We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Chapter 10

Computational Studies and Biosynthesis of Natural Products with Promising Anticancer Properties

Aurélien F.A. Moumbock, Conrad V. Simoben, Ludger Wessjohann, Wolfgang Sippl, Stefan Günther and Fidele Ntie-Kang

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/67650

Abstract

We present an overview of computational approaches for the prediction of metabolic pathways by which plants biosynthesise compounds, with a focus on selected very promising anticancer secondary metabolites from floral sources. We also provide an overview of databases for the retrieval of useful genomic data, discussing the strengths and limitations of selected prediction software and the main computational tools (and methods), which could be employed for the investigation of the uncharted routes towards the biosynthesis of some of the identified anticancer metabolites from plant sources, eventually using specific examples to address some knowledge gaps when using these approaches.

Keywords: anticancer, biosynthesis, computational prediction, natural products, plant metabolism

1. Introduction

An immense number of secondary metabolites (SMs) exist in nature, originating from plants, bacteria, fungi and marine life forms, serving as drugs for the treatment of many life-threatening diseases, including cancer [1–4]. Taxol, vinblastine, vincristine, podophyllotoxin and camptothecin, for example, are typically well-known drugs used in cancer treatment, which are of plant origin. The search for drugs against cancer has often resorted to plants and marine life for lead compounds. To illustrate this, Newmann and Cragg published a recent study in which it was shown that ~49% of drugs used in cancer treatment were either natural products

INTECH
open science | open minds

© 2017 The Author(s). Licensee InTech. This chapter is distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. CC BY

(NPs) or their derivatives [5]. We would henceforth refer to SMs and NPs interchangeably, since NPs are the products of secondary (or specialised) metabolism, as opposed to primary metabolism, which results in molecules playing a key role in physiological processes of the organism and are thus necessary for the plant's survival. It should be mentioned that SMs are important for the plant's defence against attacks by other organisms. Several efforts have also been made towards the collection of data on naturally occurring plant metabolites showing anticancer properties. As an example, Mangal and co-workers published the naturally occurring plant-based anti-cancer compound activity-target database (NPACT), containing about 1,500 NPs [6]. In addition to the experimentally verified *in vitro* and *in vivo* data for these NPs, the authors also include biological activities (in the form of IC_{50}s, ED_{50}s, EC_{50}s, GI_{50}s, etc.), along with physical, elemental and topological properties of the NPs, the tested cancer types, cell lines, protein targets, commercial suppliers and drug likeness of the NPACT compounds. A similar effort was published the following year, for NPs from African flora, resulting in a dataset of about 400 compounds, named AfroCancer [7]. A further study showed that the NPACT and AfroCancer datasets showed little intersection, thus providing us a combined dataset of about 2,000 NPs [8]. The anticancer properties of some of the most promising AfroCancer compounds have been described in detail in recent reviews [9–12]. Further curation of data from Northern African species has recently resulted in the Northern African Natural Products Database (NANPDB), a web accessible and completely downloadable vast database of NPs, with a significant proportion of anticancer metabolites [13]. The NANPDB effort was founded on the observation that the Northern Africa region is particularly highly endowed with diverse vegetation types, serving as a huge reservoir of bioactive natural products [14–16].

For decades, NPs were identified exclusively by using chemical identification based on bioactivity-guided screening approaches. Recently, it has been postulated that genomics and bioinformatics would transform the approach of natural products discovery, even though genome mining has had only little influence on the advancement of natural product discovery until now [17]. Several algorithms have been developed for the mining of the (meta)genomic data, which continue to be generated. Computational methods and tools for the identification of biosynthetic gene clusters (BGCs, which are physically clustered groups of a few genes in a particular genome that together encode a biosynthetic pathway for the production of a specialised metabolite) in genome sequences and the prediction of chemical structures of their products have been developed [18]. BGCs for SM biosynthetic pathways are important in bacteria and filamentous fungi, with examples being recently discovered in plants [19, 20], although some metabolic processes in plants, for example, the thalianol pathway for triterpene synthesis in *Arabidopsis thaliana* has been suggested to be controlled by operon-like (clusters of unrelated) gene clusters [21]. This, coupled with the rapid progress in sequencing technologies has led to the development of new screening methods, which focus on whole genome sequences of the organisms producing the NPs. Genome mining approaches for NP discovery basically focus on:

- identifying the genes of the organism involved in the biosynthesis of the NPs,
- identifying the metabolic pathways by which the NPs are biosynthesised and
- predicting the products of the identified pathways (**Figure 1A**).

http://dx.doi.org/10.5772/67650

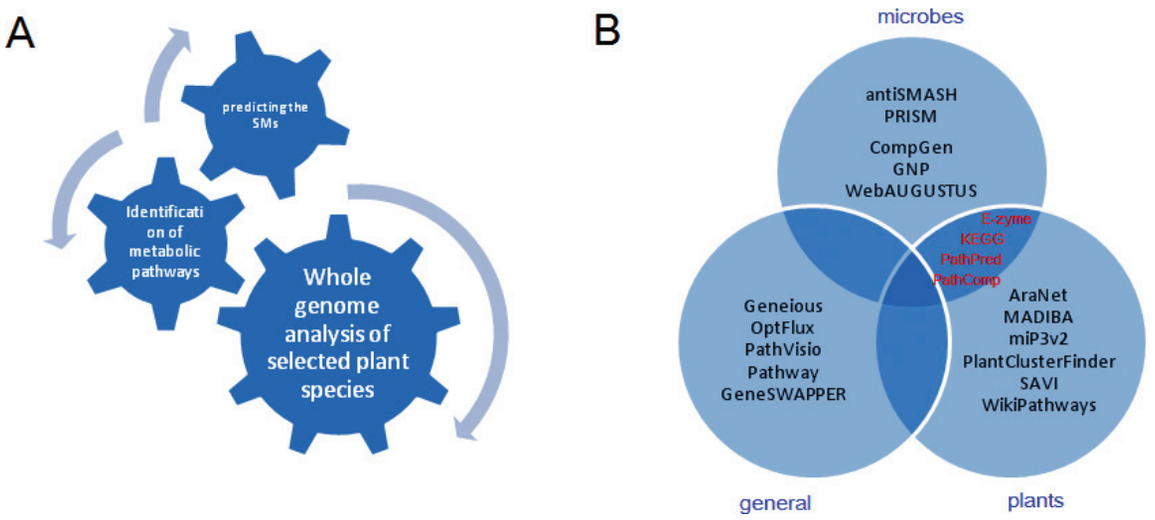


Figure 1. (A) Summary of genome mining approaches for the discovery of SMs and (B) classification of tools by applicability domain.

The four main strategies that are mostly employed to identify such pathways are based on processes involved in the production of plant secondary metabolites, for example, physical clustering, co-expression, evolutionary co-occurrence and epigenomic co-regulation of the genes [22–25]. Such approaches have been successfully applied for the investigation of fungal and microbial metabolites [26–28]. Since the discovery of the first gene cluster for secondary metabolism in *Zea mays*, the corn species [29], BGCs for plant secondary metabolism have become an emerging theme in plant biology [30]. It is even believed that synthetic biology technologies will eventually lead to the effective functional reconstitution of candidate pathways using a variety of genetic systems [25]. A knowledge of BGCs and their manipulation is therefore important in understanding how to activate a number of 'silent' gene clusters observed from the investigation of whole-genome sequencing of organisms. This would make available a wealth of new chemical entities (NCEs), which could be evaluated as drug leads and biologically active compounds [20].

This chapter aims at discussing the metabolic pathways by which plants biosynthesise compounds with anticancer activities, with a focus on selected very promising anticancer SMs from the African flora. We also aim to provide an overview of computational tools, which have been used to predict metabolic pathways and eventually address knowledge gaps when using the former. Additionally, we will present some databases for the retrieval of useful genomic data, discuss the strengths and limitations of selected computational (prediction) tools, which could be employed for the investigation of the uncharted routes towards the biosynthesis of some of the identified anticancer metabolites from plant sources, with specific examples. It is believed that properly addressing knowledge gaps that exist would lay the foundation for proper future investigations.

2. Natural products and plant genomic data

Genome data mining indicates that the vast majority of plant-based NPs have not yet been discovered [24, 25]. In addition, SMs are normally produced only at later growth stages of

plant metabolism and are frequently found only at low concentrations within complex mixtures in plant extracts, due to several factors. Some of these factors include physiological variations, geographic variations, environmental conditions and genetic factors [25, 31, 32]. The aforementioned factors are the main drawbacks in the isolation and purification of NPs in meaningful quantities for either research or commercial aims. Nowadays, BGCs can be investigated using computational methodologies and used to predict the NPs present in microbial, fungal and floral matter [18, 20, 33, 34]. It is current knowledge that more than 70 genome sequences for several plant species have been made available, along with a wealth of transcriptome data [25]. However, the interpretation of such data, for example, the translation of predicted sequences into enzymes, pathways and SMs remains challenging. Advances in bioinformatics and synthetic biology have permitted the cheap and efficient overproduction of secondary metabolites of medicinal interest in heterologous (non-native) host organisms by reengineering of BGCs [35]. This is carried out through reengineering of BGCs as well as the activation of silent BGCs to yield unreported natural products of the target chemical space [17, 36], for example, an engineered *Escherichia coli* strain was used as the heterologous host organism for the production of taxadiene (a vital precursor of paclitaxel, an anticancer agent isolated from the bark of *Taxus brevifolia*), a precursor of the anticancer agent taxol [37]. In this way, quite a number of interesting SMs of plant origin (e.g. resveratrol, vanillin, conolidin, etc.) have been objects of pathway engineering in bacteria, yeast and other plants [38]. Thus, chemical libraries of diverse and novel hybrid natural products analogues can now be generated through combinatorial biosynthesis by manipulation of biosynthetic enzymes [39], for example, several analogues of the antibiotic erythromycin were obtained *via* combinatorial biosynthesis [40]. Such bioengineered libraries of 'unnatural' natural products show promises in drug discovery campaigns against multidrug-resistant cancer cells.

