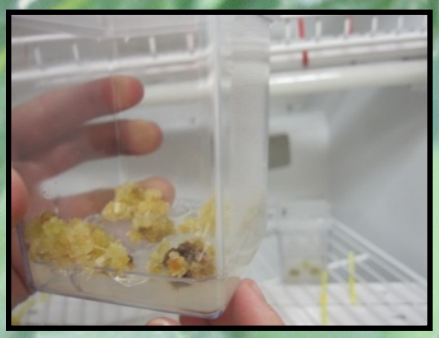
The Biotechnology of *Cannabis sativa*

2nd edition



Sam R. Zwenger, PhD

The Biotechnology of *Cannabis sativa*

2nd edition

Sam R. Zwenger, PhD

EXTREME PUBLICATIONS, INC.
New York

 ISBN: 00-000-0-420

Editorial Note

This edition of "The Biotechnology of *Cannabis sativa*" includes scientific procedures and examples of potentially controversial research ideas. In no way is our publishing company (Extreme Publications, Incorporated) or the author responsible for the actions of people who pursue cannabis biotechnology (unless something really cool results).

Cover

White Widow female flower.
Insets (left to right); cannabis calli, young cannabis plants growing in vitro (provided by Joe Oakes at Skunk Pharm Research, LLC), and hard red winter wheat (TAM111).

Photographs

Many photographs within this text were provided by Otto Terkel. Photographs that were not provided by Otto are indicated accordingly where they appear.

To those who seek to improve cannabis-

Preface

With the therapeutic effects of cannabis becoming realized by industry, some have predicted that in 2016 the market for cannabis-based drugs will be just under $30 billion. Several major companies are seriously working towards developing pharmaceutical products from cannabis, and more companies are realizing the potential and joining the cannabis rush. Undoubtedly, cannabis biotechnology will be crucial to delivering medical and recreational needs for millions of people.

This edition of "The Biotechnology of *Cannabis sativa*" is markedly different than the first edition, which I wrote during my early years of graduate school studying plant biotechnology. That edition was also prompted by furious passion and was quickly written. The result was an inspirational novella that was lacking in technical detail. In this edition, the passion has been retained while some sections become quite technical, allowing the readers to immerse themselves in a deeper level of understanding. Molecular genetics, manipulation of cannabinoid production, tissue culture, transgenic systems, molecular details of gene transfer, cannabis genomics, and bioinformatics are only a few of the emerging sciences that have been included in this edition. Although much of the presentation is in a form that should be pleasantly digestible to almost any reader, there are several sections where technical details were necessarily included.

The target reading group, as it has been argued all books must have one, is therefore quite diverse. Those concerned with how the cannabis landscape is evolving will find this book highly informative. Cannabis growers and consumers will understand why holding on to their seeds might be important. Students of plant biotechnology will find the ideas and techniques useful in guiding their own research questions. At the same time, this is not a textbook and instead delivers the study of plant biotechnology in a flavorful way.

Considering the diverse interests of researchers working around the world on cannabis and the excellent, detailed published research articles currently available, much can be said about the biotechnology of cannabis. While legalization is changing who is consuming cannabis, biotechnology is redefining cannabis. Thus, this world-famous plant has the potential to take several new paths that are filled with uncertainty. At least for the moment, the only certainty is that cannabis will never be the same.

Sam R. Zwenger
4-20

Acknowledgements

The scientific research concerning *Cannabis sativa* is extremely broad. Thousands of people are working towards understanding its attributes in research labs around the world. Although a valiant attempt to have adequately represented as much of the relevant literature and present the findings of so many scientists to a greater public was made, everyone who has dedicated time to cannabis research could not possibly be mentioned in a single text.

My hope is that at least most people were recognized and if not, my hope is that the contributions of the ones I have neglected are recognized somewhere in the world. The science of cannabis is an emergent property of so many people working hard to improve humanity.

I owe a great deal of thanks to particular individuals who really helped make this edition of "The Biotechnology of *Cannabis sativa*" a more complete text. Their help and encouragement are greatly appreciated. These individuals include Joe Oakes and his colleagues at Skunk Pharm Research, LLC; Hemant Lata, Suman Chandra, Ikhlas A. Khan and Mahmoud A. ElSohly at the National Center for Natural Products Research, School of Pharmacy, University of Mississippi USA; Aurelia Slusarkiewicz-Jarzina and her colleagues at the Institute of Plant Genetics, Polish Academy of Sciences, in Poland; David Watson at HortaPharm B.V.; David Marks at the University of Minnesota USA and his colleagues; members of the Staff at the Torrey Botanical Society; the Unipro UGENE team in Russia; Sudhir Kumar at the Center for Evolutionary Medicine and Informatics; Francisco Melo and Ignacio Ibarra and their work on the Sequence Alignment Teacher; and Francisco Ligero and his colleagues at Catedrático de Universidad.

This book would never have been possible without Nate, Will, Otto (and the clandestine GMBudz), Lianne, Felix, and Oliver whose creativity and comments led to interesting and productive discussions. Last but not least, thank you Mrs. Staab (and Linda Smith!) for your encouragement.

Table of Contents

Chapter 1

The Synthesis of Biotechnology and *Cannabis sativa*

"Let your life be guided by your imagination."
-Mrs. Staab, 5th grade teacher
Washington Elementary School, Hays, Kansas

Biotechnology and cannabis

Plant biotechnology is one of the most fascinating fields of science to have ever been discovered. Admittedly, there have been other times in history where a new science changed the world. Galileo brought the powers of the telescope into public view, Newton explained the diffraction of white light into different colors, and Darwin put forth the driving mechanism of evolution called natural selection. With the insight of these great thinkers and their countless colleagues, society changed and grew. With each discovery, manipulating nature and improving our understanding of natural laws, humans reached a higher consciousness. Thus humans have come to play a more distinctive role and interact in new ways with their environment. The genetic manipulation of cannabis might offer the ability to better direct our species while also improving the quality of human life, much like what Galileo, Newton, Darwin, and other scientific thinkers accomplished.

Biotechnology is a field that obviously includes biology, the science that encompasses the study of life. However, the technology part might be a bit more vague to some. To compound matters, not all technology is consistently recognized as good. While some people might recognize a technology as safe, others consider the same technology unsafe. With plant biotechnology, the technology part could best be described as implying that there is application or engineering of some

scientific knowledge. Therefore, plant biotechnology is simply the use of scientific knowledge or discovery to generate, by application of that knowledge, useful plants. In the context of molecular genetics, plant biotechnology often refers to the manipulation of the DNA sequence of a plant species in order to produce a more desirable end product. As we will see in the upcoming chapters, there are several ways in which the manipulation can be done, some more widely used than others.

The efforts from plant biotechnology research continue to pay out billions of dollars and change billions of lives. With imagination and long hours of laboratory work coupled with the savvy entrepreneurial skills of corporate personnel, plant biotechnology has, arguably, changed the world. However, whether the change has been welcomed by everyone is a different discussion. Certainly those who have intensely studied plant molecular biology try to accurately describe the plethora of benefits that humans have already gained and that are destined to be revealed through time with plant biotechnology.

While this book focuses on genetically enhanced cannabis, the subject of genetically modified crops should not be altogether ignored. Not only because transgenic crops are having an increasingly important impact on our world but because the same approaches for genetic modification in crop species are also necessary for genetic modification of cannabis.

Credit is often given to transgenic crops for relying on less pesticide use, but there might be several other benefits for the environment as well. Certainly, plant biotechnology has had a direct positive impact on decreasing the use of fossil fuels[1] that are needed to plant, maintain, and harvest crops. More farmers seem to be turning to the benefits of genetically modified organisms (GMOs) and their seeds to increase crop productivity and help ensure consistent yields in each harvest. Several guidelines for laws governing the use of genetically modified seeds have been proposed[2] as more farmers around the world are becoming familiar with the agricultural technologies of the 21st century. Thus, as time progresses, more and more people are learning about plant biotechnology, gene patents, and the evolution of farming.

There are extreme opinions on both ends of the plant biotechnology spectrum. The passion of each is obvious. One individual might be ready to hold a pro-GMO sign and spend an entire day in front of a building where researchers are exploring the latest genetic technology. Another person might distribute anti-GMO flyers while also spending their day in front of the same research building. In each case, is time being spent in a useful way? In what other way might people want to work towards understanding this rapidly growing scientific field?

Such questions come to mind as one might peruse the popular literature, libraries or web pages and notice the copious amounts of technical jargon and the extent of the biotechnology industry. Many plant biotech companies invite anyone to read their web pages and learn of the fascinating and beneficial applications of plant biotechnology while other pages attempt to persuade the reader about how the natural world will cease to function if plant biotechnology is allowed to continue. Which side, pro or con, should a person join? The best answer might be to sidestep the acrimony and learn the science driving the change. In other words, arguments and opinions aside, the plants that we eat, wear, and otherwise consume are changing. We should try to understand the change.

To understand how plants are changing, one must become familiar with the science of plant biotechnology. An introduction to molecular details of plant biotechnology is certainly of importance but prior to a thorough foundation of molecular details, one should have a balanced perspective on the biological sciences as a general discipline. This is important since plant biotechnology requires discussion of genes, cells, and organisms. Changing the genetic composition of a plant using the tools of biotechnology might influence the organism by altering its cellular processes, which is often considered a fundamental concept in the biological sciences.

In addition to basic biology, equally helpful in understanding and pursuing plant molecular biology is an imagination. Almost anyone can read about how organelles like chloroplasts or ribosomes function. Yet, how many of those readers close their eyes and picture themselves riding on an mRNA molecule zipping through the amino acid side groups and rRNA comprising the 50S subunit? Can those same readers envision themselves floating in the matrix of the mitochondria, dodging the suspended ions and super-huge proteins? Clearly, reading about the science of plant biotechnology is necessary to acquire knowledge but the science in and of itself cannot always open the human mind to imagination.

Of course, with cannabis, imaginative thoughts can happen quite rapidly. Almost so rapidly that people who experience such rapidity in thought and generation of ideas, quickly lose the ideas almost as fast as they were acquired! The first thing a person should do upon recognizing that their mind is something of this sort, is reach for a pen and paper. Then, after several hours have passed (or upon awakening the next morning), the toker who has come back down can read through their ideas and realize for themselves either the outlandish thoughts they were floating in their head or the incredible insight that flickered from their mind to paper. After time one will soon realize that the progress in research and

application is a result of having a solid education coupled with an imaginative vividness, the latter sometimes stimulated with the very plant they might one day genetically modify. Blending knowledge and imagination with the act of doing (science!), makes for a great recipe to propel a student, and the human species, further along an evolutionary trajectory.

The case at the present time is that technology exists to manipulate a plant's genome. With each passing year more and more people realize that plant biotechnology is not going away. Dozens of plant species have been genetically modified and nearly 10% of the global food supply is genetically modified.[3] That is not necessarily a huge concern to most consumers; one must consider the demands the human population is making on the earth. Some have argued that biotechnology is needed to increase food production as the human population grows.[4] Others have argued that research money should be directed towards getting people to buy locally produced, non-genetically modified foods.[5]

As the arguments continue, the list of plants that are being genetically modified is growing, not slowing. Unfortunately not all plants are given the same attention in the world of plant biotechnology. If one were to walk through a library and find a book on weed genomics, plant biotechnology, or crop biotechnology they might be disappointed to discover an almost complete lack of mention of the world-famous plant, *Cannabis sativa*. If they search hard enough, one might find brief references to cannabis. As of recently, a young student or curious mind can now find a well-written book chapter.[6] However, more is needed.

There are many books about other plants regarding biotechnology including the biotechnology of maize, biotechnology of ornamental flowers, genetic manipulation of the lesser-known crops, modification of cotton, virus biotechnology, genetic modification of bacteria, and biotechnology of fungi. There are even books on the biotechnology of algae and trees. Cannabis, in the context of plant biotechnology, more than deserves a place among other books pertaining to this awesome science.

Cannabis the plant

Cannabis sativa might just be the most famous plant species on the planet. While some people have ignored this mysterious yet beautiful plant, cannabis has received an overwhelming amount of attention from others. Since early civilization, humans have been artificially selecting food plants like maize, peaches,

pears, wheat, and rice to increase quality and yield. Yet, along side these popular foods, cannabis plants have also been selected, crossed, grown, and harvested. The founder of ethnobotany Richard E. Schultes suggested that cannabis has always deserved more positive recognition than the plant has received, since humans have been using cannabis products for 10,000 years or more.[7] More recently, Robert C. Clarke and Mark D. Merlin have provided an extensive description of the amazing history of cannabis in their book, *Cannabis: Evolution and Ethnobotany.*[8] Of course, the scale of importance and outright visibility that cannabis has played in societies varies across the world.

Regardless, the botany of cannabis has been studied for several millennia. Wild cannabis that was first gathered and consumed by people undoubtedly produced a pleasant taste, perhaps even a euphoric feeling-since humans have continued the breeding tradition for so many thousands of years. Over that time the overall features of the plant have changed little, still having palmately compound leaves and amazingly tough fibers. Although the use of its fiber dates back several thousand years, a recent study suggests cannabis was being consumed 2,700 years ago by shamans in central China.[9]

One can imagine the impact cannabis made on its original discoverers. In a time when plants were constantly being investigated for their medicinal properties, cannabis would surely have captured attention. The various biochemical pathways that give way to the diversity of cannabis metabolites are something of a divine quality, as must have been thought by the first discoverers. Even the uncultivated, "ruderal" variety (Figure 1) has a distinct and unforgettable odor.

Figure 1. *Cannabis sativa* 'ruderalis' (left and right) shares phytochemical properties with cultivated cannabis but has genetic variations that might be useful to plant biotechnologists.

Importantly, plant biotechnologists and breeders are now looking to wild relatives for genetic variations that might be useful in developing new varieties of plants.[10]

The physiology, genetics, ecology, biogeography, and classification of cannabis are areas that are constantly being reworked and refined. New discoveries are being made with the help of new tools that have just recently become available, some within the last few years. While the botany of cannabis might not be changing in apparently dramatic ways, the information being gained from the newest tools has added new insight. Many of the new tools revolve around the field of molecular biology, the science that seeks to understand the molecules that interact with, and macromolecules that comprise, life. Molecular research continues to expand and enhance our view of life processes and is now helping to explore the curious cellular activities of cannabis.

One of the more popular applications of plant molecular data is molecular systematics. Although important in other kingdoms of life (e.g. animals and fungi), plant material in the form of leaves, fruits, or stems is readily available in hundreds of herbaria for millions of specimens. Herbaria have traditionally served as a place where students, citizens and educators can investigate a "hard copy" of a specimen under discussion. Herbarium specimens also serve as vouchers in research projects and give tangibility to a verbal description of a species. With the advent of molecular techniques and the ability to analyze specific DNA sequences, more insight is rapidly being gained from questions with an evolutionary context that rely on samples derived from herbarium specimens.[11]

Similarly, researchers can also explore and hypothesize the evolutionary relationships of cannabis to other members of the same plant family (Cannabaceae) such as *Humulus lupulinus*, the flavoring and preservative source for many beers. Important to science, DNA sequence data analyzed with computer programs gives biologists the opportunity of evolutionary placement of a species within a group of organisms. The result is a phylogenetic tree, an evolutionary diagram that can lead to new proposals on what related species might harbor interesting secondary metabolites. Additionally, tree-constructing algorithms can help with comparisons among different cannabis strains and rapidly determine parental lines and relationships.

Flow of information and genetic components

While DNA sequences can help determine relationships to answer research questions pertaining to evolution, DNA is also crucial to the cannabis plant for production of proteins. These in turn largely dictate how the plant will develop over time, respond to environmental fluctuations, generate metabolites, and maintain important cell processes. Proteins have a diverse array of functions that include carrying out daily metabolic processes and facilitating cell reproduction. Yet, before a protein is made, the DNA sequence needs to be read in a particular frame, an orderly process where biochemical laws govern. Some proteins function to serve as machinery that will help read the DNA to make an intermediate messenger transcript, which will exit the nucleus and then couple with the ribosome. The ribosome is reminiscent of a sewing machine, stitching together little amino acids into a continuous string, which will then fold onto themselves, sometimes with help from other proteins and sometimes simply guided by biochemical interactions of the amino acids side groups.

Cannabis genetics follows similar genetic principles shared by all other species. Humans have always known traits are transferred but were unsure how, since the molecular mechanism was a mystery. Although Gregor Mendel was the first to study single-gene traits in peas, others had come before him but studied polygenic traits, something much too complicated to understand at the time since several genes are involved. In the early 1900s, years after his death, Mendel's scientific papers on pea breeding were discovered and the mystery of inheritance was slowly unraveled. Maize, wheat, and cannabis breeders are now well aware that there is importance in crossing parents of particular phenotypes (appearances) and chemotypes (chemical makeup) to produce progeny with desirable traits. Indeed any farmer or gardener knows today that seed genetics might be incredibly influential in determining the success of a crop.

