between biological systems to disease mechanisms are very encouraging (Capriotti et al. 2018). New trends in NP research include microbial genomics, synthetic biology, control of NP biosynthesis and bioinformatics. Computational approaches that can analyse large amounts of data are rapidly developing in every field of NP research (Sarker and Nahar 2018).

Limitations

This study has several limitations which are mainly linked to the composition of the DNP. Indeed, numerous biases related to the collection, isolation, and identification of biota and their chemical constituents can be found in this database. First of all, taxonomic data stems usually from the first report, hence sources can sometimes be wrongly assigned. As an example, the genus *Lyngbya* has been completely revised during the last five years. However, most of the species cited in this work are well known and readily identifiable (e.g., *Aspergillus terreus*, *Azadirachta indica*, *Helianthus annuus*, *Nicotiana tabacum*, and *Saccharomyces cerevisiae*) and so we expect to have very few incidences of misinformation. Moreover, for most of the analyses we have intentionally limited our study to the family level or higher taxonomic rank. Another limitation is that some biological sources cited in this work (e.g., *Bacillus subtilis, Escherichia coli*, and *Streptomyces* spp.) are used as hosts for the production of drugs and small molecules. Therefore, compounds reported in our study as being produced by these species may not originate from these species. Given the restricted number of species used as hosts and the recent development of this technique for some species (i.e., *Streptomyces* spp.), this limitation does not significantly affect our results. In addition, not all compounds reported in the DNP to have biological activities (3,882 NPs in total) have been used as drugs. Therefore, for some compounds, preclinical and clinical data may be limited, and their biological activities need to be confirmed. Finally, this dictionary is not exhaustive, and our study cannot be taken as a descriptive study of all known NP.

Despite these limitations, our results correlate with the findings of Bérdy (2005) who aimed to provide a summary of bioactive compounds from microbial sources, as well

as those from Blunt et al. (2015), Mehbub et al. (2014), and Hu et al. (2011) who reviewed the NPs derived from marine sources.

Conclusion

This review aimed to characterize the global diversity of compounds isolated from natural sources. We based our analysis on the DNP which is one of the most comprehensive libraries of NPs, and one of the few to link chemical entities to their natural sources.

Our study of the diversity of natural compounds shows that a high proportion of NPs are from plant origin, with terpenoids being the most represented chemical class (except in the Bacteria kingdom), and antibacterial as well as antineoplastic activities are the two main biological activities reported. Besides the importance of plant-derived compounds, we also highlight the large number of NP isolated from *Streptomyces* spp. as well as from the Ascomycota phylum and marine sources (i.e., sponges and soft corals).

Although a large number of compounds have been isolated from natural sources, few are used as medicines today. More integrative and comprehensive approaches should unravel the limitations in drug discovery from NPs outlined in this review. Drug discovery from NPs is still in its infancy, and NPs remain relevant in modern drug discovery.

Acknowledgements

We would like to thank the French National Research Institute for Sustainable Development (IRD) documentation service for subscribing to the DNP. We would also like to thank Fiona Macdonald, Taylor & Francis group, for the clarifications she provided concerning the information given in the DNP.

Supplementary Materials

Supplementary Material S1: Most represented families, genera and species for each kingdom of life in the DNP

Conflict of interest

The authors declare that they have no conflict of interest.

References

Adams JD, Wang R, Yang J, Lien EJ (2006) Preclinical and clinical examinations of *Salvia miltiorrhiza* and its tanshinones in ischemic conditions. Chinese Medicine 1:3. doi: 10.1186/1749-8546-1-3

Agrawal S, Adholeya A, Deshmukh SK (2016) The pharmacological potential of non-ribosomal peptides from marine sponge and tunicates. Front Pharmacol 7. doi: 10.3389/fphar.2016.00333

Aly AH, Debbab A, Kjer J, Proksch P (2010) Fungal endophytes from higher plants: a prolific source of phytochemicals and other bioactive natural products. Fungal Divers 41:1–16. doi: 10.1007/s13225-010-0034-4

Antunes AM de S, Guerrante RDS, Ávila J de PC, et al (2016) Case study of patents related to captopril, Squibb's first blockbuster. Expert Opin Ther Pat 26:1449–1457. doi: 10.1080/13543776.2016.1227321

Appeltans W, Ahyong ST, Anderson G, et al (2012) The magnitude of global marine species diversity. Curr Biol 22:2189–2202. doi: 10.1016/j.cub.2012.09.036

Atanasov AG, Waltenberger B, Pferschy-Wenzig E-M, et al (2015) Discovery and resupply of pharmacologically active plant-derived natural products: A review. Biotechnology Advances 33:1582–1614. doi: 10.1016/j.biotechadv.2015.08.001

Baker D, Chu M, Oza U, Rajgarhia V (2007) The value of natural products to future pharmaceutical discovery. Nat Prod Rep 24:1225–1244. doi: 10.1039/B602241N

Banerjee P, Erehman J, Gohlke B-O, et al (2015) Super Natural II—a database of natural products. Nucleic Acids Res 43:D935–D939. doi: 10.1093/nar/gku886

Bérdy J (2005) Bioactive microbial metabolites. J Antibiot 58:1–26. doi: 10.1038/ja.2005.1

Bérdy J (2012) Thoughts and facts about antibiotics: Where we are now and where we are heading. J Antibiot 65:385–395. doi: 10.1038/ja.2012.27

Bernardini S, Tiezzi A, Masci VL, Ovidi E (2017) Natural products for human health: an historical overview of the drug discovery approaches. Nat Prod Res 0:1–25. doi: 10.1080/14786419.2017.1356838

Bills G, Spatafora JW, Blackwell M (2005) Phylogeny of the fungal kingdom and fungal-like eukaryotes. In: Handbook of Industrial Mycology. Marcel Dekker, New York, pp 27–47

