Optimisation potential of the cannabis plant through sequencing

Dr. Lisann Eßer and Rainer Krüger

J&K Consulting GmbH, Rommerskirchen

The prescription of medicinal cannabis in Germany has been rising constantly in the last few years, which means that both the demand and the market are growing equally. However, medicinal cannabis is a drug whose active ingredient composition and concentration depend on the genetics of the plant and thus also achieve a different effect on the patient based on the genetics. By sequencing the genomes and comparing the sequences of different cannabis varieties, characteristic genes are to be identified and gene databases created in the future. Sequencing is a molecular biological procedure that can be used to decode the nucleotide sequence of the genetic code. With the help of this technique, producers of medicinal cannabis can decide before the beginning of cultivation which genetics and thus which combination and concentration of active ingredients of the cannabis plants to aim for. By sequencing the produced plants, manufacturers can clearly define their product and thus ensure unambiguous traceability. In addition, the expression of certain genes can be enhanced through targeted breeding, thereby optimising the yield. Sequencing generates large amounts of product-specific data, which can be successfully managed and accessed through blockchain technology.

This article explains the methods of sequencing plant genomes and the benefits that result for manufacturers and patients. It describes how large amounts of data from sequencing are managed by digital tools and how the resulting potential can be used by pharmaceutical companies.

Use of the cannabis plant as a medicine

The cannabis plant was underestimated for a long time, also because of its controversial image. In fact, however, the cannabis plant has many uses, whether as a building material or as a medicinal plant because of its ingredients. Already thousands of years ago, cannabis was used to produce ropes and textiles and later also for paper because of its stalks and fibres. Nowadays, commercial hemp is again increasingly cultivated to produce hemp fibres but also for hemp seeds and oil for the food market [1]. In addition to being used as a source of raw materials, the cannabis plant is also cultivated for its healing and pain-relieving effects. Many clinical studies have already confirmed the positive effect of cannabis on various diseases [2], although cannabis research is only just beginning. Cannabis products are often prescribed for chronic pain, Tourette's disease, multiple sclerosis (MS), or ADHD. Due to the broad therapeutic spectrum, cannabis can be used for the treatment of a variety of diseases. By law, there is no restriction as to which indication is necessary to prescribe medicinal cannabis. Medical doctors can basically prescribe cannabis for any serious illness, provided that other treatment methods cannot be used or there is no chance of a successful therapy.

However, due to its misuse as an addictive substance, the use of cannabis as medicine was banned in Germany for a long time. Only since 2011 cannabis may also be used in Germany in the form of finished medical products due to changed narcotic legal requirements. With a legal change in 2017, medical products made from cannabis became distributable and prescribable, so doctors are now also allowed to prescribe cannabis flowers and preparations as prescription medicine. This makes it possible to prescribe medicinal cannabis flowers or extracts with pharmaceutical grade quality on a narcotic prescription.

So far, patients have only been supplied with imported medicinal cannabis. Then, on 7th July 2021, the official sale of medicinal cannabis to pharmacies from German production started [3], which ensures an additional supply for patients. Since the law was changed in 2017, the number of patients accessing cannabis preparations has more than doubled and the demand for medicinal cannabis products continues to rise steadily. For example, in Sep. 2018, monthly gross sales of finished cannabinoid products and preparations were approximately 6.5 mio euros, whereas one year later in Sep. 2019, sales almost doubled [4]. In 2019, gross sales of 54 mio euros could be achieved in Germany with the billing of cannabis flowers in unchanged condition alone, and gross sales of 61 mio euros in 2020. Only figures from the national health insurance funds were recorded, private prescriptions are not included [5]. Driven by the increasing demand, the market for cannabis products offers great potential.

