NATURAL PRODUCTS AND DRUG DISCOVERY

AN INTEGRATED APPROACH

Edited by

Subhash C. Mandal

Vivekananda Mandal

Tetsuya Konishi

Natural Products and Drug Discovery

This page intentionally left blank

Natural Products and Drug Discovery

An Integrated Approach

Edited By

Subhash C. Mandal
Professor
Division of Pharmacognosy
Department of Pharmaceutical Technology
Jadavpur University
Kolkata, India

Vivekananda Mandal
Assistant Professor
Division of Pharmacognosy
Institute of Pharmaceutical Sciences
Guru Ghasidas University (A Central University)
Bilaspur, India

Tetsuya Konishi
Professor Emeritus
Niigata University of Pharmacy & Applied Life Sciences (NUPALS)
Tojima, Akiha-ku, Niigata, Japan
&
Director, Office HALD Food Function Research
Sakai, Nishi-ku, Niigata, Japan

Elsevier
Radarweg 29, PO Box 211, 1000 AE Amsterdam, Netherlands
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, United Kingdom
50 Hampshire Street, 5th Floor, Cambridge, MA 02139, United States

Notices
Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-0-08-102081-4

Publisher: Mica Haley
Acquisitions Editor: Anneka Hess
Editorial Project Manager: Michelle W. Fisher
Production Project Manager: Poulouse Joseph
Designer: Matthew Limbert

Typeset by TNQ Books and Journals

Contents

Section II
Leads From Natural Products

9. Phytodrugs and Immunomodulators for the Therapy of Leishmaniasis

C. Benjamin Naman, Ciro M. Gomes, and Gaurav Gupta

Section III
Herbal Drug Research

17. Herb and Drug Interaction

Nilanjan Ghosh, Rituparna C. Ghosh, Anindita Kundu, and Subhash C. Mandal

18. Toxicity Studies Related to Medicinal Plants

Kavimani Subramanian, Divya Sankaramourthy, and Mahalakshmi Gunasekaran

This page intentionally left blank

List of Contributors

Sugato Banerjee, Department of Pharmaceutical Sciences and Technology, Birla Institute of Technology, Ranchi, India

Souvik Basak, Dr. B.C. Roy College of Pharmacy and Allied Health Sciences, Durgapur, India

Tuhin K. Biswas, Department of Kayachikitsa (Medicine), J.B. Roy State Ayurvedic Medical College and Hospital, Kolkata, India

Anand K. Chaudhari, Centre of Advanced Study in Botany, Institute of Science, Banaras Hindu University, Varanasi, India

Jihang Chen, Hong Kong University of Science & Technology, Hong Kong SAR, China

Kavi B.S. Chouhan, Institute of Pharmacy, Guru Ghasidas Central University, Bilaspur, India

Somenath Das, Centre of Advanced Study in Botany, Institute of Science, Banaras Hindu University, Varanasi, India

Nawal K. Dubey, Centre of Advanced Study in Botany, Institute of Science, Banaras Hindu University, Varanasi, India

Anirban Dutta, ICAR-Indian Agricultural Research Institute, New Delhi, India

Abhishek K. Dwivedy, Centre of Advanced Study in Botany, Institute of Science, Banaras Hindu University, Varanasi, India

Mostafa Elachouri, Mohammed first University, Oujda, Morocco

Geetanjali, Department of Chemistry, Kirori Mal College, University of Delhi, Delhi, India

Nilanjan Ghosh, Dr. B.C. Roy College of Pharmacy and Allied Health Sciences, Durgapur, India

Rituparna C. Ghosh, Dr. B.C. Roy College of Pharmacy and Allied Health Sciences, Durgapur, India

Ciro M. Gomes, Hospital Universitário de Brasília, Brasília, Brazil

Mahalakshmi Gunasekaran, Department of Pharmacology, Mother Theresa Post Graduate and Research Institute of Health Sciences, Puducherry, India

Gaurav Gupta, Department of Immunology, University of Manitoba, Winnipeg, MB, Canada

Dharmik Joshi, Department of Pharmaceutical Sciences and Technology, Birla Institute of Technology, Ranchi, India

Hirokazu Kawagishi, Shizuoka University, Shizuoka, Japan

Kam M. Ko, Hong Kong University of Science & Technology, Hong Kong SAR, China

Sudhir Kumar, Medicinal and Process Chemistry Division, Central Drug Research Institute, Lucknow, India

Anindita Kundu, Department of Pharmaceutical Technology, Jadavpur University, Kolkata, India

Sukalyan K. Kundu, Department of Pharmacy, Jahangirnagar University, Dhaka, Bangladesh

Namrita Lall, Department of Plant and Soil Sciences, University of Pretoria, Pretoria, South Africa

Pou K. Leong, Hong Kong University of Science & Technology, Hong Kong SAR, China

Rachel W. Li, ANU Medical School, The Australian National University, Canberra, ACT, Australia

G. David Lin, Research School of Chemistry, The Australian National University, Canberra, ACT, Australia

Alex Loukas, Centre for Biodiscovery and Molecular Development of Therapeutics, Australian Institute of Tropical Health and Medicine, James Cook University, QLD, Australia

Anuradha S. Majumdar, Department of Pharmacology, Bombay College of Pharmacy, Mumbai, India

Abhishek Mandal, ICAR-Indian Agricultural Research Institute, New Delhi, India

Vivekananda Mandal, Institute of Pharmacy, Guru Ghasidas Central University, Bilaspur, India

Subhash C. Mandal, Department of Pharmaceutical Technology, Jadavpur University, Kolkata, India

Rakesh Maurya, Medicinal and Process Chemistry Division, Central Drug Research Institute, Lucknow, India

Dhrubojyoti Mukherjee, Department of Pharmaceutical Technology, Jadavpur University, Kolkata, India

C. Benjamin Naman, Center for Marine Biotechnology and Biomedicine, University of California, San Diego, California, United States

Ajay G. Namdeo, Poona College of Pharmacy, Bharati Vidyapeeth Deemed University, Pune, India

Sunday O. Otimenyin, Department of Pharmacology, University of Jos, Plateau State, Nigeria

Partha Palit, Department of Pharmaceutical Sciences, Assam University, Silchar, India

Somdatta Roy, Department of Pharmaceutical Sciences and Technology, Birla Institute of Technology, Ranchi, India

Shubhadeep Roychoudhury, Department of Life Science & Bio-Informatics, Assam University, Silchar, India

Supradip Saha, ICAR-Indian Agricultural Research Institute, New Delhi, India

Divya Sankaramourthy, Department of Pharmacology, Mother Theresa Post Graduate and Research Institute of Health Sciences, Puducherry, India

Ram Singh, Department of Applied Chemistry, Delhi Technological University, Delhi, India

Sahil J. Somani, Department of Pharmacology, RK University, Rajkot, India

Kavimani Subramanian, Department of Pharmacology, Mother Theresa Post Graduate and Research Institute of Health Sciences, Puducherry, India

Roshni Tandey, Institute of Pharmacy, Guru Ghasidas Central University, Bilaspur, India

Danielle Twilley, Department of Plant and Soil Sciences, University of Pretoria, Pretoria, South Africa

Phurpa Wangchuk, Centre for Biodiscovery and Molecular Development of Therapeutics, Australian Institute of Tropical Health and Medicine, James Cook University, QLD, Australia

Sujata Wangkheirakpam, Department of Chemistry, NIT Manipur, Imphal, India

This page intentionally left blank

Foreword

I feel genuinely honored in writing this foreword for the book *Natural Products and Drug Discovery: An Integrated Approach*, edited by three excellent scholars from the area of natural products research: Subhash C. Mandal, Vivekananda Mandal, and Tetusya Konishi. Natural products have evidently been one of the major sources of new drugs, and will continue to be so in the years to come. This very reason has prompted a huge body of research exploring natural products for new drugs to combat various ailments. There are several books and excellent review articles available to date covering various areas relating to natural products research, particularly the area of natural products drug discovery, but this book will stand out from the crowd probably because of its inclusive approach to integrating several aspects of natural products drug discovery processes in one book.

This book offers 23 chapters organized in three distinct sections: traditional medicine and drug discovery (six chapters), leads from natural products (nine chapters), and herbal drug research (eight chapters). All these chapters are written by experts from relevant areas of natural products drug discovery.

Natural Products and Drug Discovery: An Integrated Approach integrates several classical and modern aspects of drug discovery, from Chinese traditional medicine to Ayurvedic medicine, as well as modern aspects of drug discovery strategies, e.g., natural products lead discovery, and will act as an outstanding reference book for natural products researchers.

I wholeheartedly recommend this book to all who are interested in natural products drug discovery and related areas.

Professor Satyajit D. Sarker
Editor-in-Chief, Phytochemical Analysis
Director, School of Pharmacy and Biomolecular Sciences
Liverpool John Moores University
Liverpool L3 3AF
United Kingdom

This page intentionally left blank

Preface

Natural product research has become the leading force in the drug discovery sector. This fact has been further triggered due to the enormous risk and time involved in the synthetic route of drug discovery. Natural product research, though more complicated due to the complex mixtures involved, still offers a more successful rate when compared to synthetic drug discovery. From ancient histories it becomes evident that traditional medicine (Ayurveda, the Indian traditional system, traditional Chinese medicine, traditional Japanese medicine, etc.) has always been there to reduce the sufferings of human ailments, even before the advent of antibiotics. Today's drug discovery is no longer just a case of trial and error or mere serendipity but rather has become a more programmed and strategized venture. Drug discovery these days has become an integrated approach of modern biology and traditional medicine using a holistic approach. The modern tools of chemistry and biology—in particular, the various "-omics" technologies—now allow scientists to detail the exact nature of the biological effects of natural compounds on the human body, as well as to uncover possible synergies, which hold much promise for the development of new therapies against many devastating diseases. Henceforth, we cannot deny the shift of the scientific community more toward traditional medicines involving complementary and alternative therapies. Well-strategized ethnobotanically inspired natural product research can provide vital leads with the potential for developing them as future drug candidates. Henceforth, this is the perfect time to bring out a book that can act as a fuel to this driving force of drug discovery. This book serves as a "one-stop solution" for all beginners in the field of botanical research leading to drug discovery and is committed to fulfilling the needs of herbal drug researchers. The book is an amalgamation of 23 scientifically crafted chapters prioritized judiciously into three major groups. Through the various chapters, the book acts as a vital support system for natural product researchers where all issues pertaining to drug discovery from botanicals are dealt with under a single umbrella system. The book aims to dig deep into our cultural roots and extract the ancient science of different traditional systems of medicine practiced worldwide to try to integrate ancient knowledge with modern approaches for empowering the drug discovery process. Application of ethnopharmacology in developing preventive and clinical medicine is emphasized upon. On the other hand, the book also amalgamates different strategies and ideologies under one roof, presents a simplified approach of bioassay-guided fractionation and

isolation, and showcases important traditional leads that can be explored for future drug discovery. Recent developments in the science of enzyme substrate reactions are highlighted and the role of in vitro techniques is exemplified in the process of drug discovery.

We humbly express our gratitude to our national and international funding agencies and home universities who have supported us in our journey of natural product research. We are also thankful to our peer review team for timely reviewing the manuscripts and providing valuable inputs. Finally, we express our deep gratitude to our family members for their constant support, particularly during the busy days of compiling this book.

Chapter 21

Digitization of Traditional Knowledge

Souvik Basak
Dr. B.C. Roy College of Pharmacy and Allied Health Sciences, Durgapur, India

1. INTRODUCTION

In the last few decades research involving natural products has been a prime focus across the globe since natural products have been found to be toxicologically safer to the human community than synthetic products. The prior utilization of natural products over synthetic or semisynthetic products is the basis for lower toxicity of naturally obtained molecules yet befits the particular receptor for drug action. In addition, natural products are structurally more diverse than synthetic or semisynthetic compounds, hence they are more promising for new drug discovery. Thus increasing attention has been focused on this arena to utilize a broader set of natural products for drug discovery.

2. WHY DIGITIZATION OF NATURAL PRODUCTS IS NECESSARY

Over a time period of 4 billion years evolution has taken place to create complex biodiversity across the globe [1] and to indulge in fundamental life forms over the earth. Of 7−20 million species around the world, only 1−2 million have been identified and scientifically named [2,3]. In addition, tropical countries are the storehouses of biodiversity because of their temperate climate as well as humidity [4]. Concomitant with this biodiversity evolution, plants and animals have also evolved and were used in the past as traditional medicinal plants across the globe [5]. However, with the increase in natural product usage, an emerging need has grown over the years to organize or compile knowledge for the effective utilization of the data. There are almost 300,000 medicinal plants worldwide [6] of which 75,000 plants have been used as medicinal plants [7]. Thus attempts have been made to compile this knowledge in an organized form, which has been initiated through digital

Natural Products and Drug Discovery. https://doi.org/10.1016/B978-0-08-102081-4.00021-6

embodiment of the relevant information. Digitization has occurred predominantly via four approaches:

1. Digital databases on traditional knowledge (web based).
2. Bioinformatics-guided approach of traditional knowledge.
3. Virtual screening of natural products.
4. In-silico approach for natural product-guided drug discovery

2.1 Digital Databases on Traditional Knowledge (Web Based)

To compile, store, and organize medicinal plant knowledge, attempts have been made in several countries to code traditional medicinal plant knowledge in several databases. These databases not only help researchers to find specific information about medicinal plants, it also helps them to locate, identify, and collate data about certain plants in specific places followed by utilization in their system. The most commonly used databases are summarized in Tables 21.1 and 21.2.

