Metagenomics for Drug Discovery

Amit Jethwa, Jayesh Bhagat, Jacinta Teresa George, and Sagar Shah

Abstract In the twenty-first century, approximately 100 years since the discovery and use of the first drug penicillin, antimicrobial resistance has become the major global threat to human health, thereby, demanding the need to discover new candidate metabolites exhibiting therapeutic potential to meet current medical treatment demands. Newer emerging diseases, relatively long duration of discovery and development process of drugs among others also necessitate identification of novel drug candidates. Till date thousands of drugs have been discovered, two thirds of which have their origins from the culturable microbes present in soil or diverse habitats around the globe. Since the advancements in high-throughput technologies in genomics, it has become possible to discover novel candidate metabolites having therapeutic potential from complete microbiomes chosen in a culture independent manner leading to an exponential increase in novel drug discoveries. The present chapter, thus, aims to bring to focus the role of metagenomics—a biotechnological tool that involves extraction of DNA from communities of microbial populations circumventing the need of culturing the organism followed subsequently by screening and/or sequencing of DNA—in drug discovery. The chapter will begin with a background of drugs including microbial-derived ones followed by a section on various approaches currently available for drug discovery from microbes. This will lead to the field of metagenomics, various techniques involved and how metagenomics of microbes can help bioprospect their potential to synthesize drugs as well as how

All authors have contributed equally to the chapter formulation.

A. Jethwa · J. Bhagat · J. T. George (✉) · S. Shah
Department of Biotechnology, S K Somaiya College, Somaiya Vidyavihar University, Mumbai 400 077, India
e-mail: jacintateresa@somaiya.edu

A. Jethwa
e-mail: amit.jethwa@somaiya.edu

J. Bhagat
e-mail: jayesh@somaiya.edu

S. Shah
e-mail: sagar.shah3@somaiya.edu

S. Kulkarni et al. (eds.), *Novel Technologies in Biosystems, Biomedical & Drug Delivery*,
https://doi.org/10.1007/978-981-99-5281-6_6

metagenomics can assist in selection of microbes having the ability to synthesize drugs of higher therapeutic potential.

Keywords Bioprospect · Drug discovery · High-throughput technologies · Metagenomics · Therapeutic potential

1 Introduction to Drugs

1.1 What Are Drugs?

The words 'drugs', 'natural products' and 'pharmaceuticals' are being quite interchangeably used worldwide (Taylor 2015). The word 'drugs' can also be used to refer to an illegal substance banned for consumption. However, in the medical or pharmaceutical context, drugs refer to majorly secondary metabolites, but could also be primary metabolites, sourced from various types of living forms on earth including plants, animals (terrestrial and marine vertebrates or invertebrates) or microorganisms that have the potential to treat ailments or diseases (Demain and Sanchez 2009; Dias et al. 2012; Pham et al. 2019; Abdel-Razek et al. 2020). Primary metabolites refer to intermediates or terminal products of primary metabolic pathways that are crucial for growth and survival of the living organism. Contrastingly, secondary metabolites refer to compounds not crucial for development, reproduction and/or growth of a given organism. The obvious question that riddles our minds is, 'Why do these organisms then synthesize such compounds?' Majority of these compounds synthesized act as defence molecules for these organisms especially even more crucial for living systems such as plants that cannot evade predators or stresses that easily. Some are also produced because of an adaptation mechanism of the organism to its changing environment (Dias et al. 2012). This does seem logical since surviving under constant biotic and abiotic stresses would be a key housekeeping chore for organisms throughout the span of evolution. With the assistance of modern screening programs and analysis techniques, over 1 million natural compounds have been discovered till date with approximately 20–25% of the natural products showing some biological activity. An interesting fact is that about 10% of this established activity are from natural products obtained from microbes (Demain and Sanchez 2009).

However, not all drugs under survey for pharmaceutical application or those in use are strictly natural products. Many of the drug candidates and drugs in present times are derivatised molecules from natural products using semi-synthetic methods or are completely synthesised in a chemical fashion (Demain and Sanchez 2009; Mathur and Hoskins 2017). Drugs have been further categorised as natural products, biologics referring to usually large peptides (>50 amino acid residues) or proteins, natural products botanicals (NB) for those derived from plants, natural products mimics, totally synthetic drugs, synthetic drugs but having pharmacophore based on natural products and vaccines (Newman and Cragg 2020). Based on the latest

review by Newman and Cragg (2020), >50% of drugs that have obtained clearance for use from the United States Food and Drug Administration (USFDA) in the span of 1981–2019 have been procured from natural sources either in a derivative form or as inspired from natural structures.

Based on their pharmacological effects, drugs can further be categorized as antibiotic, antifungal, antiparasitic, antitumor/cancer, anti-inflammatory and immunosuppressive, biofilm inhibitory compounds among others. Antibiotics are drugs that combat bacterial infection where some are narrow-range antibiotics having limited or specific target bacteria whereas some are broad-spectrum that exert their effects against a large range of bacteria (examples: penicillin, erythromycin, streptomycin) (Pham et al. 2019). Antifungal and antiparasitic agents, likewise, are drugs that target fungal (examples: nystatin, amphotericin B) (Pham et al. 2019) and parasitic (examples: chloroquine, metronidazole and tinidazole) (Campbell and Soman-Faulkner 2022) infections respectively. Antitumor and anticancer drugs are those that inhibit the proliferation of cancer/tumour cells or those that target these cells for killing (examples: bleomycin, actinomycin D, rapamycin) (Pham et al. 2019). Anti-inflammatory drugs are those used to reduce inflammation found in infections as well as non-infectious cases (examples: rapamycin, strepsesquitriol, diclofenac, ibuprofen) (Pham et al. 2019; Ghlichloo and Gerriets 2022). Immunosuppressive agents are used for treating autoimmune and immune-mediated diseases as well as employed to prevent the rejection of grafts and/or organs during transplantation by suppressing the immune reactivity towards the transplant (examples: rapamycin, murine anti-CD3 mAb Muromonab-CD3 (OKT3), voclosporin) (Wiseman 2016; Pham et al. 2019). Yet another category of drugs is that of biofilm inhibitors which target biofilm formation especially crucial to prevent biofilm formation as well as disrupt biofilms that have been formed on hospital devices and prevent nosocomial infections (examples: aminoglycosides, quinolones, deacylated lipopolysaccharide, actinomycin D) (Pham et al. 2019; Ghosh et al. 2020). These are just a few that have been specified; the entire list of pharmacological classification is quite enormous and can be viewed at https://www.fda.gov/drugs/investigational-new-drug-ind-application/general-drug-categories.

1.2 Historical Background of Drugs

Drugs have been used by human beings for treatment of ailments and/or diseases since time immemorial with records of such use being reported in China, India, Greece, Egypt among other nations (Taylor 2015; Pham et al. 2019; Karmakar et al. 2020; Abdel-Razek et al. 2020). These drugs were primarily of plant and/or animal origin with little or no knowledge about the mechanics of the drug involved in the treatment being rendered; the entire treatment was primarily based on empiricism (Taylor 2015). This concept, however, underwent a dramatic change in the 18th century with the foundations of pharmacology being laid down wherein research into the mechanism of drug action slowly began to gain popularity. One of the first compounds

to be isolated and studied in detail was digitalis which was extracted from foxglove by William Withering in 1780s (Taylor 2015). Similar works continued in the next century too with the contribution of Sertüner towards chemical studies of opium and extraction of morphine from opium in 1815 (Pina et al. 2010) and the pioneering works of Oswald Schmiedeberg (1838–1921) who is considered the father of pharmacology (Taylor 2015). This paved way for interdisciplinary research to contribute to drug discovery and development. Majority of the drugs being discovered, however, were still primarily from plants up to this time point.

One of the hallmarks of drug discovery came as a happenstance event in 1928 with an extraordinary discovery by Alexander Fleming that a mold, *Penicillium notatum*, secreted a compound potent enough to kill a bacteria *Staphylococcus aureus* that he had grown in a petri-dish (Fleming 1929). The active ingredient was identified as being penicillin which following isolation was used widely as an antibiotic during World War II (Demain and Sanchez 2009). This became the first microbial drug to be recognized worldwide and initiated the hunt for more such antibiotics and pharmaceuticals from microbes. This was followed by several other success stories such as the discovery of streptomycin from *Streptomyces griseus* (Waksman et al. 1946), cephalosporin-C from *Cephalosporium acremonium* (Newton and Abraham 1955) and vancomycin from *Amycolatopsis orientalis* (Geraci et al. 1956). The age-old method of fermentation was employed for the production of these antibiotics in bulk quantity. Moyer and Coghill (1946) of Northern Regional Research Laboratory, Illinois developed a specific medium for growing *Penicillium* using corn steep liquor and lactose that increased yields of penicillin. Culturing methods employed included surface culture and submerged fermentation (Raper and Benedict 1950).