Blackwell M (2011) The Fungi: 1, 2, 3 … 5.1 million species? Am J Bot 98:426–438. doi: 10.3732/ajb.1000298

Blunt J, Munro M, Upjohn M (2012) The role of databases in marine natural products research. In: Handbook of Marine Natural Products. Springer, Dordrecht, pp 389–421

Blunt JW, Copp BR, Hu WP, et al (2008) Marine natural products. Nat Prod Rep 25:35–94. doi: 10.1039/b701534h

Blunt JW, Copp BR, Keyzers RA, et al (2015) Marine natural products. Nat Prod Rep 32:116–211. doi: 10.1039/c4np00144c

Blunt JW, Copp BR, Keyzers RA, et al (2017) Marine natural products. Nat Prod Rep 34:235–294. doi: 10.1039/c6np00124f

Brinker AM, Ma J, Lipsky PE, Raskin I (2007) Medicinal chemistry and pharmacology of genus *Tripterygium* (Celastraceae). Phytochemistry 68:732–766. doi: 10.1016/j.phytochem.2006.11.029

Buczek O, Bulaj G, Olivera BM (2005) Conotoxins and the posttranslational modification of secreted gene products. Cell Mol Life Sci 62:3067–3079. doi: 10.1007/s00018-005-5283-0

Butler MS (2008) Natural products to drugs : Natural product -derived compounds in clinical trials. Nat Prod Rep 25:475–516. doi: 10.1039/B514294F

Cantisani C, Gado FD, Ulrich M, et al (2013) Actinic keratosis: review of the literature and new patents. Recent Pat Inflamm Allergy Drug Discov 7:168-175(8)

Capriotti E, Ozturk K, Carter H (2018) Integrating molecular networks with genetic variant interpretation for precision medicine. Wiley Interdiscip Rev Syst Biol Med in press:e1443. doi: 10.1002/wsbm.1443

Chakravarti R, Sahai V (2004) Compactin—A review. Appl Microbiol Biotechnol 64:618–624. doi: 10.1007/s00253-003-1553-7

Chang K-W, Tsai T-Y, Chen K-C, et al (2011) iSMART: An integrated cloud computing web server for traditional Chinese medicine for online virtual screening, de novo evolution and drug design. J Biomol Struct Dyn 29:243–250. doi: 10.1080/073911011010524988

Chao J, Dai Y, Verpoorte R, et al (2017) Major achievements of evidence-based traditional Chinese medicine in treating major diseases. Biochem Pharmacol 139:94–104. doi: 10.1016/j.bcp.2017.06.123

Chen BJ (2001) Triptolide, a novel immunosuppressive and anti-inflammatory agent purified from a Chinese Herb *Tripterygium wilfordii* Hook F. Leuk Lymphoma 42:253–265. doi: 10.3109/10428190109064582

Chen S-J, Lin H-H, Huang W-C, et al (2017a) Ling-Zhi-8 protein (LZ-8) suppresses the production of pro-inflammatory mediators in murine microglial BV-2 cells. Food Agric Immunol 28:1393–1407. doi: 10.1080/09540105.2017.1346062

Chen S-R, Dai Y, Zhao J, et al (2018) A mechanistic overview of triptolide and celastrol, natural products from *Tripterygium wilfordii* Hook F. Front Pharmacol 9. doi: 10.3389/fphar.2018.00104

Chen W, Li Y, Guo Y (2012) Terpenoids of *Sinularia* soft corals: chemistry and bioactivity. Acta Pharm Sin B 2:227–237. doi: 10.1016/j.apsb.2012.04.004

Chen Y, de Bruyn Kops C, Kirchmair J (2017b) Data resources for the computer-guided discovery of bioactive natural products. J Chem Inf Model 57:2099–2111. doi: 10.1021/acs.jcim.7b00341

Cheng S, Sliva D (2015) *Ganoderma lucidum* for cancer treatment: we are close but still not there. Integr Cancer Ther 14:249–257. doi: 10.1177/1534735414568721

Clardy J, Fischbach MA, Walsh CT (2006) New antibiotics from bacterial natural products. Nat Biotechnol 24:1541–1550. doi: 10.1038/nbt1266

Clardy J, Walsh C (2004) Lessons from natural molecules. Nature 432:829–837. doi: 10.1038/nature03194

Costello MJ, May RM, Stork NE (2013) Can we name Earth's species before they go extinct? Science 339:413–416. doi: 10.1126/science.1230318

Cragg GM, Grothaus PG, Newman DJ (2014) New horizons for old drugs and drug leads. J Nat Prod 77:703–723. doi: 10.1021/np5000796

Cragg GM, Newman DJ (2013) Natural products: A continuing source of novel drug leads. Biochim Biophys Acta Gen Subj 1830:3670–3695. doi: 10.1016/j.bbagen.2013.02.008

Dagenais TRT, Keller NP (2009) Pathogenesis of *Aspergillus fumigatus* in invasive aspergillosis. Clin Microbiol Rev 22:447–465. doi: 10.1128/CMR.00055-08

Daly M, Brugler MR, Cartwright P, et al (2007) The phylum Cnidaria: A review of phylogenetic patterns and diversity 300 years after Linnaeus. Zootaxa 1668:127–182

David B (2018) New regulations for accessing plant biodiversity samples, what is ABS? Phytochem Rev 17:1211–1223. doi: 10.1007/s11101-018-9573-1

David B, Ausseil F (2014) High-throughput screening of plant chemodiversity. In: Encyclopedia of Analytical Chemistry. American Cancer Society, pp 1–24

David B, Wolfender J-L, Dias DA (2015) The pharmaceutical industry and natural products: historical status and new trends. Phytochem Rev 14:299–315. doi: 10.1007/s11101-014-9367-z