3. Some database resources for retrieving secondary metabolism prediction information

A summary of databases for retrieving information on BGCs is provided in **Table 1**. A majority of them focus on microbial BGCs, for example, ClusterMine360, ClustScan, DoBISCUIT, IMG-ABC and the Recombinant ClustScan Database. Details on the utility of the aforementioned databases have been provided in excellent recent reviews [26–28, 53]. Further efforts towards the construction of plant-based BGC and genomic databases include those of the Medicinal Plants Genomics and Metabolomics Resource consortium [47]. This effort has been focused on 14 medicinal plants and includes a BLAST search module, a genome browser, a genome putative search function tool and transcriptome search tools. While the entire database is available for download, similar efforts from the Plant Metabolic Network (PMN) have the advantage of having included several plant metabolic pathway databases, mostly among food crops [49, 50]. The PMN, for example, currently houses one multi-species reference database called PlantCyc and 22 species/taxon-specific databases, providing access to manually curated and/or computationally predicted information about enzymes, pathways, and more for individual species.

Database	Description	Web accessibility	Advantages	Disadvantages	Reference
ClusterMine360	A database of microbial polyketide and non-ribosomal peptide gene clusters.	http://www.clustermine360.ca/	Users can make contributions. Automation leads to high data consistency and quality data.	Focuses only on microbial PKS/NRPS biosynthesis	[41, 42]
ClustScan Database	A database for *in silico* detection of promising new compounds.	http://csdb.bioserv.pbf.hr/csdb/ClustScanWeb.html	Allows easy extraction of DNA and protein sequences of polypeptides, modules, and domains.	Currently includes data for only 57 SMs (PKS), 51 SMs (NRPS) and 62 SMs (PKS-NRPS hybrid) biosynthesis.	[43, 44]
DoBISCUIT	A database of secondary metabolite biosynthetic gene clusters.	http://www.bio.nite.go.jp/pks/	Provides standardised gene/module/domain descriptions related to the gene clusters. Available for download	Contains mostly data relating to bacterial species, mostly of the genus *Streptomyces*.	[45]
GenomeNet	A network of databases and computational services for genome research and related research areas in biomedical sciences.	http://www.genome.jp/	Provides several web accessible tools, e.g. KEGG, E-zyme, etc. See **Table 2**.		
IMG-ABC	A knowledge base for biosynthetic gene clusters for the discovery of novel SMs.	https://img.jgi.doe.gov/cgi-bin/abc-public/main.cgi	Integrates structural and functional genomics with annotated BGCs and associated SMs.	Not available for download. Limited to data on microbes	[46]
Medicinal Plants Genomics Resource	A database for medicinal plants genome sequence data.	http://medicinalplantgenomics.msu.edu/	Available for download	Only genomic data for 14 species are currently available.	[47]
Medicinal Plants Metabolomics Resource	A database for medicinal plants metabolomics data.	http://metnetweb.gdcb.iastate.edu/mpmr_public/	Available for download	Currently limited to metabolite data for 2 medicinal plant species.	[48]

http://dx.doi.org/10.5772/67650

Database	Description	Web accessibility	Advantages	Disadvantages	Reference
Minimum Information about a Biosynthetic Gene cluster (MIBiG)	A community standard for annotations and metadata on biosynthetic gene clusters and their molecular products.	http://mibig.secondarymetabolites.org/index.html	Facilitates the standardised deposition and retrieval of biosynthetic gene cluster data. Useful for the development of comprehensive comparative analysis tools. Available for download		[18]
Plant Metabolic Network (PMN)	Several plant metabolic pathway databases.	http://www.plantcyc.org/	Includes species/taxon-specific data for more than 22 plant species.		[49, 50]
Plant Reactome/"Cyc" Pathways	A pathway database for several crops and model plant species.	http://gramene.org/pathways	Currently includes gene homology-based pathway projections to 62 plant species.		[51]
Recombinant ClustScan Database	A database of gene cluster recombinants and their corresponding chemical structures.	http://csdb.bioserv.pbf.hr/csdb/RCSDB.html	Provides a virtual compound library, which could be a useful resource for computer-aided drug design of pharmaceutically relevant chemical entities.	Currently contains only 47 cluster combinations	[44, 52]
SMBP	Secondary metabolites bioinformatics portal.	http://www.secondarymetabolites.org/	Includes hand-curated links to all major tools and databases commonly used in the field		[53]

Table 1. Summary of currently available database resources for retrieving genomic data for biosynthesis prediction.

http://dx.doi.org/10.5772/67650

It provides a broad network of plant metabolic pathway databases that contain curated information from the literature and computational analyses about the genes, enzymes, compounds, reactions and pathways involved in primary and secondary metabolism in the included plant species. The PlantCyc database also provides access to manually curated or reviewed information about shared and unique metabolic pathways present in over 350 plant species. On the other hand, Plant Reactome is a pathway database for several crops and model plant species, making use of a framework of a eukaryotic cell model. Currently, it uses rice as a reference species and gene homology-based pathway projections have been made to 62 plant species [51].

4. Some computational tools for the analysis of genomic data and specialised metabolism prediction

Some computational tools for biochemical pathway prediction have been summarised in excellent reviews [54]. We have provided a more detailed summary of the main tools that could be useful in analysing plant and microbial genomic data for metabolism prediction in **Table 2**. Some of the tools are designed for the detection and analysis of specialised metabolism in microbes (e.g. antiSMASH, CompGen, GNP, PRISM and WebAUGUSTUS). Others are specially designed for plant metabolism prediction or may only include data for some specific organisms (e.g. AraNet, MADIBA, miP3v2, PlantClusterFinder, SAVI and WikiPathways for plants), while others are more general tools, useful for both microbial and plant metabolism prediction and BGC analysis (e.g. E-zyme, KEGG, PathPred and PathComp) and others are more useful for developers (e.g. Geneious, OptFlux, PathVisio and Pathway GeneSWAPPER), **Figure 1B**. We could also classify the tools according to their respective tasks; prediction and analysis of BGCs (e.g. antiSMASH, MADIBA, Pathway GeneSWAPPER, WebAUGUSTUS), searching, visualisation and prediction of biosynthetic pathways and reaction paths (e.g. BioCyc, CycSim, FMM, GNP, KEGG, MetaCyc, PathComp, PathPred, PathSearch, PathVisio, Pathway GeneSWAPPER, PlantClusterFinder, SAVI, WikiPathways for plants), prediction of SMs (PRISM), metabolic engineering (OptFlux), other functions (miP3v2). Among the tools for specialised metabolism in plants, AraNet is a probabilistic functional gene network (with currently a total of 27,029 protein-encoding genes) of *A. thaliana*. It is based on a modified Bayesian integration of data from multiple organisms, each data type being weighted based on how well it links genes that are known to function together in *A. thaliana*. Each interaction is associated with a log-likelihood score (LLS), which is a measure of the probability of an interaction representing a true functional linkage between two genes [56]. On the other hand, MADIBA facilitates the interpretation of *Plasmodium* and plant (data currently available for *Oryza sativa* and *A. thaliana*) gene clusters [64]. This tool eases the task by automating the post-processing stage during the assignment of biological meaning to gene expression clusters. MADIBA is designed as a relational database and has stored data from gene to pathway for the aforementioned species. Tools within the GUI allow the rapid analyses of each cluster with the view of identifying the Gene Ontology terms, as well as visualising the metabolic pathways where the genes are implicated, their genomic localisations, putative common

Tool	Utility	Web accessibility	Advantage	Disadvantage	Reference
antiSMASH*	A web server and tool for the automatic genomic identification and analysis of biosynthetic gene clusters.	http://antismash.secondarymetabolites.org.	Detects putative gene clusters of unknown types. Identifies similarities of identified clusters to any of 1172 clusters with known end products, etc.	Designed for analysis of BGCs in microbes.	[55]
AraNet	Gene function identification and genetic dissection of plant traits.	http://www.functionalnet.org/aranet/	Had greater precision than literature-based protein interactions (21%) for 55% of tested genes. Is highly predictive for diverse biological pathways.	Applicability is limited to one species - *A. thaliana.*	[56]
BioCyc/CycSim/MetaCyc	Online tools for genome-scale metabolic modelling.	https://biocyc.org/http://www.genoscope.cns.fr/cycsim https://metacyc.org/	Support the design and simulation of knockout experiments, e.g. deletions mutants on specified media, etc.		[57, 58]
CompGen	Carry out *in silico* homologous recombination between gene clusters.	http://csdb.bioserv.pbf.hr/csdb/RCSDB.html		Focuses on gene clusters encoding PKSs in *Streptomyces* sp. and related bacterial genera.	[52]
E-zyme	Assignment of EC numbers.	http://www.genome.jp/tools/e-zyme/	Classifies enzymatic reactions and links the enzyme genes or proteins to reactions in metabolic pathways.		[59]
From Metabolite to metabolite (FMM)	A web server to find biosynthetic routes between two metabolites within the KEGG database.	http://FMM.mbc.nctu.edu.tw/	Both local and global graphical views of the metabolic pathways are designed.		[60]
Geneious	Organisation and analysis of sequence data.	http://www.geneious.com/basic	Includes a public application programming interface (API) available for developers. Freely available for download.		[61]
Genomes-to-Natural Products platform (GNP)	Prediction, combinatorial design and identification of PKs and NRPs from biosynthetic assembly lines.	http://magarveylab.ca/gnp/	Uses LC–MS/MS data of crude extracts to make predictions in a high-throughput manner.	Focuses on bacterial NPs.	[62]