Modern molecular genetics has helped determine gene-specific processes such as regulation of biochemical pathways and gene expression patterns. Molecular methods have resulted in the ability of eliminating the function of some genes and site-directed insertion of other genes. Particular bases can be altered within a gene to optimize expression or produce a slightly altered protein, perhaps with higher specificity for a substrate. Such molecular modifications might be exactly what Watson and Crick implied might be inevitable in their 1953 paper describing the structure of DNA.[12] Molecular genetics is now something all plant

biotechnologists become intimately familiar with as they progress through their educational path.

Thus, plant biotechnologists do not completely rely on simply crossing a male and female to result in a genetically improved plant. Further, precise manipulations of particular DNA sequences external to an organism imply a technology is needed to deliver or (re)introduce genetic material. In other words, plant biotechnologists first manipulate a gene in a small test tube by using a series of techniques and then deliver the genetic material to living plant tissue or individual cells. Such methods sometimes rely heavily on tissue culture, a method where plant tissue is grown on sterile material that functions in place of soil. The media often contains essential micro and macronutrients as well as sugars and hormones that the plant cells need to grow and develop. Tissues grown in vitro ("in glass") can therefore be coaxed into forming tissues that are sometimes a necessary part of generating a genetically modified plant.

Timing the delivery of the genes either before tissue culture or during the tissue culture process largely depends on the investigator and hypothesis being investigated. The delivery process can also be, at least to some degree, left to the discretion of the researcher. Several methods of gene delivery have been suggested and tried and not all methods require tissue culture. However, only a few methods are used with such frequency as to consider them useful to the majority of researchers. The method for gene delivery used by most plant biotechnologists working in a typical lab relies on the infectious capabilities of a bacterium known as *Agrobacterium tumefaciens*.

While the use of bacteria to deliver foreign DNA in the laboratory has been occurring for more than 30 years,[13] such gene transfer has been occurring in bacteria as well as eukaryotes long before humans.[14] Humans have simply made the process of gene transfer more efficient and precise using the tools of biotechnology.

Cannabis in the genomics era

While individual genes have been manipulated for several decades, there are now studies arising where the entire set of genes of an organism, the genome, is being sequenced, analyzed, and manipulated. There are certain advantages of single-gene investigations, genomics has its own advantages, one of which is throughput. Today a person has the option to study hundreds of genes

simultaneously and assess their expression levels at either developmental times or in response to biotic or abiotic stimuli.

Similar to classical genetics, genomics has its own tools to study the genomes of organisms. These include relatively popular gene expression assays that take a global view of the genome, where all genes can be analyzed simultaneously. Genomics also relies on methods to gather sequence data and precisely determine the exact position of genes on chromosomes, which in part leads to determining the structural architecture of a genome. Thus, genomics has been divided into more specialized areas known as functional and structural genomics.

What is very exciting is that with the application of genomics, one can now study the entire set of genes in an organism. Some have argued that such a large-scale approach to answering a biological question has drawbacks. For example, a person might become inundated with so much data that inferring meaningful results might be difficult. The "problem" considered by some is that with so much information, what does a researcher emphasize in their report? If information is available for all genes in a genome, what should be the focus and highlighted in the results portion of the report? The answer is, it depends. One person might be more interested in stress response pathways and another might be interested in a biochemical pathway that leads to a final metabolic product, a metabolite such as delta-9-tetrahydrocannabinolic acid.

Genomics, in addition to investigating metabolic pathways leading to plant products such as cannabinoids, has allowed researchers to generate novel questions about how altering a single gene might affect other genes in a genome. Such investigations into the interrelationship of multiple genes would not be possible without the rapid changes in DNA and RNA sequencing technologies of the last decade. Huge leaps have been made in sequencing technology. As a direct result of better technology, the cost to sequence a nucleotide base has dropped in the last decade by almost 100,000 fold.[15]

Researchers also now have several sequencing technologies from which to choose. With the various technologies there are benefits and drawbacks. For example, with one sequencing technology a person is able to obtain sequence data for genome regions that are considered difficult to sequence. However, the sequences generated can be accurately placed onto a genomic map, which is similar to an architect's blueprint. Using a different sequencing technology one might obtain more data that is restricted to easily-sequenced areas, such as regions that are important in coding for proteins. While the sequencing cost might be a bit

less, determining the placement of the sequences onto chromosome maps might not be as feasible. In many instances, the type of technology and methods used to obtain molecular sequence data is reflective of the goal of the research project.

While cannabis genes that are responsible for some cannabinoid pathways have been intensely studied for decades, only recently has genomic insight been gained. To the excitement of many, the first draft of the cannabis genome has been obtained, shared, and is now publically available.[16] Thus, a distinct improvement that genomics has provided over simple, single gene studies is the whole-genome view. Genomics offers a perspective from the entire set of genes, while in some instances a select group of genes can be analyzed. The selected genes can range from a few dozen to a few thousand. Since the metabolites that cannabis makes is not limited to delta-9-tetrahydrocannabinolic acid, application of genomic techniques to a particular cannabis strain might reveal which genes are up-regulated (turned on) in a metabolic path, helping to understand the complete gene expression (transcriptional) profile. Many cannabis researchers are now aware there are numerous enzymes that produce the more than 500 interesting metabolites that contribute to the flavors, aromas, and effects of cannabis.[17] Thus, genomics has given way to metabolomics, the study of the metabolic products of an organism.

Engineering cannabis metabolites

Since the new field of genomics has given researchers the ability to study entire metabolic pathways at once, products of a pathway of interest can be investigated in a broader fashion or manipulated with more meaningful results, not typically available in single gene studies. If a researcher finds that a gene belonging to a metabolic pathway has lower or greater expression, something might be inferred about its role in the pathway. This in turn provides several routes of exploration. A gene that shows low expression can be up-regulated, using the various biotechnological approaches. A particular gene can also be down-regulated using similar tools, if producing less gene product is indeed the desire.

Knowing which gene is normally down-regulated in a metabolic pathway might lead to determining a rate-limiting step of a pathway. Thus, an end product that is highly desirable might be obtained by either up-regulating a gene or adding an intermediate metabolite into growth media. Sometimes, depending on the final metabolic product desired, intermediates can be found in abundant supply since

they are sometimes chemically synthesized with ease in vitro. In addition, simply perturbing a cannabis plant genome by impregnating growth media with a particular chemical, or ensuring its absence, is of great interest to metabolomics.

Perhaps more interesting than learning about which genes are being turned on or off is the ability to transfer one metabolic pathway into another organism. The feasibility to transfer genes of a complex organism such as a plant, into a relatively simple organism such as *E. coli*, is a common question that often arises in cannabis biotechnology discussions. Certainly the question is a valid one. Researchers have been completely bypassing the use of a plant system to produce cannabinoids and have used bacterial systems for cannabinoid production for decades. The latter is faster in its growth rate, easier to grow, and much simpler to manipulate genetically.

However, there are important differences in the molecular mechanisms of protein production between eukaryotes like cannabis and prokaryotes such as bacteria. Much relies on the central dogma, where DNA makes messenger RNA, which is translated into protein. Some genes of cannabis can be made into messenger RNA and a short time later, become chopped in a certain way that leaves specific portions available for translation into a protein. Interestingly, that same messenger RNA transcript might be spliced yet another way, that is, different regions of the transcript can be removed, resulting in a different protein product. Splicing messenger RNA (mRNA) to yield different protein products is something that is largely limited to eukaryotic cells. Therefore, one consideration is to decide which spliced variant of the gene, which ultimately affects the mRNA product, should be used in bacterial expression systems.

Protein stability is yet another concern for the plant biotechnologists. While different RNA molecules have different life spans, proteins also show various rates of decay within a cell. Some proteins are made continuously while others are made only at distinct life stages of an organism. As anyone who has grown cannabis knows, proteins involved in the synthesis of distinct aromatic compounds and cannabinoids are only more highly expressed during the flowering stage of growth. Female cannabis plants produce different amounts and types of metabolites, suggesting different enzymes are involved in the pathway of those products present compared to the metabolites found in the male cannabis plant.

Modern plant biotechnology offers the ability to manipulate proteins of metabolic pathways, even those present within female flowers, so they are abundantly present in all stages of development. That is to say, plant biotechnology allows for over-production of cannabinoids by up-regulation of

genes involved in the metabolic pathways during earlier stages of the cannabis life cycle. Desired cannabis metabolites no longer need to be limited to a developmental stage or flower type. Using the tools of biotechnology, a young male cannabis plant could be engineered to be completely saturated with cannabinoids.

Cannabis in the informatics era

The first step in generating any transgenic plant is to consider the preliminary research needed. Much of the preliminary research being done, like all other fields of science, now relies heavily on the use of computers. Over the last 20 years, computers have been of increasing importance in biology. Margaret Dayhoff, who used computers to store protein sequences in the 1970s, was one of the first to implement computers in biological research.[19] The use of computers in biology has evolved into an awesome field of study called bioinformatics. A few applications include genome analysis, protein visualization, and altering individual bases in a gene.

Computers, at a very rudimentary level, consist of hardware and software. Both have influenced the rapidity in the storage, analysis, and discovery of information. Many people are familiar with the computer and its primary hardware components that include the central processing unit (CPU), random access memory (RAM), and hard drive. The software component consists of programs, some more easily understood than others. Programming in bioinformatics means giving the computer instructions, words that make sense to the computer. Biotechnologists are not always concerned with the programming side of bioinformatics, although interest is growing.

One should come to recognize how bioinformatics has directly influenced important aspects of human society such as medicine or agriculture. However, bioinformatics has also given us the tools to work with databases, which have become crucial for information retrieval and the subsequent comparison of data to lend insight into a biological question. Two sequences can be compared in pairwise sequence alignment or several sequences can be compared in multiple sequence alignment. Different algorithms embedded within software can be used to optimally align regions or entire lengths of sequences, depending on the extent of the evolutionary divergence or question of interest.

Bioinformatics software is also routinely used for predicting domains within

a protein sequence. Knowing the secondary or tertiary structure within proteins can be useful if hypothesizing something about the cellular role the protein might perform. Prediction of genes can also be done, if one is sifting through raw sequence data, and is often based on something previously discovered. As we will see, particular regions and specific sequences that flank genes often help guide protein machinery interacting with the DNA. Determining these flanking regions can be useful when considering expression of introduced genes.

Entire macromolecules, such as enzymes responsible for the biosynthesis of cannabinoids, can also be visualized and compared. The proteins can be manipulated in silico (on a computer) to view how an active site might differentially react with a substrate. Such a tool is highly valuable, since interactions between drugs and proteins are incredibly important to drug manufacturers (and drug consumers). Two proteins can also be superimposed to determine the extent of their structural similarity. Similar to tissue culture, genomics, and metabolomics, bioinformatics is only one part of a bigger piece that is changing how we understand and manipulate cannabis. New areas of biology such as systems biology and synthetic biology are also parts of a bigger picture that are changing how we view what might be possible in the life sciences. Therefore, all of these subjects are discussed in the later chapters of the book.

On a superficial level, putting genes from one organism into another organism seems relatively straightforward. However, there is little guarantee that a plant harboring a newly introduced gene will adequately express that gene. This can be due to several factors, one of which is the plant making use of coding sequences derived from another species. Thus, computers can help with changing the coding sequences of a gene so the gene is more likely to be properly expressed when in cannabis.

Although bioinformatics has helped overcome so many hurdles in biotechnology, new questions are continually arising. Will mastery of cannabis through molecular biology and biotechnology lead us towards experimenting with other, potentially more potent plants? Might there be a book or journal article entitled, "The Biotechnology of *Erythroxylum coca*" that is on the verge of being published? Might we be headed towards a slippery slope of plant modification? If so, should that stop people from generating genetically enhanced varieties of cannabis? If one person is prohibited from genetically modifying cannabis, is there any guarantee that another person will not modify cannabis? Perhaps these are questions that will be answered in the next few years.

The next generation of cannabis

There are a variety of educational avenues for acquiring biotechnology skills and knowledge on cannabis science but the effort takes time and dedication. The student of today's cannabis is not so different than the first cultivators of the wild type marijuana. Repeated observations, careful studying, persistence, and dedication are the core basic requirements.

What remains so amazing in privileged societies today, are the excellent opportunities to pursue higher education with unrestrained passion. Boundaries and limits are almost nonexistent on the bookshelves of university libraries, since many institutions offer interlibrary loan material, a way to rapidly obtain nearly any book (or research article). The Internet also allows, with increasing requirements of judiciousness, more than a lifetime supply of useful information. This is what the institutions of higher education are for; to serve as a conduit in order to increase one's knowledge in a direction that is rewarding and pleasurable. The idea is that perhaps the individual can give something back to society at a later time. Perhaps in the form of a research article or book of their own.

Indeed, every student is an investment for the future generations and although one might not realize so at the time, they can be the one who will go on to have world-changing impacts. Initially, there is an upfront investment. All students must face the decision of shunning particular social groups and embracing new ones. Those who are able to do such are in turn able to pursue knowledge, books on a diversity of subjects, fill their minds with history, and exercise their imaginations. The library patrons and dedicated students become the ones that are able to learn about the diversity of plants, seemingly invisible microbes, and the molecular underpinnings that unite all life.

Higher education allows for years of studying any subject. Who could pass up plant biotechnology? To be able to learn everything there is to know about how to genetically modify and then grow any plant to manufacture a given metabolite, which might change the world. Perhaps one could eventually manipulate a genome of a tree to yield branches that form a house and that also bear a constant supply of fruits. Branches might bear food akin to a salad bar and might regrow everyday, right in a tree house kitchen. Could a person engineer an avocado tree that yielded a thirteen kilogram avocado if the tree were given the right amount of water and sunlight? How about a banana that is adapted to grow in cold climates, peanut butter-flavored celery, or cheese-flavored broccoli? With a little imagination, one can see how plant biotechnology is set to change the world. We

simply need more people to envisage the endless possibilities and act on their ideas.

The process of obtaining the knowledge and learning to exercise one's brain might be made more complex by the use of cannabis. The combination of learning and cannabis consumption generates questions in one's mind. In a general sense, what might be made? What are the limits? What gene might be put into cannabis that would improve its flavor, aroma, stability, or appearance? Did early humans ask similar questions thousands or even millions of years ago as the first generation of growers and consumers? Evidently, humans have always continued to seek a better cannabis plant. The qualities of contemporary cannabis are remarkably different from wild type relatives, as some have argued that certain strains of cannabis should be a distinct species.

Compared to those early growers of cannabis, the subsequent generations included traditional breeders, people who were selecting for a particular metabolic profile. Such breeders still exist and are highly important to the present-day cannabis industry that has flourished. Thus, breeders are those who have especially attempted to generate a better plant. Considering the popularity of cannabis, they have obviously succeeded.

At the moment we are at a turning point in cannabis evolution. The next generation of growers has the opportunity to improve cannabis so that the plant might have a strikingly different, perhaps unrecognizable appearance. We are now able to use cannabis plants that were so carefully bred and directly manipulate the genome. We are able to dictate what enzymes of a particular metabolic pathway are produced in abundance or alter enzymes to have an increased affinity for a substrate.

Some people who pursue cannabis biotechnology are likely to be alone for a while, but others will soon understand the power of genetic enhancement. If they do not, the science of cannabis biotechnology will still move forward. The next generation of cannabis growers and consumers are likely to be the bearers of the new cannabis science.

What is to be said of any growers of cannabis several decades from now? In any area of study, the future is difficult to predict. However, clearly without the advent of cannabis biotechnology and scientific experimentation, the full potential of the cannabis plant might never be realized. Cannabis biotechnology needs young and ambitious scientists. People who understand that science itself is a process that is based on knowledge gained by several investigators, sometimes over a very long period of time. We need people who will be bringers of ideas, shared ideas that are sometimes argued about, but that nonetheless slowly change and

improve over time. Such a process has led to the present-day emergence of the biotechnology of *Cannabis sativa*.

Biotechnology considerations

There have been many misconceptions about genetically modified organisms but public attitudes, since they are often flexible, are apparently changing. Becoming familiar with the science of DNA modification has helped people realize there are many concerns about genetic modification that might be unwarranted, while others are perhaps very important. For example, some are concerned that transgenic cannabis will result in decreased varieties. Intuitively, adding new genes to a gene pool will only add to the present varieties currently in existence. Transgenic plants are generated from existing plants, but those original plants are not lost, they still exist.