Davis J, Jones A, Lewis RJ (2009) Remarkable inter- and intra-species complexity of conotoxins revealed by LC/MS. Peptides 30:1222–1227. doi: 10.1016/j.peptides.2009.03.019

de Lima Procópio RE, da Silva IR, Martins MK, et al (2012) Antibiotics produced by *Streptomyces*. Braz J Infect Dis 16:466–471. doi: 10.1016/j.bjid.2012.08.014

De Silva DD, Rapior S, Sudarman E, et al (2013) Bioactive metabolites from macrofungi: ethnopharmacology, biological activities and chemistry. Fungal Divers 62:1–40. doi: 10.1007/s13225-013-0265-2

Demain AL, Sanchez S (2009) Microbial drug discovery: 80 years of progress. The Journal of Antibiotics 62:5–16. doi: 10.1038/ja.2008.16

Denning DW (2003) Echinocandin antifungal drugs. Lancet 362:1142–1151. doi: 10.1016/S0140-6736(03)14472-8

Dhakal D, Pokhrel AR, Shrestha B, Sohng JK (2017) Marine rare Actinobacteria: Isolation, characterization, and strategies for harnessing bioactive compounds. Front Microbiol 8:. doi: 10.3389/fmicb.2017.01106

Dias DA, Urban S, Roessner U (2012) A historical overview of natural products in drug discovery. Metabolites 2:303–336. doi: 10.3390/metabo2020303

DiMasi JA, Grabowski HG, Hansen RW (2016) Innovation in the pharmaceutical industry: New estimates of R&D costs. J Health Econ 47:20–33. doi: 10.1016/j.jhealeco.2016.01.012

Dischinger J, Josten M, Szekat C, et al (2009) Production of the novel two-peptide lantibiotic lichenicidin by *Bacillus licheniformis* DSM 13. PLoS One 4:e6788. doi: 10.1371/journal.pone.0006788

Ernst M, Grace OM, Saslis-Lagoudakis CH, et al (2015) Global medicinal uses of *Euphorbia* L. (Euphorbiaceae). J Ethnopharmacol 176:90–101. doi: 10.1016/j.jep.2015.10.025

Falk H, Wolkenstein K (2017) Natural Product Molecular Fossils. In: Progress in the Chemistry of Organic Natural Products, vol 104. Springer, Cham. pp 1-126. doi: 10.1007/978-3-319-45618-8

Frisvad JC, Smedsgaard J, Larsen TO, Samson RA (2004) Mycotoxins, drugs and other extrolites produced by species in *Penicillium* subgenus *Penicillium*. Stud Mycol 49:201–241

Gao H, Li G, Lou H-X (2018) Structural diversity and biological activities of novel secondary metabolites from endophytes. Molecules 23:646. doi: 10.3390/molecules23030646

Gaudêncio S, Pereira F (2015) Dereplication: racing to speed up the natural products discovery process. Nat Prod Rep 32:779–810. doi: 10.1039/C4NP00134F

Genilloud O (2017) Actinomycetes: still a source of novel antibiotics. Nat Prod Rep 34:1203–1232. doi: 10.1039/C7NP00026J

Gershenzon J, Dudareva N (2007) The function of terpene natural products in the natural world. Nat Chem Biol 3:408–414. doi: 10.1038/nchembio.2007.5

Goldbach-Mansky R (2009) Comparison of *Tripterygium wilfordii* Hook F versus sulfasalazine in the treatment of rheumatoid arthritis: A randomized trial. Ann Intern Med 151:229. doi: 10.7326/0003-4819-151-4-200908180-00005

Goyal S, Ramawat KG, Mérillon JM (2016) Different shades of fungal metabolites: An overview. In: Fungal Metabolites. Springer, Cham, pp 1–29

Gross H, Loper JE (2009) Genomics of secondary metabolite production by *Pseudomonas* spp. Nat Prod Rep 26:1408–1446. doi: 10.1039/B817075B

Haas D, Défago G (2005) Biological control of soil-borne pathogens by fluorescent pseudomonads. Nat Rev Microbiol 3:307–319. doi: 10.1038/nrmicro1129

Harvey AL, Edrada-Ebel R, Quinn RJ (2015) The re-emergence of natural products for drug discovery in the genomics era. Nat Rev Drug Discov 14:111–129. doi: 10.1038/nrd4510

Henkel T, Brunne RM, Müller H, Reichel F (1999) Statistical Investigation into the Structural Complementarity of Natural Products and Synthetic Compounds. Angew Chem Int Ed 38:643–647. doi: 10.1002/(SICI)1521-3773(19990301)38:5<643:AID-ANIE643>3.0.CO;2-G

Henry SH, Bosch FX, Bowers JC (2002) Aflatoxin, hepatitis and worldwide liver cancer risks. In: Mycotoxins and Food Safety. Springer, Boston, MA, pp 229–233

Hu G-P, Yuan J, Sun L, et al (2011) Statistical research on marine natural products based on data obtained between 1985 and 2008. Mar Drugs 9:514–525. doi: 10.3390/md9040514

Huang M, Lu J-J, Huang M-Q, et al (2012) Terpenoids: natural products for cancer therapy. Expert Opin Investig Drugs 21:1801–1818. doi: 10.1517/13543784.2012.727395

Ivanescu B, Miron A, Corciova A (2015) Sesquiterpene lactones from *Artemisia* genus: Biological activities and methods of analysis. J Anal Methods Chem 247685. doi: 10.1155/2015/247685

Jassbi AR (2006) Chemistry and biological activity of secondary metabolites in *Euphorbia* from Iran. Phytochemistry 67:1977–1984. doi: 10.1016/j.phytochem.2006.06.030

Jose PA, Jebakumar SRD (2013) Non-streptomycete actinomycetes nourish the current microbial antibiotic drug discovery. Front Microbiol 4. doi: 10.3389/fmicb.2013.00240

Katz L, Baltz RH (2016) Natural product discovery: past, present, and future. J Ind Microbiol Biotechnol 43:155–176. doi: 10.1007/s10295-015-1723-5