Tool	Utility	Web accessibility	Advantage	Disadvantage	Reference
Gene Regulatory network inference ACcuracy Enhancement (GRACE)	An algorithm to enhance the accuracy of transcriptional gene regulatory networks.	https://dpb.carnegiescience.edu/labs/rhee-lab/software	Focuses on plant species. Available for download.	Only algorithm is available. Lacks a graphical user interface	
KEGG Mapper	A tool to search a biosynthetic pathway.	http://www.kegg.jp/kegg/tool/map_pathway1.html?rn	KEGG is applicable to all organisms and enables interpretation of high-level functions from genomic and molecular data.		[63]
MicroArray Data Interface for Biological Annotation (MADIBA)	A webserver toolkit for biological interpretation of *Plasmodium* and plant gene clusters.	http://www.bi.up.ac.za/MADIBA	It allows rapid gene cluster analyses and the identification of the relevant Gene Ontology terms, visualisation of metabolic pathways, genomic localisations, etc.	Only 2 plant species are currently considered [rice (*Oryza sativa*), and *A. thaliana*].	[64]
miP3v2	Predicts microproteins in a sequenced genome.	https://github.com/npklein/miP3	Sheds light on the prevalence, biological roles, and evolution of microProteins.	Only the algorithm is available. Lacks a graphical user interface	[65]
OptFlux	A software platform for *in silico* metabolic engineering.	http://www.optflux.org/	Open source platform. Integrates visualisation tools. Allows users to load a genome-scale model of a given organism. Wild type and mutants can be simulated. Available for download.		[66]
PathComp	Possible reaction path computation.	http://www.genome.jp/tools/pathcomp/			
PathPred	Prediction of biodegradation and/or biosynthetic pathways.	http://www.genome.jp/tools/pathpred/	Specifically designed for biosynthesis of SMs (in plants) and xenobiotics biodegradation of environmental compounds (by bacteria).		[67]
PathSearch	Search for similar reaction pathways.	http://www.genome.jp/tools/pathsearch/			

http://dx.doi.org/10.5772/67650

Tool	Utility	Web accessibility	Advantage	Disadvantage	Reference
PathVisio	A biological pathway analysis software that allows users to draw, edit and analyse biological pathways.	http://www.pathvisio.org/	Plugins are included, which provide advanced analysis methods, visualisation options or additional import/export functionality. Available for download.		[68, 69]
Pathway GeneSWAPPER	Maps homologous genes from one species onto the PathVisio pathway diagram of another species.	http://jaiswallab.cgrb.oregonstate.edu/software/PGS	Improves the functionalities of PathVisio and WikiPathways for plants.		[70]
PlantClusterFinder	Predicts metabolic gene clusters from plant genomes.	https://dpb.carnegiescience.edu/labs/rhee-lab/software	Focuses on plant species. Available for download.	Only the algorithm is available. Lacks a graphical user interface	
Prediction informatics for secondary metabolomes (PRISM)	Genomes to natural products prediction informatics for secondary metabolomes.	http://magarveylab.ca/prism/	Open-source, user-friendly web available application.	Focuses on microbial SMs.	[71]
RetroPath	A webserver for retrosynthetic pathway design.	http://www.jfaulon.com/bioretrosynth/	Integrates pathway prediction and ranking, prediction of compatibility with host genes, toxicity prediction and metabolic modeling.		[72, 73]
Semi-Automated Validation Infrastructure (SAVI)	Predicts metabolic pathways using pathway metadata (e.g. taxonomic distribution, key reactions, etc.).	https://dpb.carnegiescience.edu/labs/rhee-lab/software	Decides which pathways to keep, remove or validate manually. Available for download.	Only the algorithm is available. Lacks a graphical user interface.	
WebAUGUSTUS	Gene prediction tool.	http://bioinf.uni-greifswald.de/webaugustus	One of the most accurate tools for eukaryotic gene prediction.	Focuses on eukaryotes.	[74]
WikiPathways for plants	A community pathway curation portal.	http://plants.wikipathways.org	Freely available.	Currently limited to rice and *Arabidopsis* sp.	[70, 75, 76]

*Currently provided detection rules for 44 classes and subclasses of SMs.

Table 2. Summary of current computational tools which could be useful for the plant genomic data analysis.

http://dx.doi.org/10.5772/67650

transcriptional regulatory elements in the upstream sequences, and an analysis specific to the organism being studied.

PlantClusterFinder, SAVI and WikiPathways for plants are all purpose tools designed to assist in the prediction of metabolic gene cluster from plant genomes, although WikiPathways for plants has currently included mostly data for rice and *Arabidopsis* sp. SAVI has the added advantage of offering the user the possibility of including pathway metadata (e.g. taxonomic distribution, key reactions, etc.) and offering the possibility to decide which pathway(s) to keep and which to remove or validate manually.

5. Some computational methods for efficient production and the *de novo* engineering of natural products

Two main areas for computational tools can be distinguished: on the one hand the rational modification of genomes for the production of molecules by host organisms, and on the other hand the modification or the *de novo* design of gene clusters for the biosynthesis of novel NPs. For both genetic engineering approaches, the already known genomes of bacteria, fungi and more and more plants provide the basic datasets. A very important computational approach for a rational modification of NP-producing host organisms is the genome-scale metabolic modelling [77, 78].

Automatic assignments of functional annotations of all genes in a genome are ideally proven by manual curation and enriched by current knowledge about the metabolic network of subjected organisms. The curated genomes are then applied to a complete automatic reconstruction of the metabolic pathways of the cell. These metabolic models are normally encoded in the Systems Biology Markup Language (SBML) and are compatible with various software tools, for example, Cytoscape [79], which can be applied for static network analyses. For instance, missing enzymes (gaps) within the network become apparent by substrates that are not taken up or have not been produced by the cell, as well as products that are not consumed by other reactions and are not secreted from cell. The RAST annotation pipeline provides a full automatic server for predicting all gene functions and discovering new pathways in microbial genomes of bacteria [80]. Such models can then be used to predict the turnover rate of each reaction in a Flux Balance Analysis (FBA) [81]. Several tools have been built, which apply FBA to identify enzymes that should be either introduced or knocked-out in the organism to increase production rate in the host organisms. A widely used FBA package is the MATLAB-based COBRA Toolbox [82]. With CycSim [58], BioMet [83] and FAME [84] powerful web-based FBA applications were published that do not require any software installation.

Within the last 10 years, FBA was applied to support numerous genetic engineering approaches, for example, for the determination of minimal media in *Helicobacter pylori* [85], for growth rate predictions in *Bacillus subtilis* [86] or for the development of metabolic engineering strategies in *Pseudomonas putida* [87]. Based on FBA, it was possible to increase vanillin production in baker's yeast by twofold and enhance sesquiterpene production in the same species [88, 89].

The rational modification of a given genome to design novel molecules needs a detailed understanding of the producing gene clusters. Well-studied gene clusters such as polyketide synthases consist of specific domain types that can be identified by trained hidden Markov models that are stored in related databases, for example, PFAM [90]. Gene cluster analysis tools such as antiSMASH [55, 91] or PRISM [71] analyse a given gene cluster to predict the specific domains and to describe the architecture of a gene cluster. However, the prediction of the structure of the resulting natural products is a difficult task because substrate recognition of active sites and the correct ordering of enzymatic reactions has to be predicted. If subjected enzymes are catalysing multiple substrates, the availability of each substrate has to be predicted. Most frequently, the automatic analysis of a cluster is based on the deduction of information from gene clusters similar to the queried one. If well-annotated similar gene clusters do not exist, the prediction of the structure of the biosynthesised NP is challenging. With more and more knowledge about the structure of natural products and the encoding sequences, the relation between the composition of the active sites and substrate binding will be better understood. Existing algorithms are often based on machine-learning approaches and predict the correct substrates for a selected set of enzyme families [92]. For the prediction of NPs synthesised by non-ribosomal peptide synthetases, such a sequence-based prediction method is integrated in the related web-server NRPSpredictor2 [93]. Rational substitution of residues to generate novel molecules still requires a detailed manual analysis of the encoding gene cluster, and new software tools that propose mutations leading to novel molecules might accelerate this approach considerably in future.

6. Selected natural products with promising anticancer properties from African sources

Recent reviews on the anticancer potential of African flora have discussed the anticancer, cytotoxic, antiproferative and antitumour activities of about 500 NPs [9–12]. In this section, we focus on the most promising (recent) results for anticancer SMs from African flora (**Table 3**, **Figure 2**), published after the last reviews. The isolation of two new lignans; 3*α*-O-(β-*D*-glucopyranosyl) desoxypodophyllotoxin (**1**) and 4-O-(β-*D*-glucopyranosyl) dehydropodophyllotoxin (**2**), alongside other known lignans (**3** and **4**), have been reported from the species, *Cleistanthus boivinianus* (Phyllanthaceae), collected in Madagascar (coordinates 13°06′37″S 049°09′39″E) [94]. These compounds showed potent to moderate antiproliferative activities against the A2780 ovarian cancer cell line, with compound **1** showing potent antiproliferative activity against the HCT-116 human colon carcinoma cell line (IC_{50} = 0.03 μM). The known compounds with promising activities from this species included the lignans; (±)-β-apopicropodophyllin (**3**, PubChem CID: 6452099), (−)-desoxypodophyllotoxin (**4**, PubChem CID: 345501). The same authors also isolated a new butanolide, macrocarpolide A (**5**, PubChem CID: 122372160) and two new secobutanolides; macrocarpolides B (**6**, PubChem CID: 122372161) and C (**7**, PubChem CID: 122372162), together with other known compounds from the ethanol extract of the roots of the Madagascan

Cpd. No.*	Molecule class	Source species (Family)	Cancer cell line	IC_{50} (μM)	Biosynthetic pathway	References
1	lignan	*Cleistanthus boivinianus* (Phyllanthaceae)	HCT-116 human colon carcinoma cell line	0.03	shikimic acid pathway, via phenylalanine	[94]
			A2780 ovarian cancer cell line	0.02		
2	″	″	″	2.10	″	″
3	″	″	″	0.06	″	
4	″	″	″	0.23	″	
5	butanolide	*Ocotea macrocarpa* (Lauraceae)	″	2.57		[95]
6	secobutanolide	″	″	1.98		″
7	″	″	″	1.67		″
8	butanolide	″	″	2.43		″
9	″	″	″	1.65		″
10	polyoxygenated cyclohexene derivative	*Cleistochlamys kirkii* (Annonaceae)	MDA-MB-231 triple-negative human breast cancer cell line	0.03	Shikimic acid pathway	[96]
11	″	″	″	0.29	″	″
12	″	″	″	0.29	″	″
13	″	″	″	0.12	″	″
14	″	″	″	0.45	″	″
15	″	″	″	2.10	″	″
16	″	″	″	0.09	″	″
17	″	″	″	2.70	″	″
18	″	″	″	0.24	″	″

*Compound number.