Post-World War II dawned the Golden era of pharmacology with accelerated research into the discovery of newer antibiotics, analgesics, oral contraceptives, β-adrenergic blockers, anti-cancer drugs among others (Taylor 2015). This was possible due to advancements in the fields of chemistry, biochemistry, molecular biology, and computational biology which helped identify how drugs interact with proteins in the human systems thereby mediating their effects, and also helped design drugs and identify scaffold structures involved in pharmacological effects (Pina et al. 2010). Towards the closure of the 20th century, further revolutionary advancements in the fields of cellular and molecular biology including hybridoma technology and recombinant DNA technology revolutionized the pharmaceutical industry. This paved way for the bulk production of monoclonal antibodies for use as therapeutics as well as enabled eukaryotic and other recombinant therapeutic proteins to be synthesized in large amounts in prokaryotic cells such as *Escherichia coli*, eukaryotic cells such as *Saccharomyces cerevisiae* and even animal cell lines, for example Chinese hamster ovary cells (Pina et al. 2010; Pham et al. 2019).

In the last few decades several high throughput screening technologies that screen various types of libraries inclusive of small metabolites and peptide libraries including combinatorial chemistry have emerged as the key technologies influencing drug discovery (Szymański et al. 2011). The field of omics including genomics, metagenomics, proteomics, transcriptomics, and metabolomics has also in more recent times contributed immensely towards drug discovery.

1.3 General Pipeline for Drug Discovery and Development

The multifaceted process of discovering drugs and developing them to the marketable stage is divided into several phases. The first phase is that of drug discovery by single or combinatorial approaches. A typical workflow for microbial-based drug discovery is illustrated in Fig. 1. Once a drug is discovered, the next phase involves target identification and validation where understanding the disease target or infectious agent is critical for effective evaluation of drug effectiveness. This is followed by hit identification involving estimating which of the various drug candidates interact with the target followed by validation. The hits are then converted to leads by employing secondary assays to assess off-target effects, estimate solubility properties, and assay for the "absorption, distribution, metabolism and excretion" (ADME) properties of the lead molecules. Further, lead optimization is performed where the lead compounds are chemically modified, or structural analogues are queried to improve drug efficacy. The next phase includes in vivo studies beginning with animal models. Once toxicity and dosage responses are established in animal models, the candidate drug is subjected to clinical trials in humans including Phase I (20–80 participants), Phase II (100–300 participants) and Phase III (1000–3000 participant) trials. Following this, the drug is reviewed by FDA for safety and effectiveness. The last step is Phase IV where following drug approval the drug is tested on a larger sample set with more than 1000 participants being involved (Taylor 2015; Sinha and Vohora 2018; Sun et al. 2022).

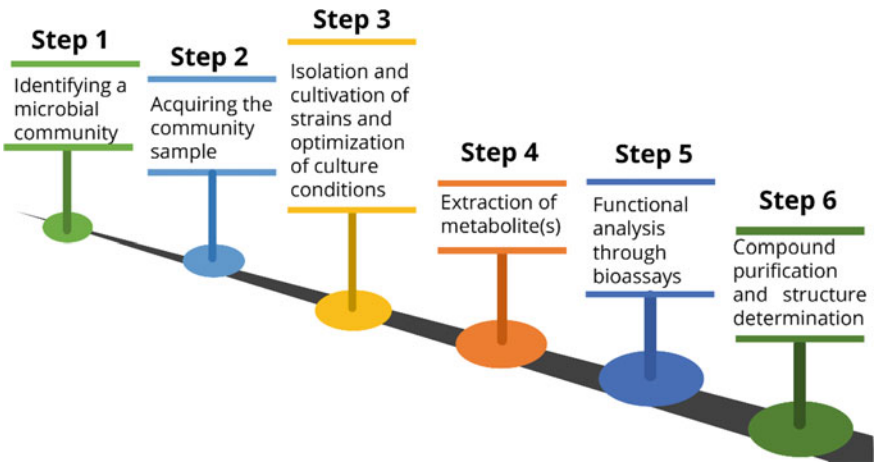

Fig. 1 Typical workflow followed during conventional drug discovery process

1.4 Need for Newer Drug Discoveries

"Two new drugs to fight superbugs available on NHS soon" (BBC news 2022), "Sunlenca® (lenacapavir) Receives FDA Approval as a First-in-Class, Twice-Yearly Treatment Option for People Living With Multi-Drug Resistant HIV" (Gilead Press Release 2022)—headlines such as these are not so common and capture our rapt attention. Indeed, as compared to the drug demand, there seems to be a large imbalance in the number of drugs emerging on the market. This surely makes one wonder regarding the reason for this large void. Some of the reasons for this void are mentioned below.

To begin with, the entire process of drug discovery and development including the various clinical trials takes several years to reach fructification. As outlined in the section above, there are several steps involved in this process. In the earlier years of drug development, these steps progressed in a linear fashion. However, considering the time consumed in the same, many pharmaceutical companies now have started running some of the phases in parallel to minimize the time taken till the final drug reaches the patient. For instance, modifications of lead drug molecules and conversion of possible drugs into more 'druggable' forms may be conducted in parallel (Taylor 2015).

Further, according to Sun et al. (2022), 90% of failure in drug development accounts from a lack of clinical efficacy, very high unmanageable toxicity, poor drug-like properties of the candidate molecules, and poor understanding of commercial and societal needs and strategic planning. Hingorani et al. (2019) also implicated a lack of efficiency in disease indication as the foremost cause of delayed drug development and high failures in the same. This necessitates the discovery of newer drug candidates with better potential especially those that are natural in origin.

The need for newer drug discoveries, especially for antibiotics and antitumor or anticancer drug, has also arisen due to the increased drug resistance found in bacterial and human cells to these agents respectively. In the case of antibiotic resistance, the problem is not restricted to just resistance to one drug but the observation of resistance to multiple drugs in bacteria. These superbugs are evolving at an alarming rate with majority of them being pathogenic in nature (Demain and Sanchez 2009). Examples of these include methicillin-resistant *Staphylococcus aureus* and multidrug resistant *Pseudomonas aeruginosa* among many others (van Duin and Paterson 2016). This resistance in bacteria to drugs has resulted from either natural resistance or acquired resistance that the bacteria have developed to combat the drug effects. These include reduction in outer membrane permeability, expression of efflux pumps that pump the drugs from the interior of the bacterial cell into the extracellular side of the cells, modifying the target of the drugs and inactivation of drugs (Nikaido 2009; Reygaert 2018). Similar evasion and tolerance mechanisms have also been exhibited by cancer cells while developing resistance to single and multiple drugs. These include increased drug efflux, reduced inflow of drug, drug compartmentalization, changes in drug metabolism, alterations in drug targets, increased resistance to apoptotic cell death and increased repair of DNA damage induced due to the drug (Mansoori et al. 2017; Emran et al. 2022).

Further, there is a need for novel classes of anti-infective compounds to be developed and released at the market level to combat newer emerging diseases. Many of these emerging infectious agents are those that the host has never encountered before, thereby, posing an even greater challenge to the host's immunity (Demain and Sanchez 2009). Furthermore, one cannot neglect the fact that majority of drugs available in the market today have side effects too such as vomiting, diarrhoea, stomach upsets, rashes, and allergic reactions among others.

Against this background, one can conclude that there is a monumental need for discoveries of newer drug molecules with better therapeutic potential and preferably lesser or absolutely no side effects. The promise of high throughput technologies and combinatorial compound libraries to deliver a diversity of new drugs within a short period of time has not been fulfilled majorly owing to unrealistic expectations (Abdelnasser et al. 2012). This has resulted in a rekindling of interest in nature-derived products for their therapeutic potential. Natural products offer special advantages with respect to features when compared with synthetic molecules. These features include exhibiting massive scaffold diversity, high molecular mass, structural complexity, higher molecular rigidity to name a few. Drugs derived from natural origin have the advantage of being optimized by evolution in terms of structure and function that permit and favour better interaction with biological molecules which is the major basis of drug action (Atanasov et al. 2021). Natural products have also been found to show lesser side effects as compared to their synthetic counterparts (Mathur and Hoskins 2017).

Amongst the various natural sources of drug candidates, microbes provide the largest contribution both in terms of numbers as well as diversity of drug molecules. Therefore, it is not surprising that a lot of present-day discovery programs are channelizing their resources back towards bioprospecting microbes for new drug molecules.

2 Current Approaches for Microbial-Derived Drug Discovery

2.1 Unique Habitats and Novel Metabolites

For the discovery of metabolites having unique and novel pharmaceutical properties and effects, the need of the hour is to identify and screen novel bio-resources or revisit the already explored habitats or the known resources. Exploration of existing known habitats by new cultivation methods combined with genomics has shown to yield more fruitful results. Further, expeditions to not so explored habitats such as deep ocean hydrothermal vents have hinted to us the existence of unique ecosystems which can offer novel drugs. This is possible due to robotic sampling vehicles, novel cultivation methods, or extensive use of recombinant DNA technology

combined with metagenomic technology. Various habitats such as soil, aquatic sediments, deserts, hot springs, marine habitats including abyssal plains, hydrothermal vents, marine organisms (corals and sponges), mangroves and coastal ecosystems have been tapped for novel metabolites by isolating associated microorganisms. Microorganisms useful for drug discovery can be free-living organisms, may exist as symbionts, or can be endophytic (Pereira 2019). Tapping the potential of these habitats for novel metabolites requires unique approaches. Discussed below are the conventional and emerging cultivation-dependent methods and their limitations.