Kayser O, Kiderlen AF, Croft SL (2003) Natural products as antiparasitic drugs. Parasitol Res 90:S55–S62. doi: 10.1007/s00436-002-0768-3

Kharwar RN, Mishra A, Gond SK, et al (2011) Anticancer compounds derived from fungal endophytes: their importance and future challenges. Nat Prod Rep 28:1208–1228. doi: 10.1039/C1NP00008J

King GF (2011) Venoms as a platform for human drugs: translating toxins into therapeutics. Expert Opin Biol Ther 11:1469–1484. doi: 10.1517/14712598.2011.621940

Koch MA, Schuffenhauer A, Scheck M, et al (2005) Charting biologically relevant chemical space: A structural classification of natural products (SCONP). Proc Natl Acad Sci 102:17272–17277. doi: 10.1073/pnas.0503647102

Kong D-X, Guo M-Y, Xiao Z-H, et al (2011) Historical variation of structural novelty in a natural product library. Chem Biodivers 8:1968–1977. doi: 10.1002/cbdv.201100156

Kurtböke Dİ (2012) Biodiscovery from rare actinomycetes: An eco-taxonomical perspective. Appl Microbiol Biotechnol 93:1843–1852. doi: 10.1007/s00253-012-3898-2

Langenheim JH (1994) Higher plant terpenoids: A phytocentric overview of their ecological roles. J Chem Ecol 20:1223–1280. doi: 10.1007/BF02059809

Lawton EM, Cotter PD, Hill C, Ross RP (2007) Identification of a novel two-peptide lantibiotic, Haloduracin, produced by the alkaliphile *Bacillus halodurans* C-125. FEMS Microbiol Lett 267:64–71. doi: 10.1111/j.1574-6968.2006.00539.x

Lazzarini A, Cavaletti L, Toppo G, Marinelli F (2000) Rare genera of actinomycetes as potential producers of new antibiotics. Antonie Van Leeuwenhoek 78:399–405. doi: 10.1023/A:1010287600557

Leal MC, Puga J, Serôdio J, et al (2012) Trends in the discovery of new marine natural products from invertebrates over the last two decades – Where and what are we bioprospecting? PLoS One 7:e30580. doi: 10.1371/journal.pone.0030580

Lewis RJ, Garcia ML (2003) Therapeutic potential of venom peptides. Nat Rev Drug Discov 2:790–802. doi: 10.1038/nrd1197

Lindequist U, Niedermeyer THJ, Jülich W-D (2005) The pharmacological potential of mushrooms. Evid Based Complement Alternat Med 2:285–299. doi: 10.1093/ecam/neh107

Locey KJ, Lennon JT (2016) Scaling laws predict global microbial diversity. Proc Natl Acad Sci 113:5970–5975. doi: 10.1073/pnas.1521291113

Mangal M, Sagar P, Singh H, et al (2013) NPACT: Naturally occurring plant-based anti-cancer compound-activity-target database. Nucleic Acids Res 41:D1124–D1129. doi: 10.1093/nar/gks1047

Mariottini GL, Grice ID (2016) Antimicrobials from cnidarians. A new perspective for anti-infective therapy? Mar Drugs 14:48. doi: 10.3390/md14030048

May RM (2010) Tropical arthropod species, more or less? Science 329:41–42. doi: 10.1126/science.1191058

Mehbub MF, Lei J, Franco C, Zhang W (2014) Marine sponge derived natural products between 2001 and 2010: Trends and opportunities for discovery of bioactives. Mar Drugs 12:4539–4577. doi: 10.3390/md12084539

Miller CM, Miller RV, Garton-Kenny D, et al (1998) Ecomycins, unique antimycotics from *Pseudomonas viridiflava*. J Appl Microbiol 84:937–944. doi: 10.1046/j.1365-2672.1998.00415.x

Mishra BB, Tiwari VK (2011) Natural products: An evolving role in future drug discovery. Eur J Med Chem 46:4769–4807. doi: 10.1016/j.ejmech.2011.07.057

Mishra HN, Das C (2003) A review on biological control and metabolism of aflatoxin. Crit Rev Food Sci Nutr 43:245–264. doi: 10.1080/10408690390826518

Mora C, Tittensor DP, Adl S, et al (2011) How many species are there on earth and in the ocean? PLoS Biol 9:e1001127. doi: 10.1371/journal.pbio.1001127

Newman DJ, Cragg GM (2016) Natural products as sources of new drugs from 1981 to 2014. J Nat Prod 79:629–661. doi: 10.1021/acs.jnatprod.5b01055

Ntie-Kang F, Zofou D, Babiaka SB, et al (2013) AfroDb: A select highly potent and diverse natural product library from African medicinal plants. PLoS One 8:e78085. doi: 10.1371/journal.pone.0078085

Oprea TI (2000) Current trends in lead discovery: Are we looking for the appropriate properties? Mol Divers 5:199–208. doi: 10.1023/A:1021368007777

Palazzolo AME, Simons CLW, Burke MD (2017) The natural productome. Proc Natl Acad Sci 114:5564–5566. doi: 10.1073/pnas.1706266114

Pascolutti M, Campitelli M, Nguyen B, et al (2015) Capturing nature's diversity. PLoS One 10:e0120942. doi: 10.1371/journal.pone.0120942

Paterson RRM (2006) Ganoderma – A therapeutic fungal biofactory. Phytochemistry 67:1985–2001. doi: 10.1016/j.phytochem.2006.07.004

Patridge E, Gareiss P, Kinch MS, Hoyer D (2016) An analysis of FDA-approved drugs: natural products and their derivatives. Drug Discov Today 21:204–207. doi: 10.1016/j.drudis.2015.01.009

Pedrós-Alió C (2006) Marine microbial diversity: can it be determined? Trends Microbiol 14:257–263. doi: 10.1016/j.tim.2006.04.007