2.2 Bioinformatics-Guided Approach for Traditional Knowledge

Although a plethora of digital databases is available for searching medical plants at a particular ecological niche or geographical region, the problem is identifying them based on their evolutionary pattern as well as their genetic makeup, which often determine the plants' properties in an applied field. To help with this, bioinformatics-based approaches have evolved that, based on specific programming-guided coding/decoding/scoring systems, aid in identification and property evaluation of certain plants in a specific database. These approaches often rely on the mutual interaction of two datasets on a common bioinformatics platform either for identification or for revelation of a common sequence space inside a genetic subset for predictive exploration of its functional properties. The general notion in this approach is that plants or animals have evolved due to changes in temperature, humidity, oxygen, or any other geographical factor/s, thus there is a genetic similarity between all these species in spite of biodiversity all over the world. Thus bioinformatics-based approaches can be classified as follows:

1. Identification of traditional components.
2. Digitization of traditional datasets.
3. Creation of biodiversity databases.

2.2.1 Identification of Traditional Components

Identification of traditional components is based on several identification tools that are basically web based. In this format, taxonomic descriptions are coded

TABLE 21.1 Database of Detailed Information on Natural Products

Database	Subdatabase	Summary	Website
African traditional knowledge database	Francophone database or PHARMEL (Pharmacopée et Medecine Traditionnelle)	A database containing 19,691 recipes from 24 African countries, 4000 medicinal records, and 51 references	http://www.ulb.ac.be/sciences/bota/pharmel.htm
	Bilingual database (PRELUDE database)	This database comprises information on traditional veterinary and human medicine plants in sub-Saharan Africa	http://pc4.sisc.ucl.ac.be/prelude.html
	PROTA (Plant Resources of Tropical Africa)	A database containing 7000 useful plants, 200,000 references, 30,000 photographs, and 6000 geographic distribution maps	http://www.prota.org/PROTAstartframes.htm
	The English database TRAMED (Traditional Medicine Database)		http://www.healthnet.org.za/tramed/gen/tramedsearch
	NTRAP	The Database of Natural and Traditional Pesticidal Materials and Pest Control in Sub-Saharan East Africa	http://www.ippc.orst.edu/ipmafrica/db/index.html
	IMRA (Institute Malgache de Recherches Appliquées)	IMRA has a database containing computerized ethnobotanical data of over 4000 Madagascan plants	

Continued

TABLE 21.1 Database of Detailed Information on Natural Products—cont'd

Database	Subdatabase	Summary	Website
	NAPRECA (NAtural Products Research Network for Eastern and Central Africa)	NAPRECA (www.naprecanetwork.net) is a network of natural products research scientists in East and Central Africa. The database tries to discover relevant chemicals used to remediate health problems and other beneficial problems all over the world	http://www.napreca.ne
	WANPRES (West African Network of Natural Products REsearch Scientists)	A sub-Saharan database for the Western Africa Network of Natural Products Research Scientists. It helps coordinate chemists and natural products all over the world	www.wannpres.org
CRISP		Computer-based retrieval of Information on Scientific Projects database, which has been government funded and has been carried out by universities, hospitals, and other research institutions	http://crisp.cit.nih.gov
Indian Medicine		A database for Indian Ayurvedic, Yoga, Siddha, Homeopathy, and Unani systems of medicine	http://indianmedicine.nic.in
NAPRALERT	NAtural PRoducts ALERT from STN International	A relational database of all products including ethnomedical information and other experimental studies including pharmacological in vitro/in vivo bioassays. Currently more than 200,000 scientific reports are in the database	http://info.cas.org/online/DBSS/napralertss.html

UK Crop Net Database			http://ukcrop.net/db.html
Database on medicinal plants, a database formed by the Government of India			http://www.nmpb-mpdb.nic.in/
Medicinal and Aromatic Plants Abstracts (MAPA)		Published by the National Institute of Science Communication, CSIR, New Delhi. 55,000 abstracts published in the first 18 volumes of the journal, which are solely made up of databases on medicinal plants	http://www.fao.org/docrep/w7261e/W7261e09.htm
GLOBinMED		A database on Malaysian traditional and complementary medicine	http://www.globinmed.com/index.php?option=com_content&view=article&id=80849&Itemid=101
Database on Indian medicinal plants			http://www.medicinalplants.in/
An Herbal/Medical Dictionary		A Glossary of terms used in Herbalism, Medicine and Physiology, Descriptions, Explanations, and Implications in Wholistic and Vitalist Therapy	http://www.swsbm.com/ManualsMM/MedHerbGloss2.pdf
Annie's Remedy		General guide to herbal medicine including common and botanical names; some records have citation information (check out the Herb Chart link—chart contains links to records)	http://www.anniesremedy.com/

Continued

TABLE 21.1 Database of Detailed Information on Natural Products—cont'd

Database	Subdatabase	Summary	Website
Arctic Science Portal		A library of links to websites where Arctic data are made publicly available. These websites contain information about indigenous plants and their uses	https://www.uspto.gov/patent/laws-and-regulations/comments-public/traditional-knowledge-and-medicine-dictionariesdatabases
Cacti Guide		Common and Latin names plus photographs	http://cactiguide.com/
Dasherb		Latin, English, PinYin, and Chinese names of medicinal herbs, descriptions, and uses	http://www.dasherb.com/database/index.html
CHEMnetBASE		170,000 natural products with names and synonyms, formulae, chemical structures, CAS Registry Numbers, extensive source data, uses and applications, physical state, melting point, boiling point, pKa, and key literature citations	http://dnp.chemnetbase.com/
Dr. Duke's Phytochemical and Ethnobotanical Databases		All references are to Duke, James A. (1992). The linked records contain plant names, alternative names for plants, and information about chemicals found in plants	https://phytochem.nal.usda.gov/phytochem/search

Eastern Chinese Medicine Export Company		Latin, English, and PinYin names, searchable by medical use or symptoms	http://tcmtreatment.com/images/herb-supply/herb-price/three-lists.htm
Find Me A Cure		Botanical names with common names in English, Chinese, and multiple Indian languages. Includes information about uses of herbs. Searchable via text input box, Herbs Glossary (botanical names), and list of Ailments and Remedies.	http://tcmtreatment.com/images/herb-supply/herb-price/three-lists.htm
Find Wild Flowers		Identification of British flora	http://www.botanicalkeys.co.uk/flora/
Fungal Database		United States Department of Agriculture. Scientific and common names, synonyms, specimens, and literature	https://nt.ars-grin.gov/fungaldatabases/
Germplasm Resources Information Network (GRIN)		United States Department of Agriculture (USDA). Scientific and common names and synonyms	https://nt.ars-grin.gov/fungaldatabases/
Gernot Katzer's Spice Pages		10,500 plant names in more than 60 different languages, with origins, constituents, etymology, images, and uses	http://gernot-katzers-spice-pages.com/engl/index.html?redirect=1
Hawaiian Ethnobotany Online Database		Database of Hawaiian names, species names, and vernacular names of plants. The linked records contain information about the uses of the plants with references.	http://data.bishopmuseum.org/ethnobotanydb/ethnobotany.php?b=list&o=1

Continued

TABLE 21.1 Database of Detailed Information on Natural Products—cont'd

Database	Subdatabase	Summary	Website
Herbal Medicine Materia Medica		Descriptions, constituents, and cited references	http://www.ichineseherbs.com/cross_ref_of_names.html
iChinese Herbs		Latin, common, and PinYin names	http://www.ichineseherbs.com/cross_ref_of_names.html
Interagency Taxonomic Information System (ITIS)		Diverse US government agencies partnership. Standardized nomenclature, taxonomic data, and hierarchical classification. Hyperlinks to diverse off-site resources	https://www.itis.gov/
Korean Traditional Knowledge Portal (KTKP)		Korean Intellectual Property Office's database service for searching traditional knowledge from old and very recent Korean and Chinese medicines, including journal articles and patents. Requires registration/subscription.	http://www.koreantk.com/ktkp2014/
Liber Herbarum		Cross-referenced herbal medicine database based fully on printed sources, inspired by the first known Danish medicine book *Liber Harbarum* written by Henrik Harpestgreng in the 13th century.	http://www.liberherbarum.net/

Malta Wild Plants		Detailed records for wild flowering plants of Malta. By Stephen Mifsud	http://www.maltawildplants.com/
Mushroom Nutrition		Scientific and common names, descriptions, and medicinal uses for mushrooms with citations to literature provided	https://www.mushroomnutrition.com/
NAPRALERT		Database of natural products, including ethnomedical information and pharmacological/biochemical information on extracts of organisms in vitro, in situ, in vivo, in human (case reports, nonclinical trials), and clinical studies. Fee required	https://www.napralert.org/
Native Plants Hawaii		A single, comprehensive, and searchable online knowledgebase of endemic and indigenous plants of Hawaii. Some records name additional reference works	http://nativeplants.hawaii.edu/
Natural Medicines Comprehensive Database		Evidence-based clinical relevance. Searchable by product name or medical condition. Links to PubMed	http://naturaldatabase.therapeuticresearch.com/home.aspx?AspxAutoDetectCookieSupport=1
Northern Ontario Plant Database		Records for some 55,000 herbarium specimens from northern Ontario educational and government institutions	http://naturaldatabase.therapeuticresearch.com/home.aspx?AspxAutoDetectCookieSupport=1

Continued

TABLE 21.1 Database of Detailed Information on Natural Products—cont'd

Database	Subdatabase	Summary	Website
Oro Verde Green Gold from Amazonia		Guide to Amazonian and Andean medicinal plants including common and botanical names; records contain citation information (click on any of the "more info" links to see records)	http://www.oroverde.cz/
Plants For A Future		Latin and common names, uses, constituents, and cited references. Registration required	http://www.pfaf.org/user/plantsearch.aspx
Plantsciencenetbase		CRC collection. Covers individual plants, from historical to modern topics, as well as transgenics and evolutionary biology	http://www.crcnetbase.com/page/plant_science_ebooks
Society for Research and Initiatives for Sustainable Technologies and Institutions		Database of medicinal plants containing uses, botanical names, common names, and Sanskrit names	http://www.sristi.org/hbnew/plant_db.php
Southwest School of Botanical Medicine		Known chemical constituents for over 250 medicinal plants	http://www.swsbm.com/Constituents/Constituents.html
The Encyclopedia of New Zealand: medicinal use of plants		Plants used in traditional methods for healing	http://www.teara.govt.nz/en/rongoa-medicinal-use-of-plants/page-1

The International Plant Names Index		A database of the names and associated basic bibliographical details of seed plants, ferns, and lycophytes	http://www.ipni.org/ipni/plantnamesearchpage.do
The Plant List		From the Royal Botanic Gardens, Kew, and Missouri Botanical Garden. Latin names, synonyms, and unresolved names for vascular plants, mosses, and liverworts. Excludes algae, fungi, and common names	http://www.theplantlist.org/
Traditional Knowledge Digital Library—India (TKDL)		TKDL is based on 148 books of Indian systems of medicine. Requires registration with the Government of India	http://www.tkdl.res.in/
University of Melbourne		Multilingual (21 languages) plant name database with links to sites in various languages	http://www.plantnames.unimelb.edu.au/Sorting/List_bot.html
University of Washington		Hyperlinks to Medline, USDA Plants Database, and Plants For A Future Database	https://staff.washington.edu/boerm/uwmhg//
USDA Plants Database		Searchable via multiple access points	http://plants.usda.gov/java/
Western United States Flora Checklists		Latin names and common names	http://www.swsbm.com/HOMEPAGE/Floras/Checklists.html
Prelude Medicinal Plants Database		The PRELUDE database concerns the use of plants in different traditional veterinarian and human medicines in Africa	http://www.africamuseum.be/collections/external/prelude

Continued

TABLE 21.1 Database of Detailed Information on Natural Products—cont'd

Database	Subdatabase	Summary	Website
Database Anti-Diabetic Medicinal Plants diversity (DADMP)		DADMP is developed and maintained by Gopinath Krishnasamy, Department of Bioinformatics, Alagappa University, Karaikudi. This database contains more than 100 medicinal plants from Valaiyans of Alagarkoil hills to treat diabetes	http://www.mkarthikeyan.bioinfoau.org/dadmp/
HerbMed		Collection of several herbal websites such as AGRICOLA, American Indian Ethnobotany Database, Carotenoid Database for US Foods, IBIDS, etc. It is basically a categorized, evidence-based resource for herbal information, with hyperlinks to clinical and scientific publications and dynamic links for automatic updating; produced by the nonprofit Alternative Medicine Foundation	http://www.herbmed.org/links.html
PLANT		Database on Brazilian medicinal plants	Manha et al. [8]
Medherb		A medicinal plant database with genetic information	Rajoka et al. [9]

AyurMedBase		An Ayurvedic medicinal database for traditional and Ayurvedic medicinal systems	www.grin.com
Cameroon 3D		Botanical database of Cameroon containing 2500 compounds of natural origin and 224 medicinal plants belonging to 55 families	Ntie-Kang et al. [10]
p-ANAPL		A collection of more than 500 natural products from African medicinal plants, which has been subjected to Lipinsky's "Rule of Five" and virtually unleashed relevant compounds with pharmacological properties	Ntie-Kang et al. [11]