The active ingredients of the cannabis plant are cannabinoids and the best known is tetrahydro cannabidiol (THC). THC is psychoactive and has a very diverse effect. Another well-known cannabinoid is cannabidiol (CBD), which, unlike THC, has no psychoactive effect. When selling cannabis flowers and other cannabis products, the concentration of the two cannabinoids is usually indicated. This depends on the species and is between <1 % and 25 % for THC and between <0.5 % and 14 % for CBD in the most common cannabis flowers [6]. In addition to the 2 active substances described above, more than 100 different cannabinoids have been identified. This results in a characteristic cannabinoid profile for each species and thus a different spectrum of effects. Besides cannabinoids, the cannabis plant contains other substances such as terpenes and flavonoids. Terpenes are natural compounds that are responsible for the scent and taste of plants. Classic terpenes in the cannabis plant are for example myrcene, limonene or humulene. Myrcene and limonene smell and taste lemony, whereas humulene smells rather spicy and earthy and has a peppery, hoppy taste. Flavonoids are also natural compounds that give the cannabis plant its intense aroma and flavour. Studies have already shown that both terpenes and flavonoids can have an anti-inflammatory, calming and antioxidant effect on the body [7, 8]. As with cannabinoids, the concentration of terpenes and flavonoids varies from species to species.

Sequencing opens new possibilities

So far, the production of medicinal cannabis has mainly focused on the THC and CBD content but looking into the genome of the plant can also be very rewarding. The method for decoding the genome is sequencing. Sequencing is a molecular biological procedure that helps to decode the nucleotide sequence, i.e., the DNA sequence. Nowadays, various methods are available for sequencing, which are far superior to classical methods in terms of efficiency and speed. The efficient next-generation sequencing, a high-throughput method for decoding DNA, has revolutionised genome research. More than 30 years ago, the Human Genome Project was initiated as a research project to completely decode the human genome. The project cost around 3 bn dollars [9] and it took 13 years until the human genome was completely decoded in 2003 [10]. Nowadays, thanks to new technology, entire genomes can be decoded in days instead of years, and at a fraction of the cost.

The decoding of the cannabis genome could now help to cultivate targeted plants with specific active ingredients in the future. In 2011, Canadian researchers decoded the genome of the THC-rich cannabis plant *Cannabis sativa* for the first time [11]. They reported their findings in the scientific journal Genome Biology. For the sequencing of the plant genome, the scientists used the cannabis species *Purple Kush*, which is known for a particularly high THC content. In the study, they compared the genome and activity of the THC-rich *Purple Kush* plant species with 2 THC-free cannabis strains. By comparing the transcriptome of *Purple Kush* with that of the THC-free commercial cannabis strains, they found that many genes encoding proteins involved in cannabinoid and prodrug pathways were more highly expressed in *Purple Kush*. In addition, they found that an enzyme essential for THC production is only switched on in the *Purple Kush* species, but not in the THC-free commercial hemp varieties. This knowledge can be used in the breeding of cannabis plants to generate plants with increased THC content. It is not necessary to genetically modify the plants, but this can be done conventionally through targeted cross-breeding. In this way, the plants can be optimised specifically for their intended use.

The discovery of the mechanism just described is only the beginning. As the genome of the cannabis plant is further decoded and better understood, there will be more and more methods to adapt the plant according to the desired requirements. The possible applications of this are various. For example, targeted breeding can be used to adapt the plant architecture or the growth rate to generate more yield. In addition, an optimisation option would be to develop plants with a defined concentration of THC and CBD. Varieties that are completely free of THC but have special terpene or flavonoid profiles are possible. This has the advantage that these species are no longer regulated by the Narcotics Law and the patient, as far as he is not reliant

Pharm. Ind. 83, Nr. 12, 1664–1669 (2021)

on THC, can be treated with cannabis products without being restricted by the high effect of THC.

Once the genome of the cannabis plant has been sufficiently analysed, there is the option of creating gene databases with the genetic information. In addition, specific information on the species and their optimisation could be made available to producers. Producers could select the characteristics of their plants according to the requirement of customers or patients even before the first plant is cultivated.