Peláez F (2005) Biological activities of fungal metabolites. In: Handbook of Industrial Mycology. Marcel Dekker, New York, pp 49–92

Pimm SL, Jenkins CN, Abell R, et al (2014) The biodiversity of species and their rates of extinction, distribution, and protection. Science 344:1246752. doi: 10.1126/science.1246752

Pimm SL, Joppa LN (2015) How many plant species are there, where are they, and at what rate are they going extinct? Ann Mo Bot Gard 100:170–176. doi: 10.3417/2012018

Pouny I, Batut M, Vendier L, et al (2014) Cytisine-like alkaloids from *Ormosia hosiei* Hemsl. & E.H. Wilson. Phytochemistry 107:97–101. doi: 10.1016/j.phytochem.2014.07.022

Pye CR, Bertin MJ, Lokey RS, et al (2017) Retrospective analysis of natural products provides insights for future discovery trends. Proc Natl Acad Sci 114:5601–5606. doi: 10.1073/pnas.1614680114

Quinn RJ, Carroll AR, Pham NB, et al (2008) Developing a drug-like natural product library. J Nat Prod 71:464–468. doi: 10.1021/np070526y

Rask-Andersen M, Masuram S, Schiöth HB (2014) The druggable genome: evaluation of drug targets in clinical trials suggests major shifts in molecular class and indication. Annu Rev Pharmacol Toxicol 54:9–26. doi: 10.1146/annurev-pharmtox-011613-135943

Reed JL, Palsson BØ (2003) Thirteen years of building constraint-based in silico models of *Escherichia coli*. J Bacteriol 185:2692–2699. doi: 10.1128/JB.185.9.2692-2699.2003

Rocha J, Calado R, Leal M (2015) Marine bioactive compounds from cnidarians. In: Springer Handbook of Marine Biotechnology. Springer, Berlin, Heidelberg, pp 823–849

Rocha J, Peixe L, Gomes NCM, Calado R (2011) Cnidarians as a source of new marine bioactive compounds—An overview of the last decade and future steps for bioprospecting. Mar Drugs 9:1860–1886. doi: 10.3390/md9101860

Rosén J, Gottfries J, Muresan S, et al (2009) Novel chemical space exploration via natural products. J Med Chem 52:1953–1962. doi: 10.1021/jm801514w

Roskov Y, Abucay L, Orrell T, et al (2018) Species 2000 & ITIS catalogue of life, 2018 Annual Checklist. www.catalogueoflife.org/col. Accessed 4 Jun 2018

Russo M, Russo GL, Daglia M, et al (2016) Understanding genistein in cancer: The "good" and the "bad" effects: A review. Food Chem 196:589–600. doi: 10.1016/j.foodchem.2015.09.085

Sanchez JF, Somoza AD, Keller NP, Wang CCC (2012) Advances in *Aspergillus* secondary metabolite research in the post-genomic era. Nat Prod Rep 29:351–371. doi: 10.1039/C2NP00084A

Sarker SD, Nahar L (2018) Chapter 1 - An introduction to computational phytochemistry. In: Sarker SD, Nahar L (eds) Computational Phytochemistry. Elsevier, pp 1–41

Scheffers BR, Joppa LN, Pimm SL, Laurance WF (2012) What we know and don't know about Earth's missing biodiversity. Trends Ecol Evol 27:501–510. doi: 10.1016/j.tree.2012.05.008

Schmidt U, Struck S, Gruening B, et al (2009) SuperToxic: a comprehensive database of toxic compounds. Nucleic Acids Res 37:D295–D299. doi: 10.1093/nar/gkn850

Schoch CL, Sung G-H, López-Giráldez F, et al (2009) The Ascomycota tree of life: A phylum-wide phylogeny clarifies the origin and evolution of fundamental reproductive and ecological traits. Syst Biol 58:224–239. doi: 10.1093/sysbio/syp020

Schulz B, Boyle C, Draeger S, et al (2002) Endophytic fungi: a source of novel biologically active secondary metabolites. Mycol Res 106:996–1004. doi: 10.1017/S0953756202006342

Sertuerner F (1817) Ueber das Morphium, eine neue salzfähige Grundlage, und die Mekonsäure, als Hauptbestandtheile des Opiums. Annalen der Physik 55: 56-89. doi: 10.1002/andp.18170550104

Shi Q-W, Su X-H, Kiyota H (2008) Chemical and pharmacological research of the plants in genus *Euphorbia*. Chem Rev 108:4295–4327. doi: 10.1021/cr078350s

Stadler M, Hoffmeister D (2015) Fungal natural products—the mushroom perspective. Front Microbiol 6:. doi: 10.3389/fmicb.2015.00127

Stein T (2005) *Bacillus subtilis* antibiotics: structures, syntheses and specific functions. Microbiol Mol 56:845–857. doi: 10.1111/j.1365-2958.2005.04587.x

Stierle A, Strobel G, Stierle D (1993) Taxol and taxane production by *Taxomyces andreanae*, an endophytic fungus of Pacific yew. Science 260:214–216. doi: 10.1126/science.8097061

Stone JK, Bacon CW, White JF (2000) An overview of endophytic microbes: Endophytism defined. In: Microbial Endophytes. Marcel Dekker, New York, pp 3–29

Strebhardt K, Ullrich A (2008) Paul Ehrlich's magic bullet concept: 100 years of progress. Nat Rev Cancer 8:473–480. doi: 10.1038/nrc2394

Strobel G, Daisy B, Castillo U, Harper J (2004) Natural products from endophytic microorganisms. J Nat Prod 67:257–268. doi: 10.1021/np030397v

Su C-Y, Ming Q-L, Rahman K, et al (2015) *Salvia miltiorrhiza*: Traditional medicinal uses, chemistry, and pharmacology. Chin J Nat Med 13:163–182. doi: 10.1016/S1875-5364(15)30002-9