2.2 Conventional and Emerging Cultivation Techniques

2.2.1 Conventional Cultivation Techniques

The conventional techniques for bioprospecting microbes for drugs include first enriching and culturing the microorganisms under investigation from varied sources. Enrichment culture techniques are routinely employed microbiology techniques which are used for increasing low counts of the targeted organisms to detectable and cultivable levels (Bari and Yeasmin 2022). Enrichment employs the principle of natural selection, in which a collected sample having a heterogenous population of microorganisms is transferred into a medium having well-defined chemical composition. This inoculated sample is then allowed to proliferate under controlled conditions which are either the primary requirements for growth or natural conditions from where the sampling was done. These include temperature, light, oxygen or carbon dioxide supply, pH, pressure, etc. Such naturally simulated conditions may only favour the growth of one or more particular type of microbes. The enrichment culturing can be done in submerged (liquid medium) or in solid-state fermentation where solid substrates are used as either raw or processed material.

Once the microorganisms' presence is detected, the sample can be subjected to isolation on agar plates using the same nutritional and environmental conditions provided during enrichment. Often to reach this stage, the enriched sample may need to be serially diluted multiple times before being plated onto agar medium. In the case of microalgae, serial dilution may itself suffice without isolation on a solid plate.

However, the conventional enrichment cultivation approach suffers major limitations due to one or multiple reasons leading to "the great plate count anomaly"—a term introduced in 1985 by Staley and Konopka. It refers to the anomaly in numbers of colonies successfully grown on media as against the numbers present in the environmental sample. The "great plate anomaly" phenomenon is widely observed in marine ecosystems where only 0.01–0.1% of marine oceanic bacterial cells could be isolated as colonies by standard bacterial isolation techniques (Connon and Giovannoni 2002).

This anomaly is attributed to the lack of knowledge on specific physiological needs, oligotrophic requirements, overgrowth of fast growers or too slow growth,

poor cell density in the sample, dormant stage of an organism, special chemical or physical needs, requirement of solid surface for attachments, loss of interactions between interdependent microorganisms and complexity of microbial communities. Together these contribute to "uncultivable" communities. It has been understood that of 40 known prokaryotic phyla, only 50% could be cultured in the laboratory. The remaining uncultured organisms are believed to offer possibility of novel metabolites of medical interest.

2.2.2 Emerging Cultivation Techniques

To harness this unexplored possibility from uncultivable microbes, novel cultivation techniques have emerged with the potential to recover the uncultivable organisms several folds greater than the conventional approaches. Emerging cultivation methods addresses the limitations of conventional cultivation strategies by permitting the growth of a wide range of microorganisms (Lozada and Dionisi 2015).

These novel cultivation methods are high throughput techniques able to handle several samples and having the ability to isolate a significant proportion of microorganisms from natural environments. However, these methods require basic biology knowledge along with information and understanding of the ecological background of the sample. Detailed below are some of the novel cultivation methods that have been developed and employed successfully.

Dilution to Extinction Method (DTE)

By this approach, ubiquitous, however, not so easy to cultivate or isolate bacteria and archaea can be successfully isolated. In one of the studies, Rappé et al. (2002) isolated marine bacterioplanktons by inoculating freshly obtained sample from Oregon coast seawater into microtiter wells by dilution (to achieve approximate 1–20 cells per well). The authors employed media prepared by adding necessary salts and a defined formulation of organic carbon compounds in sterile Oregon coast sea water and incubated the same at 15 °C for 23 days either under dark or 14 h light/10 h dark cycle. The technique is assisted with epifluorescence microscopy technique. Before cultivation it allows rapid detection of number of cells present in the sample whereas during cultivation it allows detection of growing cells (Rappé et al. 2002).

Benítez et al. (2021) have also reported the use of DTE methodology for isolating non-filamentous bacteria associated with marine sponges having the capacity for production of cytotoxic compounds that can be antineoplastic.

Diffusion Chambers

Diffusion chambers makes use of an apparatus in which microbial cells from a sample are combined with agar layer which in turn is bracketed on either side by semi-permeable membrane of porosity ranging from 0.2 μm to 0.3 μm. The assembled chamber is incubated in natural environment for one to several weeks. The membrane serves to isolate the caged cells from the environment; however, it allows the passage of growth factors and nutrients from the media. This favours the growth and thus isolation of otherwise uncultivable bacteria from diverse environments (Kaeberlein et al. 2002; Jung et al. 2021).

Microbial Trap Technique

Microbial trap technique is an alteration of the diffusion chamber technique to capture unusual and rare actinomycetes in situ. This is possible due to their ability to produce hyphae and the penetrability of these hyphae through solid environments. It involves a trap formed by sandwiching sterile agar between two semi-permeable membranes with the base having a pore size of 0.2 to 0.6 μm and the top membrane which has a size of 0.03 μm. The trap assembly is incubated on top of the microbial source where the hyphae of filamentous microbes selectively penetrate through the membranes and colonize the agar layer (Gavrish et al. 2008).

Microdroplet Encapsulation

This is yet another novel high-throughput cultivation method which combines encapsulation of single cell with flow cytometry. Sufficiently diluted sample is mixed with molten hydrogel such as agarose followed by emulsification to form microcapsules usually of size 50 and 80 μm. The microcapsule thus formed is expected to entrap a single cell, which when cultivated in natural like media and simulated growth conditions leads to the formation of microcolonies. The porous nature of agarose permits the diffusion of nutrients and signalling molecules into the capsule providing nutrition to the expanding colony while simultaneously permitting the waste metabolites to diffuse out. Growth of colonies can be detected using flow cytometry. Microcapsules can easily be separated, and the cultivated microbes can further be grown in rich media in microtiter plates and can thus be isolated with ease. Using the microdroplet encapsulation, several bacteria and fungi have been isolated from natural samples (Zengler et al. 2005).

In Situ Cultivation by Tip (I-tip)

Jung et al. in (2014) developed an in-situ cultivation method which targeted isolating symbionts from aquatic invertebrates. This approach uses a commercial device

comprised of a micropipette tip. The upper part of the tip is blocked with the help of an adhesive to prevent entry of contaminants. The pointed end of the tip is filled with agar layers and microbeads which is directly positioned onto the surface of the target environment. The microbead layer does not permit the entry of organisms larger in size while the layer of agar supports the growth of microbes due to the permeability of nutrients entering from the natural habitat. Using this approach uncultivated sponge-associated bacteria have been efficiently isolated (Jung et al. 2021).

Isolation Chip (iChip)

Automation of first-generation high throughput methods have further led to emergence of second-generation methods. Isolation chip (iChip) is one such example. iChip device is composed of numerous miniature diffusion chambers ranging in counts of hundreds. These chambers are inoculated with single cells, followed by incubation in an environment that mimics its natural niche conditions which can lead to several fold increased isolation of uncultivable microorganisms (Palma Esposito et al. 2018).

Hollow-Fibre Membrane Chamber (HFMC)

The HFMC approach is quite like iChip wherein a chamber for cultivating the microorganisms in isolation is created using a piece of hollow-fibre polyvinylidene fluoride (PVDF) porous membrane. A single HFMC system comprises of approximately 48- to 96-chamber units made up of porous hollow membranes of approximately around 30 cm in length, 67–70% porosity. Further, these chambers are equipped with syringes for injection and sampling. The environmental sample is first serially diluted, then injected into a chamber, which is then placed in a natural system or environment simulated in laboratory. Due to the porosity of the membrane, free exchange of nutrients, signalling molecules and metabolites can occur while preventing the movement of microbes thereby resulting in establishment of pure cultures of different types in each chamber (Aoi et al. 2009).

Once the enriched or unculturable microorganism is recovered, the organism is further transferred onto agar plates and multiplied. However, it may show the presence of mixed cultures or culture surviving only till limited division cycles. The cultivated microbes are then subjected to bioprospecting using analytical techniques or other appropriate methods to determine the presence of unique or significance secondary metabolites of medical importance.

3 Role of Metagenomics in Drug Discovery

3.1 Metagenomics Approach for Bioprospecting

Despite timely troubleshooting combined with recent advances, majority of the microorganisms from the natural world are still unculturable and marine microorganisms top this list. The hallmark studies of Woese and Pace combining molecular biology and molecular ecology concepts revolutionized the understanding of microbial diversity. The advent of metagenomics brought to limelight the potential of different ecosystems in harbouring a diversity of microorganisms with unknown functions which hitherto had remained masked and undiscovered due to the tag of uncultivability (Abdelnasser et al. 2012; Lozada and Dionisi 2015). This approach relies on the information obtained from DNA sequences regarding the genetic and, therefore, molecular diversity of a given organism or groups of organisms growing in a particular niche. This bypasses the need of cultivation by directly extracting the DNA from the natural niche (Lozada and Dionisi 2015).

Metagenomics is the study of community genetics based on next generation sequencing technology. This methodology aids researchers in learning about the variety, purposes, and evolution of uncultivated microorganisms found in a variety of habitats or ecosystems. This upcoming approach and its coexisting technology can reveal the capacities of microbial communities which drive the earth's energy and nutrition cycles, maintain the health of its people, and determine life's elaboration (Chopra et al. 2020). Metagenomics also attempts to improve our understanding of the interactions between microbes to enhance people's health, food production, and energy production. Although originally metagenomics was developed for understanding microbial ecology, it soon has become a tool for bioprospecting new habitats as well as earlier explored habitats for microbes that could produce novel drugs.

The metagenomics approach commences with genomics DNA extractions from the sample as community DNA, followed by construction of metagenomic libraries leading to library screening using either the function-driven or sequencing-based approach for identifying genes or gene clusters that will implicate biosynthetic pathways of novel drug molecules or drugs with better potential (Abdelnasser et al. 2012). Outlined below are each of these techniques in detail.