Table 3. Summary of recently published selected promising anticancer SMs from African flora.

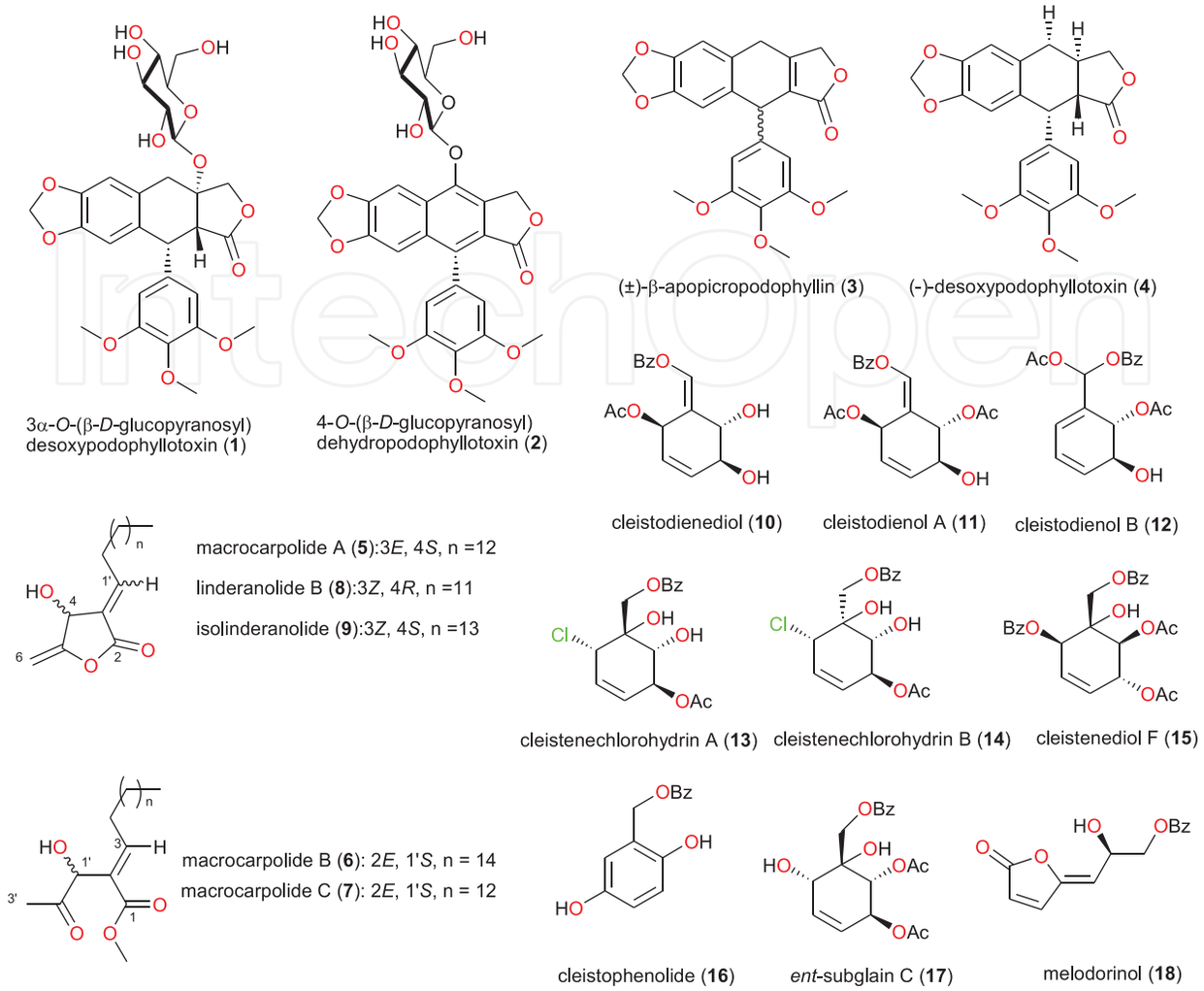


Figure 2. Chemical structures of selected anticancer SMs from African flora.

species *Ocotea macrocarpa* (Lauraceae), which showed antiproliferative activities against the A2780 ovarian cell line [95]. The known isolates included the butanolides; linderanolide B (**8**, PubChem CID: 53308122) and isolinderanolide (**9**, PubChem CID: 44576054). The anticancer activities showed IC_{50} values of 2.57 (**5**), 1.98 (**6**), 1.67 (**7**), 2.43 (**8**) and 1.65 μM (**9**) against A2780 ovarian cancer cell lines. Additionally, the leaves of *Cleistochlamys kirkii* (Annonaceae) from Tanzania have been recently shown to be a rich source of polyoxygenated cyclohexene derivatives with antiplasmodial activities, along with very potent activities against MDA-MB-231 triple-negative human breast cancer cell line [96]. The isolates; cleistodienediol (**10**), cleistodienol A (**11**), cleistodienol B (**12**), cleistenechlorohydrin A (**13**), cleistenechlorohydrin B (**14**), cleistenediol F (**15**), cleistophenolide (**16**), *ent*-subglain C (**17**) and melodorinol (**18**, PubChem CID: 6438687) showed some activities as low as IC_{50} = 0.09 μM against the aforementioned cancer cell lines. To the best of our knowledge, mode of action studies have not yet been conducted for the SMs **1** to **18** and *in vivo* activity data is currently unavailable.

http://dx.doi.org/10.5772/67650

7. Case studies

In this section, we shall discuss specific examples of the investigation of biosynthesis of anticancer plant-based SMs by (computational) analysis of genomic data.

7.1. Biogenesis of several anticancer metabolites by *Ocimum tenuiflorum* (Lamiaceae)

Species from the genus *Ocimum* are well known for their high medicinal values and are therefore used to cure a variety of ailments in Ayurveda, an Indian system of medicine [97, 98]. About 30 SMs have been reported from the genus *Ocimum*, with a variety of biological properties [99]. Only 14 of these SMs belong to the five basic groups of compounds having a complete biosynthetic pathway information in the PMN database [49, 50], thereby leaving us with ~15 medicinally relevant metabolites from *Ocimum* sp. with unknown pathways. This has prompted further investigation on SMs with uncharted biosynthetic pathways. Several bioactive SMs, including the anticancer compounds; apigenin (**19**, PubChem CID: 5280443), rosmarinic acid (**20**, PubChem CID: 5281792), taxol (**21**, PubChem CID: 36314), ursolic acid (**22**, PubChem CID: 64945), oleanolic acid (**23**, PubChem CID: 10494) and the plant steroid sitosterol (**24**, PubChem CID: 222284) have been identified from the herb Krishna Tulsi (*O. tenuiflorum*, Lamiaceae), with the mature leaves retaining the medicinally relevant metabolites [100]. Upadhyay et al. carried out a draft genome analysis of the species and generated paired-end and mate-pair sequence libraries for the whole sequenced genome, together with transcriptomic analysis (RNA-Seq) of two subtypes of *O. tenuiflorum* (Krishna and Rama Tulsi) and reporting the relative expression of genes in the both varieties. The authors further investigated the pathways, which lead to the biosynthesis of the identified SMs, with respect to similar pathways in *A. thaliana* and other model plants (e.g. *Oryza sativa japonica*). Six important genes (including *Q8RWT0* and *F1T282*) were expressed and identified from analysis of genome data. These were validated by q-RT-PCR on the different studied tissues (e.g. roots, mature leaves, etc.) of five closely related species (e.g. *O. gratissimum*, *O. sacharicum*, *O. kilmund, Solanum lycopersicum* and *Vitis vinifera*), which showed a high extent of urosolic acid-producing genes in young leaves. The other identified anticancer metabolites included eugenol and ursolic acid. As an example, the authors employed sequence search algorithms to search for the three enzymes of the three-step synthetic pathway of ursolic acid from squalene in the Tulsi genome. Each of these enzymes in Tulsi (squalene epoxidase, α-amyrin synthase and α-amyrin 2,8 monoxygenase) were queried from the PlantCyc database, starting from their protein sequences. The search for analogous enzymes in the model plants *O. sativa japonica* and *A. thaliana*, showed sequence identity covering from 50 to 80% of the query length. The whole genome and sequence analysis of *O. tenuiflorum* suggested that small amino acid changes at the functional sites of genes involved in metabolite synthesis pathways could confer special medicinal (particularly anticancer) properties to this herb.