There also exists a belief that plant biotechnology might generate the potential for genetic pollution. That is to say, the unintended introduction of a gene not normally found in a particular species or environment. One should certainly remain objective and consider the possibility that an introduced gene might indeed introgress into a wild population. What are the chances and what might follow if such an event were to occur?

One of the most interesting feats that biotechnology has provided is the demonstration that plants have the ability to integrate genes derived from completely different species, including animals. Although many agrobiotechnologists have been interested in transferring genes responsible for herbicide tolerance, an experiment that captured public attention was the "glowing" tobacco plant. Might a glowing cannabis plant also capture the attention of the public? Would a fluorescent cannabis plant be highly undesirable?

In spite of the concerns of decreased varieties, the potential for genetic pollution, and strange phenotypes such as glowing plants, cannabis biotechnology is ready to move forward. Although some ethical questions are certainly unavoidable and should be addressed, the reality is that often scientific advancements are made first and reflection follows. There are always costs and benefits with any scientific technology. Certainly this is the case with genetically modified plants. Towards the later chapters, we will look more closely at some of the concerns of genetically modified cannabis.

Whatever opinions exist, our current abilities to manipulate life fused with the wonders of cannabis, leave ample room for countless new experiments and explorations. Little difference exists in how the genetically enhanced cannabis is generated, whether in academia, by governments that facilitate such research, under the guidance of private companies, or in secret laboratories. What seems most important is for all growers and consumers of cannabis to be aware of the changing landscape that has previously defined *Cannabis sativa*. Let us begin our investigative journey of cannabis biotechnology by looking at some botanical characteristics of the cannabis plant that might be chosen as candidates for genetic modification.

Chapter 2

The Botany of *Cannabis sativa*

"It is a plant-this thing that we are about to discuss: a green plant, a very abundant and ubiquitous plant, an unusually valuable economic plant, possibly a dangerous plant, certainly in many ways a mysterious plant."
-Richard Evans Schultes[1]

Water and early life

The plant sciences include genetics, physiology and anatomy, biochemistry ecology, systematics, biogeography, economics, and, of course, biotechnology. With so many subdisciplines, one can easily get overwhelmed with information and suffer from a sort of botanical overload. However, a person appreciates a plant species all the more if they have an understanding of the complex series of events, random yet sometimes seemingly directed, which have led to its current position among the millions of other species on the planet.

Cannabis, a complex eukaryotic organism, arose long after other life began and evolved on our 4.5 billion year old planet. The first simple prokaryotes, organisms that lack a nucleus and include the often single-celled organisms like bacteria, took about a billion years to evolve after the formation of the earth.[2] Prior to bacteria there were simple "protocells" and even earlier were self-replicating molecules, freely floating in aqueous solutions.[3] Perhaps the solutions were puddles, stranded on young, barren rock jutting over a gigantic ocean. Or, perhaps, the aqueous solution was an enormous body of water; a single, super oceanic body of water.

Whatever size the body of water was, almost all people who study the formation of early life on earth agree, water was essential. Water allows dissolved solutes to interact with one another. At the molecular level water is a three-atom molecule comprised of two hydrogens and oxygen, which has the well-known molecular formula of H_2O. If one could view a water molecule and get up close and personal, it would look somewhat like a triangle, have well-rounded edges due to the electron clouds, and the two hydrogens would symmetrically bend towards one another, away from the unpaired electrons of oxygen.

A water molecule has the unique ability to bond with other water molecules due to the slight charges on each of the three atoms. The oxygen is slightly negatively charged and the hydrogens are slightly positively charged. The charge difference, or polarity, is caused by oxygen having a higher affinity for the electrons it shares with the two hydrogen atoms, which is referred to as electronegativity. The partial negative charge on one water molecule can interact with the partial positive charge on a different water molecule, resulting in a hydrogen bond. Billions and billions of hydrogen bonds are occurring in a single drop of water at any given moment. The duration and stability of the bonds is a function of temperature and are inversely related; higher temperature water has less stable hydrogen bonds.

In some plant cells, water comprises 90% of the volume and for good reason; water has several features that are important to life processes. Water has cohesive and adhesive forces, a high heat capacity and heat of vaporization, and is less dense as a solid than when a liquid. Water is also fundamental as a solvent. Solutions, important to early life on earth and present-day life, are (often) comprised of water with dissolved solutes. Sodium chloride, table salt, is one of the most common examples of a solute. On early earth, ions such as sodium, chloride, and organic molecules that were able to diffuse into water were all present. Different gases were also present in different concentrations, with oxygen being far less abundant than carbon dioxide.

Understanding the origin and evolution of cannabis really began with experiments in the 1950s that sought to investigate the origin of life. During that time, researchers showed that when mixtures of water vapor, hydrogen gas, methane (CH_4), and ammonia (NH_3) are placed into a sterile vessel and given an electric spark, within several days simple organic molecules that are typically associated with life are generated.[4] Early on, when such experiments began, several amino acids were easily generated. Later, when different starting ingredients were added and the experiment was performed again, additional types

of amino acids were detected. Through still more variations on the theme of early life experiments, results have shown how lipids, specifically a variety called phospholipids that help form cell membranes, also might have formed.[5] More fascinating was that other experiments with yet still different starting molecules, showed that nucleotides could be synthesized. Nucleotides are part of the DNA and RNA used in cells to store and transfer information, respectively.

Presently, using tools that can help identify chemical signatures in interstellar clouds, researchers have identified adenine as a highly abundant molecule in space. Adenine is one of five well-known nitrogenous bases of DNA and RNA but adenine can also function as a cell signaling molecule called cyclic adenine monophosphate (cAMP). Adenine is also a central component in the energy molecule known as adenosine triphosphate (ATP), which is the energy currency of all living organisms. The adenine in interstellar clouds is likely formed from hydrogen cyanide (HCN) with the help of cosmic radiation. Adenine has the molecular formula of $H_5C_5N_5$, implying that five HCN molecules can be chemically rearranged to form adenine. Indeed, heating HCN for an extended period (refluxing) results in adenine. If adenine is so abundant in interstellar space, this nucleotide might have also been abundantly supplied on early earth and thus, have been present very early in the history of life. Such a situation might explain why adenine is so abundant in cells today.

Knowing how some of the first prebiotic molecules formed lead to help us understand how ancestral cells of cannabis might have arisen. Based on what we know about chemistry and biochemical laws, the origin of life via organic chemistry is likely, especially if given enough time. Since we know the earth is roughly 4.5 billion years old, time is a component that has been in abundant supply.

A few hundred thousand years after the formation of the solar system and earth, the planet was much cooler and water vapor that was present had largely condensed and formed oceans. Under the fertile energy provided by constant lightening storms and another billion years, molecules had even more time to react. After even more time, simple cells evolved, which replicated and made new cells. Early cells might have been simple sacks of lipids that contained organic molecules and nucleotides. Lipids naturally associate into spheres when in aqueous solution, something easily seen in oil and vinegar salad dressing. Inside early lipid spheres, catalytic strings of nucleotides were the early progenitors of today's polymers of nucleotides but functioned much like enzymes function; they helped

reactions move forward. Descendents of early catalytic nucleotide strings can be observed in ribosomes, which have an important role in protein synthesis.

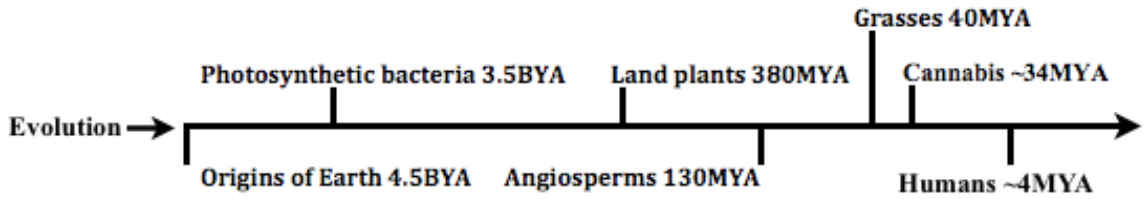


Figure 1. Timeline showing cannabis evolution, hypothesized to have evolved about 34 million years ago.[6] Recent molecular data might suggest an earlier evolutionary time for plants that use flowers to reproduce (angiosperms).

Certainly part of the role of such catalytic molecules residing in small lipid sacks was replication, generating slightly more complex cells. Natural laws suggest that simplicity combined with more simplicity results in complexity. Such was the case with early cells, which led to simple cells (bacteria) being firmly established. Bacteria multiplied and filled the oceans, stripping electrons from inorganic and organic materials. The electrons were used in generating a primitive proton motive force across their lipid membranes, which functioned as a type of potential energy for the cell. Thus, more and more molecules that were present on the early earth such as organic molecules, methane, and ammonia could be recombined in countless ways, giving way to even more complex molecules.

With countless bacterial cells floating in the ocean, evolutionary principles were well in effect. Only the best-adapted bacteria that could exploit their surroundings, obtain energy, and grow were able to persist and reproduce. Some of the best survivors were mutants who had synthesized a molecule that serendipitously interacted with sunlight. As time progressed, early photosynthesizing bacteria adapted and were able to remove electrons from water. They also used their lipid bilayer to keep protons concentrated, generating a charge difference (potential energy), which could then be used to drive cellular reactions. The oxygen that resulted from oxidized water came to be present on earth in great quantities as early photosynthesizers known as cyanobacteria became more abundant.

Similar to early photosynthesizers, cannabis chloroplasts use energy from sunlight to reduce carbon dioxide and integrate the carbon into a form of "fixed" carbohydrates. Initially the fixed carbon is in the form of glyceraldehyde-3-phosphate, which can be joined together in further reactions to generate larger carbohydrates. Alternatively, glyceraldehyde-3-phosphate can be used in

constructing plant secondary metabolites such as terpenes, precursors to cannabinoids. Early cells that could use the sun to split a water molecule for its electrons became what are known today as oxygenic photosynthesizers since oxygen is generated as a byproduct. Today a person can use various techniques to observe cannabis chloroplasts performing photosynthesis much like their ancestors first did, billions of years ago.

Macromolecules of cannabis-and all life

Photosynthesis was not the only metabolic process that was under evolutionary pressure. Organisms produced defense compounds, signaling molecules, and several other metabolites. The diversification of complex metabolic pathways eventually gave rise to the tens of thousands of different metabolites characterized to date. Metabolites are sometimes divided into primary and secondary metabolites. Secondary metabolites are what many cannabis smokers are interested in since they include terpenoid derivatives such as cannabinoids. Primary metabolites are more crucial for cell survival and general functions and can be segregated into four major categories based on their cellular function. The categories, which include proteins, nucleic acids, lipids, and carbohydrates, are the macromolecules of life and serve as a basis for the biological sciences.

About twenty amino acids can usually be detected in almost all cells and resemble one another in regards to having an amino group and acidic group but each having a slightly different side chain. Bacteria show a bit more diversity in their amino acid profile compared to eukaryotes while meteorites from space show even higher diversity. An amino acid is considered the monomeric unit of proteins. Through loss of hydrogen from one amino acid and a loss of an oxygen and hydrogen from another, water is formed in a condensation reaction and two amino acids are joined together. Strings of amino acids that constitute proteins might function as enzymes important in cannabinoid biosynthesis pathways, form cell structures like the spindle fibers that help in cell division, and form more obvious structures like your fingernails and hair, which are made from a particular protein called keratin.

Cannabis and other plants have important proteins called phytochromes. Phytochromes have evolved with the important task of holding a chromophore, a light-absorbing molecule. Phytochromes interact with light to help signal to the plant when to initiate flowering, for example. Subsequently, some phytochromes

can be activated and deactivated by certain wavelengths of light. When phytochromes (P) receive red light (660nm) they become the P_{fr} type, which is active and allow flowering to proceed. If far-red light (730nm) is detected the phytochrome becomes the P_r type, which is a biologically inactive form and so flowering cannot proceed.

An indoor gardener can use this principle to initiate flowering. During the dark period of a plant's life, they can be given a brief pulse of red light. This changes the P_r type into the P_{fr} form and allows flowering to begin. Interestingly, these same phytochrome proteins play a crucial role in seed germination; the P_{fr} form of phytochrome allows germination to proceed. Therefore, if one is having difficulty germinating cannabis seeds, exposure to a short period of red light before sowing might increase germination. Since phytochromes have several additional functions, phytochrome genes might be excellent targets of cannabis biotechnology.[7] However, as we will see there are several other proteins worth targeting in cannabis as well.

The second category of macromolecules is nucleic acids. Nucleic acids include two major forms, deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). The two macromolecules differ in a simple lack of an oxygen atom in the former (hence the "deoxy") but vary quite remarkably in their functions. Both are made of a phosphate, either deoxyribose or ribose, and a nucleotide base. Nucleotide bases come in five flavors and are often seen as the abbreviations A (adenine), T (thymine), G (guanine), C (cytosine), and U (uracil). Nucleic acids are of high importance among the macromolecules since they direct so much of what happens in a cell. The ability to manipulate nucleic acids with molecular precision forms the backbone of plant biotechnology. Much more will be said of nucleic acids.

Lipids are insoluble macromolecules and include fats, phospholipids, and steroids. Fats are comprised of a three-carbon molecule called glycerol and a fatty acid attached to each carbon of the glycerol. Electrons are highly abundant in each fatty acid and can be used in energy-capturing reactions if they are moved along a cell membrane to generate a proton motive force, similar to primitive cells. The cell membrane, also known as the phospholipid bilayer, is made of orderly aggregations of phospholipids that provide a separation of the cell's internal environment to the exterior. Phospholipids also have a glycerol backbone but only have two fatty acids. The third carbon of the glycerol is occupied by a polar phosphate group.

Saturated fatty acids lack double bonds between the carbon chains that make up the fatty acid while unsaturated fatty acids have at least one double bond between carbons in the fatty acid chain. Carbon atoms are stable when they have four bonds, two electrons in each bond. Since some carbons are not all saturated with hydrogens they might share two pairs of electrons in a double bond, and three pairs of electrons in a triple bond. Cell membrane fluidity is directly related to the degree of saturation and one goal of plant biotechnology has been to engineer plants to better cope with temperature by modifying their fatty acid profile.

Carbohydrates are the fourth group of macromolecules and can be monomers called monosaccharides. The most common example is glucose. Disaccharides, like sucrose, have two sugar units. When the number of monomers increases, carbohydrates are generally referred to as polysaccharides. Polysaccharides help make up the rigid cell wall of plants in the form cellulose and store energy in the form of starch. Cellulose is claimed to be the most abundant organic molecule on the planet.

Macromolecule	**Monomers**	**Polymers**
proteins	**amino acids**	**THCA synthase**
nucleic acids	**nitrogenous base, sugar, phosphate**	**DNA, RNA**
lipids	**fatty acids, glycerol**	**fat, phospholipid**
carbohydrates	**glucose, monosaccharides**	**starch, cellulose**

Table 1. Summary of the macromolecules of life with example monomers and polymers of each.

One thing all macromolecules have in common is that they are all generated from monomers to form polymers in anabolic reactions (Table 1). The anabolic reactions are facilitated by enzymes that can be manipulated using the tools of plant biotechnology. Thus, understanding the macromolecules of life is crucial for understanding more complex concepts in cannabis biotechnology.

Cell organelles

Macromolecules are assembled in various ways to comprise organelles, which in turn carry out major cellular activities. The importance of understanding the fundamental levels of the cannabis cell, such as macromolecules and organelles cannot be overstressed, since macromolecules are where plant biotechnology

begins. Some structures, like the cell membrane, are not strictly considered organelles, although the importance is obviously critical to cell stability. The major plant organelles include the nucleus, ribosomes, endoplasmic reticulum, Golgi apparatus, lysosomes, vacuole, chloroplasts, and mitochondria. While plant biotechnology is concerned with each of these in various ways, only a few will be introduced for the moment.

The nucleus is where much of the genetic material for the cell is stored. Inside the nucleus is not just DNA, there are also several different proteins at work, some of which closely interact with the DNA and help keep the lengthy macromolecule tightly compacted. The plant's genes that code for proteins are sequences along the strands of DNA. The DNA does not leave the nucleus. Instead, small bits of an intermediary molecule called messenger RNA (mRNA) leaves the nucleus through small pores in the nuclear membrane.

The mRNA itself is made from a template of DNA in a process called transcription. The message on mRNA is used to make proteins in a process called translation. However, prior to protein production the mRNA must bind with a ribosome. Ribosomes are complexes of proteins and a type of RNA called ribosomal RNA (rRNA). Upon interacting with the ribosome, the mRNA strand can begin slipping through a small opening in the ribosome. As it does so, sections of the mRNA will dictate the amino acid that is allowed to be transferred in such a way that a polypeptide is synthesized.