3.2 Techniques Involved in Metagenomics

3.2.1 Extraction of DNA From a Selected Environmental Niche

Selection of the environmental niche for bioprospecting is a prerequisite to ensure maximum bioprospecting for novel natural products. For instance, if the aim of the metagenomics approach is to screen for anti-tumour agents, it would be advisable to bioprospect in marine ecosystems as compared to terrestrial ones since it has been

shown that terrestrial systems offer low diversity of such agents (0.01%) as compared to marine ecosystems that show higher diversity (1%) (Abdelnasser et al. 2012). Microbial communities being targeted for drug discovery are no longer restricted to terrestrial and marine ecosystems. Metagenomics has made it possible to even bioprospect inaccessible and difficult to culture microbes from intestinal tract of humans and other animals.

Once the environmental niche has been carefully selected to yield maximum bioprospecting success, the next step is that of extracting the environmental DNA. Two approaches are employed for the same, viz. direct, and indirect lysis methods. In the direct lysis method, various lytic strategies such as freeze-thaw, ceramic bead beating, and use of lytic buffers aids the lysis of microbial cells resulting in extraction of DNA into the lysis buffer. This is followed by purification of extracted DNA from the remnant cellular components. However, this method faces a few limitations such as the inability to discriminate between genetic material of different origins and co-extraction of humic acid. This results in a high fraction of extracted environmental DNA containing unwanted sequences, thereby, complicating the analysis. Contrastingly, the indirect method employs separation of microbial cells from the remaining cells in the environmental sample first which is then subjected to lysis to achieve extraction of DNA. This offers the advantage of eliminating unwanted DNA sequences in the extracted DNA. Another way for circumventing low DNA yield in the sample and unwanted sequences is to employ amplification strategies for whole genome sequences (Abdelnasser et al. 2012; Wydro 2022). One cannot dispute the fact that this step is the most critical step in bioprospecting a niche for newer drugs and requires improved methods of DNA extraction to maximize recovery of diversity of genes and functions especially from high humic content soils, low nutrient content samples and extremophilic conditions.

3.2.2 Construction of Metagenomic Libraries

The extracted and isolated environmental DNA is subjected to fragmentation followed by creation of clones by inserting the fragmented DNA into vectors such as plasmids, cosmids and fosmids, thereby, generating metagenome libraries. The first choice of vectors is that of plasmids; however, plasmid-based vectors are riddled with the limitation of small size insert restriction. For bioprospecting, one may require querying with larger inserts. Hence, it is desirable to use cosmids (35–40 kbp), fosmids (35–40 kbp) and bacterial artificial chromosome (BAC) vectors (100 kbp) for library preparation. Once the libraries have been created, the next step involves maintaining them in a suitable host, namely *Escherichia coli*, *Pseudomonas* spp. and *Streptomyces* spp. among others (Abdelnasser et al. 2012; Danhorn et al. 2012; Lozada and Dionisi 2015).

3.2.3 Screening of Metagenomic Libraries

Mining of metagenomic libraries for biological data employ two key approaches, viz. function-driven and sequence-driven analyses. Function-driven approach initially screens the clones expressing a trait of interest following which the selected clones are subjected to sequencing and biochemical analyses. This approach swiftly locates clonal fragments with various commercial and medicinal uses by concentrating on natural compounds with relevant activity. In contrast, sequence-driven analysis searches the metagenomic libraries for clones of regions of interest by creating PCR primers for hybridization or probes from DNA sequences that are conserved (Schloss and Handelsman 2003).

Function-Driven Analysis

By using a function-based approach, it is possible to find novel medications whose properties cannot be predicted only by DNA sequence (Lam et al. 2015). It involves phenotypic-based detection techniques that make use of reagent dyes and enzyme substrates, which are frequently connected to chromophores. The products are detected with the help of spectrophotometric analysis following the reaction of compounds of the individual metagenomic clones with specific reagents (Quintero et al. 2022). Functional analysis also entails assaying for the anti-cancer, anti-proliferative, anti-inflammatory, antibiotic potentials among others of the extracts derived from the clones. These assays are performed in respective susceptible cells to understand the drug potential of the novel compounds being identified (Abdelnasser et al. 2012). Functional screening-based methodologies maximize the probability of discovering new classes of genes that encode for activities which have been characterized before or those that are novel in nature. This is possible since there is no bias involved in the screening process due to the absence of requirement of sequence-dependent information.

There are few drawbacks that are observed in this metagenomic approach of searching novel genes or gene products. Choice of host and selection of appropriate methodology for the screening of functions is one of the major limitation of metagenomic clone libraries as the expression of gene/s of interest in the host is affected owing to differential preference in codon usage, lack of gene expression due to failure of recognizing promoter region, reduced transcription rate of genes due to non-recognition of transcription factors and altered end product due to improper protein folding and failure in exporting gene products out of host cell. It was also observed in few cases that foreign gene products were toxic to host (Taylor 2015). Cloning of target genes is also affected due to presence of large quantity of target genes in the environmental samples and longer target gene and library size. Also, it is difficult to search an appropriate host that will be able to conduct correct post-translational modifications and expression of single Biosynthetic Gene cluster (BGC) (Alam et al. 2021).

Another problem with functional based approach is with respect to the intracellular accumulation of the desired products irrespective of the choice of host system. This problem can be resolved by enabling the release of the product from the cells with the aid of lysis compounds such as detergents which allows better secretion of compounds without affecting the native conformation (Johnson 2013). This strategy can be specifically employed to screen for novel compounds using the microtiter plate approach co-incubated with specific substrate/s. Additional aid of specific robots along with plate readers for the handling of liquids and microbial colonies enhances the efficiency of this process (Ngara and Zhang 2018).

Sequence-Driven Analysis

In sequence-based metagenomic approach, the detection of a particular gene encoding product of interest is achieved by using specific primers of the genes and its amplification by polymerase chain reaction or by hybridization techniques using specific DNA probes. Probe design is based on conserved sections of characterized genes or proteins. This method involves searching existing metagenomic data sets and/or nucleic acid databases for regions of interest instead of using heterologous expression of any genes or clones. This is followed by synthesis of target genes using chemical-based methods (Bayer et al. 2009).

Sequence-based metagenomics includes collection of genomic information of the microbes without culturing them. High throughput sequencing platforms like Sanger's sequencing, Illumina/Solexa, Roche 454 pyrosequencing, PacBio, Oxford Nanopore and others are used to generate high quality sequence reads which are analysed for the prediction of their putative functions using bioinformatic tools. The introduction of cloning-independent "next-generation sequencing technologies", for instance pyrosequencing, has made the need for metagenome library creation redundant especially if the sequence-based approach is to be employed (Danhorn et al. 2012). Table 1 summarizes the sequencing techniques available and their basic principles of working.

Vast number of datasets obtained from sequencing projects from diverse environments is now catalogued in various databases which are publicly available. Sequence-based approaches will be more insightful about microbial communities as trends in environmental sequences begin to appear. Some specific studies use a target gene as an anchor to identify the clones for the further generation of metagenomic libraries. With the help of PCR-based techniques, target genes are amplified and sequenced to construct and screen metagenomic libraries. Anchors used to construct such libraries are often an rRNA gene or a BGC so that researchers can easily identify the clone of interest.

The sequence-driven approach itself can be categorized into three types, viz. metabarcoding, whole metagenome analysis and single-cell genome analysis.

Table 1 Sequencing techniques used for metagenomic analysis

Technique	Principle	Read lengths	Reference
Sanger's sequencing method	DNA sequencing based on chain termination method due to inability of DNA polymerase to elongate a chain following incorporation of a dideoxynucleotide	400–900 bp	Heather and Chain (2016), Totomoch-Serra et al. (2017)
Illumina/Solexa	Sequencing by synthesis involving immobilization of DNA fragment on a surface followed by PCR amplification to generate clusters of amplified DNA which are sequenced by labelled reversible terminators	2 × 150 bp, 300 bp	Sharpton (2014), Escobar-Zepeda et al. (2015)
454/Roche Pyrosequencing	Application of emulsion polymerase chain reaction (ePCR) to generate clones of DNA fragments within microscopic beads which are individually subjected to pyrosequencing. All four nucleotide triphosphates are sequentially added; nucleotide incorporation causes pyrophosphate release, detection of which is coupled to sulfyrylase and luciferase systems that emit light	~100–150 bp	Escobar-Zepeda et al. (2015), Heather and Chain (2016)
Ion Torrent	Microscopic beads containing clonally amplified DNA are loaded into picowell plate, followed by nucleotide triphosphate incorporation. The release of protons results in change in pH which is then measured with the aid of a "complementary metal–oxide–semiconductor" (CMOS) technology	~400 bp	Sharpton (2014), Escobar-Zepeda et al. (2015), Heather and Chain (2016)
Oxford Nanopore	The technology employs a nanopore (nanoscale protein pore) embedded in an electrically resistant membrane which is then positioned in an electrolyte solution with constant current applied across the membrane. Single stranded DNA is allowed to pass through this pore. When a single nucleotide exits on the other side of the membrane the change in voltage is measured which is distinctive for each of the four nucleotides thus permitting accurate sequencing of a DNA strand	>100 kb	Singh and Roy (2020), Wang et al. (2022)
Single molecule real time (SMRT)/ PacBio sequencing	It involves detection of fluorescent signal emitted on incorporation of a fluorescently labelled nucleotide by a DNA polymerase which is bound to a Zero Mode Waveguide well	>10 kb	Escobar-Zepeda et al. (2015), Ardui et al. (2018)

Metabarcoding

Metabarcoding approach, also referred to as amplicon sequencing or targeted sequencing, entails targeting a particular region of the genome from the environmental community as an indicator of diversity and phylogenetic relationships. SSU rRNA genes such as 16S for prokaryotes and 18S for eukaryotes and internal transcribed spacer (ITS) regions for both prokaryotes and eukaryotes are most frequently used as markers for identification. Specific primers designed to target these regions are used to explicitly amplify only these regions followed by diversity analysis. Additionally, from bioprospecting angle, gene clusters that have been identified as being involved in the synthesis of potential drug molecules can also be used in the targeted approach. Following marker-based amplification and sequencing, the generated reads are subjected to basic bioinformatics pipeline including demultiplexing of barcoded samples, performing pair-end assembly, elimination of chimeric reads, checking quality of reads and performing read filters, clustering followed by sequence alignment with a reference database (Trindade et al. 2015; Maghembe et al. 2020; Francioli et al. 2021). However, this approach does have a limitation of low resolution of species identity due to relative short length of the markers.