7.2. Biosynthesis of the anticancer alkaloid noscapine by *Papaver somniferum* (Papaveraceae)

Noscapine (**25**, PubChem CID: 275196) is an antitumour phthalideisoquinoline alkaloid from opium poppy (*Papaver somniferum*, Papaveraceae). Compound **25** is known to bind

stoichiometrically to tubulin, alters its conformation, affects microtubule assembly (promotes microtubule polymerisation), hence arresting metaphase and inducing apoptosis in many cell types [101]. It has been demonstrated that the compound has potent antitumour activity against solid murine lymphoid tumours (even when the drug was administered orally). This drug has also shown potency against human breast, ovarian and bladder tumours implanted in nude mice and in dividing human cells [102, 103]. Although the compound is water-soluble and absorbed after oral administration, its chemotherapeutic potential in human cancer could not be fully exploited for drug discovery projects because, like most SMs, this has been limited by the typically small amounts produced in the slow-growing plant species [104]. The quest to improve production levels of the NP is essential for drug discovery. However, such would require a proper understanding biological processes underlying the biosynthesis of this SM, known from isotope-labelling experiments to be derived from scoulerine since the 1960s [105]. Winzer et al. have carried out a transcriptomic analysis, with the aim of elucidating the biosynthetic pathway of this important metabolite for the improvement of its commercial production in both poppy and other systems [106]. The analysis of a high noscapine-producing poppy variety, HN1, showed the exclusive expression of 10 genes encoding five distinct enzyme classes, whereas five functionally characterised genes (*BBE*, *TNMT*, *SaIR*, *SaIAT* and *T6ODM*) were present in all three of the studied poppy varieties, respectively, rich in morphine, thebaine and noscapine (HM1, HN1 and HT1). The authors analysed the expressed sequence tag (EST) abundance and discovered some previously uncharacterised genes expressed in HN1, which were completely absent from the other (HM1 and HT1) EST libraries. This led to the identification of the corresponding enzymes as three *O*-methyltransferases (*PSMT1*, *PSMT2*, *PSMT3*), four cytochrome P450s (*CYP82X1*, *CYP82X2*, *CYP82Y1* and *CYP719A21*), an acetyltransferase (*PSAT1*), a carboxylesterase (*PSCXE1*) and a short-chain dehydrogenase/reductase (*PSSDR1*). Further analysis of an F2 mapping population, using HN1 and HM1 as parents, indicated that these genes are tightly linked in HN1. Moreover, bacterial artificial chromosome sequencing confirmed the existence of a complex BGC for plant alkaloids. Based on the knowledge derived from the investigation, the authors could make suggestions for the improved production of noscapine and related bioactive molecules by the molecular breeding of commercial poppy varieties or engineering of new production systems, for example, by virus-induced gene silencing, which resulted in the accumulation of pathway intermediates, thus allowing gene function to be linked to noscapine synthesis [104, 106].

7.3. Biosynthesis of vinblastine and vincristine by *Catharanthus roseus* (Apocynaceae)

Vinblastine (**26**, PubChem CID: 13342) and vincristine (**27**, PubChem CID: 5978) are chemotherapy drugs used to treat a number of cancer types. These are among the >120 known terpenoid indole alkaloids from the medicinal plant *C. roseus*, also known as the Madagascar periwinkle [107]. Since these two very important anticancer compounds have only been produced in very low amounts in *C. roseus*, as opposed to the fairly high levels of several monomeric alkaloids (e.g. ajmalicine and serpentine) [108], attempts to improve the yields of compounds **26** and **27** have led to the genome-wide transcript profiling of elicited *C. roseus* cell cultures, by cDNA-amplified fragment-length polymorphism combined with metabolic

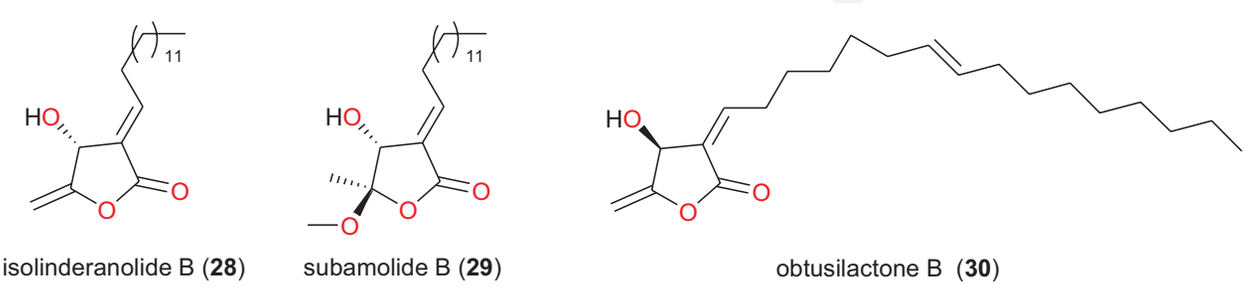
http://dx.doi.org/10.5772/67650

profiling [107]. This resulted in the identification of several gene-to-gene and gene-to-metabolite networks obtained by an attempt to establish correlations between the expression profiles of 417 gene tags and the accumulation profiles of 178 metabolite peaks. The results proved that different branches of terpenoid indole alkaloid biosynthesis and various other metabolic pathways are affected by differences in hormonal regulation. Thus, the investigations of Rischer et al. provided the foundations for a proper understanding of secondary metabolism in *C. roseus*, thereby enhancing the applicability of metabolic engineering of Madagascar periwinkle. This study provided the possibility of exploring a select number of genes (e.g. *STR*, *10HGO*, *T16H* and *DAT*) involved in biosynthesis of terpenoid indole alkaloids [107].

8. The way forward

The case studies show that the detailed computational analysis of the transcriptomic and metabolomic data of a plant species could reveal its metabolic capacity and hence help identify candidate genes involved in the biosynthesis of the important SMs it contains. Thus, modifying the plant genes could represent a premise for improving metabolite yield. It should be mentioned that other compounds from some of the aforementioned compound classes (**Table 3**), from both floral and microbial sources, have shown promising anticancer activities [109–113], e.g. isolinderanolide B (**28**, PubChem CID: 53308122) (**Figure 3**), a butanolide from the stems of *Cinnamomum subavenium* (Lauraceae) had shown antiproliferative activity in T24 human bladder cancer cells by blocking cell cycle progression and inducing apoptosis [112]. In addition, subamolide B (**29**, PubChem CID: 16104907), another butanolide from this same species, is known to induce cytotoxicity in human cutaneous squamous cell carcinoma through mitochondrial and CHOP-dependent cell death pathways [113]. Meanwhile, obtusilactone B (**30**, PubChem CID: 101286261), from *Machilus thunbergii* (Lauraceae), is known to target barrier-to-autointegration factor to treat cancer [111].

From the African flora, apart from the Lauraceae, Phyllanthaceae and Annonaceae, known to be rich in anticancer metabolites, the genus *Tacca* of the yam family (Dioscoreaceae) is known for the abundant presence of taccalonolides, which are microtubule stabilisers with clinical potential for cancer treatment [114]. Additionally, the genus *Tamarix* (e.g. *T. aphylla*



isolinderanolide B (**28**) subamolide B (**29**) obtusilactone B (**30**)

Figure 3. Chemical structures of selected anticancer butanolides from Lauraceae.

and *T. nilotica* from Northern Africa), together with the genus *Reaumuria* (Tamaricaceae) are known for the abundant presence of tannins (gallo-ellagitannin, gallotannins) with remarkable cytotoxic effects. The high salt content of the leaves of *Tamarix* species, rendering them useful locally as a fire barrier, and their adaptability to drought and high salinity are of equal interest. It therefore becomes urgent to investigate the genomics of some of the aforementioned plant species, particularly those from the *Cinnamomum* sp., *Ocotea* sp. and *Machilus* sp. (Lauraceae), *Tacca* sp. (Dioscoreaceae), *Cleistanthus* sp. (Phyllanthaceae), *Cleistochlamys* sp. (Annonaceae), *Tamarix* sp. (Tamaricaceae) and so on, and hence further investigate the genes or BGCs responsible for secondary metabolism with the view of understanding and better exploring the biosynthetic pathways of the anticancer SMs.

9. Conclusions

It has been our intention in this chapter to provide a detailed overview of the important computational tools and resources for the analysis of plant genomic data and for the prediction of biosynthetic pathways in plants. We have taken a few case studies of anticancer SMs to illustrate this. Even though it is unclear how widespread plant genes are clusters, genes that encode the biosynthesis of several small plant SMs are well known, including the vital genes for the production of some highly potent anticancer drugs. With the use of the tools and databases described, along with the drop in the cost of whole genome sequencing in plant species, the future for the discovery of new plant-based anticancer metabolites would involve the identification of one or more genes or BGCs encoding the enzymes in the biosynthetic pathway for the target compound(s), followed by the co-expression analysis, also exploiting the knowledge of the chemical structure of the target compound, for the identification of other enzymes that might be involved in this pathway. As an example, the exploration of the pathway for podophyllotoxin biosynthesis by the use transcriptome mining in *Podophyllum hexandrum* led to the identification biosynthetic genes, 29 of which were combinatorially expressed in the tobacco plant (*Nicotiana benthamiana*), leading to the identification of six pathway enzymes, among which is oxoglutarate-dependent dioxygenase responsible for closing the core cyclohexane ring of the aryltetralin scaffold [115]. An alternative approach could be, if the metabolic pathway and nature of SMs are unknown, then the identified co-expressed genes encoding the enzymes for secondary metabolism could be subjected to untargeted metabolomics for the elucidation of unknown pathways and chemical structures. As an example, a single pathogen-induced P450 enzyme, CYP82C2, with a combination of untargeted metabolomics and co-expression analysis was used to uncover the complete biosynthetic pathway, which leads to the metabolite 4-hydroxyindole-3-carbonyl nitrile, previously unknown to *Arabidopsis* sp. This rare and hitherto unprecedented plant metabolite, with a cyanogenic functionality revealed a hidden capacity of *Arabidopsis* sp. for cyanogenic glucoside biosynthesis. This was confirmed by expressing 4-OH-ICN engineering biosynthetic enzymes in *Saccharomyces cerevisiae* and *Nicotiana benthamiana*, to reconstitute the complete pathway *in vitro* and *in vivo*, thus validating the functions of the enzymes involved in the pathway [116].

http://dx.doi.org/10.5772/67650

Acknowledegments

FNK acknowledges a Georg Forster fellowship from the Alexander von Humboldt Foundation, Germany. CVS is currently a doctoral candidate financed by the German Academic Exchange Services (DAAD), Germany.