Ribosomes of eukaryotic organisms like cannabis are much different than ribosomes of prokaryotes such as *E. coli*. Both have ribosomes that function to help in translating an mRNA transcript to protein but they differ in what is known as their sedimentation coefficient, a reflection of size and density. Based on the differences, biologists have determined that prokaryotes have 70S ribosomes and eukaryotes have 80S ribosomes. The S after each number indicates the sedimentation units, in honor of Svedberg who helped characterize the difference. The difference in ribosomes is very important to plant biotechnology since some antibiotics that target bacteria specifically bind to only 70S ribosomes. Eukaryotic 80S ribosomes are left largely unaffected. Thus, a researcher can hinder growth of *Agrobacterium tumefaciens* or *E. coli* by addition of antibiotics to culture media and usually not inhibit the plant from growing.

If a protein is successfully generated the mRNA might be reused or be diced up and its monomeric units later reassembled. The ribosome, with its catalytic function, is available for a new mRNA transcript. The abundance of a newly synthesized protein depends on many factors. Sometimes proteins might move to

the outside of the cell, which often happens if ribosomes are associated with the endoplasmic reticulum. The proteins might also be chemically altered by addition of functional groups (phosphates, sulphydryl groups, carboxylic acids, etc.), within the endoplasmic reticulum.

Areas of genomic DNA that harbor genes are organized a little differently in prokaryotes (i.e. *Agrobacterium*) than in eukaryotes (*Cannabis*). For instance, prokaryotes lack a nucleus and therefore the mRNA might be translated as transcription occurs. Bacteria, in the strict sense, lack non-coding intervening sequences in their genes. Eukaryotic DNA might not only contain intervening sequences within a gene, their DNA is often arranged on histones, small proteins that the DNA is tightly wound around. Modification of the histones can modify enzyme access to the DNA.

Prokaryotes also have their genomic DNA arranged in a different way than eukaryotes. Prokaryotes usually have circular DNA while eukaryotes have linear DNA. Prokaryotes can also have other small, extrachromosomal pieces of DNA within their cells. Such pieces are called plasmids and they come in a variety of sizes and can harbor a range of genetic information. Plasmids are highly important in biotechnology, since smaller pieces of DNA are much easier to work with compared to larger pieces of DNA. Bacteria also readily take up plasmids under the right conditions.

Other proteins that are synthesized using ribosomes not associated with the first part of the internal membrane system, the endoplasmic reticulum, are synthesized using ribosomes freely floating inside the cell, in the cytosol. Proteins made with free ribosomes are often retained by the cell and used for internal processes.

Many students of biology ask how a cell can determine where a protein is sent after its synthesis. This is an excellent question important to plant biotechnology and brings forth several new pieces of biochemical information. One way to target where a protein is transported is to add on little signals, much like an address is put on a piece of mail. Biochemically, the signals can be short peptides or carbohydrates that interact with special carrier proteins that will deliver the newly synthesized protein. Some proteins come with a built-in signal sequence based on their mRNA transcript. Other proteins can be tagged with additions in the endoplasmic reticulum, Golgi apparatus, the cytoplasm, or chloroplasts. Alternatively, proteins can be targeted to these locations as well.

Chloroplasts and mitochondria, both double-membrane bound organelles, have their own ribosomes and their own DNA. That is, these two organelles do

not completely rely on the plant cell for making proteins or holding their genetic information. They also divide on their own. The ribosomes of chloroplasts and mitochondria also resemble prokaryotic 70S ribosomes. Similar to bacterial cells, mitochondria and chloroplasts have circular DNA, not linear DNA like most eukaryotes. For all of the above reasons, biologists have come to realize that chloroplasts and mitochondria were once free-living cells that took up residency within larger cells. Chloroplasts deserve a bit more attention, since they serve to capture and transform solar energy for use in cell activities and also offer several interesting opportunities for genetic manipulation.

Chloroplasts, photosynthesis, and pigments

Nearly all the energy provided for the cannabis plant comes from the chloroplast in a process known as photosynthesis. The process of photosynthesis in green plants and some other organisms is a biochemically complex series of events in which carbon dioxide is incorporated ("fixed") into a reduced sugar such as glucose. Many stages of photosynthesis offer themselves to genetic manipulation and so knowing the core details of photosynthesis is important.

The structure of the chloroplast is also important to know in order to adequately describe how photosynthesis works. A typical plant mesophyll cell will have approximately 50-100 chloroplasts. The number of chloroplasts is higher for those cells at the surface of a leaf (palisade mesophyll cells) compared to those near middle (spongy mesophyll cells). Increasing chloroplast numbers might be a way to enhance the photosynthetic activity of the cannabis plant, although often plant biotechnologists consider localizing protein products to the chloroplast.[8] The chloroplasts, since they are ancestors of free-living cyanobacteria that were engulfed by a larger cell, have an outer and inner membrane, each membrane a remnant of the ancestral endocytotic event. The permeability of each membrane of the chloroplast differs, suggesting the presence of a specialized function.

In the broadest sense, photosynthesis has two components; a light dependent component in which ATP and NADPH are made and a light-independent reaction in which ATP and NADPH are used to later generate a molecule of glyceraldehyde-3-phosphate, a sugar. Within the inner membrane of the chloroplast is the aqueous stroma and is the site of the light-independent reactions, also called the Calvin cycle. The stroma contains stacks of another membrane system called grana, which are made of individual pancake-looking

structures called thylakoids. The thylakoids host the large photosystems of the light reactions and large synthesizing complexes that make ATP and NADPH.

The light reactions require antenna complexes, which are an array of chlorophyll molecules (and often accessory pigments) arranged to funnel excitation energy into a reaction site, or photosystem. Although there are other routes for the energy to go, the ultimate "goal" of photosynthesis is transfer energy from a photon to the reaction site, the center of an antenna complex harboring a special chlorophyll molecule that has electrons ready for excitation. The electrons were previously removed from a water molecule (i.e. water is oxidized) with the help of a protein complex called photosystem II. Upon excitation, the electrons are raised to a higher energy level and are transported along a series of membrane proteins that function to simultaneously move hydrogen ions into the lumen, or space, of the thylakoid.

The initial hydrogen ion transfer is dependent upon electrons passing from photosystem II to plastoquinone, to cytochrome b6f complex (which pumps hydrogen ions), and then to plastocyanin. This is the so-called plastocyanin pool that photosystem I can use over and over in cyclic phosphorylation. After plastocyanin the electrons might be given to photosystem I, also containing a reaction center. The electrons can be further excited with energy from new photons and can be passed to the NADP+ reducing complex to generate NADPH. Another option is for the electrons to return to plastocyanin for cyclic phosphorylation. Due to the concentration of hydrogen ions, the ions will have a natural tendency to move to the stroma but can only do so through an enzyme complex known as ATP synthase, which acts as a transport channel. As the hydrogen ions flow through the channel, a molecular crank is turned and a phosphate is attached to adenosine *di*phosphate, or ADP. Thus photosystem I can produce ATP without oxidizing water.

The ATP and NADPH generated in the light reactions are used to reduce (donate electrons to) atmospheric CO_2 in the Calvin cycle. For 3 molecules of CO_2 entering the Calvin cycle one molecule of glyceraldehyde-3-phosphate (a three-carbon sugar) is produced. Two of these three-carbon compounds are used to make one molecule of glucose, which can leave the cell and be used in anabolic reactions. The Calvin cycle has a CO_2 fixation stage and a regeneration phase.

For CO_2 to first enter the cycle, a huge protein complex called ribulose 1,5-bisphosphate oxygenase/carboxylase (conveniently shortened to rubisco) functions to fix, or attach the CO_2 molecule. The small CO_2 gas molecule is attached to a carbohydrate called ribulose-1,5-bisphosphate (RuBP), which is then split to form

two molecules of 3-phosphoglycerate. Each of the two molecules of 3-phosphoglycerate is converted to 1,3-bisphosphoglycerate and then to glyceraldehyde-3-phosphate (G3P). The ribulose-1,5-bisphosphate is regenerated from molecules of glyceraldehyde-3-phosphate and ready to be carboxylated again. For three turns of the cycle, an "extra" molecule of glyceraldehyde-3-phosphate results while 9 ATP and 6 NADPH are required, the latter being produced from the photosystems.

A problem can occur with carbon fixation, which is something plant biotechnology has attempted to address. When cells first evolved on earth, the atmosphere contained extremely small amounts of atmospheric oxygen. Today oxygen is abundant as an atmospheric component (~20%) thanks to oxygenic photosynthesis. Thus, early life on earth was not concerned about oxygen taking electrons-much the way oxygen is "greedy" for electrons in the water molecule. However, in the Calvin cycle, the negative effects of the O_2 molecule are seen in the first stages. Although rubisco will react with atmospheric CO_2, it will also react with O_2, an unfortunate phenomenon of accidental cross-reactivity. Since rubisco is unable to distinguish CO_2 from O_2, sometimes O_2 fixation can happen and is known as photorespiration, a process that results in a net loss of carbon from the plant; the O_2 is not fixed but does react. A great example of photorespiration is during the summer months in northern climates, when lawns of green grass fade to brown. The lawns require more water to stay alive, since their efficiency in photosynthesis has decreased.

Indeed cells on the inside of some leaves have evolved to be arranged in such a way to sequester initial products of fixation. In plants that typically grow in cooler regions rubisco generates an initial three-carbon compound called glyceraldehyde-3-phosphate (found in C3 plants like fescue grass and cannabis); the majority of plants are in this category. However in other plants, mostly plants that evolved in warmer climates, a different pathway is used and the first detectable product is a four-carbon compound called oxaloacetate (found in C4 plants like Bermuda grass). Rubisco does not initially function to fix the CO_2 in the latter example, rather a different enzyme called phosphoenol pyruvate carboxylase is at work. The oxaloacetate is transferred and eventually delivered to rubisco residing in a different cell, where atmospheric O_2 is almost absent. The CO_2 is released directly to rubisco, where the Calvin cycle can continue undisturbed.

Considering this brief overview of photosynthesis, there are several areas worth investigating for genetic manipulation. Chlorophyll absorbs photons that are in the frequency of red and blue light while plants themselves are mostly green.

That is simply because green light from the visible spectrum is not used but instead is reflected. Could a cannabis plant be generated that has an underproduction of chlorophyll or an overproduction of accessory pigments? Production of accessory pigments might require engineering of cannabis that has a complete metabolic pathway (multiple genes), which is possible as we will see. Would an overproduction of orange carotenoids result in an orange cannabis plant? If so, cannabis would be placed along with golden rice and golden mustard[9] in not only appearing as hues of orange, but would also contain higher levels of provitamin A. Golden cannabis would not only be aesthetically pleasing but might also improve eyesight upon ingesting. Lycopene, the red pigment found in tomatoes and watermelon could also be produced so that foods made with lycopene-loaded cannabis might have increased protection against free radicals.

The light and dark reactions both offer opportunities for genetic modification as well.[10] Could the genes responsible for the ATP synthase be up-regulated? If so, would a plant show increased levels of ATP production and have enhanced levels of photosynthesis? Might a better approach be to up-regulate the pathways that lead to establishing the photosystems? Perhaps one might be able to produce a faster growing or higher yielding cannabis plant by up-regulating the genes that produce enzymes in the Calvin cycle. Knowing the basics of the light and dark reactions provides an excellent starting point for asking questions, which is often the first step in any scientific investigation.

Photosynthetic pathways of cool season (C3) plants and warm season (C4) plants are already being investigated.[11] Such an engineering feat might allow for a C3 plant like cannabis to grow more efficiently in warmer conditions. To be truthful, there are several other factors to consider due to the complexities of the biochemistry and anatomical differences (e.g. stomatal influences) between C3 and C4 plants. Thus far, properly engineering C3 plants with C4 photosynthetic pathways has been difficult. Obviously, more plant biotechnologists are needed in order to find reliable and realistic solutions to engineering a C4 cannabis plant.

Cells and tissues

The basic unit of life is the cell. Cannabis has several types of cells and each has specific, yet overlapping functions. All cells work to maintain homeostasis, a type of equilibrium in order for proteins and other macromolecules to properly function. Homeostasis is actually the result of several parts of the cell working

together. Should any part cease to function or fail in its capacities, cell integrity is compromised and homeostasis might be lost. The result might be death. There are several types of cells including epidermal cells, guard cells, and mesophyll cells. Each is highly important to the cannabis plant and sometimes, shared features between several cell types can be observed.

Plant cell walls are one of the most important structures for maintaining homeostasis and are comprised mostly of cellulose. However cell walls can also contain other carbohydrates such as hemicellulose, pectin, or non-carbohydrate derived polymers like lignin. In some cases, researchers need to work with plant cells that lack a cell wall and therefore must treat cells with cellulase, an enzyme that degrades cellulose. Stripped of their cell wall, the plant cells are called protoplasts and have only a cell membrane to protect them from their surroundings. Usually treatment with cellulase occurs in an isotonic solution, where the solute concentration matches those of the cell interior. Protoplasts have the tendency to uptake foreign DNA when a small pulse of electricity is applied and so, have been useful in plant transformation experiments.

When the same cells work together with a common function they are considered to be a tissue. Three main types of tissues are found in plants; ground, vascular, and dermal tissue. Ground tissue makes up the bulk of the plant. Vascular tissue functions in transport of water and solutes and dermal tissue forms the outer layer and is in contact with the surrounding environment. Genetic pathways that lead to each type of tissue in plant development might be considered targets worth investigating by plant biotechnologists.[12]

Vascular tissue includes xylem and phloem. The xylem is involved in water movement from the roots to the leaves and is hollow, nonliving tissue. Different cells exist within the phloem, primarily vessel elements and tracheids. Vessel elements are much wider than tracheids and are not found in gymnosperms (e.g. pines and cedars). Flowering plants like cannabis have tracheids but they also have vessel elements that help them grow much faster because of the increase in flow of water and solutes. One might be curious as to what would happen if a cannabis plant were engineered to be able to manufacture wider vessel elements or simply, generate a plant with more vessel elements in the vascular tissue.

Plants also have dermal tissues. The dermal tissues act as a protective barrier and are often compared to the skin of animals. Epidermal cells, which are found on the outermost layer of the leaves and stem, also help the plant retain water since they exude a waxy substance. The waxy substance can be either imbedded between cell walls or can be deposited external to the cell wall as a

cuticle. Production of a thicker cuticle, perhaps through genetic modification, might lead to a plant that can withstand warmer temperatures and drier climates.

An important component of the epidermis is the presence of stomata, pores that permit diffusion of gasses from the ambient air to reach the interior mesophyll cells of the leaf. The number of stomata per unit area varies by species and in some cases, the abundance of stomata can fluctuate depending on environmental conditions. Each stoma functions in gas exchange and as more stomata are present, more gas exchange is allowed. The trade-off is that with more stomata more water is allowed to exit the plant leaf. Thus, well-watered organisms can afford to produce more stomata. In general, plants in drier habitats tend to have fewer stomata while some organisms like cacti only open their stomata at night when the temperature is lower. Cannabis with fewer stomata or stomata that are triggered to open at night might increase drought tolerance, but growth rates might be hindered due to limiting the rate of gas exchange for photosynthesis. Similar to anatomical differences between C3 and C4 plants, engineering a cannabis plant that fixes CO_2 at night might be difficult (but might still be possible!).

Of specific interest to many cannabis growers are modified epidermal cells known as trichomes (Figure 2). There are a variety of shapes that trichomes can take on and range from small, single-celled (unicellular) trichomes to dendritic trichomes, which look like miniature trees.[13] Some trichomes might help reduce water loss from the stomata while others can function to store excess salts.

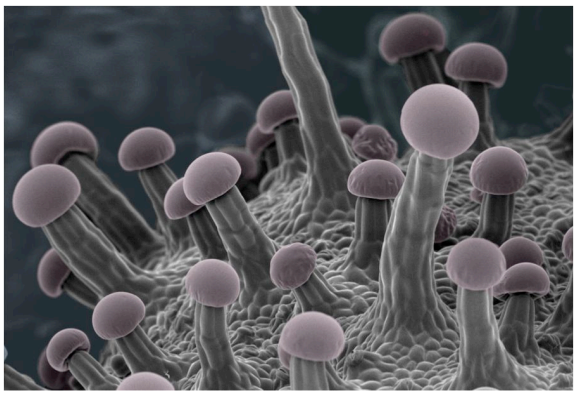


Figure 2. Electron microscopy image of glandular trichomes of *C. sativa* courtesy of David Marks, University of Minnesota.

Depending on how they are classified, there are two major types of trichomes recognized in plants, which include simple trichomes and glandular secreting trichomes. As their name implies, glandular trichomes act as glands, secreting various substances. Some glandular trichomes secrete fluids, which might contain high amounts of sugars in order to attract pollinators. Other trichomes secrete substances that can act as a protectant against the harmful effects of ultraviolet light.