Whole Metagenome Sequencing (WGS)

Whole metagenome sequencing (WGS) refers to sequencing the entire environmental DNA obtained from the niche under study. With the help of shotgun sequencing, WGS has become a faster tool as compared to earlier sequencing that required cloning as a prior step. The entire metagenome is subjected to random fragmentation to produce fragments ranging in size from 2 to 300 kb. These fragments are then sequenced followed by assembly and binning using computerized programs to identify overlaps and build contigs. Using gene annotation tools, putative gene functions can be identified that may correspond to novel metabolite synthesis (Sharpton 2014; Pereira 2019). Since the amount of DNA involved in whole metagenome analysis is enormous, it is but evident that it would contribute to greater insights into niche diversity (microbial and functional). It can also aid in elucidation of candidate pathways involved in novel secondary metabolite synthesis as well as improve the understanding of microbial interactions amongst themselves as well as with the surrounding environment that might influence improved and higher metabolite production (Maghembe et al. 2020).

Single-Cell Metagenomics (SCM)

SCM performs analysis on single cells separated from like cells as well as from their attachment to the surrounding natural matrix. Additionally, cell sorting techniques can be employed to identify and distinguish target cells from associated unwanted populations through cell-sorting techniques. It provides a new avenue for working with uncultured organisms and bioprospecting them at individual levels (Lozada and Dionisi 2015; Alam et al. 2021).

Overall, both the approaches of metagenomics (Fig. 2) highlight the potential of discovering essential genes from environmental samples and provides opportunities for developing novel drugs using these strategies.

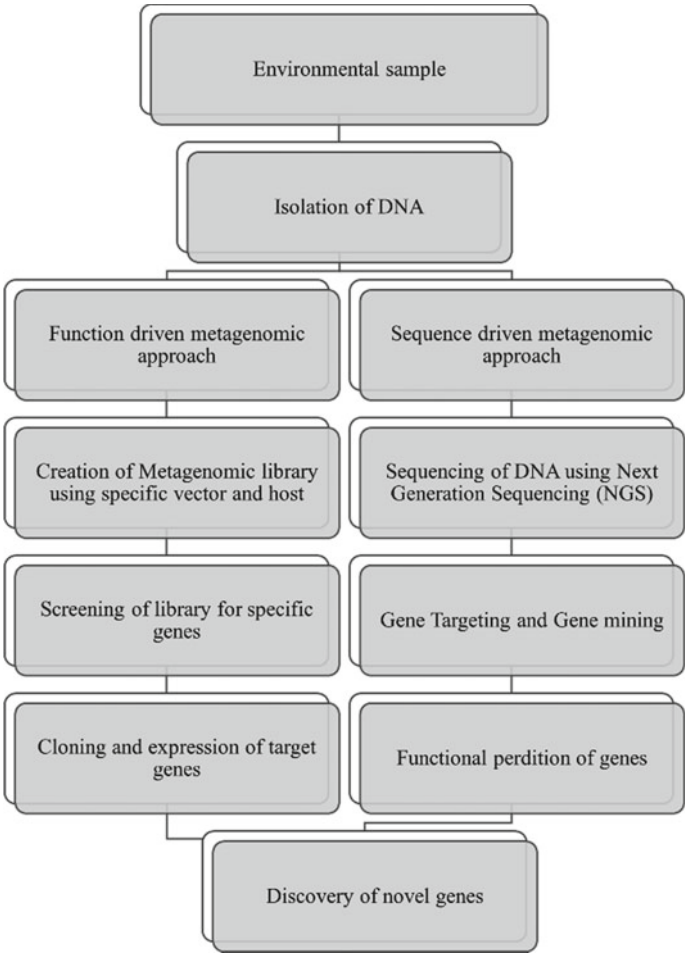

Fig. 2 Overview of gene discovery by metagenomic approach

3.3 *Metagenomics for Bioprospecting Drug Synthesizing Potential of Microbes*

3.3.1 Genome Mining

The process of querying and finding candidate genes that are active in the biosynthetic pathways for natural products or drugs from metagenome data is referred to as genome mining. This process involves the use of bioinformatics approaches to identify natural product gene clusters which encode for previously uncharacterized natural compounds from a known class of biosynthetic pathways or to predict an entirely new class of gene cluster/s involved in new biosynthetic pathways culminating in the identification of novel natural products. The process also involves analysing gene sequences that encode for enzymes encoded by genes and determination of gene products following experimental analysis (Albarano et al. 2020; Atanasov et al. 2021).

Identification of these biosynthetic genes is facilitated by screening for two characteristics namely,

(1) Presence of gene clusters and
(2) Presence of genes expressing enzymes (for instance, non-ribosomal peptide synthetases (NRPSs) and polyketide synthases (PKSs)).

A group of genes found in the genome which are responsible for expression of proteins involved in the metabolic pathway producing products which are found to be chemically diverse is referred to as "biosynthetic gene clusters" (BGCs). These diverse chemical products are reported to exhibit varied bioactive properties ranging from antibacterial to cholesterol reducing properties to being an insecticide (Santana-Pereira et al. 2020).

There are many classes of BGCs which can be structurally classified. Bacteriocins, terpenes, β-lactam, NRPS, PKS, indole, and furans are some examples of the many (Belknap et al. 2020). Amongst these, NRPS and PKS are very widely used as targets to discover natural products since they are classically known to be involved in synthesis of antibiotics, immunosuppressants and molecules exhibiting excellent pharmaceutical value.

The polyketide synthases are enzyme complexes which catalyze the condensation of CoA fatty ester precursors to create the carbon skeleton of polyketide molecules (Gomes et al. 2013). Polyketides are metabolites found in nature that are characterized by numerous β-hydroxyketone or β-hydroxyaldehyde (–$H_2C(=O)CH_2CH(OH)CH_2C(=O)$–) functional groups. These compounds form the backbone of several anticancer, anticholesteremic, antifungal compounds as well as antibiotics, immunomodulators and antiparasite agents (Gomes et al. 2013). Streptomycin is a classic example of a polyketide.

Identification of gene clusters exhibiting enzyme activity comprises of finding the conserved domain sequence for putative function of the genes. Condensation domains from NRPS clusters and ketosynthase domains in PKS clusters are such

examples used as suitable targets for genomic analysis due to their conserved status. They may be exploited to differentiate between various NRPS/PKS pathways (Chen et al. 2020). An example of such a gene cluster was described by Qiao et al. (2011) found a BGS in the genome of *Aspergillus clavatus* that correlated to production of cytochalasin E and K. NaPDoS2, the second-generation Natural Product Domain Seeker, rapidly screens genomic, metagenomic or PCR amplicon-based sequences to identify and categorize condensation and ketosynthase domains.

Genome mining in the last decade has developed to the extent where it now facilitates better identification and prediction of SMBGCs "secondary metabolite biosynthetic gene clusters" through computational modes for in silico synthesis of putative predicted chemical drugs. Integration of numerous databases and web tools have been performed to improve the implementation of this approach. The following web tools are widely used in metagenome-based drug discovery.

(i) antiSMASH, "antibiotics and Secondary Metabolite Analysis SHell", (available at https://antismash.secondarymetabolites.org/) is a completely automated pipeline used for mining fungal and bacterial genome data for identification, annotation and SMBGCs analysis. It integrates and uses different tools for analysis such as NCBI BLAST+ , Muscle 3, HMMer 3, PySVG, FastTree and JQuery SVG. The most recent updated tool, antiSMASH v5, contains an elaborate group of manually curated and authenticated discovery guidelines for >50 types of BGCs (Medema et al. 2011, Palaniappan et al. 2020).

(ii) PRISM; "PRediction Informatics for Secondary Metabolomes", is an tool freely available on the web (available at https://bio.tools/prism_3), that performs prediction of metabolomes of secondary metabolites from the genomes/metagenomes. It predicts genes encoding natural product structures or BGCs from microbial genomes using BLAST for sequence homology, MUSCLE for Multiple Sequence alignment and HMMER tools for hidden Markov model generation and searches, and generates combinatorial libraries of structure predictions using Chemistry development kit and RDKit (Skinnider et al. 2015). This approach makes it possible to compare known natural drugs versus new ones. Most notable example is that of Malacidins and Humimycins which have been discovered using synthetic bioinformatics approach (Atanasov et al. 2021).