Competing interests

The authors declare that they have no competing interests.

Abbreviations

AfroCancer	African Anticancer Natural Products Database
BGC	Biosynthetic gene clusters
EC_{50}	Half maximal effective concentration, that is, the concentration of a drug, antibody or toxicant, which induces a response halfway between the baseline and maximum after a specified exposure time
ED_{50}	The median effective dose, a dose that produces the desired effect in 50% of a population
FBA	Flux Balance Analysis
GI_{50}	The growth inhibition of 50%, drug concentration resulting in a 50% reduction in the net protein increase.
IC_{50}	The drug concentration causing 50% inhibition of the desired activity
IMG-ABC	The Integrated Microbial Genomes Atlas of Biosynthetic gene Clusters
NANPDB	Northern African Natural Products Database
NP	Natural product
NPACT	Naturally Occurring Plant-based Anti-cancer Compound Activity-Target Database
NRP	Nonribosomal peptide
NRPS	Nonribosomal peptide synthase
PK	Polyketide
PKS	Polyketides synthase
PMN	Plant Metabolic Network
PRISM	PRediction Informatics for Secondary Metabolomes
SM	Secondary metabolite

Author details

Aurélien F.A. Moumbock[1], Conrad V. Simoben[2], Ludger Wessjohann[3], Wolfgang Sippl[2], Stefan Günther[4,] and Fidele Ntie-Kang[1,2]*

*Address all correspondence to: ntiekfidele@gmail.com

1 Department of Chemistry, University of Buea, Buea, Cameroon

2 Department of Pharmaceutical Chemistry, Martin-Luther University of Halle-Wittenberg, Halle, Germany

3 Leibniz Institute of Plant Biochemistry, Halle, Germany

4 Pharmaceutical Bioinformatics, Albert-Ludwig-University Freiburg, Freiburg, Germany

References

[1] Cragg GM, Newman DJ. Plants as a source of anti-cancer and anti-HIV agents. Ann Appl Biol. 2003;143:127-133. doi:10.1111/j.1744-7348.2003.tb00278.x

[2] Cragg GM, Grothaus PG, Newman DJ. Impact of natural products on developing new anti-cancer agents. Chem Rev. 2009;109:3012-3043. doi:10.1021/cr900019j

[3] Lamari FN, Cordopatis P. Exploring the potential of natural products in cancer treatment. In: Missailidis S, editor. *Anticancer therapeutics*. West Sussex: Wiley-Blackwell; 2008, pp. 3-16.

[4] Pan L, Chai HB, Kinghorn AD. Discovery of new anticancer agents from higher plants. Front Biosci (Schol Ed). 2013;4:142-156.

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

[6] Mangal M, Sagar P, Singh H, Raghava GPS, Agarwal SM. NPACT: naturally occurring plant-based anti-cancer compound activity-target database. Nucleic Acids Res. 2013;41:D1124-D1129. doi:10.1093/nar/gks1047

[7] Ntie-Kang F, Nwodo JN, Ibezim A, Simoben CV, Karaman B, et al. Molecular modeling of potential anticancer agents from African medicinal plants. J Chem Inf Model. 2014;54:2433-2450. doi:10.1021/ci5003697

[8] Ntie-Kang F, Simoben CV, Karaman B, Ngwa VF, Judson PN, et al. Pharmacophore modeling and *in silico* toxicity assessment of potential anticancer agents from African medicinal plants. Drug Des Dev Ther. 2016;10:2137-2154. doi:10.2147/DDDT.S108118

[9] Beutler JA, Cragg GM, Iwu M, Newman DJ, Okunji C. Anticancer potential of African plants: the experience of the United States National Cancer Institute and National

Institutes of Health. In: Gurib-Fakim A, editor. Novel plant bioresources: applications in food, medicine and cosmetics, 1st ed. Oxford: John Wiley & Sons Ltd; 2014, pp. 133-149. doi:10.1002/9781118460566.ch10

[10] Nwodo JN, Ibezim A, Simoben CV, Ntie-Kang F. Exploring cancer therapeutics with natural products from African medicinal plants, part II: alkaloids, terpenoids and flavonoids. Anticancer Agents Med Chem. 2016;16:108-127. doi:10.2174/1871520615666150520143827

[11] Simoben CV, Ibezim A, Ntie-Kang F, Nwodo JN, Lifongo LL. Exploring cancer therapeutics with natural products from African medicinal plants, part I: xanthones, quinones, steroids, coumarins, phenolics and other classes of compounds. Anticancer Agents Med Chem. 2015;15:1092-1111. doi:10.2174/1871520615666150113110241

[12] Simoben CV, Ntie-Kang F. African medicinal plants: an untapped reservoir of potential anticancer agents. In: Prasad S, Tyagi AK, editors. Cancer preventive and therapeutic compounds: gift from mother nature. Beijing: Bentham Science Publishers; 2016. p. 78-95.

[13] Ntie-Kang F, Telukunta KK, Döring K, Simoben CV, Moumbock, et al. The Northern African Natural Products Database (NANPDB), 2016. www.african-compounds.org/nanpdb

[14] Ntie-Kang F, Yong JN. The chemistry and biological activities of natural products from Northern African plant families: from Aloaceae to Cupressaceae. RSC Adv. 2014;4:61975-61991. doi:10.1039/C4RA11467A

[15] Yong JN, Ntie-Kang F. The chemistry and biological activities of natural products from Northern African plant families: from Ebenaceae to Solanaceae. RSC Adv. 2015;5:26580-26595. doi:10.1039/C4RA15377D

[16] Ntie-Kang F, Njume LE, Malange YI, Günther S, Sippl W, et al. The chemistry and biological activities of natural products from Northern African plant families: from Taccaceae to Zygophyllaceae. Nat Prod Bioprospect. 2016;6:63-96. doi:10.1007/s13659-016-0091-9

[17] Medema MH, Fischbach M. Computational approaches to natural product discovery. Nat Chem Biol. 2015;11:639-648. doi:10.1038/nchembio.1884

[18] Medema MH, Kottmann R, Yilmaz P, Cummings M, Biggins JB, et al. Minimum information about a biosynthetic gene cluster. Nat Chem Biol. 2015;11:625-631. doi:10.1038/nchembio.1890

[19] Nützmann HW, Osbourn A. Gene clustering in plant specialized metabolism. Curr Opin Biotechnol. 2014;26:91-99. doi:10.1016/j.copbio.2013.10.009

[20] Osbourn A. Secondary metabolic gene clusters: evolutionary toolkits for chemical innovation. Trends Genet. 2010;26:449-457. doi:10.1016/j.tig.2010.07.001

[21] Osbourn AE, Field B. Operons. Cell Mol Life Sci. 2009;66:3755-3775. doi:10.1007/s00018-009-0114-3

[22] Rhee SY, Mutwil M. Towards revealing the functions of all genes in plants. Trends Plant Sci. 2014;19:212-221. doi:10.1016/j.tplants.2013.10.006

[23] Xu M, Rhee SY. Becoming data-savvy in a big-data world. Trends Plant Sci. 2014;19:619-622. doi:10.1016/j.tplants.2014.08.003

[24] Chae L, Lee I, Shin J, Rhee SY. Towards understanding how molecular networks evolve in plants. Curr Opin Plant Biol. 2012;15:177-184. doi:10.1016/j.pbi.2012.01.006

[25] Medema MH, Osbourn A. Computational genomic identification and functional reconstitution of plant natural product biosynthetic pathways. Nat Prod Rep. 2016;33:951-962. doi:10.1039/c6np00035e

[26] Weber T. *In silico* tools for the analysis of antibiotic biosynthetic pathways. Int J Med Microbiol. 2014;304:230-235. doi:10.1016/j.ijmm.2014.02.001

[27] Li YF, Tsai KJ, Harvey CJ, Li JJ, Ary BE, et al. Comprehensive curation and analysis of fungal biosynthetic gene clusters of published natural products. Fungal Genet Biol. 201689:18-28. doi:10.1016/j.fgb.2016.01.012

[28] van der Lee TA, Medema MH. Computational strategies for genome-based natural product discovery and engineering in fungi. Fungal Genet Biol. 2016;89:29-36. doi:10.1016/j.fgb.2016.01.006

[29] Frey M, Chomet P, Glawischnig E, Stettner C, Grun S, et al. Analysis of a chemical plant defense mechanism in grasses. Science. 1997;277:696-699. doi:10.1126/science.277.5326.696

[30] Osbourn A. Gene clusters for secondary metabolic pathways: an emerging theme in plant biology. Plant Physiol. 2010;154:531-535. doi:10.1104/pp.110.161315

[31] Figueiredo AC, Barroso JG, Pedro LG, Scheffer JJC. Factors affecting secondary metabolite production in plants: volatile components and essential oils. Flavour Fragr J. 2008;23:213-226. doi:10.1002/ffj.1875

[32] Leal MC, Hilario A, Munro MHG, Blunt JW, Calado R. Natural products discovery needs improved taxonomic and geographic information. Nat Prod Rep. 2016;33:747-750. doi:10.1039/c5np00130g

[33] Luo Y, Enghiad B, Zhao H. New tools for reconstruction and heterologous expression of natural product biosynthetic gene clusters. Nat Prod Rep. 2016;33:174-182. doi:10.1039/c5np00085h

[34] Carbonell P, Currin A, Jervis AJ, Rattray NJW, Swainston N, et al. Bioinformatics for the synthetic biology of natural products: integrating across the Design-Build-Test cycle. Nat Prod Rep. 2016;33:925-932. doi:10.1039/c6np00018e