The way ground, vascular, and epidermal tissues develop is dependent upon genetics. New plant cells that form tissues arise from meristematic tissues, a type of undifferentiated tissue that has similar characteristics to animal stem cells. Ultimately, meristematic regions rely on a process called mitosis to generate new cells. Mitosis is part of a larger process called the cell cycle.

The cell cycle has an interphase that consists of several subphases and is the phase of the cell cycle that takes the longest amount of time to complete. The cell carries out normal growth and will double its organelles (G_1), begin to replicate its DNA (S phase), and continues its growth (G_2) to prepare for mitosis and cytokinesis. At all other times cells are considered to be in a G_O phase of interphase and at this point are able to carryout cellular functions.

Although mitosis takes less time to complete compared to other stages of the cell cycle, more biology students are familiar with the stages of mitosis than the stages of interphase. The stages (prophase, metaphase, anaphase, telophase, and cytokinesis) are a reflection of the activities of the chromosomes but are also concerned with the breakdown and reassembly of the nuclear membrane. The underlying theme behind mitosis is that one cell divides to result in two genetically identical (barring mutations) daughter cells. In meiosis four genetically distinct reproductive cells result. Mitosis shares similarities and has differences from the type of cell division in prokaryotes called binary fission. Since prokaryotes lack a nucleus and often have one circular piece of DNA, the process is seemingly less complicated.

The cell cycle is important to all plant tissues but is perhaps most studied in the plant meristematic regions, since they have a high amount of mitosis occurring. Meristematic regions can occur either in the shoot tips or root tips and growth in these areas is known as primary growth. Additionally, just below the surface of the stem in a region known as the vascular cambium (a meristematic tissue), where growth might also occur and results in widening of the stem (secondary growth).

Another region of meristematic tissue (the intercalary meristem) is found in grasses and has contributed to their success across the globe. The grass family

(Poaceae) is the third largest family (~18,000 species) and members can be found on every continent, even Antarctica.[14] Grasses do not grow from the tips, rather they grow from a region near the middle of the plant, just above the soil, called the intercalary meristem. Such an adaptation might seem strange until one realizes that grasses are loved by so many herbivores, which eat the tops off if given the chance. When most plants are chopped down and lose all apical meristems, they suffer and might even die. However, grasses not only continue to grow, in some cases they are stimulated to grow at a faster rate.

Perhaps, the genetic instructions to generate an intercalary meristem could be transferred to cannabis. Such an engineered plant might show regrowth after being chopped down. That said, any meristematic region might be of interest for genetic modification. Genetic pathways could be altered to reduce secondary growth and result in a plant with a thinner stem. Perhaps this in turn would lead to more plant energy being spent on growing in other ways, such as larger flowers. Perhaps a different target, one that has been already done in plants,[15] is to generate more trichomes. Genetically engineered cannabis could be made to have all plant parts covered with trichomes at such high densities to leave the plant looking like one big sticky trichome. As we will see, there are other interesting metabolites of cannabis that are not found in the trichome that some people might prefer to modify. However, the beauty of plant biotechnology is that preferences are allowed to vary.

Cannabis organs

Organs are parts of the plant that carry out a specific function. Leaves, stems, and roots are all easily recognizable plant organs. Although each can be discussed separately, their functions are directly related, since food is made in the leaves and transferred to the roots via the stems.

Perhaps the most recognizable feature of the cannabis plant, besides the smell, is the leaves. The green leaves of plants carry out the majority of photosynthesis, at least in most cases. In cannabis plants, as in some other plants, the stem is also green. The green indicates presence of chlorophyll and thus, all green plant parts are usually photosynthetic. Leaves can be simple or compound and to determine if this is the case, one must look for small buds (shoot systems that are immature) that subtend the petiole, the structure that connects the leaf to the stem. A compound leaf will have smaller leaflets that comprise the overall leaf

and the leaflets will lack a bud at the base. Cannabis is said to have a palmately compound leaf, since all of the leaflets arise from a central point.

Leaves develop from leaf buds. There are several types of buds besides leaf buds including those that will develop into flowers or branches. Some leaves are deciduous, turning brown and falling off at the completion of a growing season while other leaves are evergreen and tend to stay green year-round. Although several genes are involved, one might take pleasure in seeing an engineered evergreen cannabis plant in their snow-covered yard.

Depending on their genetic program, leaves can develop into other recognizable plant features. Cactus spines are really modified leaves. The cactus pads are modified stems but are responsible for photosynthesis. The developmental pathway for cacti to grow spines instead of leaves has been characterized. Growing a cannabis plant with spines in addition to normal leaves might be beneficial in several ways. For example, if one were required to grow their plants in a field, herbivores would be less likely to eat the plants. The spines might also deter other people from attempting to rob a grower of their harvest. However, some people might not want to go through the pains of trimming colas if they had to contend with spines.

Perhaps a better modification of cannabis leaves might be to alter the genetic pathway leading to leaves and instead permit some leaves to develop into tendrils. Plant tendrils are the winding parts that can be easily seen on bindweed and many species of the cucumber family (Cucurbitaceae). Tendrils would allow a cannabis plant to climb up a lattice in a garden or grow along side of houses. Many plants with tendrils also have a vine-like, climbing nature so these traits might also need to be genetically transferred in addition to the ability to form tendrils.

Since the function of leaves is to collect solar energy, perhaps engineering a cannabis plant with broader leaves might allow it to produce more photosynthetic products and grow faster and bigger. For added attractiveness, cannabis plant leaves could be genetically modified to develop simple broad leaves with wavy margins (edges of the leaf) instead of palmately compound leaves with serrate margins. The developmental pathways that result in spines, tendrils, and even leaf shape have been studied intensively in plants.[16] A person wanting to determine how to genetically modify such features in cannabis, therefore, has an excellent starting point.

Leaves often attach at nodes of a plant. In between each node is a region of the stem, appropriately called the internode. The length between each node is

under hormonal control partly by a hormone called gibberellic acid. Hormones are naturally produced endogenously (within) but can also be applied to growing plants in order to induce certain effects. Since gibberellins can increase internode length, one might be curious to see what effects this has on a plant. While a transgenic cannabis plant could be generated with increased levels of gibberellins, elongated internode regions can also be induced by direct application of hormones (Figure 3).

Figure 3. Two genetically identical cannabis plants. The picture on the left is a container-grown cannabis plant. On the right are two people standing next to a female cannabis plant with elongated internodes induced by gibberellic acid. At harvest, the plant was nearly three meters tall.

Besides cannabis leaves, there are other organs that might be of interest to genetically modify, such as the stem. The stem of plants is largely comprised of ground tissue (parenchyma) to physically hold plant leaves to the sun. The other main purpose is to transport products between the leaves and the roots. The molecular forces of water help it travel through the stem and up through the plant. Cohesive forces keep the water in contact with itself as positive and negative charges interact. Adhesive forces allow the water to interact with the inner walls of the xylem, much like water interacts with the sides of a drinking glass.

Some plants have below ground stems called rhizomes, which travel beneath the soil surface and eventually sprout above ground. Often, the rhizomes will generate an entire new plant but will still be attached to the original plant. There are some plants that actually produce more rhizomes when cut back. In a neighborhood along a creek, there was a small infestation of poison ivy (*Toxicodendron radicans*). One citizen claimed the ivy could be mowed down, which

would lead to its death. The people of the neighborhood let the man mow the ivy down but instead of dying, the poison ivy sprouted from the ground in even more places with enhanced vigor. Perhaps if cannabis had the genetic capacity to generate rhizomes one could mow their cannabis in order to sprout new plants with enhanced vigor.

The cannabis plant is considered an herbaceous annual since the stem and associated parts die at the end of one growing season. However, plant biotechnology might be able to change cannabis into a woody perennial. Non-woody stems have primary growth (elongation at root and branch tips) whereas woody stems undergo primary and secondary growth (lateral growth, or thickening of the stem). Wood is really secondary xylem and is due to the seasonal activity of the vascular cambium, the meristematic region just interior to the bark but which remains exterior to the newly produced xylem cells. While trees and cannabis both have vascular cambium, typically perennials continue production of new xylem cells since their vascular cambium is genetically programmed to continue mitosis.

Cannabis plants with the genetically modified ability to make increased amounts of secondary xylem might allow growth to continue in a perennial fashion. That is to say, if one could genetically modify cannabis to grow as a tree does, one might be guaranteed a yearly harvest, a harvest that might increase with the age of the tree. Instead of picking apples each year, cannabis growers would be picking sticky flowers from their tree. Entire forests might consist of nothing but cannabis trees. The unfortunate event of a forest fire might result in adjacent populations of people, possibly entire cities, inadvertently getting high. There might be other downsides as well. For example, if cannabis trees were grown in a front yard, they might attract unwanted attention.

To be less conspicuous, a person might desire to grow a cannabis bush instead of a cannabis tree. Developing an annual cannabis plant into a perennial bush might also yield more flowers, since there are more areas for floral buds to develop. Currently cannabis growers sometimes trim back and then help the new growth spread, sometimes by adding weights to allow light to penetrate the interior buds. A genetically modified cannabis plant could be generated that had a shrubby habit, with spreading branches so that a grower could spend more time tending to other needs of the plant and less time arranging the stems of their plant.

Although cannabis trees and bushes that grow back every year are not yet a reality, cannabis has always been known to have excellent fibers (Figure 4). Fibers are part of the sclerenchyma tissue in plants that make up the ground tissue. The male cannabis plant is also known as hemp, and is recognized world-wide for fiber

durability, tensile strength, and out performance of tree fibers in many additional areas. Hemp seeds also provide a valuable resource for many essential fatty acids that help maintain cell membrane stability in organisms that consume the seeds. Hemp has been crucial to sustaining civilizations for thousands of years and was likely the reason cannabis was first cultivated.

Interestingly, some people have considered industrial hemp for biomass and energy production.[17] Using a fast growing plant that requires little maintenance such as hemp instead of food crops like maize for bioenergy might decrease energy production costs. Perhaps hemp could be genetically modified to increase its fiber production (as if that is even necessary) in order to increase the bioenergy potential.

Figure 4. Cross section (left) and longitudinal view (right) showing cannabis fibers.

While stem tissues have several attributes worth trying to genetically modify, the root is equally attractive to study. The roots function to provide water and nutrients to the vascular tissue for transport to the organism (Figure 5). The root is what the stem and other parts of the plant are anchored by and is crucial in the transfer of substances from soil to stem. Root hairs, outgrowths of the dermal tissue, are special cells that enable the plant to increase the root surface area and thus, uptake more water and nutrients. Since cell differentiation is genetically programmed, one might be interested in generating a transgenic cannabis plant with increased root hairs. Such a cannabis plant might show enhanced rates of growth or display a significant increase in size upon reaching maturity.

Other factors help uptake water and nutrients too. Root associations with fungi called mycorrhizae increase surface area and help with water uptake, but can be very species specific. Interestingly, as they catabolize the breakdown of carbohydrates in the soil, mycorrhizae provide monomeric sugars to the plant.

Depending on the species and circumstances, sometimes amino acids can also be delivered to the plant. The plant can also supply carbohydrates back to the mycorrhizae, resulting in a cooperation that gives both organisms advantages in growth. Plants will recognize, based on molecular signaling events, different species of fungi and bacteria to allow their penetration of the root epidermis. Since the plant root system functions much better with root hairs and mycorrhizae, plants engineered to recognize a wider array of mycorrhizal species might result in a more robust cannabis plant.

There are other features of some plant roots that could provide a blueprint for generating an interesting cannabis plant. Cannabis might be able to be engineered to have roots like a potato, a sort of cannabis tuber. Cannabinoids could be directed to the tuber and instead of using female flowers to cook edibles for cannabis patients, tubers could be used instead. Of course, one would not want to mistake a truckload of Idaho potatoes for cannabis potatoes. Although fast-food restaurants serving cannabis fries might see an immediate spike in profits!

Figure 5. Plant roots of cannabis grown in a hydroponic system (left) and in soil (right).

A final interesting plant to use as a model to improve cannabis roots is *Cuscuta*, a parasitic plant that has evolved roots called haustoria that literally penetrate other plants and rob them of their photosynthetic products. *Cuscuta*, a member of the morning glory family (Convolvulaceae), is a yellow-brown color, an indication that chlorophyll is lacking. If engineered correctly, cannabis plants might photosynthesize *and* have haustorial roots to penetrate other plants that might in turn supply extra photosynthetic products, thus increasing cannabis growth and success.

Typically, plants that grow for only one growing season are noted for their rapid primary growth (many annual weeds). Perennials seemingly die back, but

sprout new leaves from buds in the following spring. There are some plants, like celery (Apiaceae), that are biannual; they grow vegetatively for one year and flower the following year. A goal of plant biotechnology might be to generate a perennial cannabis plant. Perennials have been developed from annuals through hybridizing experiments. The Land Institute in the USA (landinstitute.org) is an agricultural research facility that has promoted and actively researched growing perennial polycultures (mixed species) in order to increase environmental sustainability while simultaneously producing a stable food supply. They suggest that the land used in agriculture should not be used for large-scale monocultures, large areas where only a single crop is planted. Instead, agricultural systems should plant several crops together. In addition perennial crops might be an alternative to annual crops so as to preserve the soil structure and decrease erosion.

Inserting the genetic mechanisms that leads to the perennial cannabis might result in a cannabis plant with the ability to regrow each year, after each harvest. New seeds would not be required and every harvest would be, genetically speaking, identical to the previous harvest.

The flowers of cannabis

Flowers are also considered organs and are of "high" interest to cannabis growers and thus, they deserve their own section when considering cannabis botany. Flowers have adapted and evolved over millions of years and have resulted in the spread of not only angiosperm species, but of hundreds of thousands of animal species as well. The evolution of the flower occurred by chance when leaves of ancient non-flowering plants happened to fold in on themselves along the bottom margins, which was advantageous in protecting the spore bearing structures, now tucked within the folds.[18] Eventually, as time progressed the folds became more closely associated with the spore bearing structures. As yet more time passed, opportunities for mutations were available and at some point the plant lost control over developing other leaves that were closely associated with the spore-bearing portion. Mutational serendipity led to some of these leaves evolving into bracts and with even more time, those bracts produced pigments other than chlorophyll. Thus, we observe flowers today with colorful petals that have radiated into the great diversity we recognize in the angiosperm phylogeny.

Flower development has been studied intensively and several genetic models have been proposed. The most popular model for flower development is

known as the ABC model.[19] This is based on the fact that there are four whorls in a typical flower. The first whorl is the sepals. The petals make up the second whorl and as one travels towards the center of the flower the next two whorls are the stamens (male parts) and pistil (female part), respectively. The ABC model describes transcription factors, which are proteins that act like an on/off switch for the genes that allow development of flower parts.

In many cases, transcription factors are needed to give the signal for turning a DNA sequence into a functioning protein. Alternating periods of light and dark, called the photoperiod, normally induces genetic pathways to produce transcription factors. The exact function of the A, B, and C transcription factors has been determined through mutating plants in order to lose various functions of each of the genes. By carefully tracking the mutational defects and the observable trait, the function of each gene has been deciphered (Table 2).

Transcription Factor	Resulting Growth
A	**Sepals and petals**
B	**Petals and stamens**
C	**Stamens and carpels**

Table 2. Summary of transcription factors involved in flowering. Each transcription factor is a small protein that activates genes, stimulating different flower parts to develop.

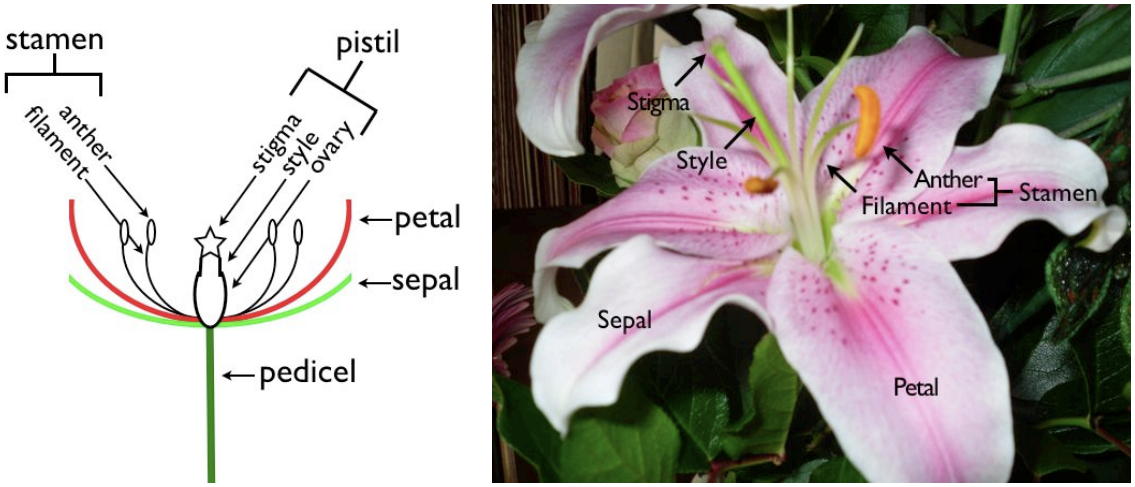


Figure 6. Diagram with basic parts of a flower (left) and a lily flower with labeled parts (right).