(iii) "Integrated Microbial Genomes Atlas of Biosynthetic gene Clusters" (IMG/ABC), is the world's biggest database which is publicly available and that contains experimentally determined as well as predicted SMBGCs. Since its launch in 2015, it has been updated in 2020 to incorporate the power of antiSMASH v5 for better prediction of BGCs within the IMG database. Its exceptional feature is that it incorporates both computationally predicted and empirically validated (through rigorous experimentation) SMBGCs in genomes and additionally also includes scaffold bins obtained from metagenome. It, thus, can reveal BGCs in populations and rare taxa which cannot be cultured (Palaniappan et al. 2020).

(iv) eSNaPD (environmental Surveyor of Natural Product Diversity) found at http://esnapd2.rockefeller.edu has been created to evaluate "Natural Product Sequence Tags" derived from metagenome using PCR. Profiles of diversity of BGCs hidden inside the (meta)genome is created by sequence comparison of sequence tags to referenced data of characterized BGCs in the eSNaPD data analysis pipeline (Reddy et al. 2014; Hover et al. 2018).

3.3.2 Novel Approaches: "Synthetic Bioinformatic Natural Product Approach" (Syn-BNP)

Once the genome of the microorganisms from the environmental samples is sequenced through high throughput techniques, the reads then are subjected to functional analysis pipeline. Steps in the pipeline may include BLAST+ sequence based searches for homology in gene cluster databases; further downstream analysis in this includes multiple sequence alignment, conserved domain analysis, phylogenetic tree construction, functional annotation which can be performed with the aid of numerous biological databases and bioinformatics tools (Araujo et al. 2018).

In Syn-BNP approach, functional annotation analysis (based on sequence homology) is used extensively to predict the putative function of the genes encoding enzymes in the SMBGCs. Further, metabolic pathway elucidation, bioproduction targets, biosynthetic routes and probable chemical reactions between substrates and the products formed may be performed manually or through computational prediction models like those used in biotransformation. Popular cheminformatic tools for the same are GEM-Path, BNICE.ch, ReactPRED, enviPath, NovoStoic, RetroPath2.0, Transform-MinER and novoPathFinder. These tools are used in recreating biosynthetic pathways using chemical reaction rules (Hafner et al. 2021).

In conventional natural product discovery approach, the natural product is purified through heterologous gene expression of the SMBGCs in a suitable host cell. However, in Syn-BNP the natural product is chemically synthesized using synthetic chemistry due to the silent nature of most natural product BGCs in the laboratory (Wang et al. 2022).

3.3.3 Success Stories of Drug Discoveries Using Metagenomics

Metagenomics has indeed revolutionized the process of drug discovery adding a new dimension to this field which desperately needed a boost. Table 2 provides a glimpse into some of the success stories of metagenomic-based drug discoveries.

The latest approach using Syn-BNP has also resulted in some monumental success stories. The study by Wang et al. (2022) stated the application of Syn-BNP to discover Lapcin (Fig. 3) which shows inhibitory effect on both topoisomerases I and II, exhibiting potent activity in distinct cancer cell lines. Their work entailed cloning of DNA from soil that would contain PABA-specific adenylation-domain sequences since they predicted that this sequence would result in a product

Table 2 Metagenomics-based drug discoveries

Drug	Source	Activity	Reference
Malacidin	Soil microbes probably unculturable *Myxobacterium* spp	Calcium dependent antibiotic	Hover et al. (2018)
Flouroquinolone	Polluted aquatic environment	–	Boulund et al. (2017)
Minimide	*Didemnum molle*	–	Donia et al. (2011)
Erdacin	Desert sand soil	Pentacyclic Polyketide activity not found	King et al. (2009)
Mycalamide And Pederin	Non culturable *Psuedomonans* linked with *Paederusfuscipes*	Anti-cancer	Singh and Roy (2020)
Bryostatin	Bacterial symbiont *Bugula neritina*	Anti-cancer	Hildebrand et al. (2004), Singh and Roy (2020)
Indirubin Violacein Deoxyviolacein Turbomycin A And B	Soil metagenome	Antibiotics	Lim et al. (2005), Singh and Macdonald (2010)
Diazepinomicin and Eco-7942	*Micromonospora* and *Streptomyces* spp	Anti-cancer	Singh and Macdonald (2010)
E-637 and E492	*Streptomyces* spp.	Anti-helminthic drug	Singh and Macdonald (2010)
Ecteinascidin ET-743, Polytheonamides Calyculin A	Bacterial endosymbiont of marine tunicates	Anti-cancer	Trindade et al. (2015), Singh and Roy (2020)
Psymberin	Different marine sponges	High cytotoxicity and selective antitumor polyketide	Trindade et al. (2015)
Divamides	Bacterial endosymbiont of marine tunicates	Anti-HIV	Smith et al. (2018)
Lapcin	Soil microbe	Dual topoisomerase inhibitor I/II	Wang et al. (2022)
Crocagins and Crocadepsins	*Chondromyces crocatus* Cm c5	Moderate inhibitory activity on the interaction between RNA and CsrA of *Yersinia pseudotuberculosis*	Surup et al. (2018)

(continued)

Table 2 (continued)

Drug	Source	Activity	Reference
Humimycin	Human gut Microbiome	Antibiotic active against MRSA	Chu et al. (2016, 2018)
Novel antitumor polyketides	Microbial symbionts of sponges *Discodermia dissolute, Pseudoceratina clavata* and *Theonella swinhoei*	Anti-tumor activity	Ferrer et al. (2009)

having N-acylated PABA and thiazole. Following cloning and identification, they used chemical synthesis to produce the final lapacin molecule. Their study also showed the advantage of combining metagenomics, bioinformatics prediction tools and databases, and total chemical synthesis of the compound to discover a novel metabolite from those BGCs which are uncharacterized. This opens opportunities to develop new pipelines and incorporate the use of Artificial Intelligence in discovering new drugs.

Another successful example of genome mining using the above approach was reported by Ueoka et al. (2022). Their study discovered structurally unprecedented polyketide alkaloids named janustatins with uncommon biological activity against cancer cells. Janustatins were found to kill the cells at sub-nanomolar concentrations but in a delayed and synchronized fashion. Their approach involved chemical prediction of PKS cluster from the genome of rhizhospheric marine bacteria *Gynuella sunshinyii* which had hitherto not been acknowledged with the aid of an *in silico* approach.

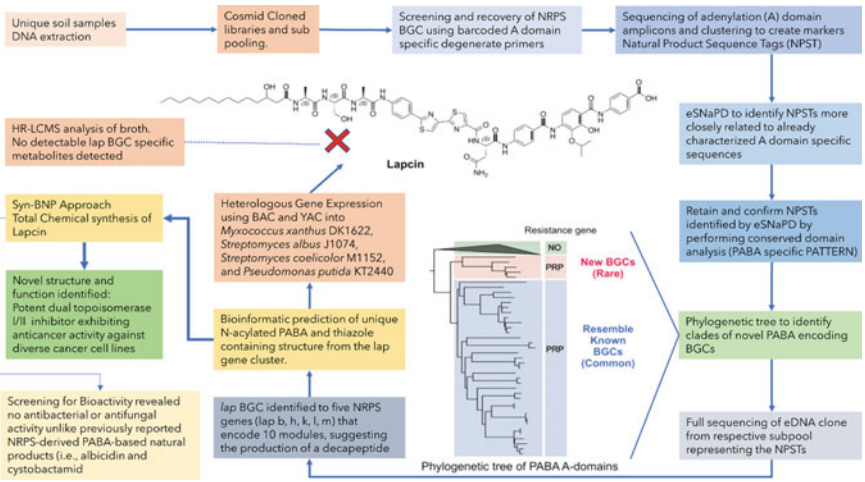

Fig. 3 Strategy used by Wang et al. (2022) to discover Lapcin using Syn-BNP approach

In one of the latest research findings, genome mining was successfully used for mining of anabolic pathways riddled with no identifiable enzymes resulting in identification of what is termed as "unknown-unknown natural products". The two unknowns refer to unidentified pathways and unidentified molecular structure of products. This approach recently identified a hypothetical fungal protein which in turn was found to be a novel arginine-containing cyclopeptidase which resulted in production of a novel natural product, cyclo-arginine-tyrosine-dipeptide (Yee et al. 2023).

4 Conclusion

The application of metagenomics for the identification and discovery of novel natural products has seen a revolutionizing enhancement in recent years due to the high speed and lesser expensive features of the newer sequencing technologies. Unculturable microbial symbionts and commensals of all life forms from different biomes are being studied using this approach and many have been reported to exhibit previously unknown metabolic pathways and SMBGCs leading to discovery of novel drugs. The success of naturally derived metabolite in drug discovery can be correlated to better understanding of the structural diversity of the molecules. Structural diversity in turn greatly influences how the drug interacts with biomolecules thereby mediating their mode of action (Mathur and Hoskins 2017). Metagenomics offers great potential in this regard and additionally shows great promise in identifying best lead candidates with low toxicity.