[35] Song MC, Kim EJ, Kim E, Rathwell K, Nama SJ, et al. Microbial biosynthesis of medicinally important plant secondary metabolites. Nat Prod Rep. 2014;31:1497-1509. doi:10.1039/c4np00057a

[36] Zhao H, Medema MH. Standardization for natural product synthetic biology. Nat Prod Rep. 2016;33:920-924. doi:10.1039/c6np00030d

[37] Ajikumar PK, Xiao WH, Tyo KE, Wang Y, Simeon F, et al. Isoprenoid pathway optimization for Taxol precursor overproduction in *Escherichia coli*. Science. 2010;330:70-74. doi:10.1126/science.1191652

[38] De Luca V, Salim V, Atsumi SM, Yu F. Mining the biodiversity of plants: a revolution in the making. Science. 2012;336:1658-1661. doi:10.1126/science.1217410

[39] Kim E, Moore BS, Yoon YJ. Reinvigorating natural product combinatorial biosynthesis with synthetic biology. Nat Chem Biol. 2015;11:639-659. doi:10.1038/nchembio.1893

[40] Menzella HG, Reid R, Carney JR, Chandran SS, Reisinger SJ, et al. Combinatorial polyketide biosynthesis by *de novo* design and rearrangement of modular polyketide synthase genes. Nat Biotechnol. 2005;23:1171-1176. doi:10.1038/nbt1128

[41] Conway KR, Boddy CN. ClusterMine360: a database of microbial PKS/NRPS biosynthesis. Nucleic Acids Res. 2013;41:D402-D407. doi:10.1093/nar/gks993

[42] Tremblay N, Hill P, Conway KR, Boddy CN. The use of ClusterMine360 for the analysis of polyketide and nonribosomal peptide biosynthetic pathways. Methods Mol Biol. 2016;1401:233-252. doi:10.1007/978-1-4939-3375-4_15

[43] Starcevic A, Zucko J, Simunkovic J, Long PF, Cullum J, et al. ClustScan: an integrated program package for the semi-automatic annotation of modular biosynthetic gene clusters and *in silico* prediction of novel chemical structures. Nucleic Acids Res. 2008;36:6882-6892. doi:10.1093/nar/gkn685

[44] Diminic J, Zucko J, Ruzic IT, Gacesa R, Hranueli D, et al. Databases of the thiotemplate modular systems (CSDB) and their *in silico* recombinants (r-CSDB). J Ind Microbiol Biotechnol. 2013;40:653-659. doi:10.1007/s10295-013-1252-z

[45] Ichikawa N, Sasagawa M, Yamamoto M, Komaki H, Yoshida Y, et al. DoBISCUIT: a database of secondary metabolite biosynthetic gene clusters. Nucleic Acids Res. 2013;41:D408-D414. doi:10.1093/nar/gks1177

[46] Hadjithomas M, Chen IA, Chu K, Ratner A, Palaniappan K, et al. IMG-ABC: a knowledge base to fuel discovery of biosynthetic gene clusters and novel secondary metabolites. mBio. 2015;6:e00932-15. doi:10.1128/mBio.00932-15

[47] Kellner F, Kim J, Clavijo BJ, Hamilton JP, Childs KL, et al. Genome-guided investigation of plant natural product biosynthesis. Plant J. 2015;82:680-692. doi:10.1111/tpj.12827

[48] Hur M, Campbell AA, Almeida-de-Macedo M, Li L, Ransom N, et al. A global approach to analysis and interpretation of metabolic data for plant natural product discovery. Nat Prod Rep. 2013;30:565-583. doi:10.1039/c3np20111b

[49] Chae L, Kim T, Nilo-Poyanco R, Rhee SY. Genomic signatures of specialized metabolism in plants. Science. 2014;344: 510-513. doi:10.1126/science.1252076

[50] Dreher K. Putting the plant metabolic network pathway databases to work: going offline to gain new capabilities. Methods Mol Biol. 2014;1083:151-171. doi:10.1007/978-1-62703-661-0_10

[51] Naithani S, Preece J, D'Eustachio P, Gupta P, Amarasinghe V, et al. Plant Reactome: a resource for plant pathways and comparative analysis. Nucleic Acids Res. 2017;45:D1029-D1039. doi:10.1093/nar/gkw932

[52] Starcevic A, Wolf K, Diminic J, Zucko J, Ruzic IT, et al. Recombinatorial biosynthesis of polyketides. J Ind Microbiol Biotechnol. 2012;39:503-511. doi:10.1007/s10295-011-1049-x

[53] Weber T, Kim HU. The secondary metabolite bioinformatics portal: computational tools to facilitate synthetic biology of secondary metabolite production. Synth Syst Biotechnol. 2016;1:69-79. doi:10.1016/j.synbio.2015.12.002

[54] Medema MH, van Raaphorst R, Takano E, Breitling R. Computational tools for the synthetic design of biochemical pathways. Nat Rev Microbiol. 2012;10:191-202. doi:10.1038/nrmicro2717

[55] Weber T, Blin K, Duddela S, Krug D, Kim HU, et al. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Res. 2015;43:W237-W243. doi:10.1093/nar/gkv437

[56] Lee I, Ambaru B, Thakkar P, Marcotte EM, Rhee SY. Rational association of genes with traits using a genome-scale gene network for *Arabidopsis thaliana*. Nat Biotechnol. 2010;28:149-156. doi:10.1038/nbt.1603

[57] Caspi R, Billington R, Ferrer L, Foerster H, Fulcher CA, et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res. 2016;44:D471-D480. doi:10.1093/nar/gkv1164

[58] Le Fèvre F, Smidtas S, Combe C, Durot M, d'Alché-Buc F, et al. CycSim—an online tool for exploring and experimenting with genome-scale metabolic models. Bioinformatics. 2009;25:1987-1988. doi:10.1093/bioinformatics/btp268

[59] Yamanishi Y, Hattori M, Kotera M, Goto S, Kanehisa M. E-zyme: predicting potential EC numbers from the chemical transformation pattern of substrate-product pairs. Bioinformatics. 2009;25:i179-i186. doi:10.1093/bioinformatics/btp223

[60] Chou CH, Chang WC, Chiu CM, Huang CC, Huang HD. FMM: a web server for metabolic pathway reconstruction and comparative analysis. Nucleic Acids Res. 2009;37:W129-W134. doi:10.1093/nar/gkp264

[61] Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, et al. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28:1647-1649. doi:10.1093/bioinformatics/bts199

[62] Johnston CW, Skinnider MA, Wyatt MA, Li X, Ranieri MRM, et al. An automated Genomes-to-Natural Products platform (GNP) for the discovery of modular natural products. Nat Commun. 2015;6:8421. doi:10.1038/ncomms9421

[63] Kanehisa M. KEGG bioinformatics resource for plant genomics and metabolomics. Methods Mol Biol. 2016;1374:55-70. doi:10.1007/978-1-4939-3167-5_3

[64] Law PJ, Claudel-Renard C, Joubert F, Louw AI, Berger DK. MADIBA: a web server toolkit for biological interpretation of *Plasmodium* and plant gene clusters. BMC Genomics. 2008;9:105. doi:10.1186/1471-2164-9-105

[65] de Klein N, Magnani E, Banf M, Rhee SY. microProtein Prediction Program (miP3): a software for predicting microProteins and their target transcription factors. Int J Genomics. 2015;2015:734147. doi:10.1155/2015/734147

[66] Rocha I, Maia P, Evangelista P, Vilaça P, Soares S, et al. OptFlux: an open-source software platform for *in silico* metabolic engineering. BMC Syst Biol. 2010;4:45. doi:10.1186/1752-0509-4-45

[67] Moriya Y, Shigemizu D, Hattori M, Tokimatsu T, Kotera M, et al. PathPred: an enzyme-catalyzed metabolic pathway prediction server. Nucleic Acids Res. 2010;38:W138-W143. doi:10.1093/nar/gkq318

[68] Kutmon M, van Iersel MP, Bohler A, Kelder T, Nunes N, et al. PathVisio 3: an extendable pathway analysis toolbox. PLoS Comput Biol. 2015;11:e1004085. doi:10.1371/journal.pcbi.1004085

[69] van Iersel MP, Kelder T, Pico AR, Hanspers K, Coort S, et al. Presenting and exploring biological pathways with PathVisio. BMC Bioinformat. 2008;9:399. doi:10.1186/1471-2105-9-399

[70] Hanumappa M, Preece J, Elser J, Nemeth D, Bono G, et al. WikiPathways for plants: a community pathway curation portal and a case study in rice and *Arabidopsis* seed development networks. Rice. 2013;6:14. doi:10.1186/1939-8433-6-14

[71] Skinnider MA, Dejong CA, Rees PN, Johnston CW, Li H, et al. Genomes to natural products PRediction Informatics for Secondary Metabolomes (PRISM). Nucleic Acids Res. 2015;43:9645-9662. doi:10.1093/nar/gkv1012

[72] Carbonell P, Planson AG, Fichera D, Faulon JL. A retrosynthetic biology approach to metabolic pathway design for therapeutic production. BMC Syst Biol. 2011;5:122. doi:10.1186/1752-0509-5-122

[73] Planson AG, Carbonell P, Grigoras I, Faulon JL. A retrosynthetic biology approach to therapeutics: from conception to delivery. Curr Opin Biotechnol. 2012;23:948-956. doi:10.1016/j.copbio.2012.03.009

[74] Hoff KJ, Stanke M. WebAUGUSTUS - a web service for training AUGUSTUS and predicting genes in eukaryotes. Nucleic Acids Res. 2013;41:W123-W128. doi:10.1093/nar/gkt418

[75] Kelder T, van Iersel MP, Hanspers K, Kutmon M, Conklin BR, et al. WikiPathways: building research communities on biological pathways. Nucleic Acids Res. 2012,40:D1301-D1307. doi:10.1093/nar/gkr1074