Flowers sometimes sit atop a small floral stalk called a pedicel (Figure 6) or the pedicel can be absent, as observed in sessile flowers. The major floral parts of a perfect, complete flower include the perianth, androecium, and gynoecium. The perianth consists of all the sepals (collectively called the calyx) and the petals (collectively termed the corolla). Within the perianth are the male parts called stamens (collectively called the androecium), which consist of a filament and anther.

Central to all the floral parts is the female portion called the gynoecium. The gynoecium is the collective term for the carpels. A carpel consists of a stigma, style, and ovary. The stigma is the platform for which the pollen lands and germinates and is often seen as the hair-like structure sticking out of cannabis flowers (Figure 7). In several cannabis strains, the stigmas might become darker with age.

Figure 7. Stigmas appearing from the cluster of flowers on a female cannabis plant (left) and attachment of the stigmas to the immature ovary (right).

Flowers that have been pollinated by pollen from the anther result in a developing fruit, which is also a ripening ovary. The fruit of cannabis is often simply called a seed by many but this is an improper botanical reference. They are fruits called achenes, which appear to be seeds since the ovary wall is fused with the seed. As many growers know, if cannabis is not pollinated then achenes do not develop. High-quality cannabis usually results from un-pollinated female plants, often called sinsemilla (*sin* without, *semilla* seed).

Figure 8. A cluster of male flowers. The anthers are ready to be exposed once the bracts open so that wind (if present) can facilitate pollination.

Male flowers (Figure 8) are known as staminate while female flowers are known as either carpellate or pistillate. Whole plants are considered dioecious if male and female flowers are on separate plants or they are monoecious if one plant has both male and female flowers present. With so much terminology, plant taxonomists have generated floral formulas to classify plants to the family level. A floral formula is a description of the floral arrangement of a flower and provides a shorthand notation for the floral description. Each plant family has a slightly different formula, and in some cases there is slight variation among the members of the same plant family. Members of the Cannabaceae have the following floral formula

staminate: * Ca5 Co0 A5 G0

pistillate: * Ca5 Co0 A0 G(2)

The floral formula describes the flowers of cannabis. Two different floral formulas are used for the Cannabaceae since members of this family have two different flowers, staminate and pistillate. Flowers are actinomorphic (have radial symmetry) and so the * symbol precedes the description of the floral whorls. In the male flower there are five sepals (Ca5), absence of petals (Co0), five stamens (A5),

and female parts are absent (G0). The female (pistillate) flower is also actinomorphic and has five sepals (Ca5), lacks petals (Co0), male parts are absent (A0), and the gynoecium (G) has two carpels that are fused, indicated by the 2 with a circle. The line beneath the 2 indicates the ovary is superior. In other words, the calyx is attached below the position of the ovary (Figure 6).

Due to the diversity of the plant kingdom and the fact that there are (by some estimates) more than 400 plant families, there are several variations and additions one might encounter with floral formulas. Sometimes the name of the type of fruit (e.g. drupe, berry, etc.) is also provided in the formula but not always. Additionally, a person might see a floral formula with a K in place of the Ca, which also represents the calyx. When the sepals and petals are indistinguishable or if one of these whorls is missing, the parts are referred to as the tepals and a "T" might be present in the floral formula in place of the "Ca Co." Additional loops and curves can be added to a floral formula to show positions of various floral parts. For example, there are some flowers with alike fused parts (all the petals) or fused parts that are different (filaments are fused to the corolla).

Floral formulas help the cannabis botanist understand floral diversity and think of new ways flowers of cannabis might be altered using biotechnology. For example, the cannabis biotechnologist might want to develop a plant that has a corolla, since natural cannabis plants lack petals. Petals could then be engineered to be a variety of colors, enhancing the attractiveness. Deposition of some pigments usually occurs in the vacuoles, so pathways or metabolites might need to be targeted to the vacuole. Additionally, cannabis plants that lack the ability to develop anthers could be engineered. Since resin production is greatly reduced upon pollination, a genetically modified line of cannabis unable to produce anthers would be highly beneficial in keeping potency of female plants high. Likewise, female cannabis plants that completely lack the ability to generate a gynoecium could grow among several pollen-releasing males and never become pollinated. Thus, potency for female plants would also be ensured.

Sometimes growers have plants that develop into monoecious plants (as is the case for many hemp varieties), where both male and female flowers are present on the same individual (Figure 9). Such a hermaphroditic condition in cannabis is sometimes attributed to environmental stress, which signals to the plant that conditions are less than favorable. If conditions are not good for growth, the plant might benefit by ensuring that pollination will occur. One way to increase the chances of pollination is to produce both male and female flowers. Although less genetic diversity results without cross pollination, the ability to develop both

staminate and pistillate flowers on the same individual increases the chances that fruits (containing seeds) will be produced. In biological terms, there is an increase in fitness at the cost of decreased genetic diversity.

Figure 9. A hermaphroditic cannabis plant with male flowers just below the female flowers.

Understanding the botany of *Cannabis sativa* opens the doors to several areas worth investigating using the tools of plant biotechnology. A basic knowledge on the hierarchies of organization are important in thinking about how a genetic modification might result in a positive or negative impact for the organism. A researcher is able to understand how monomers form macromolecules, which come together to form organelles. One can also see that cells rely on organelles and other structures to maintain homeostasis. But cells also form tissues that comprise the cannabis organs.

All levels of this hierarchical order serve as starting points for asking questions on how to improve the cannabis organism. However, one must also become familiar with the biotechnology lab, which is where we will find the tools to genetically engineer cannabis.

Chapter 3

The Plant Biotechnology Lab

"In the fields of observation chance favors only the prepared mind."
-Louis Pasteur

Lab basics

There are dozens of research tools currently available for cannabis biotechnology and some are used more than others. Biotechnology tools range from having basic applications such as measuring pH of solutions and media to those with molecular precision such as restriction enzymes, which are similar to a molecular scissors. A person working on genetically modifying cannabis will rely on several different techniques and reagents to adequately attain his or her research goal. Thus, an introduction to the plant biotechnology lab and familiarization with the core set of tools is needed.

Safety should be the number one priority in any lab. Too many times careless researchers do not fully understand the dangers of reagents or cells they might be working with, leaving ample room for unwanted accidents. Rarely are professional molecular biologists susceptible to becoming sloppy in their workspace if they fully understand the dangerous aspects of a lab. However, there are exceptions, which result in an unfortunate experience for those working alongside the neglectful researcher. Still, the lab is usually not dangerous if one is conscientious about their work habits and lab space.

Several reagents are used in lab and some can be quite harmful if accidental exposure occurs, so one should take note of the protocols they are working with and be aware of any potential dangers. A person should always know exactly what reagents they are working with and how to properly handle and dispose of the reagents. In fact, local waste inspectors might fine labs for neglecting to properly

handle organic and inorganic waste.

Among the shelves of potentially dangerous reagents (Figure 1) will be containers with relatively harmless qualities such as sodium chloride, sucrose, glucose, agar, powder for media, amino acids, starch, etc. Many labs have special areas for storing large quantities that are often shared among several different labs.

Figure 1. Plant biotech labs require several different reagents, some of which are kept in stock rooms like the one shown here.

Figure 2. Common tools in a plant biotechnology lab are a pH meter, scale, and pipettes.

Solutions, mixtures of solutes and solvent, are constantly being prepared in the molecular biology lab. Thus two common devices one will almost always find are a pH meter and a scale to weigh dry powders (Figure 2). Another common tool is the pipette, which come in a variety of sizes and are used to transfer solutions. They allow for transfer of microliter (µl) quantities, although pipettes for larger

quantities exist. For the purposes of molecular biology, one will commonly find pipette sets that range in the amount of volume from 0.5 μl up to 1,000 μl.

Working with small volumes is very typical in the plant biotechnology lab. A common misconception is that one DNA molecule is manipulated at a time. The presentation of direct manipulation of single DNA molecules in textbooks is simply an artifact of education and in reality several pieces of DNA are simultaneously incorporated in a sample. The DNA is often manipulated in a small plastic tube called a microcentrifuge tube. Additionally, pipettes require plastic, sterile pipette tips that are disposed of as they are used.

While becoming more familiar with the molecular biology lab, one realizes the dependency on plastics. Moreover, one realizes the enormous volume of plasticware that is discarded every day. A very useful tool to minimize pipette tip disposal might be the construction of an apparatus that is something akin to a pipette tip washing machine. Tips with nonhazardous residues could be tossed in, vigorously washed, and emptied into a container, ready to be loaded into pipette tip boxes that are then autoclaved. Some researchers have already come up with other ways to collect and recycle plastics in order to minimize the volume of waste.[1]

Cloning genes

One of the most common techniques learned in the molecular biology lab is cloning a gene. From a broad perspective, gene cloning is the process of taking a DNA molecule, making a copy, then joining the copy with another, larger piece of DNA. Gene cloning is different than animal cloning, which is also known as somatic cell nuclear transfer. For our purposes here, one is not cloning whole organisms, rather one is focused on copying and manipulating a specific DNA sequence.

To copy a gene one must already have a template. Therefore, most experiments begin with extracting DNA from tissue. In the case of cannabis, the plant cell wall might be difficult to crush and so a mortar and pestle can be used (Figure 3). Much less than a gram of plant tissue is often needed and can be placed into a mortar with liquid nitrogen. The ultra cold temperature quickly freezes the cells and crushing with a pestle ensures breaking of the cell wall and thus release of genomic DNA.

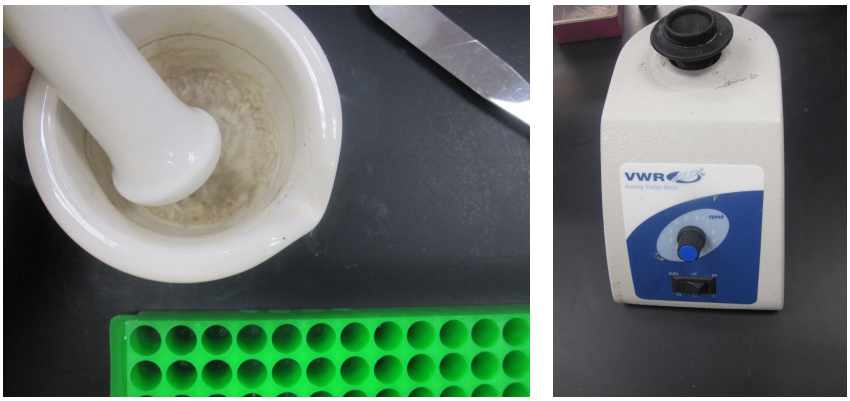


Figure 3. Mortar and pestle with plant sample obscured by liquid nitrogen (left) and vortexer (right) used for agitating and samples.

In other methods of DNA extraction, detergents and small beads or miniature plastic pipettes are used in a microcentrifuge tube in order to help disrupt the cell membrane and break the cell wall. In the case of beads, the tubes are shaken on a vortex machine for several minutes. When using a small plastic pestle, manual crushing is needed, similar to using a larger mortar and pestle. Several plant DNA extraction kits are available for purchase and are streamlined for ease and speed. One of the most common DNA extraction techniques uses cetyltrimethylammonium bromide and is called the CTAB method.[2] The method works great but needs to be carried out in a fume hood.

Part of any DNA extraction process must inherently involve separating DNA from the rest of the plant material, which is usually performed using a lysis buffer followed by a centrifugation step. The centrifuge is another tool indispensable to the molecular biologist, since spinning cells helps separate DNA from other cell debris. Centrifuges (Figure 4) come in a variety of sizes that range from "quick spin" to ultracentrifuges that can reach 100,000 rotations per minute.

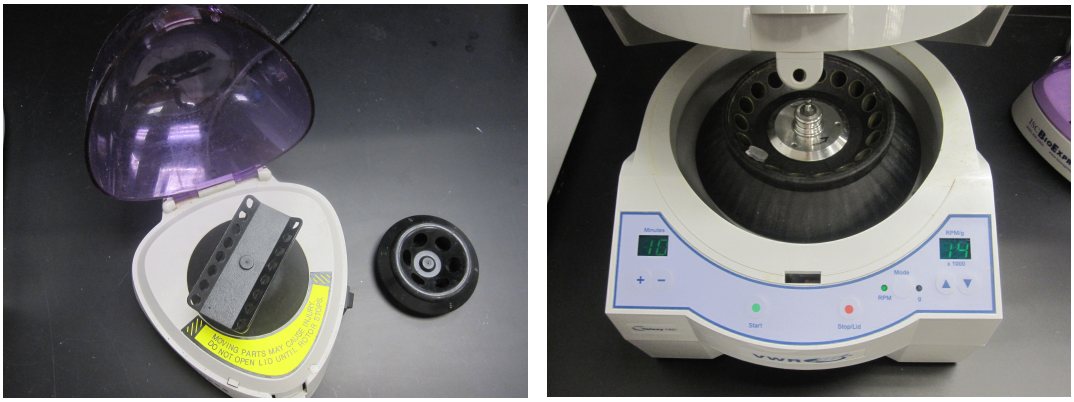


Figure 4. A minicentrifuge for quick spins (left) and microcentrifuge (right) for microcentrifuge tubes. These are two common centrifuges in most molecular biology labs.

After centrifuging the microcentrifuge tube, much of the genomic DNA remains on the top layer, which is the layer that is transferred to a new tube. The new tube can sometimes have a smaller tube with a membrane that acts to trap DNA and allow the liquid portion to seep through. The small tube with the membrane is then placed inside of a new microcentrifuge tube, sterile water is added, and the tubes are both centrifuged. As the water goes through the membrane, the DNA is also removed from the membrane and remains in the water.

The different types of water in a molecular biology lab can be its own topic. Water can be deionized, meaning all of the ions (dissolved salts) have been removed. Some water is strictly sterile meaning deionized water has been placed in an appropriate container, autoclaved, and is placed on the shelf and labeled as "sterile water." Water can also be treated with reverse osmosis (RO), which is usually referred to as "ultra pure" water. Some "ultra pure" water might have an added chemical known as diethylpyrocarbonate (DEPC) and is used when working with RNA and is simply called RNase-free water.

The various types of water one might encounter in a molecular biology lab are used for different applications. For example, one important difference between RNA and DNA is that RNases, which are enzymes that chop up pieces of RNA, do not require a cofactor. That is to say, RNases will catalyze the breakdown of RNA without requiring the presence of small ions to help. This is different than DNases, which require a cofactor to function. Such an atomic difference might not seem relevant until one works with DNA or RNA. While DNases and RNases are ubiquitous in the lab environment and elsewhere, RNases are active when present in solvents such as water. Thus, a researcher who works with RNA suspended in deionized water might suddenly find, upon analysis of their sample, that their RNA has vanished seemingly into thin air. If proper precautions are taken, a researcher should have no problem obtaining and working with an RNA sample; however, they are encouraged to use DEPC-treated water since RNases have been eliminated.

Quantification of cannabis DNA

A researcher will want to determine the efficiency of the DNA extraction and thus, will need to quantify the amount of DNA eluted from the membrane filter. Several years ago methods for DNA quantification relied on making

dilutions of a DNA solution then using a portion of the dilution in a process of spectrophotometry. A spectrophotometer passes light through a sample and the amount of light that passes through the sample tells a researcher something about the density, or concentration of the sample. A relatively new method of quantification is known as the NanoDrop (nanodrop.com). The NanoDrop is also based off of spectrophotometry but far exceeds previous methods in that only 1 µl (1/1,000 of a ml) of a homogenous sample is required. Thus, dilutions are not needed. The sample is loaded onto a small pedestal with a lens at the base where light enters to generate spectroscopic data.

The NanoDrop is connected to a computer and so one can quickly obtain an accurate estimate of the DNA, RNA, or protein concentration in a sample. Purity is estimated in ratios since nucleotide bases of DNA absorb at 260 nanometers (nm) and proteins absorb at 280 nm. An absorbance curve is also generated based off the absorbance properties of the sample (Figure 5).