Despite the promising future that metagenomics portrays, microbial production of drugs still faces several challenges. These include low production titers, problems related to isolation of the natural products and/or structural identification. These, however, can be circumvented using approaches such as strain improvement, engineering precursor (primary metabolite) supply, pathway engineering, combinatorial biosynthesis using genetic engineering and mutasynthesis for novel product generation (Pham et al. 2019). Combining these approaches with metagenomics, proteomics, transcriptomics and metabolomics as well as the newer culturing techniques will surely enhance and enrich the manner in which drug discovery and development will proceed in the years to come.

References

Abdel-Razek AS, El-Naggar ME, Allam A et al (2020) Microbial natural products in drug discovery. Processes 8:470. https://doi.org/10.3390/pr8040470

Abdelnasser SSI, Ali AAS, Ashraf AH et al (2012) Tapping uncultured microorganisms through metagenomics for drug discovery. Afr J Biotechnol 11:15823–15834. https://doi.org/10.5897/ajb12.2544

Alam K, Abbasi MN, Hao J et al (2021) Strategies for natural products discovery from uncultured microorganisms. Molecules 26:2977. https://doi.org/10.3390/molecules26102977

Albarano L, Esposito R, Ruocco N, Costantini M (2020) Genome mining as new challenge in natural products discovery. Mar Drugs 18:1–17. https://doi.org/10.3390/md18040199

Aoi Y, Kinoshita T, Hata T et al (2009) Hollow-fiber membrane chamber as a device for in situ environmental cultivation. Appl Environ Microbiol 75:3826–3833. https://doi.org/10.1128/AEM.02542-08

Araujo FA, Barh D, Silva A et al (2018) GO FEAT: a rapid web-based functional annotation tool for genomic and transcriptomic data. Sci Rep 8:1794. https://doi.org/10.1038/s41598-018-20211-9

Ardui S, Ameur A, Vermeesch JR, Hestand MS (2018) Single molecule real-time (SMRT) sequencing comes of age: applications and utilities for medical diagnostics. Nucleic Acids Res 46:2159–2168. https://doi.org/10.1093/nar/gky066

Atanasov AG, Zotchev SB, Dirsch VM, Supuran CT (2021) Natural products in drug discovery: advances and opportunities. Nat Rev Drug Discov 20:200–216. https://doi.org/10.1038/s41573-020-00114-z

Bari ML, Yeasmin S (2022) Microbes culture methods. In: Encyclopedia of infection and immunity. Elsevier, pp 77–98

Bayer TS, Widmaier DM, Temme K et al (2009) Synthesis of methyl halides from biomass using engineered Microbes. J Am Chem Soc 131:6508–6515. https://doi.org/10.1021/ja809461u

Belknap KC, Park CJ, Barth BM, Andam CP (2020) Genome mining of biosynthetic and chemotherapeutic gene clusters in Streptomyces bacteria. Sci Rep 10:2003. https://doi.org/10.1038/s41598-020-58904-9

Benítez X, Gonzalez EG, García J et al (2021) Detection of a pederin-like compound using a dilution-to-extinction-based platform for the isolation of marine bacteria in drug discovery strategies. Microb Biotechnol 14:241–250. https://doi.org/10.1111/1751-7915.13679

Boulund F, Berglund F, Flach C-F et al (2017) Computational discovery and functional validation of novel fluoroquinolone resistance genes in public metagenomic data sets. BMC Genomics 18:682. https://doi.org/10.1186/s12864-017-4064-0

Campbell S, Soman-Faulkner K (2022) Antiparasitic drugs. StatPearls

Chen R, Wong HL, Kindler GS et al (2020) Discovery of an abundance of biosynthetic gene clusters in shark bay microbial mats. 11:1–13. https://doi.org/10.3389/fmicb.2020.01950

Chopra RS, Chopra C, Sharma NR (eds) (2020) Metagenomics: techniques, applications, challenges and opportunities. Springer, Singapore

Chu J, Vila-Farres X, Inoyama D et al (2016) Discovery of MRSA active antibiotics using primary sequence from the human microbiome. Nat Chem Biol 12:1004–1006. https://doi.org/10.1038/nchembio.2207

Chu J, Vila-Farres X, Inoyama D et al (2018) Human microbiome inspired antibiotics with improved β-Lactam synergy against MDR *Staphylococcus aureus*. ACS Infect Dis 4:33–38. https://doi.org/10.1021/acsinfecdis.7b00056

Connon SA, Giovannoni SJ (2002) High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates. Appl Environ Microbiol 68:3878–3885. https://doi.org/10.1128/AEM.68.8.3878-3885.2002

Danhorn T, Young CR, DeLong EF (2012) Comparison of large-insert, small-insert and pyrosequencing libraries for metagenomic analysis. ISME J 6:2056–2066. https://doi.org/10.1038/ismej.2012.35

Demain AL, Sanchez S (2009) Microbial drug discovery: 80 years of progress. J Antibiot (tokyo) 62:5–16. https://doi.org/10.1038/ja.2008.16
Dias DA, Urban S, Roessner U (2012) A historical overview of natural products in drug discovery. Metabolites 2:303–336. https://doi.org/10.3390/metabo2020303
Donia MS, Ruffner DE, Cao S, Schmidt EW (2011) Accessing the hidden majority of marine natural products through metagenomics. ChemBioChem 12:1230–1236. https://doi.org/10.1002/cbic.201000780
van Duin D, Paterson DL (2016) Multidrug-resistant bacteria in the community. Infect Dis Clin North Am 30:377–390. https://doi.org/10.1016/j.idc.2016.02.004
Emran TB, Shahriar A, Mahmud AR et al (2022) Multidrug resistance in cancer: understanding molecular mechanisms, immunoprevention and therapeutic approaches. Front Oncol 12. https://doi.org/10.3389/fonc.2022.891652
Escobar-Zepeda A, Vera-Ponce de León A, Sanchez-Flores A (2015) The road to metagenomics: from microbiology to dna sequencing technologies and bioinformatics. Front Genet 6. https://doi.org/10.3389/fgene.2015.00348
Ferrer M, Beloqui A, Timmis KN, Golyshin PN (2009) Metagenomics for mining new genetic resources of microbial communities. Microb Physiol 16:109–123. https://doi.org/10.1159/000142898
Fleming A (1929) On the antibacterial action of cultures of *Penicillium*, with special reference to their use in the isolation of *B. influenzae*. Br J Exp Pathol 10:226–236
Francioli D, Lentendu G, Lewin S, Kolb S (2021) DNA metabarcoding for the characterization of terrestrial microbiota—pitfalls and solutions. Microorganisms 9:361. https://doi.org/10.3390/microorganisms9020361
Gavrish E, Bollmann A, Epstein S, Lewis K (2008) A trap for in situ cultivation of filamentous actinobacteria. J Microbiol Methods 72:257–262. https://doi.org/10.1016/j.mimet.2007.12.009
Geraci JE, Heilman FR, Nichols DR, Wellman EW, Ross GT (1956) Some laboratory and clinical experiences with a new antibiotic, vancomycin. Antibiot. Annu, pp 90–106
Ghlichloo I, Gerriets V (2022) Nonsteroidal Anti-inflammatory Drugs (NSAIDs) [Updated 2022 May 19]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan. Available from: https://www.ncbi.nlm.nih.gov/books/NBK547742/
Ghosh A, Jayaraman N, Chatterji D (2020) Small-molecule inhibition of bacterial biofilm. ACS Omega 5:3108–3115. https://doi.org/10.1021/acsomega.9b03695
Gomes ES, Schuch V, de Macedo Lemos EG (2013) Biotechnology of polyketides: new breath of life for the novel antibiotic genetic pathways discovery through metagenomics. 1034:1007–1034
Hafner J, Payne J, Mohammadi Peyhani H et al (2021) A computational workflow for the expansion of heterologous biosynthetic pathways to natural product derivatives. Nat Commun 12. https://doi.org/10.1038/s41467-021-22022-5
Heather JM, Chain B (2016) The sequence of sequencers: The history of sequencing DNA. Genomics 107:1–8. https://doi.org/10.1016/j.ygeno.2015.11.003
Hildebrand M, Waggoner LE, Liu H, et al (2004) bryA. Chem Biol 11:1543–1552. https://doi.org/10.1016/j.chembiol.2004.08.018
Hingorani AD, Kuan V, Finan C et al (2019) Improving the odds of drug development success through human genomics: modelling study. Sci Rep 9:18911. https://doi.org/10.1038/s41598-019-54849-w
Hover BM, Kim S, Katz M et al (2018) Culture-independent discovery of the malacidins as calcium-dependent antibiotics with activity against multidrug-resistant Gram-positive pathogens. Nat Microbiol 3. https://doi.org/10.1038/s41564-018-0110-1
Johnson M (2013) Detergents: Triton X-100, Tween-20, and more. Mater Methods 3. https://doi.org/10.13070/mm.en.3.163
Jung D, Liu L, He S (2021) Application of in situ cultivation in marine microbial resource mining. Mar Life Sci Technol 3:148–161
Kaeberlein T, Lewis K, Epstein SS (2002) Isolating "Uncultivable" microorganisms in pure culture in a simulated natural environment