[76] Kutmon M, Riutta A, Nunes N, Hanspers K, Willighagen EL, et al. WikiPathways: capturing the full diversity of pathway knowledge. Nucleic Acids Res. 2016;44:D488-D494. doi:10.1093/nar/gkv1024

[77] Durot M, Bourguignon PY, Schachter V. Genome-scale models of bacterial metabolism: reconstruction and applications. FEMS Microbiol Rev. 2009;33:164-190. doi:10.1111/j.1574-6976.2008.00146.x

[78] Feist AM, Herrgård MJ, Thiele I, Reed JL, Palsson BØ. Reconstruction of biochemical networks in microorganisms. Nat Rev Microbiol. 2009;7:129-143. doi:10.1038/nrmicro1949

[79] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13:2498-2504. doi:10.1101/gr.1239303

[80] Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, et al. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res. 2014;42:D206-D214. doi:10.1093/nar/gkt1226

[81] Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? Nat Biotechnol. 2010;28:245-248. doi:10.1038/nbt.1614

[82] Schellenberger J, Que R, Fleming RM, Thiele I, Orth JD, et al. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nat Protoc. 2011;6:1290-1307. doi:10.1038/nprot.2011.308

[83] Garcia-Albornoz M, Thankaswamy-Kosalai S, Nilsson A, Väremo L, Nookaew I, Nielsen J. BioMet Toolbox 2.0: genome-wide analysis of metabolism and omics data. Nucleic Acids Res. 2014;42:W175-W181. doi:10.1093/nar/gku371

[84] Boele J, Olivier BG, Teusink B. FAME, the flux analysis and modeling environment. BMC Syst Biol. 2012;6:8. doi:10.1186/1752-0509-6-8

[85] Schilling CH, Covert MW, Famili I, Church GM, Edwards JS, Palsson BO. Genome-scale metabolic model of *Helicobacter pylori* 26695. J Bacteriol. 2002;184:4582-4593. doi:10.1128/JB.184.16.4582-4593.2002

[86] Oh YK, Palsson BO, Park SM, Schilling CH, Mahadevan R. Genome-scale reconstruction of metabolic network in *Bacillus subtilis* based on high-throughput phenotyping and gene essentiality data. J Biol Chem. 2007;282:28791-28799. doi:10.1074/jbc.M703759200

[87] Puchałka J, Oberhardt MA, Godinho M, Bielecka A, Regenhardt D, et al. Genome-scale reconstruction and analysis of the *Pseudomonas putida* KT2440 metabolic network facilitates applications in biotechnology. PLoS Comput Biol. 2008;4:e1000210. doi:10.1371/journal.pcbi.1000210

[88] Henry CS, Broadbelt LJ, Hatzimanikatis V. Thermodynamics-based metabolic flux analysis. Biophys J. 2007;92:1792-1805. doi:10.1529/biophysj.106.093138

[89] Asadollahi MA, Maury J, Patil KR, Schalk M, Clark A, Nielsen J. Enhancing sesquiterpene production in *Saccharomyces cerevisiae* through *in silico* driven metabolic engineering. Metab Eng. 2009;11:328-334. doi:10.1016/j.ymben.2009.07.001

[90] Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, et al. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 2016;44:D279-D285. doi:10.1093/nar/gkv1344

[91] Blin K, Medema MH, Kottmann R, Lee SY, Weber T. The antiSMASH database, a comprehensive database of microbial secondary metabolite biosynthetic gene clusters. Nucleic Acids Res. 2017;45:D555-D559. doi:10.1093/nar/gkw960

[92] Röttig M, Rausch C, Kohlbacher O. Combining structure and sequence information allows automated prediction of substrate specificities within enzyme families. PLoS Comput Biol. 2010;6:e1000636. doi:10.1371/journal.pcbi.1000636

[93] Röttig M, Medema MH, Blin K, Weber T, Rausch C, et al. NRPSpredictor2: a web server for predicting NRPS adenylation domain specificity. Nucleic Acids Res. 2011;39:W362-W367. doi:10.1093/nar/gkr323

[94] Liu Y, Young K, Rakotondraibe LH, Brodie PJ, Wiley JD, et al. Antiproliferative compounds from *Cleistanthus boivinianus* from the Madagascar dry forest. J Nat Prod. 2015;78:1543-1547. doi:10.1021/np501020m

[95] Liu Y, Cheng E, Rakotondraibe LH, Brodie PJ, Applequist W, et al. Antiproliferative compounds from *Ocotea macrocarpa* from the Madagascar dry forest. Tetrahedron Lett. 2015;56:3630-3632. doi:10.1016/j.tetlet.2015.01.172

[96] Nyandoro SS, Munissi JJE, Gruhonjic A, Duffy S, Pan F, et al. Polyoxygenated cyclohexenes and other constituents of *Cleistochlamys kirkii* leaves. J Nat Prod. 2016. doi:10.1021/acs.jnatprod.6b00759. PMID: 28001067

[97] Prakash P, Gupta N. Therapeutic uses of *Ocimum sanctum* Linn (Tulsi) with a note on eugenol and its pharmacological actions: a short review. Indian J Physiol Pharmacol. 2005;49:125-131. PMID: 16170979

[98] Willis JC. A dictionary of the flowering plants and ferns. Cambridge: The University Press; 1919

[99] Khare CP. Indian medicinal plants: an illustrated dictionary. Heidelberg: Springer; 2007, p. 443

[100] Upadhyay AK, Chacko AR, Gandhimathi A, Ghosh P, Harini K, et al. Genome sequencing of herb Tulsi (*Ocimum tenuiflorum*) unravels key genes behind its strong medicinal properties. BMC Plant Biol. 2015;15:212. doi:10.1186/s12870-015-0562-x

[101] Ke Y, Ye K, Grossniklaus HE, Archer DR, Joshi HC, et al. Noscapine inhibits tumor growth with little toxicity to normal tissues or inhibition of immune responses. Cancer Immunol Immunother. 2000;49:217-225. PMID: 10941904

[102] Ye K, Ke Y, Keshava N, Shanks J, Kapp JA, et al. Opium alkaloid noscapine is an anti-tumor agent that arrests metaphase and induces apoptosis in dividing cells. Proc Natl Acad Sci USA. 1998;95:1601-1606. PMID: 9465062

[103] Zhou J, Gupta K, Yao J, Ye K, Panda D, et al. Paclitaxel-resistant human ovarian cancer cells undergo c-Jun NH_2-terminal kinase-mediated apoptosis in response to noscapine. J Biol Chem. 2002;277:39777-39785. doi:10.1074/jbc.M203927200

[104] DellaPenna D, O'Connor SE. Plant gene clusters and opiates. Science. 2012;336:1648-1649. doi:10.1126/science.1225473

[105] Battersby AR, Hirst M, McCaldin DJ, Southgate R, Staunton J. Alkaloid biosynthesis. XII. The biosynthesis of narcotine. J Chem Soc Perkin 1. 1968;17:2163-2172. PMID: 5691486

[106] Winzer T, Gazda V, He Z, Kaminski F, Kern M, et al. A *Papaver somniferum* 10-gene cluster for synthesis of the anticancer alkaloid noscapine. Science. 336:1704-1708. doi:10.1126/science.1220757

[107] Rischer H, Oresic M, Seppänen-Laakso T, Katajamaa M, Lammertyn F, et al. Gene-to-metabolite networks for terpenoid indole alkaloid biosynthesis in *Catharanthus roseus* cells. Proc Natl Acad Sci USA. 2006;103:5614-5619. doi:10.1073/pnas.0601027103

[108] Noble RL. The discovery of the vinca alkaloids-chemotherapeutic agents against cancer. Biochem Cell Biol. 1990;68:1344-1351. doi:10.1139/o90-197

[109] Dong HP, Wu HM, Chen SJ, Chen CY. The effect of butanolides from *Cinnamomum tenuifolium* on platelet aggregation. Molecules. 2013;18:11836-11841. doi:10.3390/molecules181011836

[110] Hoshino S, Wakimoto T, Onaka H, Abe I. Chojalactones A-C, cytotoxic butanolides isolated from *Streptomyces* sp. cultivated with mycolic acid containing bacterium. Org Lett. 2015;17:1501-1504. doi:10.1021/acs.orglett.5b00385

[111] Kim W, Lyu HN, Kwon HS, Kim YS, Lee KH, et al. Obtusilactone B from *Machilus thunbergii* targets barrier-to-autointegration factor to treat cancer. Mol Pharmacol. 2013;83:367-376. doi:10.1124/mol.112.082578

[112] Shen KH, Lin ES, Kuo PL, Chen CY, Hsu YL. Isolinderanolide B, a butanolide extracted from the stems of *Cinnamomum subavenium*, inhibits proliferation of T24 human bladder cancer cells by blocking cell cycle progression and inducing apoptosis. Integr Cancer Ther. 2011;10:350-358. doi:10.1177/1534735410391662

[113] Yang SY, Wang HM, Wu TW, Chen YJ, Shieh JJ, et al. Subamolide B isolated from medicinal plant *Cinnamomum subavenium* induces cytotoxicity in human cutaneous squamous cell carcinoma cells through mitochondrial and CHOP-dependent cell death pathways. Evid Based Complement Alternat Med. 2013,2013:630415. doi:10.1155/2013/630415

[114] Risinger AL, Mooberry SL. Taccalonolides: novel microtubule stabilizers with clinical potential. Cancer Lett. 2010;291:14-19. doi:10.1016/j.canlet.2009.09.020

[115] Lau W, Sattely ES. Six enzymes from mayapple that complete the biosynthetic pathway to the etoposide aglycone. Science. 2015;349:1224-1228. doi:10.1126/science.aac7202

[116] Rajniak J, Barco B, Clay NK, Sattely ES. A new cyanogenic metabolite in *Arabidopsis* required for inducible pathogen defence. Nature. 2015;525:376-379. doi:10.1038/nature14907

IntechOpen

IntechOpen

IntechOpen

IntechOpen