TABLE 21.2 Botanical Image Database

Type	Database	Description	Website
General plants worldwide	Albion College Vascular Plant Image Gallery	A database to support organizational botany courses; images mainly collected from North America and the Caribbean basin	http://www.albion.edu/plants/
	An *Array of Botanical Images*	Over 24,000 botanical images, arranged alphabetically by genus	http://www.plantsystematics.org/reveal/pbio/RevealSlides/slideindex.html
	Anthos project	Mainly identifies Spanish flora, including photos, distribution maps, and nomenclature	http://www.anthos.es/
	Atrium (biodiversity information developed by the University of Texas)		http://www.atrium-biodiversity.org/
	Botanical Society of America Online Image Collection	Has photos categorized into a number of groups, including plant anatomy, botany, and plant science (many subcategories), people, places, and events, and plant morphology	http://pix.botany.org/index.php?module=simplemedia&type=user&func=view&ot=collection&tpl=tree
	Botanique.org	Images with botanical and biodiversity information	http://www.botanique.org/
	BotIT (includes fungi too)		http://botit.botany.wisc.edu/
	Digital Flowers	Images of angiosperms	http://www.life.illinois.edu/help/digitalflowers/
	Flora of the World	Flowering plants	http://www.floraoftheworld.org/

	Internet Directory for Botany	Over 100,000 images	http://www.ou.edu/cas/botany-micro/idb-alpha/botany.html
	Life Web Site		http://www.nic.funet.fi/pub/sci/bio/life/plants/magnoliophyta/index.html
	Noble Foundation Plant Image Gallery		https://www.noble.org/imagegallery/
	PhytoImages		http://www.phytoimages.siu.edu/
	PLANTS Database	United States Department of Agriculture National Resources Conservation Service. Searchable database with images, distribution maps, nomenclatural information, and more	http://plants.usda.gov/gallery.html
	Plantillustrations.org		http://plantillustrations.org/
	PlantSystematics.org	Keys, cladograms, and over 45,000 images of vascular plants	http://www.plantsystematics.org/
	Raintree Tropical Plant Database	Images of tropical raintree plants	http://www.rain-tree.com/plantimages.htm#.WGpEIlV97IW
	Scott's botanical links		http://www.ou.edu/cas/botany-micro/botlinx/subject/sub-pict.shtml

in specific electronic languages, which are open for scientists and accessible for exploration of unknown taxonomic datasets. Such biological identification methodology can be classified into three categories: the field guide method, dichotomous paper keys (which is the mostly used method), and computer-based methods. The latter can also be accomplished by four major techniques: hypertext keys, multiaccess keys, expert systems, and neural networks. Multiaccess keys based on a species−character matrix are used predominantly for the identification of biological databases (http://www.borealis.nu/exjobb/Index_en.html).

A good interactive key bears three fundamental attributes: (1) unrestricted character use, (2) ranking of the best character at any stage of the identification, and (3) opportunity to easily reach explanations of characters or more information about species. In a comparison of 14 identification programs and six interactive keys on the internet, the best keys, according to the three fundamental attributes and other important criteria, were selected: the programs Intkey, Linnaeus II, Lucid, Taxis, XID, and the internet key PollyClave 2 (http://www.borealis.nu/exjobb/Index_en.html). The various identification tools in this regard are summarized in Table 21.3.

In 1988, DELTA was adopted by the International Working Group on Taxonomic Databases for Plant Sciences as a standard language for compilation, analysis, and recognition of taxonomic data. In 2005, FreeDELTA was taken over by sourceforge.net, the largest global web-based platform for developing free software. FreeDELTA is built on program languages such as Python, C++, and Object Pascal libraries and uses an open source code that allows users to develop the program by themselves according to their needs. Currently there are 68 datasets in FreeDELTA and 22 datasets in the NaviKey server.

2.2.2 Digitization Tool

Digitization is used to select the identifying information of the plant under investigation. The identifying information is the detailed pro forma of the plant including nomenclature, genus, species, and other information. These databases are intensively utilizable to search for plants in a particular geographical region and collate information about those plants. The major digitization tools for the creation of databases are summarized in Table 21.4.

2.2.3 Biodiversity-Based Databases

Biodiversity-based databases ensemble and detail information regarding the spectrum of biotic organisms in a particular ecological niche. They actually encompass detailed data about the species, specimen, taxonomic distribution, or phylogenetic hierarchy of the biomass in the particular niche. These kinds of databases collate data regarding either superficial biodiversity of flora or fauna in that particular ecological habitat (e.g., species or taxonomic distribution of biotic lives) or consummate molecular biological information extracted from different species of that particular habitat. The significance of the second algorithm is that it is particularly conducive to searching, exploring,

TABLE 21.3 Various Digitization Tools for Identification of Natural Products

Main Identification Tool	Type of Natural Product	Brief Description	URL
Open Identification API	Plant, animal	A computer-assisted program to identify flora or fauna with free software techniques	http://wwbota.free.fr/Identification/
OpenKey	Plant	A Delta database encrypted by 200 characters to identify any plant in North Carolina Piedmont	http://www.ibiblio.org/openkey/intkey/web/intro.html
PANKEY	Plant, animal	A software program to describe taxonomy of an unknown flora or fauna; also describes numerical taxonomy (clustering and cladistics). Prepared by version 3.0 of Delta	http://www.exetersoftware.com/cat/pankey/pankey.html
PHPKey	Plant	A new interactive key to describe calicioid lichens and fungi of the Nordic countries. The database comprises 83 species, 27 characters, 216 character states, and related information. PHPKey is written in the HTML embedded programming language PHP	http://www.borealis.nu/exjobb/Index_en.html
PollyClave	Plant, animal	A multiple entry identification tool created by the University of Toronto	http://prod.library.utoronto.ca:8090/polyclave/
Rachis	Plant, animal	Software to allocate biological entities in a hierarchical system (both LINUX and MS Windows support) and an interactive key for retrieving data from them	http://rachis.sourceforge.net/
Scratchpads	Plant, animal	A database created by the Natural History Museum (UK)	http://editwebrevisions.info/

Continued

TABLE 21.3 Various Digitization Tools for Identification of Natural Products—cont'd

Main Identification Tool	Type of Natural Product	Brief Description	URL
Stinger's Lightweight Interactive Key Software (SLIKS)	Plant	A Javascript program for biological identification	http://www.stingersplace.com/SLIKS/
Taxy	Fungi		http://www.collectivesource.com/taxy/taxy.html
TeleNature	Plant		http://www3.isrl.uiuc.edu/~TeleNature/projects/telenature.html
X:ID	Plant, animal		http://uio.mbl.edu/services/key.html
Xper	Plant	Free Delta-based software for taxonomical identification and other trait analyses	http://lis-upmc.snv.jussieu.fr/lis/?q=en/resources/software/xper2
Barcode of Life Database (BOLD) (identification by DNA barcoding)	Plant, animal	An identification portal based on sequence search and matching analysis using DNA barcoding	http://www.boldsystems.org/index.php/IDS_OpenIdEngine
EDIT's cybertaxonomy platform	Plant, animal	A common platform for digital analysis of cybertaxonomy, used for data storage and exchange; collections and specimens; descriptions; fieldwork; literature; and geography	http://wp5.e-taxonomy.eu/
Electronic field guide	Plant, animal	A web-based version structured by the Department of Computer Science and Biology at the University of Massachusetts, Boston, with funding from the National Science Foundation	http://wiki.cs.umb.edu/

TABLE 21.3 Various Digitization Tools for Identification of Natural Products—cont'd

Main Identification Tool	Type of Natural Product	Brief Description	URL
FreeDELTA[a]	Plant, animal		http://freedelta.sourceforge.net/
Idenature Guides	Plant, animal	General web-based identification tool for plant, animal, fungi, insects, and other related species	http://www.discoverlife.org/mp/20q
LucID	Plant, animal	An interactive key-based software platform to identify or diagnose biological entities	http://www.lucidcentral.com
Medical fungi identification website	Fungi	Identification tool for filamentous fungi of medical importance except the genera *Aspergillus* and *Penicillium*	http://www.cbs.knaw.nl/medical/DefaultPage.aspx
Meka	Plant		http://ucjeps.berkeley.edu/keys/downloads/Meka31.exe

[a]*FreeDELTA: DELTA stands for DEscription Language for TAxonomy. FreeDELTA is the world's largest software tool that is utilized by taxonomic scientists for the compilation and accumulation of taxonomic data all over the world (http://freedelta.sourceforge.net/). FreeDELTA is a language that comprises both qualitative (binary or multistate, ordered or unordered) or quantitative (integer or real) characters. Although the software was created by Mike Dallwitz at CSIRO Division of Entomology, Canberra, Australia, in the mid-1970s, it was later used by various other taxonomic program developers such as Eric Gouda at the Botanic Gardens of Utrecht University (TAXASOFT) in the Netherlands, Nicholas Lander at the Western Australian Herbarium (DMSWIN) in Australia, Antonio Valdecasas at the Museo Nacional de Ciencias Naturales (EDEL) in Spain, Gregor Hagedorn at the Institute of Microbiology, Federal Biological Research Center (DELTAAccess) in Germany, Michael Bartley and Noel Cross at the Arnold Arboretum of Harvard University (NaviKey) in the United States, Claudio Rivetti and Riccardo Percudani at the University of Parma (WebDelta) in Italy, and Mauro J. Cavalcanti at Museu Nacional/Universidade Federal do Rio de Janeiro (DIANA) in Brazil.*

or interconnecting different biological entities in that specific domain. This unveiling, in turn, helps to predict biological properties of a cluster of entities before performing any operation on them. These can be divided into four major databases

1. Biodiversity databases for all classes.
2. Plant-based biodiversity databases.
3. Animal-based biodiversity databases.
4. DNA barcode-based biodiversity databases.

TABLE 21.4 Major Digitization Tools for the Creation of Databases

Name of the Database	Coverage	Brief Description	Website
Bauble	Flora	Free web interface software. It generates reports through XSL and Mako formatters.	http://bauble.belizebotanic.org/
Bibmaster	Flora, fauna	Database application	http://www.gbif.es/bibmaster/bibmaster_Inphp
Biota	Flora, fauna	Do	http://viceroy.eeb.uconn.edu/biota
Biotica	Flora, fauna		http://www.conabio.gob.mx/biotica/cms/index.php
Brahms	Flora	Digitization tool to collate all the data of botanical species especially in the United States	http://dps.plants.ox.ac.uk/bol/BRAHMS/Home/Index
Herbar	Flora	General digitization program involved in showcasing, storing, and indexing information of traditional products	http://www.gbif.es/herbar/herbar_Inphp
KE Emu-Electronic Museum	Flora, fauna	Do	http://www.kesoftware.com/emu-home.html
Pandora	Flora	Do	http://www.ibiblio.org/pub/academic/biology/ecology+evolution/software/pandora/
Specify 6	Flora, fauna	Do	http://specifysoftware.org/
Zoorbar	Flora, fauna	Do	www.gbif.org/resource/81736

2.2.3.1 Biodiversity Databases for All Classes

The biodiversity databases for all classes include databases where diverse species of a biotic community irrespective of flora or fauna have been created inside the database. A brief description of the databases is summarized in Table 21.5.

TABLE 21.5 Biodiversity Databases for All Classes

Database	Brief Description	Website
Biodiversity Heritage Library (BHL)	Collection of digitized literature including images in open access form	www.biodiversitylibrary.org
Integrated TaxonomicInformation System (ITIS)	Taxonomic information about plants, animals, fungi, and microbes especially in North America	http://www.itis.gov/
Species 2000	A global database of biodiversity of all known species in the world. It is basically a collaborative program between CODATA (International Council for Science: Committee on Data for Science and Technology), IUBS (International Union of Biological Sciences), and the IUMS (International Union of Microbiological Societies) of the early 1990s. It is an associate member in the Global Biodiversity Information Facility (GBIF) , a data provider to EC LifeWatch, and is acknowledged by the United Nations Environment Program (UNEP) together with the Convention on Biological Diversity (CBD)	http://www.sp2000.org/
Tree of Life (TOL)	A biodiversity database compiled by biologists all over the world containing more than 10,000 web pages regarding biodiversity of organisms together with their phylogenetic information	http://tolweb.org/tree/phylogeny.html
TreeBASE	Basically deals with phylogenetic information about trees formed by the Phyloinformatics Research Foundation, Inc.	http://www.treebase.org/treebase-web/home.html
Barcode of Life Databases (BOLD)	Basically integrates biodiversity information all over the world by DNA barcodes. Since the DNA barcode of certain genes (such as *matK, rbcL*, mitochondrial cytochrome oxidase c subunit-I [COI], and its genes) shows variation from species to species, it is used as a reference for species identification. BOLD comprises such DNA barcodes to identify unknown species	http://www.barcodinglife.org/views/logInphp