Karmakar P, Trivedi A, Gaitonde V (2020) Introductory chapter: The modern-day drug discovery. In: Drug discovery and development—new advances. IntechOpen
King RW, Bauer JD, Brady SF (2009) An environmental DNA-Derived Type II polyketide biosynthetic pathway encodes the biosynthesis of the pentacyclic polyketide erdacin. Angew Chem Int Ed 48:6257–6261. https://doi.org/10.1002/anie.200901209
Lam KN, Cheng J, Engel K et al (2015) Current and future resources for functional metagenomics. Front Microbiol 6. https://doi.org/10.3389/fmicb.2015.01196
Lim HK, Chung EJ, Kim J-C et al (2005) Characterization of a forest soil metagenome clone that confers indirubin and indigo production on *Escherichia coli*. Appl Environ Microbiol 71:7768–7777. https://doi.org/10.1128/AEM.71.12.7768-7777.2005
Lozada M, Dionisi HM (2015) Microbial bioprospecting in marine environments. In: Hb25_ Springer Handbook of Marine Biotechnology. Springer, Berlin, pp 307–326
Maghembe R, Damian D, Makaranga A et al (2020) Omics for bioprospecting and drug discovery from bacteria and microalgae. Antibiotics 9:229. https://doi.org/10.3390/antibiotics9050229
Mansoori B, Mohammadi A, Davudian S et al (2017) The different mechanisms of cancer drug resistance: a brief review. Adv Pharm Bull 7:339–348. https://doi.org/10.15171/apb.2017.041
Mathur S, Hoskins C (2017) Drug development: lessons from nature. Biomed Rep 6:612–614. https://doi.org/10.3892/br.2017.909
Medema MH, Blin K, Cimermancic P, Jager V De, Zakrzewski P, Fischbach MA, Weber T, Takano E, Breitling R (2011) antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. 39(June), 339–346. https://doi.org/10.1093/nar/gkr466
Moyer AJ, Coghill RD (1946) Penicillin: IX. The Laboratory Scale Production of Penicillin in Submerged Cultures by *Penicillium notatum* Westling (NRRL 832). J Bacteriol. Jan;51(1):79–93 PMID: 16561059; PMCID: PMC518023.
Newman DJ, Cragg GM (2020) Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. J Nat Prod 83:770–803. https://doi.org/10.1021/acs.jnatprod.9b01285
Newton GGF, Abraham EP (1955) Cephalosporin C, a new antibiotic containing sulphur and D-α-aminoadipic acid. Nature 175:548–548. https://doi.org/10.1038/175548a0
Ngara TR, Zhang H (2018) Recent Advances in function-based metagenomic screening. Genomics Proteomics Bioinform 16:405–415. https://doi.org/10.1016/j.gpb.2018.01.002
Nikaido H (2009) Multidrug resistance in bacteria. Annu Rev Biochem 78:119–146. https://doi.org/10.1146/annurev.biochem.78.082907.145923
Palaniappan K, Chen IA, Chu K et al (2020) IMG-ABC v . 5 . 0 : an update to the IMG/Atlas of Biosynthetic Gene Clusters Knowledgebase 48:422–430. https://doi.org/10.1093/nar/gkz932
Palma Esposito F, Ingham CJ, Hurtado-Ortiz R et al (2018) Isolation by miniaturized culture chip of an antarctic bacterium aequorivita sp. with antimicrobial and anthelmintic activity. Biotechnol Rep 20. https://doi.org/10.1016/j.btre.2018.e00281
Pereira F (2019) Metagenomics: a gateway to drug discovery. In: Advances in biological science research: a practical approach. Elsevier, pp 453–468
Pham JV, Yilma MA, Feliz A et al (2019) a review of the microbial production of bioactive natural products and biologics. Front Microbiol 10. https://doi.org/10.3389/fmicb.2019.01404
Pina AS, Hussain A, Roque ACA (2010) An historical overview of drug discovery, pp 3–12
Qiao K, Chooi Y-H, Tang Y (2011) Identification and engineering of the cytochalasin gene cluster from Aspergillus clavatus NRRL 1. Metab Eng 13:723–732. https://doi.org/10.1016/j.ymben.2011.09.008
Quintero IJ, Castillo AM, Mejía LC (2022) Diversity and taxonomy of soil bacterial communities in urban and rural mangrove forests of the panama bay. Microorganisms 10:2191. https://doi.org/10.3390/microorganisms10112191
Rappé MS, Connon SA, Vergin KL, Giovannoni SJ (2002) Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. Nature 418:630–633. https://doi.org/10.1038/nature00917

Raper KB, Benedict RG (1950) The Drugs of Microbial origin. In: Crops in Peace and War, 1950-1951 Yearbook of Agriculture. United States Department of Agriculture, Washington, DC. United States Government Printing Office. pp 734–741 Retrieved from: https://naldc.nal.usda.gov/download/IND43894154/PDF

Reddy, BVB, Milshteyn A, Charlop-Powers Z, Sean FB (2014) eSNaPD: a versatile, web-based bioinformatics platform for surveying and mining natural product biosynthetic diversity from metagenomes. Chem Biol 21:1023–1033. https://doi.org/10.1016/j.chembiol.2014.06.007.eSNaPD

Reygaert WC (2018) An overview of the antimicrobial resistance mechanisms of bacteria. AIMS Microbiol 4:482–501. https://doi.org/10.3934/microbiol.2018.3.482

Santana-pereira ALR, Sandoval-powers M, Monsma S et al (2020) Discovery of novel biosynthetic gene cluster diversity from a soil metagenomic library. 11:1–17. https://doi.org/10.3389/fmicb.2020.585398

Schloss PD, Handelsman J (2003) Biotechnological prospects from metagenomics. Curr Opin Biotechnol 14:303–310. https://doi.org/10.1016/S0958-1669(03)00067-3

Sharpton TJ (2014) An introduction to the analysis of shotgun metagenomic data. Front Plant Sci 5. https://doi.org/10.3389/fpls.2014.00209

Singh BK, Macdonald CA (2010) Drug discovery from uncultivable microorganisms. Drug Discov Today 15:792–799. https://doi.org/10.1016/j.drudis.2010.07.002

Singh B, Roy A (2020) Metagenomics and Drug-Discovery. In: Metagenomics: Techniques, Applications, Challenges and Opportunities. Springer Singapore, pp 133–145

Sinha S, Vohora D (2018) Drug discovery and development. In: Pharmaceutical medicine and translational clinical research. Elsevier, pp 19–32

Skinnider MA, Dejong CA, Rees PN et al (2015) Genomes to natural products PRediction Informatics for Secondary Metabolomes (PRISM). 43:9645–9662. https://doi.org/10.1093/nar/gkv1012

Smith TE, Pond CD, Pierce E et al (2018) Accessing chemical diversity from the uncultivated symbionts of small marine animals. Nat Chem Biol 14:179–185. https://doi.org/10.1038/nchembio.2537

Sun D, Gao W, Hu H, Zhou S (2022) Why 90% of clinical drug development fails and how to improve it? Acta Pharm Sin B 12:3049–3062. https://doi.org/10.1016/j.apsb.2022.02.002

Surup F, Viehrig K, Rachid S et al (2018) Crocadepsins—depsipeptides from the Myxobacterium *Chondromyces crocatus* found by a genome mining approach. ACS Chem Biol 13:267–272. https://doi.org/10.1021/acschembio.7b00900

Szymański P, Markowicz M, Mikiciuk-Olasik E (2011) Adaptation of high-throughput screening in drug discovery—toxicological screening tests. Int J Mol Sci 13:427–452. https://doi.org/10.3390/ijms13010427

Taylor D (2015) The pharmaceutical industry and the future of drug development. In: Pharmaceuticals in the environment. The royal society of chemistry, pp 1–33

Totomoch-Serra A, Marquez MF, Cervantes-Barragán DE (2017) Sanger sequencing as a first-line approach for molecular diagnosis of Andersen-Tawil syndrome. F1000Res 6:1016. https://doi.org/10.12688/f1000research.11610.1

Trindade M, van Zyl LJ, Navarro-Fernández J, Abd Elrazak A (2015) Targeted metagenomics as a tool to tap into marine natural product diversity for the discovery and production of drug candidates. Front Microbiol 6. https://doi.org/10.3389/fmicb.2015.00890

Ueoka R, Sondermann P, Leopold-Messer S et al (2022) Genome-based discovery and total synthesis of janustatins, potent cytotoxins from a plant-associated bacterium. Nat Chem 14:1193–1201. https://doi.org/10.1038/s41557-022-01020-0

Waksman SA, Reilly HC, Johnstone DB (1946) Isolation of Streptomycin-producing Strains of *Streptomyces griseus*. J Bacteriol 52:393–397. https://doi.org/10.1128/jb.52.3.393-397.1946

Wang Z, Forelli N, Hernandez Y et al (2022) Lapcin, a potent dual topoisomerase I/II inhibitor discovered by soil metagenome guided total chemical synthesis. Nat Commun 13:842. https://doi.org/10.1038/s41467-022-28292-x

Wiseman AC (2016) Immunosuppressive Medications. Clin J Am Soc Nephrol 11:332–343. https://doi.org/10.2215/CJN.08570814

Wydro U (2022) Soil Microbiome study based on DNA extraction: a review. Water (basel) 14:3999. https://doi.org/10.3390/w14243999

Yee DA, Niwa K, Perlatti B et al (2023) Genome mining for unknown–unknown natural products. Nat Chem Biol. https://doi.org/10.1038/s41589-022-01246-6

Zengler K, Walcher M, Clark G et al (2005) High-throughput cultivation of microorganisms using microcapsules. Methods Enzymol 397:124–130