Tan D, Wu J, Zhang X, et al (2018) Sodium tanshinone II a sulfonate injection as adjuvant treatment for unstable angina pectoris: A meta-analysis of 17 randomized controlled trials. Chin J Integr Med 24:156–160. doi: 10.1007/s11655-017-2424-x

Tan RX, Zou WX (2001) Endophytes: a rich source of functional metabolites. Nat Prod Rep 18:448–459. doi: 10.1039/B100918O

Tao X, Lipsky PE (2000) The Chinese anti-inflammatory and immunosuppressive herbal remedy *Tripterygium wilfordii* Hook. f. Rheum Dis Clin 26:29–50. doi: 10.1016/S0889-857X(05)70118-6

Thacker RW, Starnes S (2003) Host specificity of the symbiotic cyanobacterium *Oscillatoria spongeliae* in marine sponges, *Dysidea* spp. Mar Biol 142:643–648. doi: 10.1007/s00227-002-0971-x

Thomas TRA, Kavlekar DP, LokaBharathi PA (2010) Marine drugs from sponge-microbe association—A review. Mar Drugs 8:1417–1468. doi: 10.3390/md8041417

Tiwari K, Gupta RK (2012) Rare actinomycetes: a potential storehouse for novel antibiotics. Crit Rev Biotechnol 32:108–132. doi: 10.3109/07388551.2011.562482

Valli M, dos Santos RN, Figueira LD, et al (2013) Development of a natural products database from the biodiversity of Brazil. J Nat Prod 76:439–444. doi: 10.1021/np3006875

Van Soest RWM, Boury-Esnault N, Vacelet J, et al (2012) Global diversity of sponges (Porifera). PLoS One 7:e35105. doi: 10.1371/journal.pone.0035105

Vasas A, Hohmann J (2014) *Euphorbia* diterpenes: Isolation, structure, biological activity, and synthesis (2008–2012). Chem Rev 114:8579–8612. doi: 10.1021/cr400541j

Veitch NC (2010) Flavonoid chemistry of the Leguminosae. In: Recent Advances in Polyphenol Research. Wiley-Blackwell, pp 23–58

Wang J, Zhang L, Teng K, et al (2014) Cerecidins, novel lantibiotics from *Bacillus cereus* with potent antimicrobial activity. Appl Environ Microbiol 80:2633–2643. doi: 10.1128/AEM.03751-13

Watve MG, Tickoo R, Jog MM, Bhole BD (2001) How many antibiotics are produced by the genus *Streptomyces*? Arch Microbiol 176:386–390. doi: 10.1007/s002030100345

Whittle M, Willett P, Klaffke W, van Noort P (2003) Evaluation of similarity measures for searching the Dictionary of Natural Products database. J Chem Inf Comput Sci 43:449–457. doi: 10.1021/ci025591m

Wilson MC, Mori T, Rückert C, et al (2014) An environmental bacterial taxon with a large and distinct metabolic repertoire. Nature 506:58–62. doi: 10.1038/nature12959

Wink M (2013) Evolution of secondary metabolites in legumes (Fabaceae). S Afr J Bot 89:164–175. doi: 10.1016/j.sajb.2013.06.006

World Health Organization (2008) WHO | "Beijing declaration." In: WHO. http://www.who.int/medicines/areas/traditional/congress/beijing_declaration/en/. Accessed 16 Dec 2018

Wu Y-B, Ni Z-Y, Shi Q-W, et al (2012) Constituents from *Salvia* species and their biological activities. Chem Rev 112:5967–6026. doi: 10.1021/cr200058f

Xue R, Fang Z, Zhang M, et al (2013) TCMID: traditional Chinese medicine integrative database for herb molecular mechanism analysis. Nucleic Acids Res 41:D1089–D1095. doi: 10.1093/nar/gks1100

Ye H, Ye L, Kang H, et al (2011) HIT: linking herbal active ingredients to targets. Nucleic Acids Res 39:D1055–D1059. doi: 10.1093/nar/gkq1165

Yuen JWM, Gohel MDI (2005) Anticancer effects of *Ganoderma lucidum*: A review of scientific evidence. Nutr Cancer 53:11–17. doi: 10.1207/s15327914nc5301_2

Zambelli VO, Pasqualoto KFM, Picolo G, et al (2016) Harnessing the knowledge of animal toxins to generate drugs. Pharmacol Res 112:30–36. doi: 10.1016/j.phrs.2016.01.009

Zhu F, Qin C, Tao L, et al (2011) Clustered patterns of species origins of nature-derived drugs and clues for future bioprospecting. Proc Natl Acad Sci 108:12943–12948. doi: 10.1073/pnas.1107336108