Continued

TABLE 21.5 Biodiversity Databases for All Classes—cont'd

Database	Brief Description	Website
Global Invasive Species Database (GISD)	A database about invasive species across the globe. It was formed by the Invasive Species Specialist Group (ISSG) of the Species Survival Commission of the IUCN (International Union for Conservation of Nature). The development of GISD was initiated by the Global Invasive Species Programme and other statutory bodies such as the National Biological Information Infrastructure, Manaaki Whenua-Landcare Research, the Critical Ecosystem Partnership Fund, the University of Auckland, and private donations	http://www.issg.org/database/welcome/
Invasive and Exotic Species	Deals with invasive and exotic species of North America	http://www.invasive.org/
Invasive Species in Canada	Deals with invasive species of Canada	http://www.invasivespecies.gc.ca/english/view.asp?x=1
Biodiversity of Mexico VN	A Mexican database	http://www.vivanatura.org/About%20VN.html
Natural History Museum	Biodiversity museum, i.e., database of the United Kingdom. Has also been developed in America and other countries	http://www.nhm.ac.uk/nature-online/life/index.html
South African National Biodiversity Institute database	A South African biodiversity database	http://www.sanbi.org/frames/infofram.htm
Arctos	A database for finding information about biodiversity using different identifier tools	http://www.arctos.database.museum
ASEAN Biodiversity Information database (BISS)	A biodiversity system of ASEAN flora and fauna; the database is maintained to be accessible by all the users across the globe	http://www.aseanbiodiversity.org

TABLE 21.5 Biodiversity Databases for All Classes—cont'd

Database	Brief Description	Website
Convention on International Trade in Endangered Species (CITES) species database	A database regarding the Convention on International Trade in Endangered Species of Wild Fauna and Flora. This database maintains the list so that the existence of such kinds of species is not threatened over the world.	All species ever listed in CITES Appendices I, II, and III
iNATURALIST	A site for reporting personal observations of any plants or animals in the world.	www.inaturalist.org
iSpot	Free international database about ecosystems and associated flora and fauna	www.iSpotnature.org
Natural History Information System	A collection of databases of natural products. This is basically a DINA project (DIgital Information System for NAtural History Collections) involving all types of collections such as zoological, botanical, geological, and paleontological collections, living collections, biodiversity inventories, observation records, and molecular data	www.dina-project.net
NatureServe	A natural biodiversity system	www.natureserve.org
WikiSpecies	A free database formed by the Wikimedia foundation	https://species.m.wikimedia.org
Pan European Species directories Infrastructure	A taxonomic database for Europe	www.eu-nomen.eu
NatureDATA	A natural products database in the United Kingdom	naturedata.org.uk
Georgia (country) biodiversity website	A database maintained in Georgia University, USA	biodiversity-georgia.net
Biodiversity Heritage Library (BHL)	A database containing over 2 million volumes of biodiversity literature involving two centuries and is jointly maintained by the United States and the United Kingdom. As of October 2010, BHL had coded 31,397,395 pages from 83,616 volumes, and from 43,140 titles.	www.biodiversitylibrary.org

2.2.3.2 Plant-Based Biodiversity Databases

There are several databases based only on plants or flora as summarized in Table 21.6.

2.2.3.3 Animal-Based Biodiversity Database

There are several databases that contain digitized information about animals, as summarized in Table 21.7.

TABLE 21.6 Digitized Databases Based on Plants or Flora

Database	Brief Description	Website
Algae Base	A botanical database especially focused on algae belonging to aqueous, land, and marine organisms	http://www.algaebase.org/
Australian Biological Resources Study Flora online	A specifically designed database compiling information on Australian biodiversity involving plants	http://www.environment.gov.au/biodiversity/abrs/online-resources/flora/main/
DiaMedBase	A database particularly covering information about the plants to cure diabetes	http://www.progenebio.in/DMP/DMP.html
Encyclopedia of Indian Medicinal Plants	Database containing information on Indian medicinal plants	www.medicinalplants.in
Plants For A Future (edible and medicinal plants)	A database comprising more than 7000 medicinal plants where each plant is categorized on the basis of its edibility and therapeutic use	http://www.pfaf.org/user/plantsearch.aspx
Royal Botanic Garden, Edinburgh	A database containing information about plants inside the Royal Botanical Garden as well as plants involved in specific research projects such as ADIAC Diatom Image Database, DIADIST Website, Apiales Resource Centre, Southeast Asian Begonia Database, etc.	http://www.rbge.org.uk/databases
United States Department of Agriculture (USDA) Plants Database	A database about vascular plants, mosses, liverworts, hornworts, and lichens of the United States and its territories. It contains more than 50,000 images of such plants	http://plants.usda.gov/

TABLE 21.6 Digitized Databases Based on Plants or Flora—cont'd

Database	Brief Description	Website
Fungal Records Database of Britain and Ireland	Mostly focused on the fungal databases of Britain and Ireland; however, it contains information about more than 2 million fungal records	http://www.fieldmycology.net/
Index Fungorum	A project containing information with all formal names (scientific names) of the fungal kingdom. It is a joint collaborative project partnered by the Royal Botanic Gardens, Kew, Landcare Research, and the Institute of Microbiology, Chinese Academy of Sciences	http://www.indexfungorum.org/Names/Names.asp
MycoBank	A fungal database run by the International Mycological Association	http://www.mycobank.org/
USDA fungal database	A database based on US national fungus collections	http://nt.ars-grIngov/fungaldatabases/

TABLE 21.7 Databases About Animals Across the Globe

Database	Brief Description	Website
Amphibian Species of the World	This database contains three types of amphibian species: anura (frogs), salamanders (Caudata), and gymnophiona (caecilians). This database contains 7645 amphibian species, of which 6745 are frogs and toads, 695 are newts and salamanders, and 205 are caecilians	http://research.amnh.org/herpetology/amphibia/index.php
Antbase	Database based on all ant species around the world	http://antbase.org/
Australian Biological Resources Study Fauna Online	A zoological database based on Australian biodiversity.	http://www.environment.gov.au/biodiversity/abrs/onlineresources/fauna/index.html
Biosystematic Database of World Diptera	Knowledge database about dipterian animals. More than 150,000 species of Diptera are described under more than 250,000 names	http://www.sel.barc.usda.gov/Diptera/

Continued

TABLE 21.7 Databases About Animals Across the Globe—cont'd

Butterflies and Moths of the World	Database about butterflies and moths.	http://www.nhm.ac.uk/research-curation/projects/butmoth/
CephBase	Knowledge database about cephalopods around the world	http://www.cephbase.utmb.edu/
Fishbase	Knowledge database about fishes around the world	Fishbase Fauna http://www.fishbase.org/home.htm
Mammal Species of the World	Knowledge database about mammalian species around the world	http://vertebrates.si.edu/mammals/msw/
The Reptile Database	Knowledge database about reptiles around the world	http://www.reptile-database.org/
Universal Chalcidoidea Database	Knowledge database about Chalcidoidea group of wasps around the world	http://www.nhm.ac.uk/research-curation/projects/chalcidoids/index.html
Zoology: Extinct and Endangered database	Organized and integrated information database about extinct and endangered species of the world	http://www.oum.ox.ac.uk/database/zoology/extinct.html

2.2.3.4 DNA Barcode-Based Databases

2.2.3.4.1 Barcode of Life Database Barcode of Life Database (BOLD) is a DNA barcode-based biodiversity database that has four portals. The first is a public data portal that contains 1.7 million DNA barcode sequences, which is freely searchable and categorized by geographical, taxonomical, and depository databases. The second is Barcode Index Numbers, which comprise several numbers signifying specific barcodes, a DNA barcode education portal, which is explorable by students and scientists and could be enriched by the latter through submitting new barcodes, the third portal being Barcode Index numbers and the fourth portal is the workbench that allows scientists to work and analyze DNA barcodes on a common platform. The current coverage of BOLD is given by Table 21.8.

2.2.3.4.2 Korean Barcode of Life This is a barcode database designated to elucidate barcodes of all Korean species. Currently, the database contains 5531 barcode sequences from 2429 Korean species.

TABLE 21.8 Database Having Barcode of Life Database (BOLD) Coverage

Items	Number
Barcode clusters for animals (Barcode Index Numbers)	495,328
All sequences	6,175,187
Barcode sequences	5,339,196
Animals	176,400
Plants	65,732
Fungi and other life forms	20,838

2.3 Metadata Portals

Since the biological databases are interconnected and provide full information upon being integrated in one common data portal, several metadata portals in the digitization of traditional knowledge have been created to access primary databases by searching through secondary search engines. Such metadata engines are provided in Table 21.9.

3. BIODIVERSITY ANALYSIS

Several interdisciplinary approaches have been emerging over the last few decades to organize, narrate, collate, and then use biodiversity data for various purposes such as phylogenetic analysis, evolutionary analysis, metabolic pathway analysis, and many others [11a]. Various bioinformatics, molecular biology, pharmacogenomics, cheminformatics, and other approaches have emerged to process the biodiversity data available across various databases. Thus to cope with this, a plethora of analytical tools has evolved to analyze the biodiversity database. Selected analytical tools based on their usage in this process are summarized in Table 21.10.

4. VIRTUAL SCREENING OF NATURAL PRODUCTS FROM DATABASES

Virtual screening of natural products is the in silico process of screening a large database of natural products obtained in a particular or diverse ecological niche to achieve a specific pharmacological response. The in silico process depends on various bioinformatics, involving docking and network pharmacology as described next.

TABLE 21.9 Metadata Portals Having Barcode of Life Database (BOLD) Coverage

Database	Brief Description	Website
Atlas of Living Australia (ALA)	An online repository of Australian biodiversity including Australian flora, fauna, and fungi. This national database of Australian biodiversity provides a platform to access and search information on specific components	http://www.anbg.gov.au/cpbr/program/hb/index.html
Australian Virtual Herbarium (AVH)	Same as above; however, the major coverage is plant species	http://www.ersa.edu.au/avh/index.jsp
Encyclopedia of Life (EOL)	A metadatabase aimed at collating information about each and every living being of 1.9 million species discovered so far on earth. It is proposed to link information on each species with an infinitely long page containing all the information including images, videos, etc. about the species	http://www.eol.org/home.html
Global Biodiversity Information Facility (GBIF)	Single data management portal through which all taxonomical, biogeographical, hierarchical, and genomic information about various species of the world can be accessed	http://www.gbif.org
iSpecies	A zoological species database created by Glasgow University, Scotland	http://darwInzoology.gla.ac.uk/~rpage/ispecies/index.php
LifeWatch	A European biodiversity database specially constructed to preserve all environmental biodiversity by accessing such information in Europe	http://www.lifewatch.eu/index.php?id=411
Ocean Biogeographic Information System (OBIS)	This is mainly focused on collecting, compiling, and processing biodiversity present in oceans. Currently OBIS possesses 27.7 million pieces data from 126,000 species from 849 databases. It is jointly run with various other digitization agencies such as GBIF, Consortium for Biodiversity of Life (CBOL), and Taxonomic Database Working Group (TDWG)	http://v2.iobis.org

TABLE 21.9 Metadata Portals Having Barcode of Life Database (BOLD) Coverage—cont'd

Database	Brief Description	Website
SpeciesBase	A species database supported by the Reference Center on Environmental Information (CRIA) to collate and share data regarding various botanical and zoological species across the world. It is structured over Visual Basic for Application (VBA) and Microsoft Access. The user interface is designed on the BONABIO information taxonomic database adopted by the Federal University of Parana system	http://www.speciesbase.org/
Universal Biological Indexer and Organizer	This is a combinatory database where biological data from different resources are collected and presented in a meaningful, legible, and organized way. Web services such as XML and SOAP are used for processing the data. It is basically known as Taxonomic Name Server, interconnected as Name Bank (11,106,374 records) and Classification Bank (90 classifications)	http://www.ubio.org

4.1 Screening Through Network Pharmacology

Different biodiversity databases, in addition to organizing information regarding flora, fauna, or microbiome within a particular area or across the globe, also help in drug discovery with the bioinformatics approach. Over the last few decades attempts have been made to organize the mammoth data of biodiversity in drug discovery processes by the virtual screening method. The most common method of such virtual screening is combining network pharmacology or polypharmacology [12–14] with molecular docking. Since network pharmacology suggests that multiple genes or proteins are involved in a particular phenotype or disease, responsible proteins are searched for first while considering a particular disease. The protein structures are then downloaded from a protein databank and reported compounds from various biodiversity databases are docked onto the particular set of proteins. The "best hit" compounds are then taken as leads for subsequent drug discovery [15,16]. The flow chart for performing virtual screening is summarized in Fig. 21.1.