However, plant sequencing can be interesting for the pharmaceutical industry not only because of better yields and an optimised spectrum of activity but also from a regulatory point of view. Through sequencing, each plant can be clearly traced back to its origin. Cannabis is usually cultivated from mother plants that are used to produce clones. The so-called cuttings not only have the same properties in terms of yield and active substance concentration as the mother plant but also have identical genetic material. It is also possible to sequence one or more plants from a batch to identify the DNA of the plants and clearly assign it to the mother plant. If there is a problem with the cannabis flowers later, the flowers can be sequenced again during the recall and the origin of the plants can be clearly determined. Sequencing provides additional security and clarity in traceability and thus supports the documentation required by regulations during production. However, when sequencing for traceability purposes, it is not always necessary to sequence the entire genome. It makes sense to sequence characteristic DNA sections to clearly identify the plants. In this way, costs and time can be saved.

Dealing with sequencing data and new technology

Great potential is offered when the sequence data is expanded with a range of other data and data packages are generated for each individual plant. The data sets can be linked using modern track-and-trace solutions, for example with data from monitoring during cultivation, such as temperature, humidity, and nutrient supply. Data from the supply chain, such as location and temperature during transport, can also be added. In this way, the producer receives specific information about his product and at the same time succeeds in fully tracing each plant back to its origin. If sequencing is operated on a larger scale, millions of data points are generated within a very short time. A single sequencing run of the sequencing systems Illumina NextSeq 1000 and 2000 can generate up to 1.1 trillion reads [12]. Especially when the sequencing data is supplemented with details from monitoring or track-and-trace systems, huge data sets are created that different parties have an interest in. The producer of the cannabis products needs the information for his regulatory verification obligations. The more data he has about his product, the better the proof of origin and quality. The data sets are not only completed by the producer himself but also by other companies, e.g., the supplier of the raw materials, who uses the information to prove the quality of his products, or the transport company, which generates tracking data. In addition, the producer can make the collected data available to the authorities during auditing or in the event of a recall. The information generated during the value chain should therefore be available to selected persons and, if necessary, be able to be processed by them. However, since not all users should be granted the same access to the data, selective data access is necessary. At the same time, data integrity must be guaranteed, and the data should be stored securely against unauthorised access and data falsification.

These features are particularly important for the highly regulated pharmaceutical industry. Thus, managing and accessing these volumes of data is a challenge that can only be solved by modern data storage.

The new blockchain technology offers a possible solution to the described problem. With the blockchain, data records, the so-called blocks, are stored decentrally. This means that the data set is not located in a cloud system, as is the case with a jointly used sharepoint, but locally on the end devices of the users. When a new block is generated, it is verified and stored by all computers in the network.
Each computer represents a network node through which the data is transmitted. The blocks are linked together with the help of a cryptographic procedure and encrypted by an algorithm. In addition, the data records have a digital fingerprint consisting of a timestamp and transaction data. The data blocks are then attached to a chain of data, the blockchain, where they can be expanded chronologically and linearly. In this way, many unique, interlinked data sets are created. With the help of this technology, the data from the value chain can be shared with and processed by a selected group of people while maintaining data integrity. In this way, the manufacturer can gain access to a wide range of quality-relevant information about its product and pass it on to selected circles. The blockchain is therefore a tool that supports the sharing and provision of sequence data. The current challenge is no longer the sequencing itself, but the analysis and interpretation of the results. The analysis of the sequences is time-consuming and, if carried out manually, requires large human resources. With the help of the blockchain, the data could be made available to several people or even automated systems, thus speeding up the analysis and interpretation of the sequences. The combination of

next-generation sequencing methods with artificial intelligence (AI) offers potential.