Supplementary Materials

A. Most represented families, genera and species from Plantae Kingdom (by decreasing order)

Biological source[a]				Total NPs[b]		Total NPs bioactive[c]	
Kingdom	Family	Genus	Species	*N*	%	*N*	%
PLANTAE				133,881	67.3	1,163	0.9
	Compositae			17,544	12.0	151	0.9
		Artemisia		1,299	7.4	21	1.6
			A. annua	83	6.4	2	2.4
			A. herba-alba	73	5.6	1	1.4
			A. absinthium	50	3.8	0	0.0
		Senecio		865	4.9	6	0.7
			S. nemorensis	28	3.2	1	3.6
			S. scandens	20	2.3	0	0.0
			S. caudatus	15	1.7	0	0.0
		Ligularia		604	3.4	0	0.0
			L. virgaurea	79	13.1	0	0.0
			L. sagitta	40	6.6	0	0.0
			L. dentata	39	6.5	0	0.0
			*Helianthus annuus**	263	69.8	0	0.0
			*Carthamus tinctorius**	123	72.4	0	0.0
			*Petasites japonicus**	103	45.6	0	0.0
	Leguminosae			10,751	7.4	138	1.3
		Cassia		622	5.8	8	1.3
			C. fistula	66	10.6	0	0.0
			C. siamea	58	9.3	0	0.0
			C. obtusifolia	52	8.4	0	0.0
		Glycyrrhiza		481	4.5	20	4.2
			*G. uralensis***	162	33.7	7	4.3
			G. glabra	149	31.0	11	7.4
			G. inflata	45	9.4	1	2.2
		Erythrina		479	4.5	2	0.4
			E. variegata	77	16.1	0	0.0
			E. abyssinica	74	15.4	1	1.4
			E. lysistemon	36	7.5	0	0.0
			*Glycine max**	174	95.1	4	2.3
			*Pisum sativum**	172	92.0	5	2.9
	Labiatae			6,049	4.1	46	0.8
		Salvia		1,166	19.3	9	0.8
			*S. miltiorrhiza***	132	11.3	4	3.0
			S. officinalis	49	4.2	1	2.0
			S. prionitis	36	3.1	0	0.0
		Isodon		895	14.8	4	0.4
			*I. rubescens***	126	14.1	1	0.8
			*I. eriocalyx***	98	10.9	0	0.0
			I. japonicus	47	5.3	2	4.3
		Scutellaria		421	7.0	7	1.7
			S. barbata	78	18.5	0	0.0
			S. baicalensis	64	15.2	4	6.3
			S. rivularis	25	5.9	0	0.0

[a]Only the three most represented taxonomic rank is shown for each level of classification
[b]Percentage shown in the table compare the fraction of NPs into the taxonomic rank reported (e.g. for Plantae Kingdom, 67.3% of NPs belong to the Plantae Kingdom compare to other Kingdoms of life)
[c]Percentage shown in the table compare the fraction of bioactives NPs with total NPs (e.g. In Artemisia genus, there is 1,299 total NPs, among which 1.5% (20) are bioactive)
*Most represented species in the family
**Biological sources which also belong to the three most represented species in the family

B. Most represented families, genera and species from Bacteria Kingdom (by decreasing order)

Biological source[a]					Total NPs[b]		Total NPs bioactive[c]	
Kingdom	Division	Family	Genus	Species	*N*	%	*N*	%
BACTERIA					17,531	8.8	1,006	5.7
	Actinobacteria				9,261	52.8	685	7.4
		Streptomycetaceae			7,969	86.0	599	7.5
			Streptomyces		7,951	99.8	597	7.5
				S. hygroscopicus	353	4.4	44	12.5
				S. griseus	204	2.6	20	9.8
				S. galilaeus	122	1.5	4	3.3
		Rare Actinomycetales			1,278	13.8	101	7.9
			Micromonospora		436	34.1	84	19.3
				*M. chalcea***	38	8.7	9	23.7
				M. griseorubida	33	7.6	13	39.4
				M. olivoasterospora	30	6.9	7	23.3
			Nocardia		312	24.4	32	10.3
				N. mediterranei	31	9.9	2	6.5
				N. brasiliensis	26	8.3	1	3.8
				N. asteroides	12	3.8	0	0.0
			Actinomadura		267	20.9	27	10.1
				A. madurae	17	6.4	1	5.9
				A. roseoviolacea	17	6.4	0	0.0
				A. verrucosospora	12	4.5	0	0.0
				*Saccharopolyspora erythraea**	63	45.3	5	7.9
				*Amycolatopsis orientalis**	39	28.1	6	15.4
		Mycobacteriaceae			136	1.5	7	5.1
			Mycobacterium		136	100.0	7	5.1
				M. tuberculosis	38	27.9	1	2.6
				M. phlei	18	13.2	1	5.6
				M. smegmatis	18	13.2	1	5.6
	Eubacteria***				7,157	40.8	365	5.1
			Pseudomonas		952	13.3	40	4.2
				P. fluorescens	169	17.8	10	5.9
				P. aeruginosa	150	15.8	6	4.0
				P. putida	78	8.2	0	0.0
			Bacillus		849	11.9	39	4.6
				*B. subtilis***	252	29.7	14	5.6
				B. cereus	68	8.0	3	4.4
				B. circulans	45	5.3	4	8.9
			Escherichia		491	6.9	13	2.6
				*E. coli***	487	99.2	12	2.5
				E. freundii	2	0.4	0	0.0
				E. alcalescens	1	0.2	1	100.0
				*Sorangium cellulosum**	214	97.3	2	0.9
	Cyanobacteria				1,521	8.7	13	0.9
			Lyngbya		353	23.2	3	0.8
				*L. majuscula***	220	62.3	2	0.9
				L. confervoides	20	5.7	0	0.0
				L. bouillonii	19	5.4	0	0.0
			Microcystis		240	15.8	0	0.0
				*M. aeruginosa***	141	58.8	0	0.0
				M. viridis	6	2.5	0	0.0
				M. ichthyoblabe	2	0.8	0	0.0
			Nostoc		142	9.3	1	0.7
				N. commune	16	11.3	0	0.0
				N. insulare	8	5.6	0	0.0
				N. spongiaeforme	6	4.2	0	0.0
				*Nodularia spumigena**	43	86.0	0	0.0

[a]Only the most represented taxonomic rank is shown for each level of classification

[b]Percentage shown in the table compare the fraction of NPs into the taxonomic rank reported (e.g. for Plantae Kingdom, 67.3% of NPs belong to the Plantae Kingdom compare to other Kingdoms of life)

[c]Percentage shown in the table compare the fraction of bioactives NPs with total NPs (e.g. In Artemisia genus, there is 1,299 total NPs, among which 1.5% (20) are bioactive)