TABLE 21.10 Tools for Biodiversity Analysis

Software	Use	Category	URL
ADE4	Ecological analysis	S, F	http://cran.rproject.org/src/contrib/Descriptions/ade4.html
ADAPTS	Paleobiological analysis	S, F	http://www.paleodb.org/paleosource/code.php?stage=download&project_no=4
APE	Phylogenetic and diversification analyses	S, F	http://pbil.univ-lyon1.fr/R/ape/
CODA	Nature conservation and planning	S, F	http://members.ozemail.com.au/~mbedward/coda/coda.html
DIVA-GIS	Mapping and ecological modeling	S, F	www.diva-gis.org/
Ecopath with Ecosim (EwE)	Ecological modeling (marine environment, including the effects of fishing)	S, F	http://www.ecopath.org/index.php?name=About
GARP	Ecological modeling	S, F	http://nhm.ku.edu/desktopgarp/
GRASS GIS	GIS is used for geospatial data management and analysis	S, F	http://grass.itc.it/
LAMARC	Population studies	S, F	http://evolution.genetics.washington.edu/lamarc.html
MAXENT	Ecological modeling (species distribution)	S, F	http://www.cs.princeton.edu/~schapire/maxent/
MEGA	Phylogenetic analysis	S, F	http://www.megasoftware.net/
Mesquite	Evolutionary analysis	S, F	http://mesquiteproject.org/mesquite/mesquite.html

Molphy	Phylogenetic analysis	W, F	http://bioweb.pasteur.fr/seqanal/interfaces/prot_nucml.html
MrBayes	Phylogenetic analysis	S, F	http://mrbayes.csit.fsu.edu/index
PAST (PAlaeontological STatistics)	Paleontological statistics	S, F	http://folk.uio.no/ohammer/past/
PATN	Pattern analysis	S, C	http://www.patn.com.au/
PopTools	Population dynamics and ecological model analysis	S, F	http://www.cse.csiro.au/poptools/
PAUP	Phylogenetic analysis	S, C	http://paup.csit.fsu.edu/
PHYLIP	Phylogenetic analysis	S, F	http://evolution.genetics.washington.edu/phylip.html
Rarefaction calculator	Diversity estimation and indices	W, F	http://www2.biology.ualberta.ca/jbrzusto/rarefact.php
TNT	Phylogenetic analysis	S, F	http://www.cladistics.com/aboutTNT.html
TreeAlign	Phylogenetic analysis	W, F	http://bioweb.pasteur.fr/seqanal/interfaces/treealign-simple.html
TreeView X	Phylogenetic tree visualization	S, F	http://darwinzoology.gla.ac.uk/~rpage/treeviewx/

C, Commercial; *F*, free; *S*, standalone application; *W*, web based application.
Reprinted with permission from J. Gaikwad, P.D. Wilson, S. Ranganathan, Ecological niche modeling of customary medicinal plant species used by Australian Aborigines to identify species-rich and culturally valuable areas for conservation, Ecol. Modell. 222 (2011) 3437–3443.

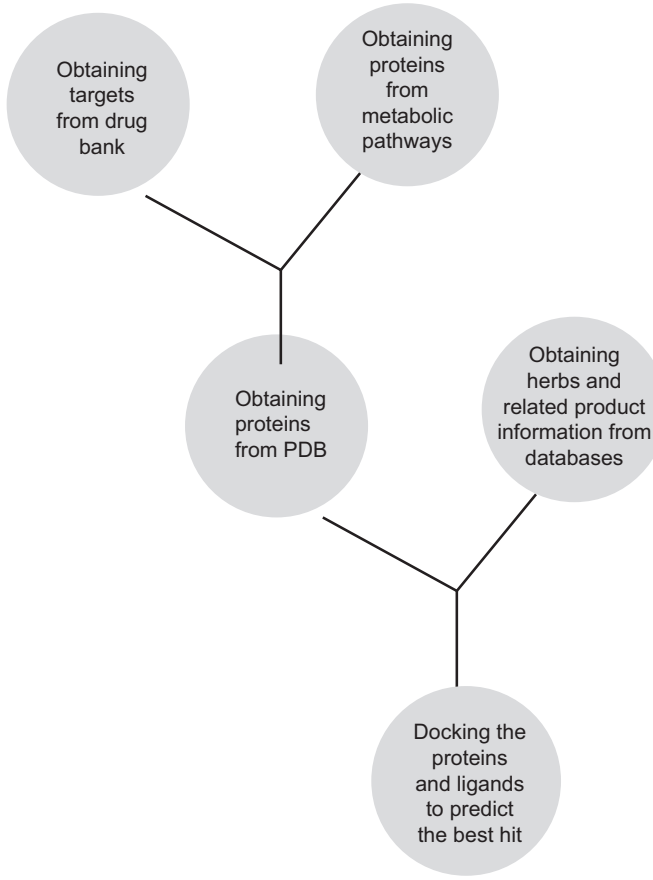


FIGURE 21.1 Tentative methodology for virtual screening of natural products.

For example, Gu et al. [15] reported that Universal Natural Products Database (UNPD)-derived natural products screening yielded five medicinal plants, namely, *Hypericum perforatum*, *Ganoderma lucidam*, *Holarhena antidysenteria*, *Celastrus orbiculatus*, and *Marraya eucherestifolia* as having antidiabetic activity. The authors used the drug target networks (DTN) methodology to explore the new set of plants against the aforementioned disease from a library of 208,000 natural products [15].

4.2 Screening Through Cheminformatics

In a review, Medina-Franco [17] reported various natural products databases, cheminformatics methods of their screening, and ultimately lead findings for various pharmacological responses thereof. For example, he acknowledged

that the database ZINC containing more than 19 million molecules, traditional Chinese medicine (TCM) database, UNPD containing 197,201 molecules, UNIIQUIM database (Mexico), and NuBBE database (Brazil) were significantly large databases. These databases have been reported to be used for drug discovery purposes. For example, a web server-based docking of TCM followed by de novo ligand design has been acknowledged by Tsai et al. [18]. Moreover, Chen et al. [19] reported discovery of pancreatic triacylglycerol inhibitors through computational approaches in TCM.

4.2.1 Analysis of Structural Diversity and Complexity

Structural complexity is the hallmark signature of natural product molecules. However, drug discovery in such conditions is aided via two digitized approaches. One is application of structural fingerprints and the other is using chemical scaffolds [20,21]. Apart from benzene and acyclic molecules, flavones, coumarins, and flavanones have been identified as the most frequent scaffolds across the various natural products databases [12].

4.2.2 Structure Promiscuity Index Difference

Dandapani and Marcaurelle [22] in a study reported that the structural diversity of natural products eventually leading to generation of diverse pharmacological activities is due to diverse fraction of unsaturation in various natural products [22]. In continuation, Clemons et al. [23] screened a library of 15,000 compounds, both natural and synthetic, over 100 diverse proteins involved in various metabolic pathways. They later acknowledged that structural diversity actually leads to specificity in substrate-protein binding, finally converging in specific pharmacological activity. To design this in silico, they created an index, namely, Structure Promiscuity Index Difference, to calculate changes in protein binding due to small changes in structure [23].

4.2.3 Chemical Space—Importance and Evaluation

One of the significant approaches to the digitized evaluation of natural products is evaluation of chemical space. It can be defined as defined by Dobson: "the total descriptor space that encompasses all the small carbon-based molecules that could in principle be created" [23a]. In another concept, Lipinsk and Hopkins mentioned that "chemical space can be viewed as being analogous to the cosmological universe in its vastness, with chemical compounds populating space instead of stars." [24]. The evaluation of chemical space has been extensively used by various authors [24a], [15,25–27]. The analysis mainly relies on ChemGPS-NP_{web}, an online tool for chemical space analysis. Web analysis is basically reliant on principle component analysis, which divides it into four dimensions and maintains specific compound descriptors in each dimension.

4.2.4 Application of Cheminformatics to Drug Discovery

The cheminformatics approach has been applied to drug discovery to successfully unfold various natural products for a set of pharmacological responses. For example, Cao et al. [28] screened more than 4000 natural products from 100 medicinal plants against estrogen receptors (ERα) and (ERβ), which eventually led to the discovery of 11 selective nonsteroidal estrogen receptor modulators.

Guasch et al. [30] discovered five new drug leads from 89,000 natural products for peroxisome-activated receptors [30]. In continuation, Ngo and Li [31] developed molecules for Alzheimer's disease from a pool of natural products [31]. The authors screened a library of 342 compounds from Vietnamese plants and docked them subsequently against a set of amyloid ($A\beta_{1-40}$ and $A\beta_{1-42}$) peptides to reveal five compounds showing promising potential against Alzheimer's disease. Also Gu et al. [32] performed virtual screening of 676 compounds from a TCM database with 37 proteins related to type 2 diabetes mellitus [32].

5. BIOINFORMATICS APPROACH TO THE DIGITIZATION OF KNOWLEDGE ON NATURAL PRODUCTS

5.1 Quality Control of Herbals Using Next Gen Sequencing

Herbal products are often supplied with supplements from various other natural products. Hence a proper, defined, quality-controlled approach to evaluate these other products is still a difficult task. Ivanova et al. [33] proposed that next gen sequencing followed by DNA barcoding could elucidate the quality of herbal supplements. In this study the authors demonstrated a DNA sequencing approach for taxonomic authentication of herbal supplements from five medicinal plants: *Echinacea purpurea*, *Valeriana officinalis*, *Ginkgo biloba*, *H. Perforatum*, and *Trigonella foenum-graecum*. Using DNA barcoding of *rbcL* and *ITS2* regions the authors successfully accomplished the identification of the foregoing medicinal plants. In addition, the authors also claimed to detect adulterants mixed with the herbal supplements in these food formulations. Interestingly, the amount of contaminants as well as products due to plant−fungi interactions could also be detected by quantitative analysis of next gen sequencing [33].

5.2 Expressed Sequence Tags

In 2012 Sharma and Sarkar described various bioinformatics approaches to discover natural products from various resources, e.g., genomics and transcriptomics data to categorize phylogenetic information about medicinal plants. The authors reported the contribution of "expressed sequence tags" (ESTs) for transcriptomic data organization in universal data portals such as

National Center for Biotechnology Information (NCBI). In addition, they also reported the EGENES database for more authentic information on plant transcriptomic data with better organization of ESTs to correlate genetic information with functional information [34]. In the Medicinal Plants Genomic Resource Database, such complete plant transcriptomic data have been created.

5.3 Simple Sequence Repeats

Apart from ESTs, the authors also acknowledged utilization of simple sequence repeats (SSRs) to compile transcriptomic information of medicinal plants. SSR markers have been shown to be most advantageous because of their multiallelic nature, reproducibility, codominant inheritance, high abundance, and extensive genome coverage [86]. SSRLocator is an example of a computational approach for detection and characterization of SSRs and minisatellite motifs [35].

5.4 Constructing Network Biology Through Chemogenomics

Network biology is an important tool to construct networking maps to unlock the role of various genes in multiple biological functions inside the body. Since body metabolic pathways are usually constructed of various genes or proteins in an orchestrated way, which often involve a spectrum of genes mutually overlapping in nature, perturbing the functional outputs of those genes often elucidates various metabolic pathways inside the body [36]. The Kyoto Encyclopedia of Genes and Genomes (KEGG) is a reliable database that provides information on such metabolic proteins as well as the pathways [37]. In addition, there is a web server called Path Pred [38] that predicts pathways of multistep reaction for a given query compound, starting with a similarity search against the KEGG COMPOUND database. With the help of these chemogenomics databases, network biology is constructed, which on in silico screening leads to the path of drug discovery. For example, the *Catharanthus roseus* gene-metabolite coexpression network was dissected and ultimately led to the discovery of genes associated with the biosynthesis of terpenoid indole alkaloids [39].

5.5 Network Biology Models—Distance-Based Mutual Information Model

In this model, a mutual information entropy and herb distance metric is used to score herb interactions [40] and constitutes an herb network with the combination rules of TCM. Thus, network-lined herb−herb interaction could produce therapeutic activity, which has already been reported to produce angiogenesis activity.

5.6 Quantitative Composition–Activity Relationship Study

Since all herbal components and constituent structures have not yet been discovered or elucidated, quantitative structure–activity relationship (QSAR) studies relating plant constituents and their bioactivities are difficult to perform. Hence the quantitative composition–activity relationship (QCAR) study has been invented to predict plant extract activity in silico [41]. Although this method is still not very accurate and needs full experimental design or bioassays to predict the correct score [42], it could be justifiably used to foretell bioactivity of plant-based extract activity on a computer chip. With approaches such as artificial neuronal network and support vector machines (SVM), prediction with QCAR has been simpler than previously encountered. For example, Nayak et al. [43] reported that adjusting components of a TCM-based herb Qi-Xue-Bing-Zhi-Fang with the aforementioned methods significantly reduced the blood cholesterol level in rats.

5.7 Network Target-Based Identification of Multicomponent Synergy

This methodology of bioinformatics-based digitization of traditional knowledge consists of two components: topology score and agent score. In the first approach, a total network of plant species is analyzed based on their contribution to different diseases and drug actions, and a topology score is assigned. The agent score is given based on their plant phenotypes. Afterward, based on these scores, a synergistic score is given to evaluate the synergistic action between two medicinal plants [44].

5.8 Application of the Bioinformatics Approach for Drug Discovery From Traditional Plants

Phylogenetic analysis of natural products revealed that it has direct correlation with biological activity. Such analysis coupled with gas chromatography-mass spectrometry studies of an alkaloidal fraction of *Phaedranassa dubia* revealed a direct correlation of its acetylcholinesterase inhibitory activity with alkaloids such as galanthamine or lycorine [45]. Anticancer drugs were developed by building network models using a bioinformatics-guided approach [46]. Furthermore, the QSAR-based approach with natural products evolved immunomodulatory compounds; cleomiscosin molecules (A, B, C) were discovered using the QSAR approach [47]. In addition, virtual screening of natural products led to the evolution of peroxisome proliferator-activated receptors (PPARs). Petersen et al. constructed a pharmacophore-based model of 13 PPAR-based partial agonists from *Pistacia lentiscus* from the Chinese Natural Products Database. Virtual screening revealed an oleoresin from the aforementioned plant to have a potential PPAR activator effect [48].