Obtaining false positives with the NanoDrop might happen and might be disastrous for any step of an experiment. Phenol, part of the original mixture used for isolation of total RNA, can give a falsely high reading and lead a researcher to believe high quantities of nucleic acid are present in the sample, since the phenol ring absorbs in a similar range as nucleotide bases. Additionally, protein contamination will compound the confusion, although this type of contamination can be estimated with 260/280 ratio that should range from 1.5-2, but can depend on the sample. DNA often gives a 260/280 ratio of 1.8 and RNA gives 2.0.

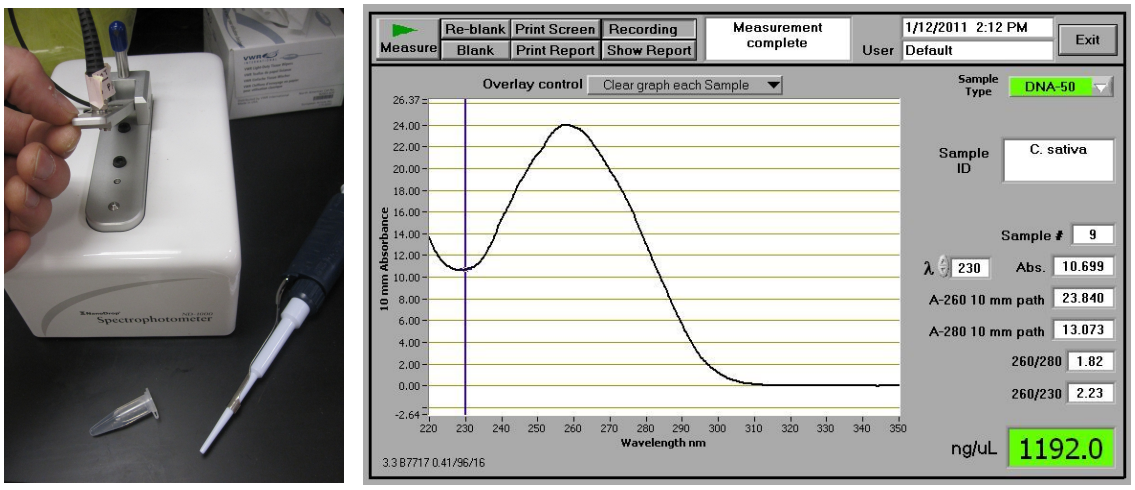


Figure 5. A NanoDrop machine (left) to estimate nucleic acid concentration. A microcentrifuge tube is in front of the machine. The computer screen (right) reports the cannabis DNA concentration and in this example it is at 1,192 nanograms (ng)/µl.

Importantly, values obtained with the NanoDrop are estimates and an "ultra-pure" sample is not always needed. In short, the sample quality needed is highly dependent on the downstream application (e.g. particular enzymes might only interact with a sample with higher purity). For amplifying DNA, concentrations and purity are not always of primary concern.

DNA amplification and gel electrophoresis

The process of DNA amplification is where much of the power of molecular biology becomes apparent. Amplifying a region of DNA is necessary in order to generate millions (sometimes billions) of copies for several reasons. In the case of cloning a gene, a copy of the gene will be joined with a carrier molecule of DNA called a plasmid. These reactions happen in small tubes were several pieces of DNA fragments are present along with several plasmids. Before plasmids are considered, discussion on the process of DNA amplification is necessary.

The polymerase chain reaction (PCR), which now there are several variations of, is used in nearly every molecular biology lab. The enzyme used in PCR is called DNA polymerase, which is the molecular workhorse that will copy a new strand of DNA in the presence of magnesium ions, deoxynucleotide triphosphates (triphosphorylated adenine, cytosine, guanine and thymine), and single stranded DNA. These ingredients, except for the sample of DNA to be amplified, are often premixed and sold as a "master mix." Also needed are short pieces of single stranded DNA that flank a region of interest. The short pieces are called primers, since they prime or prepare DNA polymerase to synthesize a complimentary strand of DNA.

The PCR machine (Figure 6) is a holding device for the sample tube while imposing different temperatures. Using the pipette, the primers, DNA sample, and master mix are placed in a small plastic tube called a PCR tube and carefully mixed. The tubes are placed in a PCR machine (also called a thermocycler), which is programmed to continually cycle between hot and cool temperatures. One cycle results in double-stranded DNA melting (that is to say, the DNA becomes single-stranded) when at 94°C, primers annealing to complimentary sites (usually 55°C but this can vary), and finally to the optimum temperature that the DNA polymerase functions (72°C). The cycle is repeated again and again. After 30 continuous cycles the theoretical result is more than a billion (2^{30}) copies of a piece of DNA might be produced. At this point the tube with the DNA sample can be

removed from the PCR machine and stored in the fridge or freezer. Alternatively, the PCR product can be used immediately for cloning (inserting) into a plasmid.

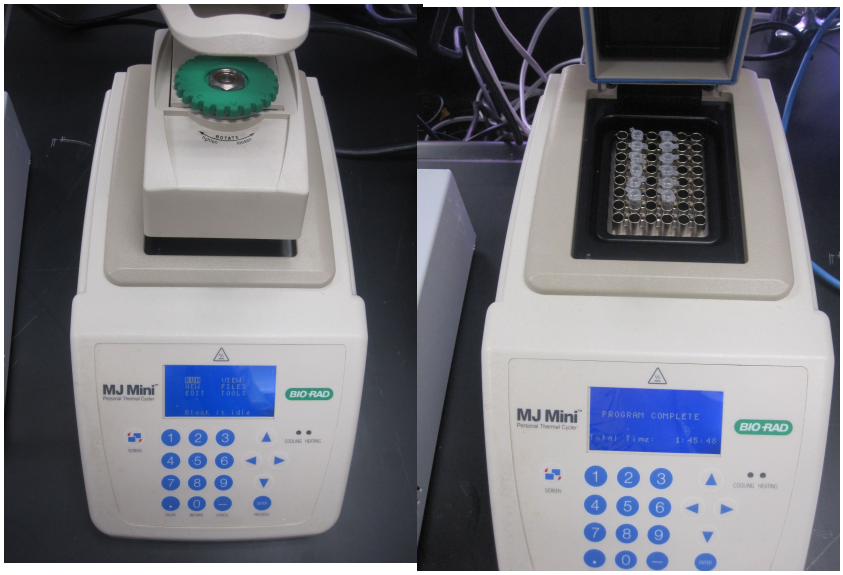


Figure 6. A thermocycler with its lid almost closed and ready to start (left). After the PCR has completed the lid is opened (right) and the PCR tubes are removed.

To verify a gene of interest was indeed amplified, another method called gel electrophoresis is performed (Figure 7). A polysaccharide called agarose is mixed with water, briefly brought to a boil in a microwave, poured into a gel box, and a small plastic comb is placed on one end of the molten gel. The gel eventually cools, hardens, and the comb is removed. The gel is flipped lengthwise and a buffer solution is poured into the gel box.

Each sample from a PCR reaction is mixed with a dye and loaded into its own well (depression) left by the comb. In another well a DNA ladder is loaded. A voltage is applied that runs through the gel, carrying the negatively charged DNA towards the positive pole of the gel box. Large pieces of DNA travel slowly through the gel while small pieces travel more quickly. The DNA ladder contains several different lengths (and several copies of each length) of DNA that are also pulled through the gel, along side the PCR sample.

After about an hour the voltage is turned off and the gel is removed from the gel box, immersed in a solution with a DNA binding chemical for several minutes, then removed and viewed with either UV light or blue light. A picture can be taken of the gel (Figure 8) and used as evidence that a particular piece of DNA is present based on size of the resulting bands.

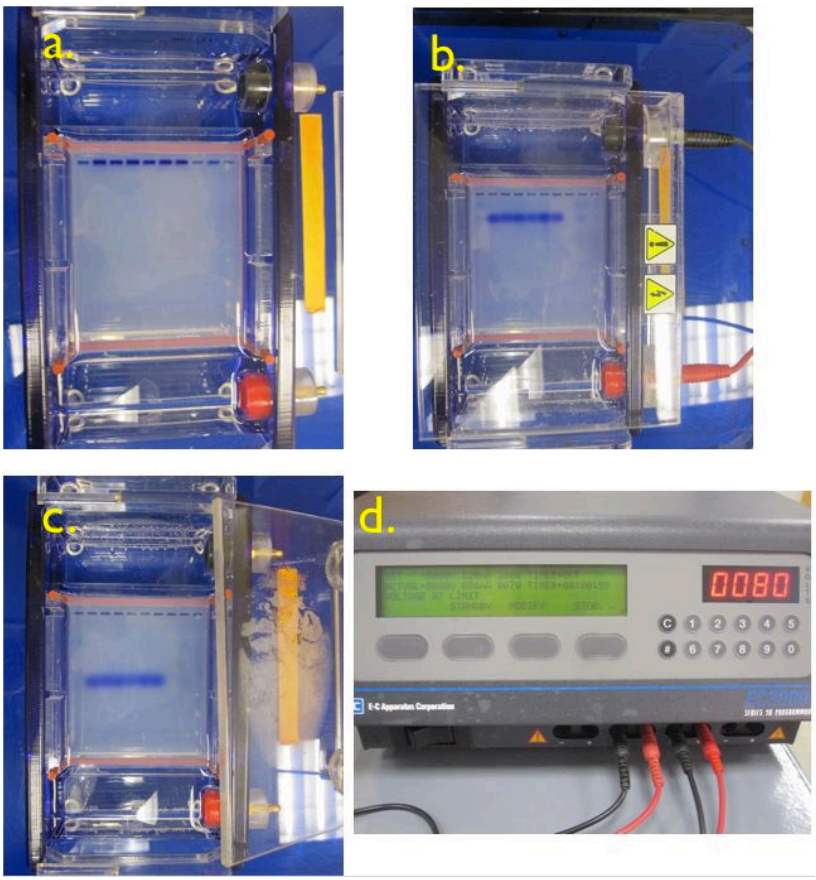


Figure 7. Gel electrophoresis (a-c) showing the samples migrating over the course of an hour. The power supply for the electric gradient and positive and negative electrodes (d) are also shown.

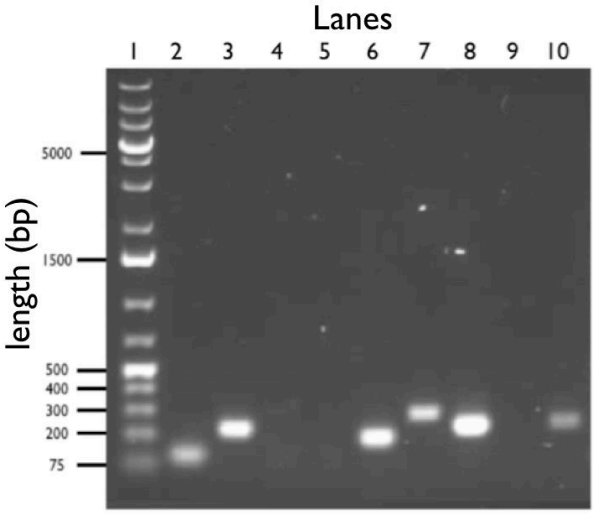


Figure 8. Sample gel image. The DNA ladder in lane 1 is used to estimate the length of different PCR products in bases. Samples in lanes 4 and 5 failed to amplify, lane 9 is a negative control, and lane 10 is a positive control.

If a band is not present, the reaction might have failed due to one of several reasons, which might include poor primer annealing or incorrect reagent concentrations. If a band is present on the gel and it matches the size of the band expected, one can smile. The next step is using part of the PCR product for cloning into a plasmid.

Plasmids can be ordered from companies or freely exchanged among researchers. Many people firmly recommend verifying a newly received plasmid is indeed the plasmid originally requested. Typically, problems do not occur with large retailers where quality is controlled. However, in rare instances incorrect plasmids are sometimes sent. The only personal experience of receiving the wrong plasmid occurred when a student had prepared the plasmid in a lab that was largely neglected by the primary investigator. Additionally, students are usually overworked and underpaid and hence, are more likely to make mistakes.

Problems aside, plasmids are similar to a small genome with a capacity to accept more DNA. Most plasmids are circular, often small (typically a few thousands bases), are of huge importance in cloning, and thus commonly manipulated in plant biotechnology. Plasmids that have fewer bases (smaller in size) are preferred to work with over larger plasmids since larger plasmids are prone to breaking when vortexing and smaller plasmids are likely to have fewer restriction (cut) sites within them. Plasmids come with a variety of interesting features and can be used in a variety of different projects. An origin of replication, selectable marker, and a multiple cloning site (MCS) are included as basic parts of a plasmid (Figure 9).

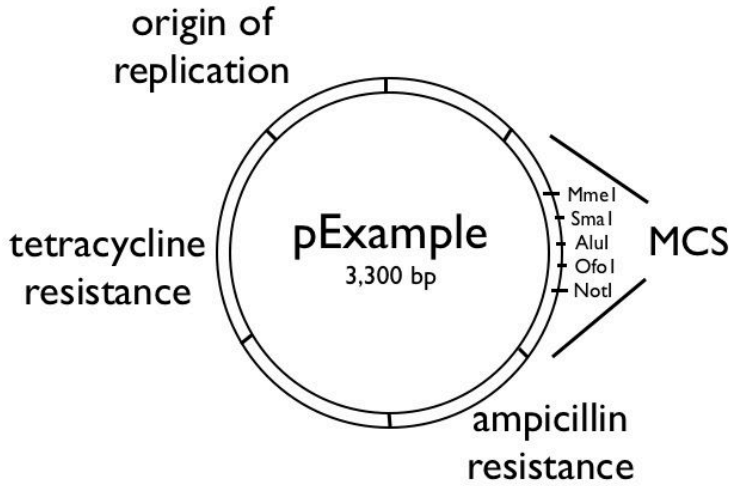


Figure 9. Hypothetical plasmid (pExample) map with 3,300 base pairs (bp). Shown are the antibiotic resistance markers (tetracycline and ampicillin), origin of replication, and multiple cloning site (MCS) with example restriction digest sites.

The origin of replication (*ori*) is a sequence of bases that is recognized by DNA polymerase. The importance of the origin of replication becomes apparent when one considers the plasmid will eventually end up in a host cell, which will function to make several more copies of the plasmid. Sometimes there can be more than one copy of the *ori* sequence, resulting in a higher copy number plasmid. There can also be more than one type of *ori* sequence so that the plasmid can be replicated in different host cells (i.e. *E. coli* and *Agrobacterium*) and still be recognized by the subtle variations in DNA replication machinery of each species.

Plasmids also carry a selectable marker, a gene that codes for an enzyme that somehow detoxifies an antibiotic present in growth media. Again, the importance becomes clear after considering that the plasmid's destination is a host cell. The presence of an ampicillin resistance gene on a plasmid residing in an *E. coli* cell will allow the transformed *E. coli* to survive on media containing ampicillin. All other *E. coli* cells that do not have a plasmid will be selected against and die. Sometimes a plasmid might have multiple selectable markers, giving the researcher a wider array of choices regarding the antibiotic added to the growth media.

Also important to have on a plasmid is a multiple cloning site. Before they are used, circular plasmids can be opened up with enzymes that behave like molecular scissors. The enzymes, called restriction enzymes, cut at very specific sites in DNA. There are hundreds of restriction enzymes that each recognize a specific sequence, usually about 4-6 bases in length. The cut site is known based on previous research and so in this way DNA can be systematically broken where desired by adding the correct restriction enzyme. Generally when working with plasmids the restriction enzyme is preferred to cut in only one spot (the MCS), in order to open the closed plasmid into a single, linear piece of DNA.

Some restriction enzymes cut nice and evenly, leaving the ends of the double-stranded DNA blunted. Other restriction enzymes cut DNA and leave one strand sticking out farther than the other strand. That is to say, there are overhangs of single stranded DNA. These overhangs are what allows a smaller, different piece of DNA (i.e. a piece generated from a PCR) to complimentary base pair with the plasmid. Thus, the plasmid and new piece of DNA, referred to as the insert, will pair up to make a complete circle. Enzymes called ligases are then added with the plasmid-insert to act as molecular glue so that two pieces of DNA (plasmid and PCR product) remain together. The newly formed plasmid-insert construct can then be put into *E. coli* or other appropriate host cell to be propagated. Restriction digests and ligations proceed best under a controlled temperature. Thus, the microcentrifuge tubes containing enzymatic reactions are

kept on ice and then placed into a device such as a heat block after the restriction enzyme and plasmid are mixed (Figure 10).

Figure 10. A heat block is often used to regulate reactions in microcentrifuge tubes such as cutting DNA. Here the heat block has been set to 42.0°C.