*Most represented species in the family

**Biological sources which also belong to the three most represented species in the family

***Eubacteria (other than Actinomycetales and cyanobacteria)

C. Most represented families, genera and species from Fungi Kingdom (by decreasing order)

Biological source[a]				Total NPs[b]		Total NPs bioactive[c]	
Kingdom	Division	Genus	Species	*N*	%	*N*	%
FUNGI				19,869	10.0	303	1.5
	Ascomycota			13,604	66.1	233	1.7
		Aspergillus		2,374	17.5	67	2.8
			*A. terreus***	244	10.3	9	3.7
			*A. fumigatus***	177	7.5	7	4.0
			A. versicolor	176	7.4	0	0.0
		Penicillium		2,104	15.5	31	1.5
			P. citrinum	150	7.1	1	0.7
			P. chrysogenum	88	4.2	0	0.0
			P. citreo-viride	73	3.5	1	1.4
		Fusarium		613	4.5	21	3.4
			F. oxysporum	82	13.4	4	4.9
			F. solani	74	12.1	10	13.5
			F. moniliforme	34	5.5	1	2.9
			*Saccharomyces cerevisiae**	220	94.8	4	1.8
	Basidiomycota			4,820	23.4	57	1.2
		Ganoderma		466	9.7	1	0.2
			*G. lucidum***	239	51.3	1	0.4
			G. applanatum	33	7.1	0	0.0
			G. sinense	29	6.2	0	0.0
		Lactarius		201	4.2	0	0.0
			L. vellereus	24	11.9	0	0.0
			L. necator	17	8.5	0	0.0
			L. scrobiculatus	16	8.0	0	0.0
		Phellinus		123	2.6	1	0.8
			P. igniarius	35	28.5	0	0.0
			P. linteus	20	16.3	0	0.0
			P. baumii	14	11.4	0	0.0
			*Hericium erinaceus**	76	78.4	0	0.0
			*Antrodia camphorata**	65	60.7	0	0.0
	Lichens			902	4.4	7	0.8
		Lecanora		89	9.9	1	1.1
			L. iseana	6	6.7	0	0.0
			L. rupicola	6	6.7	0	0.0
			L. brochaa	5	5.6	0	0.0
		Parmelia		86	9.5	3	3.5
			P. entotheiochroa	7	8.1	0	0.0
			P. notata	4	4.6	0	0.0
			P. perlata	4	4.6	0	0.0
		Cladonia		60	6.7	1	1.7
			C. cristatella	4	6.7	0	0.0
			C. furcata	4	6.7	0	0.0
			C. graciliformis	4	6.7	0	0.0
			*Evernia prunastri**	28	71.8	0	0.0
			*Umbilicaria proboscidea**	14	70.0	0	0.0
			*Graphis scripta**	12	35.3	0	0.0

[a]Only the most represented taxonomic rank is shown for each level of classification

[b]Percentage shown in the table compare the fraction of NPs into the taxonomic rank reported (e.g. for Plantae Kingdom, 67.3% of NPs belong to the Plantae Kingdom compare to other Kingdoms of life)

[c]Percentage shown in the table compare the fraction of bioactives NPs with total NPs (e.g. In Artemisia genus, there is 1,299 total NPs, among which 1.5% (20) are bioactive)

*Most represented species in the family

**Biological sources which also belong to the three most represented species in the family

D. Most represented families, genera and species from Animalia Kingdom (by decreasing order)

Biological source[a]				Total NPs[b]		Total NPs bioactive[c]	
Kingdom	Phylum	Genus	Species	*N*	%	*N*	%
ANIMALIA				25,064	12.6	403	1.6
	Porifera (sponges)			8,656	34.5	47	0.5
		Dysidea		454	5.2	1	0.2
			*D. herbacea***	88	19.4	0	0.0
			D. fragilis	69	15.2	0	0.0
			D. avara	25	5.5	1	4.0
		Xestospongia		356	4.1	5	1.4
			*X. testudinaria***	87	24.4	0	0.0
			X. exigua	21	5.9	2	9.5
			X. muta	19	5.3	0	0.0
		Plakortis		328	3.8	0	0.0
			P. simplex	82	25.0	0	0.0
			P. halichondrioides	30	9.1	0	0.0
			P. lita	26	7.9	0	0.0
			*Theonella swinhoei**	134	58.3	1	0.7
	Cnidaria (anemones, corals, jellyfish,..)			5,249	20.9	30	0.6
		Sinularia		723	13.8	0	0.0
			S. flexibilis	83	11.5	0	0.0
			S. gibberosa	55	7.6	0	0.0
			S. dissecta	52	7.2	0	0.0
		Sarcophyton		399	7.6	1	0.3
			S. glaucum	87	12.0	0	0.0
			S. trocheliophorum	56	7.7	0	0.0
			S. crassocaule	46	6.4	0	0.0
		Briareum		360	6.9	0	0.0
			*B. excavatum***	100	27.8	0	0.0
			*B. asbestinum***	91	25.3	0	0.0
			B. polyanthes	27	7.5	0	0.0
			*Clavularia viridis**	123	66.8	0	0.0
	Arthropoda			3,304	13.2	89	2.7
		Drosophila		100	3.0	4	4.0
			*D. melanogaster***	63	63.0	3	4.8
			D. subatrata	8	8.0	1	12.5
			D. arizonae	5	5.0	0	0.0
		Dilophus		96	2.9	0	0.0
			D. spiralis	39	40.6	0	0.0
			D. fasciola	14	14.6	0	0.0
			D. marginatus	14	14.6	0	0.0
		Bombyx		70	2.1	1	1.4
			*B. mori***	69	98.6	1	1.4
			B. batryticatus	1	1.4	0	0.0
			*Manduca sexta**	46	100.0	1	2.2

[a]Only the most represented taxonomic rank are shown for each level of classification

[b]Percentage shown in the table compare the fraction of NPs into the taxonomic rank reported (e.g. for Plantae Kingdom, 67.3% of NPs belong to the Plantae Kingdom compare to others Kingdoms of life)

[c]Percentage shown in the table compare the fraction of bioactives NPs with total NPs (e.g. In Artemisia genus, there is 1,299 total NPs, among which 1.5% (20) are bioactive)

*Most represented species in the family

**Biological sources which also belong to the three most represented species in the family