5.9 In Silico Docking

Docking is one of the popular in silico approaches to screen a library of compounds having medicinal interest. This approach is also employed in the case of natural products, because natural products are a diverse set of secondary metabolites and without virtual screening, isolation and subsequent bioactivity estimation are often tedious and complicated jobs. However, docking strategies have evolved a set of newer compounds that can be used as leads for emancipation of medicinal compounds. For example, Zhong et al. reported an inducible nitric oxide synthetase (iNOS) inhibitor of a quinoline derivative, which can be used as scaffold for further designing associated compounds [49]. The author reported docking of more than 90,000 natural products from the ZINC database in silico to evolve one successful compound against iNOS. Again, Li et al. [49a] reported discovery of a potential anticancer compound (breast cancer) through molecular docking by screening of 11,247 compounds from the ZINC database against human epidermal growth factor 2. Likewise, docking has been used to screen a plethora of drugs such as antiinflammatory IKKβ inhibitors [50,51], acid sphingomyelinase inhibitors [52], PPARγ partial agonists, dipeptidyl peptidase inhibitors for antidiabetic drugs [30], STAT 1 and STAT 3 inhibitors [53], multidrug efflux pump inhibitors for reducing antibiotic resistance [54], marine natural products acting on acetylcholine binding protein [55], ellagic acid derivatives on selected enzymes of *Mycobacterium tuberculosis* [56], and others. Thus docking has been a promising alternative approach to drug discovery through digitization of natural products.

6. INVERTNET

InvertNet is a database containing information about invertebrate species across the globe [57]. The database is designed and maintained by the US National Science Foundation's Advancing Digitization of Biological Collections program, and provides digital access to approximately 60 million specimens housed in 22 arthropods (primarily insects). They provide a 3D image of every insect under this category and label them with a unique digital code identifier. This database provides a unique solution to the digital database monitoring system mainly focusing on insects.

7. SCREENING FROM ACTINOBACTERIA

Doroghazi et al. [58] proposed a newer path for drug discovery with a digitized study of natural products. They undertook Actinobacteria as a model of natural product source. Since organisms classified as actinomycetes are reported for natural product biosynthetic gene clusters [59], the authors used the bioinformatics approach to combine 11,422 gene clusters with 4122 gene cluster families (GCF). Subsequent studies revealed 830 genomes from the

microorganism, which exhibited coding for hundreds of future leads. In this process, peptidogenomics tools are used to investigate new peptides from a set of mass spectrometry-based peptide fragmentation datasets [60]. Various other bioinformatics tools have been used such as NaPDoS, a natural product domain finder that works on clustering phylogenetically correlated secondary metabolite production gene clusters [61], antiSMASH, a rapid identification, annotation, and analysis of secondary metabolites producing genome sequences from bacterial and fungal origins [62], ClusterMine360, a database for microbial polyketide synthetase [63], SEARCHPKS, a program for investigating polyketide synthetase domains [64], ASMPKS, an analysis program for molecular polyketide synthase domains [85], DoBISCUIT, a database for secondary metabolite producing gene clusters [65], NORINE, a database for nonribosomal peptides [66], PKMIner, a database for exploring type-II polyketide synthetase [67], and others.

Ikram et al. [16] used a digitization screening model to isolate several compounds as neuraminidase inhibitors, i.e., active against influenza. The authors used a docking approach to hit 3000 compounds from the Malaysia Natural Products Database to find the best 12 hits as leads for antiinfluenza drugs. Lead compounds, their docking scores, and IC_{50} values are provided in Table 21.11.

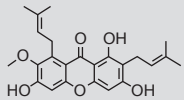
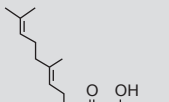
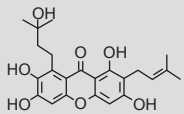
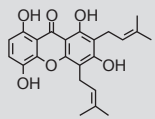
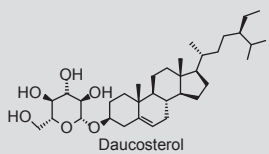
8. PREDICTION INFORMATICS FOR SECONDARY METABOLOMES

For the prediction of secondary metabolites from genetic subsets in natural products, Skinnider et al. published a report of a new web-based software design [68] that they called Prediction Informatics for Secondary Metabolomes. In this web user interface, the software has several components. One is the BLAST search program to find the homologous sequence of the gene subspace under investigation, it is then applied to hidden Markov models to identify different protein domains such as polyketide synthetase domains, transacting acyl transferase and adenylation domains, deoxysugar biosynthesis domains, β-lactam-specific domains, etc. The other components tools are HMMER (version 3.1) for hidden Markov model searches, the Chemistry Development Kit (version 1.4.19) for chemical abstractions, BioJava (version 3.0.7) for sequence translation, RDKit (version 2014.03.1) for Tanimoto coefficient computation, and Apache Batik (version 1.7) for vector image generation [68].

9. BIOINFORMATICS TO NATURAL PRODUCTS THROUGH SYNTHETIC BIOLOGY

Bioinformatics is a subject that creates a bridge between genomic data and natural product discovery. Several tools have been discovered that have been

TABLE 21.11 Neuraminidase Inhibitors From Malaysia Natural Products Database

Compounds	Autodock Score (kcal/mol)	IC_{50} (μM)	% Inhibition (at 250 μg/mL)
 α-mangostin*	−8.87	91.95 ± 0.09	93.08 ± 0.04 (at 609 μM)
 Rubraxanthone*	−9.85	89.71 ± 0.08	92.42 ± 0.12 (at 609 μM)
 Garcinone C*	−8.85	95.49 ± 0.08	90.13 ± 0.02 (at 603 μM)
 Gartanin*	−11.07	126.64 ± 0.13	80.25 ± 0.32 (at 631 μM)
 Daucosterol	−8.99	275.45 ± 0.03	60.65 ± 0.29 (at 433 μM)

Continued

TABLE 21.11 Neuraminidase Inhibitors From Malaysia Natural Products Database—cont'd

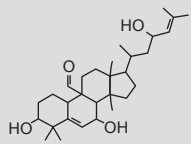
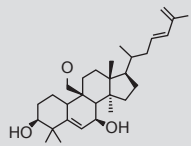
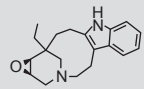
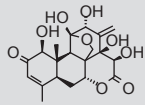
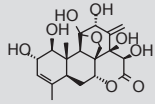
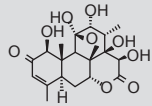
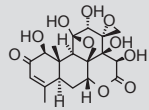
Compounds	Autodock Score (kcal/mol)	IC_{50} (μM)	% Inhibition (at 250 μg/mL)
 Momordicin I	−11.49	≥250	15.58 ± 0.36 (at 529 μM)
 Kuguacin J	−10.21	≥250	21.42 ± 0.50 (at 550 μM)
 Voaphylline	−10.49	≥250	20.95 ± 0.09 (at 800 μM)
 Eurycomanone	−10.89	≥250	20.84 ± 0.67 (at 612 μM)
 Eurycomanol	−9.83	≥250	21 ± 02.18 (at 609 μM)

TABLE 21.11 Neuraminidase Inhibitors From Malaysia Natural Products Database—cont'd

Compounds	Autodock Score (kcal/mol)	IC_{50} (μM)	% Inhibition (at 250 μg/mL)
 13α,21-dihyroeurycomanone	−9.92	≥250	2.90 ± 0.34 (at 631 μM)
 13α,21-epoxyeurycomanone	−10.45	≥250	34.50 ± 0.27 (at 589 μM)

Reprinted from N.K.K. Ikram, J.D. Durrant, M. Muchtaridi, A.S. Zalaludin, N. Purwitasari, N. Mohamed, et al., A virtual screening approach for identifying plants with Anti H5N1 neuraminidase activity. J. Chem. Inf. Model 55 (2015) 308–316.

useful to predict the coding of natural products from a set of genes or protein clusters. The tools and their applications are shown in Table 21.12.

Bioinformatics tools are also used to study natural products using synthetic biology tools. The design of natural products from biosynthetic gene clusters depends on searching relevant sequence space from a database of millions of gene sequences. Afterward, domains are located in the gene using a domain search tool, which often relies on homology match of the unknown gene sequence with established genes of known function. The most putative or conserved sequences are found and synthesizable natural products are then designed based on pharmacophore matches against a known database of secondary metabolites. The tools related to natural product discovery from genomic clusters are summarized in Table 21.13.

10. ESNAPD, A NOVEL WEB-BASED BIOINFORMATICS TOOL

Environmental Surveyor of Natural Products Diversity (eSNaPD) is a web-based bioinformatics-based platform to discover gene clusters for the discovery of natural products. This database first relies on construction of the

TABLE 21.12 Bioinformatics Tools for Natural Products

	Enzymes	Pathways	Regulatory Components	Chassis
Selection	*Mining*	*Rankling*	*Characterization*	*Genome scale modeling*
	antiSMASH	FindPath	Registry of standard biological	
		RetroPath		
		GemPath		
		Metabolic Tinker		
Prediction	*Annotation*	*Search*	*Tuning*	*Optimization*
	antiSMASH	BNICE	RBS Calculator	Optknock
	Enzymes	Pathways	Regulatory Components	Chassis
	CanOE	Route Search		EMILIO
	Enzyme Function Initiative	PathPred		SIMUP
	SymZime	RetroPath		
		GEM-Path		

Reprinted with permission from P. Carbonell, A. Currin, A.J. Jervis, N.J.W. Rattray, N. Swainston, Y. Cunyu, E. Takano, R. Breitling, Bioinformatics for the synthetic biology ofnatural products: integrating across the Design-build-test cycle. Nat. Prod. Rep. 16 (2016) 925–932.

database and then searches for any unknown gene sequence within the database. For construction of the database, first, amplification of different natural product biosynthetic gene clusters by polymerase chain reaction (PCR) is done where various biosynthetic gene clusters such as acyl carrier protein, polyketide synthetase, adenylation, acyltransferase, condensation, dehydratase, epimerization, enoyl reductase, ketoreductase, methyltransferase, peptidyl carrier protein, and thioesterase are involved. After amplification, 95% sequence identity of the PCR-amplified genes is mined and saved as consensus sequence as a unique sequence read. In search space, once an unknown gene sequence is placed after PCR and thereafter sequencing, the sequence is searched for the highest hit in the database by the NCBI BLAST algorithm and

TABLE 21.13 Computational Tools for Natural Products Discovery

Software Program or Database	URL	Last Publication or Document Update	Main Content/Function
Database Focusing on Gene Clusters			
Bactibase	http://bactibase.pfba-lab-tun.org	2011	Web accessible database of bacteriocins
ClusterMine360	http://www.clustermine360.ca/	2013	Web accessible database of biosynthetic gene clusters
ClustScanDatabase	http://csdb.bioserv.pbf.hr/csdb/ClustScanWeb.html	2013	Web accessible database of polyketide synthetase/natural product biosynthetic gene clusters
DoBISCUIT	http://www.bio.nit_e.g_o_jp/pks/	2015	Web accessible database of polyketide synthetase/natural product biosynthetic gene clusters
Integrated Microbial Genome-Atlas of Biosynthetic Gene Clusters	http://img.jgi.doe.gov/abc	2015	Web accessible database of biosynthetic gene clusters
MIBiG	http://mibig.secondarymetabolites.org	2015	Web accessible repository of biosynthetic gene clusters
Recombinant ClustScan Database	http://csdb.bioserv.pbf.hr/csdb/R CSDB.html	2013	In silico recombinant database
Database Focusing on Bioactive Compounds			
Antibioticome	http://magarveylab.ca/antibioticome	2015	Web accessible database of compounds, compound families, and mode of action

Continued

TABLE 21.13 Computational Tools for Natural Products Discovery—cont'd

Software Program or Database	URL	Last Publication or Document Update	Main Content/Function
Database Focusing on Gene Clusters			
ChEBI	https://www.ebi.ac.uk/c hebi/	2015	Web accessible database of compounds, compound families, and mode of action
ChEMBL	https://www.ebi.ac.uk/c hembl/	2015	Web accessible database of bioactive compounds with drug-like properties
Chem Spider	http://www.chemspider.com/	2015	Web accessible database of structures and properties
KNAPSAcK database	http://kanaya.aist-nara.ac.jp/ KN ApSAcK/	2015	Web accessible database of bioactive compounds with KNAPSAcK standalone database
NORINE	http://bioinfo.lifl.fr/norine	2015	Web accessible database of natural products
Novel Antibiotics Database	http://www.antibiotics.or.jp/ journal/database/database-top. htm	2008	Web accessible database of compounds
PubChem	http://pubchem.ncbi.nlm.nih. gov/	2015	Web accessible database of compounds and bioactives
StreptomeDB	http://www.pharmaceutical-bioinformatics.de/s tr ep t ome db	2015	Web accessible database of compounds produced by streptomycetes; download of compounds and metadata

Metabolomics Tools			
Cycloquest	http://cyclo.ucsd.edu	2011	Web application to correlate tandem MS data of cyclopeptides with gene clusters
GNPS	http://gnps.ucsd.edu/	2015	Generic metabolomics portal to analyze tandem mass spectrometry data (dereplication and molecular networking)
GNP/iSNAP	http://magarveylab.ca/gnp/	2015	Web application to automatically identify tandem mass spectrometry data based on genomics data
NRPquest	http://cyclo.ucsd.edu	2014	Web application to automatically identify tandem mass spectrometry data based on genomics data
Pep2Path	http://pep2path.sourceforge.net	2014	Standard replication data relating peptide sequence tags with biosynthetic gene clusters
RiPPquest	http://cyclo.ucsd.edu	2014	Web application to correlate ribosomally and posttranslationally modified peptide tandem data with gene clusters

Reprinted with permission from T. Weber, H.U. Kim, The secondary metabolite bioinformatics portal: computational tools to facilitate synthetic biology of secondary metabolite production, Syn. Sys. Biotechnol. 1 (2016) 69–79.

the most matched hit is calculated by e-value as convened by the NCBI BLAST algorithm. The search hits so far are further processed by hierarchical clustering and a phylogenetic relationship is established. Thus relevant domains cloned in the gene cluster are mined and established [70].