As already mentioned, the main challenge is the analysis and interpretation of the sequence data. This is when AI can help to analyse the flood of data. Currently, the combination of both techniques is mainly used for the interpretation of human genome data. The goal is to recognise mutations in the genetic code that lead to diseases better and faster [12]. In the future, AI technology will advance the speed and accuracy of sequencing by lengths. The decoding of genes will then only take minutes and no longer hours or even days. In addition, AI will simplify the interpretation of the data and thus also enable the development of new cannabis species or traceability down to the last base pair.

LITERATURE

[1] Anbaufläche von Hanf in der EU 2019, *https://de.statista.com/statistik/daten/studie/1174126/umfrage/anbaflaecheu-von-hanf-in-europaische-union/*

[2] Brief report of the Federal Ministry of Health „Cannabis: Potential und Risiken. Eine wissenschaftliche Analyse (CaPRis)", from 27.11.2017, *https://tinylink.net/LhURt*

[3] BfArM press release 6/21 from 07.07.2021 „BfArM startet Verkauf von Cannabis zu medizinischen Zwecken an Apotheken", *https://www.bfarm.de/SharedDocs/Pressemitteilungen/DE/2021/pm6-2021.html*

[4] Special leaflet for the ‚GKV-Arzneimittel-Schnellinformation' for Germany „Monatliche Bruttoumsätze von cannabinoidhaltigen Fertigarzneimitteln und Zubereitungen" from the German National Association of Health Insurance Funds (‚GKV-Spitzenverband'), Data status from 02.12.2019

[5] GKV GAMSI, Special leaflet for the ‚GKV-Arzneimittel-Schnellinformation' for Germany, Gross sales and prescriptions of finished cannabinoid products and preparations from January to December, *https://tinylink.net/NP4jJ*

[6] DAP Report, Medizinische Cannabisblüten, Deutsches ApothekenPortal, June 2021

[7] Limonene: Aroma of innovation in health and disease, Chemico-Biological Interactions, February 2018, *https://doi.org/10.1016/j.cbi.2018.02.007*

[8] Flavonoids: an overview, Journal of Nutritional Science, December 2016, *https://doi.org/10.1017/jns.2016.41*

[9] National Genome Research Network. *http://www.ngfn.de/de/verstehen_der_menschlichen_erbsubstanz.html*

[10] National Genome Research Network, *http://www.ngfn.de/index.php/den_dna_code_knacken.html*

[11] The draft genome and transcriptome of Cannabis sativa, H. Van Bakel et al. Genome Biology 2011, *https://doi.org/10.1186/gb-2011-12-10-r102*

[12] *https://emea.illumina.com/systems/sequencing-platforms.html*

[13] Diagnosis of genetic diseases in seriously ill children by rapid whole-genome sequencing and automated phenotyping and interpretation, M. Clark et al., Science Translational Medicine, April 2019, *https://www.science.org/doi/10.1126/scitranslmed.aat6177*

The last access to all links was on 19th Oct. 2021.

FURTHER LITERATURE

Krüger R, Eßer L. Digitale Entwicklung und Herstellung von Medizinalcannabis. Techno-Pharm 11, Nr. 1, 30–38 (2021)

Krüger R, Eßer L. Einsatz der Blockchain bei der Herstellung von Medizinalcannabis, Techno-Pharm 11, Nr. 4, 216–221 (2021)

Correspondence:
Dr. Lisann Eßer
J&K Consulting GmbH
Kirchstr. 11
41569 Rommerskirchen (Germany)
e-mail: lisann.esser@j-k-consulting.de

Either author or organization are responsible for the correct translation of this article.
Editor-in-chief: Claudius Arndt, Managing Secretary's office: Valentina Krinizki. Publisher: ECV · Editio Cantor Verlag für Medizin und Naturwissenschaften GmbH, Baendelstockweg 20, 88326 Aulendorf (Germany). Phone: +49 (0) 75 25 94 00, Fax: +49 (0) 75 25 94 01 80. e-mail: redaktion@ecv.de. www.ecv.de.
Typesetting: rdz GmbH (Germany).