11. DNA BARCODING IN NATURAL PRODUCTS

A DNA barcode is a short segment of genomic DNA (<1000 bp), which is highly variable in sequence and used to determine hierarchical and evolutionary relationships between plants and animals [71]. This is used for species identification through sequence alignment by a series of sequence alignment algorithms [72]. For DNA barcoding, the standard genomic spaces used are chloroplast ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL) and maturase K (matK) as core barcodes [73]. Together with this, other regions are also used as DNA barcodes such as the spacer between photosystem II protein D1, tRNA-His (psbA-trnH spacer), the nuclear ribosomal internal transcribed spacer 2 (ITS2) in plants, and cytochrome oxidase c subunit-I (COI) for animals [74–80]. DNA barcoding has been applied to identify the contamination of natural products such as identifying consumer relevant mushrooms [81] among poisonous and nonedible mushrooms, detecting contamination and substitution of herbal products [79], herbal medicines, and dietary supplements [82], and many others. DNA barcoding is also used for phylogenetic evolution of plants [83].

12. DISCUSSION AND CONCLUSION

The knowledge of usage of traditional medicinal plant databases has become a paradigm of immense importance due to intense utilization of natural products across the globe over the last few decades. However, attempts have been made to perturb the dataset of natural products digitally due to complexity and difficulty of exploring millions of natural products by physical sorting. In accordance with this, digitization of natural products has crept in via four major approaches. Data preservation is found in various web databases wherefrom data can be mined according to the user's demands, providing virtual screening of different DTNs or databases for drug discovery, bioinformatics-guided approaches for proper utilization of natural products knowledge for lead optimization in discovery processes, and in silico approaches such as docking or molecular modeling for drug discovery. In the first approach, several UNPDs have been created such as UNPD, CMKb, ebDB, ZINC, TCM, UNIIQUIM, NuBBE, pANAPL, InvertNet, CamMED NP, and DIVERSet, where information on 560 to more than 19 million compounds has been stored based on the database. In addition, different digitization tools have been created for various purposes such as identification

tools (EDIT's cybertaxonomy platform, Electronic Field Guide, Medical Fungi Identification Website, Free Delta, and Meka), digitization tools (Bauble, Bibmaster, Biota, and Biotica), and biodiversity together with ecological modeling tools (ADE4, APE, DIVA-GIS, GARP, LAMARC, Molphy, and others). Virtual screening-based natural product search is based on in silico chemical space and docking analyses where similar property harnessing compounds are searched for based on molecular descriptors so that new leads can be discovered from those analogous natural products. For example, using chemical space analysis and subsequent docking on estrogen receptors (ERα and ERβ) led to the discovery of 11 nonsteroidal estrogen modulators. Furthermore, screening 89,000 natural compounds from the ZINC database, five compounds as PPARs have been revealed. Bioinformatics-guided drug discovery from natural products analyzes sequence space to investigate the phylogenetic relationship, biodiversity, and ecological modeling. Some of the popular approaches for bioinformatics investigation use ESTs or SSR locators, restriction fragment length polymorphism, randomly amplified polymorphic DNA, and single nucleotide polymorphism to investigate the phylogenetic relationship between and potential gene clusters among the species in the database. For example, in one study, 11,422 natural product gene clusters from Actinobacteria were grouped into 4122 GCF. This ultimately led to the revelation of 830 genome datasets encompassing the potential for biosynthesizing newer drug leads. In another study, distance-based mutual information model and network target-based identification of multicomponent synergy approaches have been undertaken to generate synergy scores for ranking synergistic effects of agent combinations in a specific database. Statistical learning methods such as probabilistic neuronal network, k-nearest neighbor method, SVM, and decision tree have also been undertaken to elucidate similar gene clusters for new drug discovery. Most importantly, the DNA barcoding approach has also been used over the last few years to search phylogenetic and neighborhood relationships together with synergistic likeliness among diverse natural products. As a rule of thumb, *matK*, *rbcL*, and *ITS2* sequences have been undertaken as DNA barcodes for plant investigation, while for animals, mitochondrial COI has been considered as a DNA barcode. Lastly, the in silico docking approach has been a popular tool for predictive approaches in drug discovery. This approach has been successfully implemented for drug discovery in iNOS inhibitors, antineoplastic compounds such as HER-2 inhibitors in female breast cancer, and many more. Most interestingly, a study has been performed to encode entire medicinal and aromatic plants in Africa in a digital database by using programming language C++. Thus digitization of traditional knowledge is an updated, time-economic, highly investigative, and efficient strategy for studying natural products as well as for drug discovery based on these products.

REFERENCES

[1] W. Altermann, J. Kazmierczak, Archean microfossils: a reappraisal of early life on Earth, Res. Microbiol. 154 (2003) 611−617.

[2] R.T. Watson, V.H. Heywood, I. Baste, B. Dias, R. Gamez, T. Janetos, W. Reid, R. Ruark (Eds.), Global Biodiversity Assessment, 1995 ed., Panama Cambridge University Press, 1995.

[3] E.O. Wilson, The encyclopedia of life, Trends Ecol. Evol. 18 (2003) 77−80.

[4] J. Gaikwad, V. Chavan, Open access and biodiversity conservation: challenges and potentials for the developing world, Data Sci. J. 5 (2006) 1−17.

[5] C.D. Becker, K. Ghimire, Synergy between traditional ecological knowledge and conservation science supports forest preservation in Ecuador, Conserv. Ecol. 8 (2003) 1.

[6] R. Govaerts, How many species of seed plants are there? Taxon 50 (2001) 1085−1090.

[7] N.R. Farnsworth, Computerized data base for medicinal plants, in: World Health Forum, vol. 5(4), World Health Organization, Geneva, 1984, pp. 373−376.

[8] E.M. Manha, M.C. Silva, M.G.C. Alves, M.G.L. Brandão, M.B. Almeida, PLANT − a bibliographic database about medicinal plants, Braz. J. Pharmacogn. 18 (2008) 614−617.

[9] M.I. Rajoka, I. Sobia, K. Sana, E. Beenish, Medherb: an interactive bioinformatics database and analysis resource for medicinally important herbs, Curr. Bioinform. 9 (2014) 23−27.

[10] F. Ntie Kang, J.A. Mbah, L.M. Mbaz, et al., CamMedNP: building the Cameroonian 3D structural natural products database for virtual screening, BMC Complement. Altern. Med. 13 (2013) 88.

[11] F. Ntie-Kang, P.A. Onguene, G.W. Fotso, et al., Virtualizing the p-ANAPL library: a step towards drug discovery from African medicinal plants, PLoS One 9 (2014a) e90655.

[11a] J. Gaikwad, P.D. Wilson, S. Ranganathan, Ecological niche modeling of customary medicinal plant species used by Australian Aborigines to identify species-rich and culturally valuable areas for conservation, Ecol. Modell. 222 (2011) 3437−3443.

[12] A.L. Hopkins, Network pharmacology, Nat. Biotechnol. 25 (2007) 1110−1111.

[13] A.L. Hopkins, Network pharmacology: the next paradigm in drug discovery, Nat. Chem. Biol. 4 (2008) 682−690.

[14] S. Li, B. Zhang, Traditional chinese medicine network pharmacology: theory, methodology and application, Chin. J. Nat. Med. 11 (2013) 110−120.

[15] J. Gu, L. Chen, G. Yuan, X. Xu, A drug-target network-based approach to evaluate the efficacy of medicinal plants for type II diabetes mellitus, Evid. Based Complement. Altern. Med. (2013), 203614, 7 pages.

[16] N.K.K. Ikram, J.D. Durrant, M. Muchtaridi, A.S. Zalaludin, N. Purwitasari, N. Mohamed, et al., A virtual screening approach for identifying plants with Anti H5N1 neuraminidase activity, J. Chem. Inf. Model 55 (2015) 308−316.

[17] J.L. Medina-Franco, Chemoinformatics characterization of the chemical space and molecular diversity of compound libraries, in: T. Andrea (Ed.), Diversity-Oriented Synthesis: Basics and Applications in Organic Synthesis, Drug Discovery, and Chemical Biology, John Wiley & Sons Inc., 2013, pp. 325−352.

[18] T.-Y. Tsai, K.-W. Chang, C. Chen, iScreen: world's first cloud-computing web server for virtual screening and de novo drug design based on TCM database @Taiwan, J. Comput. Aided Mol. Des. 25 (2011) 525−531.

[19] K.-Y. Chen, S.-S. Chang, C.Y.-C. Chen, *In Silico* identification of potent pancreatic triacylglycerol lipase inhibitors from traditional Chinese Medicine, PLoS One 7 (2012) e43932.

[20] M.A. Koch, A. Schuffenhauer, M. Scheck, S. Wetzel, M. Casaulta, A. Odermatt, P. Ertl, H. Waldmann, Charting biologically relevant chemical space: a structural classification of natural products (SCONP), Proc. Natl. Acad. Sci. U.S.A. 102 (2005) 17272−17277.

[21] N. Singh, R. Guha, M.A. Giulianotti, C. Pinilla, R.A. Houghten, J.L. Medina-Franco, Chemoinformatic analysis of combinatorial libraries, drugs, natural products, and molecular libraries small molecule repository, J. Chem. Inf. Model. 49 (2009) 1010−1024.

[22] S. Dandapani, L.A. Marcaurelle, Accessing new chemical space for 'undruggable' targets, Nat. Chem. Biol. 6 (2010) 861−863.

[23] P.A. Clemons, N.E. Bodycombe, H.A. Carrinski, J.A. Wilson, A.F. Shamji, B.K. Wagner, A.N. Koehler, S.L. Schreiber, Small molecules of different origins have distinct distributions of structural complexity that correlate with protein-binding profiles, Proc. Natl. Acad. Sci. U.S.A. 107 (2010) 18787−18792.

[23a] C.M. Dobson, Chemical space and biology, Nature 432 (2004) 824−828.

[24] C. Lipinski, A. Hopkins, Navigating chemical space for biology and medicine, Nature 432 (2004) 855−861.

[24a] J.L. Medina-Franco, Interrogating Novel Areas of Chemical Space for Drug Discovery using Chemoinformatics, Drug Dev. Res. 73 (2012) 430−438.

[25] J. Larsson, J. Gottfries, L. Bohlin, A. Backlund, Expanding the ChemGPS chemical space with natural products, J. Nat. Prod. 68 (2005) 985−991.

[26] J. Larsson, J. Gottfries, S. Muresan, A. Backlund, ChemGPS-NP: tuned for navigation in biologically relevant chemical space, J. Nat. Prod. 70 (2007) 789−794.

[27] T.I. Oprea, J. Gottfries, Chemography: the art of navigating in chemical space, J. Comb. Chem. 3 (2001) 157−166.

[28] X. Cao, J. Jiang, S. Zhang, L. Zhu, J. Zou, Y. Diao, W. Xiao, L. Shan, H. Sun, W. Zhang, J. Huang, H. Li, Discovery of natural estrogen receptor modulators with structure-based virtual screening, Bioorg. Med. Chem. Lett. 2013 (23) (2013) 3329−3333.

[29] Deleted in review.

[30] L. Guasch, E. Sala, J. Iwaszkiewicz, M. Mulero, M.-J. Salvado, M. Pinent, V. Zoete, A. Grosdidier, G. Pujadas, S. Garcia-Vallve, Identification of PPARgamma partial agonists of natural origin (I): development of a virtual screening procedure and in vitro validation, PLoS One 7 (2012) e50816.

[31] S.T. Ngo, M.S. Li, Top-leads from natural products for treatment of Alzheimer's disease: docking and molecular dynamics study, Mol. Simul. 39 (2013) 279−291.

[32] J. Gu, H. Zhang, L. Chen, S. Xu, G. Yuan, X. Xu, Drug−target network and polypharmacology studies of a Traditional Chinese Medicine for type II diabetes mellitus, Comput. Biol. Chem. 35 (2011) 293−297.

[33] N.V. Ivanova, M.L. Kuzmina, T.W.A. Braukmann, A.V. Borisenko, E.V. Zakharov, Authentication of herbal supplements using next-generation sequencing, PLoS One 11 (5) (2016).

[34] V. Sharm, I. Neil Sarkar, Bioinformatics opprtumities for identification of medicinal plants, Brief. Bioinform. 14 (2012) 238−250.

[35] L.C. da Maia, D.A. Palmieri, V.Q. de Souza, et al., SSR locator: tool for simple sequence repeat discovery integrated with primer design and PCR simulation, Int. J. Plant Genom. (2008) 412696.

[36] K. Mochida, Y. Uehara-Yamaguchi, T. Yoshida, et al., Global landscape of a co-expressed gene network in barley and its application to gene discovery in Triticeae crops, Plant Cell Physiol. 52 (2011) 785−803.

[37] M. Kanehisa, S. Goto, M. Furumichi, et al., KEGG for representation and analysis of molecular networks involving diseases and drugs, Nucleic Acids Res. 38 (2010) D355–D360.

[38] Y. Moriya, D. Shigemizu, M. Hattori, et al., PathPred: an enzyme-catalyzed metabolic pathway prediction server, Nucleic Acids Res. 38 (2010) W138–W143.

[39] H. Rischer, M. Oresic, T. Seppanen-Laakso, et al., Gene-to-metabolite networks for terpenoid indole alkaloid biosynthesis in *Catharanthus roseus* cells, Proc. Natl. Acad. Sci. U.S.A. 103 (2006) 5614–5619.

[40] S. Li, B. Zhang, D. Jiang, et al., Herb network construction and co-module analysis for uncovering the combination rule of traditional Chinese herbal formulae, BMC Bioinform. 11 (Suppl. 11) (2010) S6.

[41] X.P. Zhao, X.H. Fan, J. Yu, et al., A method for predicting activity of traditional Chinese medicine based on quantitative composition-activity relationship of neural network model, China J. China Mater. Med. 29 (2004) 1082–1085.

[42] Y. Wang, X. Wang, Y. Cheng, A computational approach to botanical drug design by modeling quantitative composition-activity relationship, Chem. Biol. Drug Des. 68 (2006) 166–172.

[43] S.K. Nayak, P.K. Patra, P. Padhi, A. Panda, Optimization of herbal drugs using soft computing approach, Int. J. Log. Comput. 1 (2010) 34–39.

[44] S. Li, B. Zhang, N. Zhang, Network target for screening synergistic drug combinations with application to traditional Chinese medicine, BMC Syst. Biol. 5 (Suppl. 1) (2011) S10.

[45] M.M. Larsen, A. Adsersen, A.P. Davis, et al., Using a phylogenetic approach to selection of target plants in drug discovery of acetylcholinesterase inhibiting alkaloids in Amaryllidaceae tribe Galantheae, Biochem. Syst. Ecol. 38 (2005) 1026–1034.

[46] C.X. Xue, X.Y. Zhang, M.C. Liu, et al., Study of probabilistic neural networks to classify the active compounds in medicinal plants, J. Pharm. Biomed. Anal. 38 (2005) 497–507.

[47] D.K. Yadav, A. Meena, A. Srivastava, et al., Development of QSAR model for immunomodulatory activity of natural coumarinolignoids, Drug Des. Dev. Ther. 4 (2010) 173–186.

[48] R.K. Petersen, K.B. Christensen, A.N. Assimopoulou, et al., Pharmacophore-driven identification of PPARgamma agonists from natural sources, J. Comp. Aided Mol. Des. 25 (2011) 107–116.

[49] H.-J. Zhong, L. Li-J, C.-M. Chong, et al., Discovery of a natural product-like iNOS inhibitor by molecular docking with potential neuroprotective effects in vivo, PLoS One 9 (2014) e92905.

[49a] J. Li, H. Wang, J. Li, J. Bao, C. Wu, Discovery of a Potential HER2 Inhibitor from Natural Products for the Treatment of HER2-Positive Breast Cancer, Int. J. Mol. Sci. 17 (2016) 1055.

[50] C.-H. Leung, D.S.-H. Chan, Y.-W. Li, et al., Hit identification of IKKβ natural product inhibitor, BMC Pharmacol. Toxicol. 14 (2013) 1–8.

[51] E. Sala, L. Guasch, J. Iwaszkiewicz, M. Mulero, M.-J. Salvado, Pinent M., V. Zoete, A. Grosdidier, S. Garcia-Vallve, O. Michielin, G. Pujadas, Identification of human IKK-2 inhibitors of natural Origin (Part I): modeling of the IKK-2 kinase domain, virtual screening and activity assays, PLoS One 6 (2011) e16903.

[52] J. Kornhuber, M. Muelbacher, S. Trapp, S. Pechmann, A. Friedl, M. Reichel, C. Muhle, L. Terfloth, T.W. Groemer, G.M. Spitzer, R.L. Klaus, E. Gulbins, P. Tripal, Identification of novel functional inhibitors of acid sphingomyelinase, PLoS One 6 (2011) e23852.

[53] M. Szelag, A. Czerwonlec, J. Wesoly, H.A.R. Bluyssen, Identification of STAT1 and STAT3 specific inhibitors using comparative virtual screening and docking validation, PLoS One 10 (2) (2015) e0116688.

[54] V. Aparna, K. Dineshkumar, N. Mohanalakshmi, D. Velmurugan, W. Hopper, Identification of natural compound inhibitors for multidrug efflux pumps of *Escherichia coli* and *Pseudomonas aeruginosa* using in silico high throughput virtual screening and in vitro validation, PLoS One 9 (2014) e101840.

[55] D. Kudryavtsev, T. Makarieva, N. Utkina, E. Santalova, E. Kryukova, C. Methfessel, V. Tsetlin, V. Stonik, I. Kasheverov, Marine natural products acting on the cetylcholine-binding protein and nicotinic receptors: from computer modeling to binding studies and electrophysiology, Mar. Drugs 12 (2014) 1859−1875.

[56] J.A. Shilpi, M.T. Ali, S. Saha, S. Hasan, A.I. Gray, V. Seidel, Molecular docking studies on InhA, MabA and PanK enzymes from *Mycobacterium tuberculosis* of ellagic acid derivatives from *Ludwigia adscendens* and *Trewia nudiflora*, In Silico Pharmacol. 3 (2015) 1−7.

[57] C. Dietrich, J. Hart, D. Raila, U. Ravaioli, N. Sobh, O. Sobh, C. Taylor, InvertNet: a new paradigm for digital access to invertebrate collections, ZooKeys 209 (2012) 165−212.

[58] J.R. Doroghazi, J.C. Albright, A.W. Goering, K.-S. Ju, R.R. Haines, K.A. Tchalukov, D.P. Labeda, N.L. Kelleher, W.W. Metcalf, A roadmap for natural product discovery based on large-scale genomics and metabolomics, Nat. Chem. Biol. 10 (2014) 963−968.

[59] S.D. Bentley, et al., Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2), Nature 417 (2002) 141−147.

[60] R.D. Kersten, et al., A mass spectrometry-guided genome mining approach for natural product peptidogenomics, Nat. Chem. Biol. 7 (2011) 794−802.

[61] N. Ziemert, et al., The natural product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity, PLoS One 7 (2012) e34064.

[62] M.H. Medema, et al., AntiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences, Nucleic Acids Res. 39 (2011) W339−W346.

[63] K.R. Conway, C.N. Boddy, ClusterMine360: a database of microbial PKS/NRPS biosynthesis, Nucleic Acids Res. 41 (2013) D402−D407.

[64] G. Yadav, R.S. Gokhale, D. Mohanty, SEARCHPKS: a program for detection and analysis of polyketide synthase domains, Nucleic Acids Res. 31 (2003) 3654−3658.

[65] N. Ichikawa, et al., DoBISCUIT: a database of secondary metabolite biosynthetic gene clusters, Nucleic Acids Res. 41 (2013) D408−D414.

[66] S. Caboche, et al., NORINE: a database of nonribosomal peptides, Nucleic Acids Res. 36 (2008) D326−D331.

[67] J. Kim, G.-S. Yi, PKMiner: a database for exploring type II polyketide synthases, BMC Microbiol. 12 (2012) 169.

[68] M.A. Skinnider, C.A. Dejong, P.N. Rees, C.W. Johnston, H. Li, A.L. Webster, M.A. Wyatt, N.A. Magarvey, Genomes to natural products prediction informatics for secondary metabolomes (PRISM), Nucleic Acids Res. 43 (2015) 9645−9662.

[69] P. Carbonell, A. Currin, A.J. Jervis, N.J.W. Rattray, N. Swainston, Y. Cunyu, E. Takano, R. Breitling, Bioinformatics for the synthetic biology of natural products: integrating across the Design-build-test cycle, Nat. Prod. Rep. 16 (2016) 925−932.

[70] B.V.B. Reddy, A. Milshtein, Z. Charlop-Powers, S.F. Brady, eSNaPD: a versatile, web-based bioinformatics platform for surveying and mining natural product biosynthetic diversity from Metagenomes, Chem. Biol. 21 (8) (2014) 1023−1033.

[71] N.B. Zahra, Z.K. Shinwari, M. Qaiser, DNA barcoding: a tool for standardization of herbal medicinal products (HMPS) of lamiaceae from Pakistan, Pak. J. Bot. 48 (2016) 2167−2174.
[72] Z.K. Shinwari, K. Jamil, N.B. Zahra, Molecular systematics of selected genera of subfamily Mimosoideaefabaceae, Pak. J. Bot. 46 (2014) 591−598.
[73] P.M. Hollingsworth, L.L. Forrest, J.L. Spouge, A DNA barcode for land plants, Proc. Natl. Acad. Sci. 106 (2009) 12794−12797.
[74] S. Chen, H. Yao, J. Han, et al., Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species, PLoS One 5 (2010) e8613.
[75] T. Gao, H. Yao, J. Song, Y. Zhu, C. Liu, S. Chen, Evaluating the feasibility of using candidate DNA barcodes in discriminating species of the large Asteraceae family, BMC Evol. Biol. 10 (2010) 324−330.
[76] H. Yao, J. Song, C. Liu, K. Luo, J. Han, Y. Li, X. Pang, H. Xu, Y. Zhu, P. Xiao, S. Chen, Use of ITS2 region as the universal DNA barcode for plants and animals, PLoS One 5 (2010) e13102.
[77] Y.M. Fu, W.M. Jiang, C.X. Fu, Identification of species within Tetrastigma (Miq.) Planch. (Vitaceae) based on DNA barcoding techniques, J. Syst. Evol. 49 (2011) 237−245.
[78] J.P. Han, L.C. Shi, X.C. Chen, Y.L. Lin, Comparison of four DNA barcodes in identifying certain medicinal plants of Lamiaceae, J. Syst. Evol. 50 (2012) 227−234.
[79] S.G. Newmaster, M. Grguric, D. Shanmughanandhan, S. Ramalingam, S. Ragupathy, DNA barcoding detects contamination and substitution in North American herbal products, BMC Med. 11 (2013) 222.
[80] C.I. Michel, S.M. Rachel, T. Yanille, M. Jeanmaire, The nuclear internal transcribed spacer (ITS2) as a practical plant DNA barcode for herbal medicines, J. Appl. Res. Med. Aromat. (2016). https://doi.org/10.1016/j.jarmap.2016.02.002.
[81] H.A. Raja, T.R. Baker, J.G. Little, N.H. Oberlies, DNA barcoding for identification of consumer-relevant mushrooms: a partial solution for product certification? Food Chem. 214 (2017) 383−392.
[82] I. Parveen, S. Gafner, N. Techen, S.J. Murch, I.A. Khan, DNA barcoding for the identification of botanicals in herbal medicine and dietary supplements: strengths and limitations, Planta Med. 82 (2016) 1225−1235.
[83] Q.-Y. Dai, Q. Gao, C.-S. Wu, D. Chesters, C.-D. Zhu, Ai-B. Zhang, Phylogenetic reconstruction and DNA Barcoding for closely related pine moth species (Dendrolimus) in China with multiple gene markers, PLoS One 7 (2012) e32544.
[84] A.T. Peterson, D.A. Vieglais, A.G.N. Sigüenza, M. Silva, A global distributed biodiversity information network: building the world museum, Bull. Br. Ornithol. Club 123A (2003) 186−196.
[85] H. Tae, E.-B. Kong, K. Park, ASMPKS: an analysis system for modular polyketide synthases, BMC Bioinform. 8 (2007) 327.
[86] R.K. Varshney, A. Graner, M.E. Sorrells, Genic microsatellite markers in plants: features and applications, Trends Biotechnol. 23 (2005) 48−55.

Chapter 22

Good Agricultural Practices: Requirement for the Production of Quality Herbal Medicines

Supradip Saha, Abhishek Mandal, Anirban Dutta
ICAR-Indian Agricultural Research Institute, New Delhi, India

1. INTRODUCTION

Unwanted and/or tacit materials, time and again, have been present or been claimed to be present in medicinal or herbal plant medicines around the globe. The substances that have been in the news included microbes such as pathogens, pesticides, mycotoxins, radioactive particles, and heavy metals such as arsenic. The incremental demand and usage of herbal medicines around the world, coupled with the vigorous expansion of the global market demand for the medicinal plants or medicinal plant−derived active ingredients, and quality control (QC) of medicinal plant materials as well as the finished herbal medicinal products have taken center stage as issues of major concern for health agencies, herbal pharmaceutical industries, and the general public, as a whole [1].

National rules for registration and regulation of herbal medicines vary from country to country. Herbal medicines are categorized as prescription medicines or nonprescription medicines, wherever they are regulated. Herbal products as a group along with medicines, may coexist in a certain country. Due to lacunae in regulation, poor QC systems, and faulty distribution channels (which includes Internet-based sales), herbal products categorized other than as medicines and foods are inclined toward increasing potential for drastic consequences. There is a belief that GAP standards are restrictive and obstruct farmers and their agriculture processes. However, the fundamental guiding principle of GAP is the achievement of a safe and sustainable food production system for both growers and consumers. This safe production system is necessary to ensure the right of consumers to hygienic, nutritious, and affordable

Natural Products and Drug Discovery. https://doi.org/10.1016/B978-0-08-102081-4.00022-8