Plasmids that are ligated with a PCR product are electroporated into electrocompetent *E.coli* or heat shocked into chemically competent *E. coli* in a process called transformation. Electroporation involves the use of an electroporator, a special cuvette to hold the cells and plasmid, and a small pod to hold the cuvette (Figure 11). Specially prepared cells are mixed with a small amount of plasmid and a pulse of electricity is applied. The electric current destabilizes the membrane just long enough for the plasmid to enter the cells. The cells are removed and placed in fresh media where they can recover. Transforming chemically competent cells is done using cold and hot temperatures, which also destabilizes the membrane to allow entry of the plasmid. These so-called heat shock transformations are considerably more common due to lower cost.

Figure 11. An electroporator, cuvette, and pod, which are all necessary to transform electrocompetent cells.

Newly transformed cells are not all transformed, some cells refuse to take up the plasmid. Thus, all cells from a transformation event need to be grown on a selective media to purge non-transformants. Selective media will inhibit growth of cells that lack the plasmid since they also lack the resistance gene. However, transformed cells will grow since they harbor the ability to render a selective agent (e.g. antibiotic) ineffective. The result is that all cells growing on a particular media with a selective agent have been genetically transformed and carry newly inserted DNA.

Using plasmids in plant biotechnology

Plasmids, which are sometimes referred to as vectors, are of huge significance in plant biotechnology. These small, circular pieces of DNA can have interchangeable parts that are adaptable to a wide array of purposes. Different *ori* sequences, selectable markers, multiple cloning sites, promoters, and reporter genes can be used, depending on the research goal. Plasmids used in labs were first found in naturally occurring bacteria and have since been modified to suit the needs of researchers. Learning the different features of plasmids used in plant

biotechnology is easier if one understands the purpose and function of plasmids found in nature.

Bacteria found in different environmental conditions might harbor plasmids that have genes that provide a survival advantage. Some plasmids called R plasmids encode for antibiotic *r*esistance proteins that degrade or inhibit the function of antibiotics. Other plasmids encode protein machinery that helps transfer an entire plasmid from one bacteria to another in a process called conjugation. Agrobacterium carries yet another type of plasmid called a Ti plasmid that is able to deliver its DNA to a plant cell.

Biotechnology has found many uses of plasmids. Since plasmids are small extrachromosomal pieces of DNA, they can be used to carry, amplify, and manipulate foreign genes. They can also be separated from genomic DNA and other cellular components easily by using centrifugation. Producing more plasmid first requires selecting for transformed cells, transferring an isolated colony from a Petri dish to broth, and letting the cells grow to a healthy density (usually overnight culture suffices). The cells are subsequently centrifuged in a microcentrifuge tube and then resuspended in a lysis solution, which breaks the cells open and releases the plasmid.

Another centrifugation step pulls all cellular material to the bottom of the microcentrifuge tube. The upper, liquid portion contains the plasmid DNA and is transferred to a new microcentrifuge tube, similar to that of a genomic DNA extraction. Again, the new tube has a smaller, second tube inside that also has a membrane filter. A centrifugation step pulls the liquid through the membrane but traps the plasmid DNA. Finally, the small tube with the membrane filter is transferred to a new microcentrifuge tube, a small amount (~50 μl) of water is added, and the tube is spun. The DNA is eluted with the water through the membrane. Just like the genomic DNA, the plasmid DNA should be quantified. Obtaining plasmid DNA from an overnight culture is called a plasmid miniprep (plasmid maxipreps yield more plasmid) and usually provides enough DNA to be used in several downstream experiments.

Typically, once a researcher obtains a plasmid and successfully transforms a host cell, unlimited amounts of plasmid can be made by growing the transformed cells in media. Importantly, host cells cannot simply be grown in growth media. There must be a selective agent (i.e. an antibiotic) present that forces the host to carry the plasmid DNA. Since an antibiotic resistance gene and the gene of interest are both present on the plasmid, the gene of interest is copied over and

over with each generation of cells. Forgetting to add the antibiotic to selective media often results in growth of a host cell that lacks a plasmid.

There are several types of resistance genes (also called selectable markers) on plasmids and choosing the right one depends on the organism one is using as a host. In the most straightforward case, bacteria will harbor plasmids that usually have an antibiotic marker. Cells that carry the antibiotic resistance gene (i.e. the selectable marker) will persist. Plant cells that are transformed might also need to have a selectable marker in addition to a gene of interest. One of the most common selectable markers for plants is the *nptII* gene, which encodes an enzyme that essentially detoxifies kanamycin. Herbicide resistance genes (e.g. *bar*) are also used as selectable markers (resistance gene) in plant biotechnology where phosphinothricin is a common selective agent that kills non-transformed cells.

In addition to a selectable marker, a multiple cloning site (MCS) is also on a plasmid. The MCS is the site where a plasmid can be cut using a restriction enzyme. A series of cut sites are included in the MCS so that a researcher has several options of restriction enzymes to use. Choosing one restriction enzyme over another largely depends on the insert in the plasmid. Sometimes inserts will have the same series of bases that will be recognized as a cut site and so a researcher will opt to use a different restriction enzyme.

Another important plasmid feature, just upstream of (before) the multiple cloning site, is the promoter, which promotes expression of a gene. Dicotyledonous plants like cannabis will differ from monocotyledonous plants in their ability to recognize certain promoters and the sequence elements that guide the transcriptional machinery, thus not all promoter sequences are the same. Promoters are not always present upstream of a multiple cloning site, as is the case for cloning vectors that are used to solely propagate an insert.

If a researcher desires an insert to be transcribed by the host, an expression vector is required. The presence of a promoter allows RNA polymerase to bind upstream of the insert and generate mRNA, which will become translated into a protein. An expression vector transformed into a bacterial cell requires a bacterial promoter while a transgene delivered to a plant cell requires a eukaryotic promoter. This is a concern if one is attempting to deliver a bacterial gene for expression in a plant like cannabis. Expressing a gene in an organism where it is not naturally found is known as heterologous gene expression.

A common promoter used for expressing transgenes in plants is the CaMV35S, derived from the cauliflower mosaic virus.[3] The 35S promoter is a constitutive promoter in that transcription of a transgene is constantly promoted.

In other words, the mRNA of the transgene is constantly produced and often results in an abundance of protein. The 35S promoter is a good example of a promoter that is not expressed equally by both monocots and dicots, as dicots have shown higher expression levels compared to monocots.

In contrast with constitutive promoters, inducible promoters can also be incorporated into a plasmid. Under environmental conditions or application of an inducer, the gene is expressed. Antibiotics like tetracycline, light, and even temperature can be used to induce gene expression. Promoters can also be tissue-specific so that a particular tissue type expresses a transgene. A good example is if a researcher is wanting to restrict protein production to the root. Thus, an entire carrot plant does not need to waste energy making a protein above ground when the protein is intended to only accumulate in the root.

There are also anther-specific promoters that can be used to ensure male sterility. Such a promoter might be of interest to cannabis growers, since pollen produced by males can decrease the levels of cannabinoids. Experiments have shown that anther-specific promoters that are placed upstream of the gene encoding for the enzyme barnase (an RNase), results in the inability of developing anthers to form pollen.[4]

In addition to different promoters upstream of the multiple cloning site, other features might be present on expression vectors. For example, promoters are needed upstream of a transgene but sometimes terminators are needed downstream to signal RNA polymerase to halt transcription. Additionally, signal sequences can be added so they are expressed in tandem with the transgene in order to relocate a protein, perhaps to a particular organelle.

Finally, reporter genes are sometimes incorporated on the plasmid and also delivered to a plant so that a researcher can quickly assay where or when a protein is being expressed.[5] Depending on the reporter gene, protein levels (of the transgene) can also be assessed. One of the first transgenic plant studies that gained wide attention was the expression of the enzyme luciferase in tobacco.[6] Luciferase has since been used as a reporter, although to a decreasing extent. Fluorescent proteins such as green fluorescent protein (GFP) are more commonly used as reporter genes since their fluorescence can be detected more easily. The tissue potentially expressing the GFP must be subjected to short wavelength light (UV light), which interacts with the chromophore on the GFP. A photon with a longer wavelength (less energy) is released, resulting in a bright green color.

A very popular reporter gene used in plant biotechnology codes for the enzyme ß-glucuronidase (GUS). Upon addition of its substrate to a transformed

tissue, a blue color results so that a researcher can identify the location and amount of transgene expression. In contrast to GFP, a fluorescent microscope is not needed with a GUS assay.

In plant biotechnology, a researcher has many options for choosing among different combinations of *ori* sequences, multiple cloning sites, promoters, terminators, and reporter genes. Rarely does one need to build a vector from scratch. Such a method, known as vector construction, is an entirely different process compared to ligating an insert into a cloning site. Labs that dedicate their research towards generating new plasmids for distribution sometimes focus their research solely on vector construction. Plant biotechnology has been progressing long enough that there is now a large selection of plasmids from which one can choose.

An important cloning method that deserves discussion is known as Gateway cloning. Gateway cloning[7] seeks to clone a gene into a vector but avoids the cut and paste method of restriction digest followed by ligation reactions in traditional gene cloning. After the initial step of PCR the gene of interest is inserted into an 'entry vector.' To simplify the cloning reaction, the PCR product has a few bases attached at one end that are complimentary to the free ends of the entry vector. Importantly, special sites called recombination sites are present on the entry vector and flank the inserted PCR product.

The entry vector is then inserted into a host such as *E. coli* and a plasmid mini-prep generates more plasmid-insert product. A small amount of entry vector can be mixed with a second vector called a 'destination vector.' The recombination sites that flank the inserted PCR product on the entry vector facilitate the transfer of the insert into a complimentary site on the destination vector. Since there are multiple types of destination vectors, the gene in the entry vector can be transferred to several different destination vectors. Some destination vectors are binary vectors, which are used in Agrobacterium-mediated plant transformation. The Gateway method avoids the need to digest an insert out of a plasmid, which can sometimes result in digestion or loss of the insert.

While the exact methods are not given here on how to do each molecular biology technique outlined above, such information would be superfluous. Many techniques are carried out using kits, a small collection of reagents and plasticware specialized for a particular reaction that can be easily ordered through several companies. Each kit comes with a tailored set of detailed instructions. For example, there are kits for DNA and RNA extraction, plasmid minipreps, and gene cloning. In the latter case, some kits have been developed that allow direct

cloning of a PCR product with a prepared vector. Thus, details on each method are not necessary since one will have those instructions as they are needed. The aim thus far has been to provide a basic outline of the core tools and describe gene cloning, which the cannabis biotechnologist will find to be crucial.

Growing plants in lab

Besides being part of the research, plants can add a vibrant pulse to a lab but in a plant biotechnology lab they also give a unique combination of motivation and newfound curiosity. After understanding more about the plant cell, one cannot help but stand in amazement at all the processes of photosynthesis, respiration, mitotic processes, and biochemical reactions speeding through the plant. Especially unique is the feeling of being able to work with plants, manipulate and change their DNA, and propagate them.

Plants used in research are often grown in special growth chambers, specially designed for use in research labs (Figure 12). In larger research buildings, plants might occur in a completely separate room from where the molecular biology research component of research is taking place. Larger scale experiments might also require larger space and the use of greenhouses is not uncommon. Growth chambers should have proper lighting for the plants and controlled conditions should be maintained, especially when performing a research project. In vitro tissue cultures are obviously kept away from plants being grown in soil.

Micropropagation, growing plants in vitro, has the benefit of being able to vegetatively propagate a species in numbers that are only limited by the size of the space available to house the cultures. That said, in vitro cultures typically require less research space than traditional soil-grown (in solium) plants. Still, while in vitro methods are used in plant biotechnology, traditional methods are also required.

Helping a plant transition from the in vitro environment to soil environment can be challenging but also rewarding. With experience, one will find pleasure in working along side the plants they are studying. The feeling one gets from watching an experimental line of plants grow and develop over time is difficult to describe. Even their eventual production of seeds and slow senescence can be a thing of beauty.

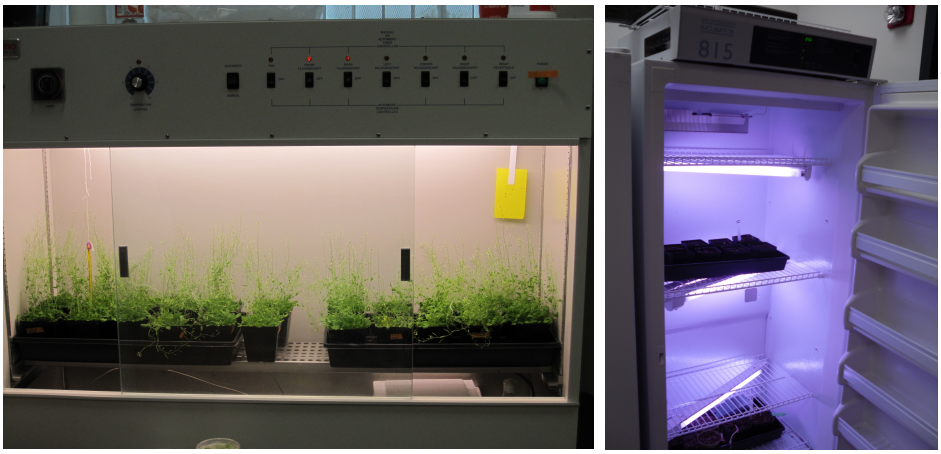


Figure 12. Two different plant growth chambers sometimes encountered in plant biotechnology labs.

With their amazing physiology and unique structures, plants might also lead to problems in a lab. Soil for the plants should be kept at a distance from any molecular biology research space. Additionally, there are several insects that can infest plants and ruin experiments. Extra caution and care must be taken when bringing in new soil or new plants into a lab since they can serve as carriers of pathogens. In a growth chamber where conditions are often ideal for plants growing in soil, conditions can also be ideal for some insects.

Once in the lab, a solution to getting rid of unwanted insects is to place the plants and the soil into a bag, put them in the autoclave, and decide who was responsible for the infestation. Completely giving up on the plants usually only happens after the researcher has tried insecticides, sticky paper or other physical means of eradication. Some common insect pathogens that can occur in lab include aphids, thrips, whiteflies, and fungus gnats. Paradoxically, generating plants with insect resistant traits is of large interest to plant biotechnologists.

A researcher should take notice of daily changes in plants growing in lab, whether for an experiment or for aesthetics. While there are many ways to deal with insects that can invade a plant biotechnology lab, perhaps prevention is the best method for avoiding catastrophes or complications with research. For the most part though, good researchers often have success in growing healthy, pest-free plants. Plants also tend to do fairly well by themselves, only requiring an occasional watering and proper light. Growing plants in vitro shares similarities to growing plants in soil but is quite different in other respects. In fact, in vitro manipulation of plants is of such high importance to plant biotechnologists that the subject deserves its own chapter.

Chapter 4

Tissue Culture

"Ich will jetzt noch die Ergebnisse der mit den isolierten Zellen anderer Pflanzen durchgeführten Culturversuche mittheilen. Allerdings handelte es sich dabei bloß um vereinzelte Versuche, die aber immerhin einige bemerkenswerte Resultate ergaben."
Gottlieb Haberlandt,[1] 1902

Making media

There are many lab techniques that are important to the plant biotechnologist. To learn the techniques, there are dozens of books with different pieces of information and nearly all of them have descriptions of tissue culture and gene delivery methods. Embedded within these subjects are several areas worth spending time covering such as making media, culturing bacteria and plants in vitro, sterile technique, and binary vectors.

One of the first and most important techniques a new lab technician is taught is how to make media. The ability to properly make sterile growth media is the foundation for maintaining microbiological cultures and plant tissue cultures. Media can be divided into two large categories, defined media and complex media. Defined media is also called synthetic media and is media where all components and ingredients are exactly known. They are defined. An example is BG-11 used for culturing cyanobacteria. Complex media might contain some ingredients that are not exactly known either in composition or quantities. As an example, peptone is a common ingredient in undefined media and is prepared by partially digesting extracts of beef or yeast. The extracts might be slightly different each time or the degree of digestion might also vary.

Media might be liquid (broth) or solid, although semi-solid media is also sometimes used, especially when roots will need to be gently removed from media. For proper cellular growth, media must contain an energy source, particular ions, and nutrients. Liquid media might be modified with the addition of a polymer called agar that is produced by algae and results in solid media. Frau Hesse, a woman who used agar in making jelly, has been given credit with first recommending its use in bacteriological media.[2] Luckily, agar is a carbohydrate that resists degradation by most microbes, so cultures grown on solid media will be unable to metabolize the agar as an energy source. Other solidifying agents are also used in plant tissue media such as phytogel.

Several types of solid and liquid media have been made and are readily available for purchase from supply companies. These include general-purpose media, which serves a wide array of organisms. Tryptic soy agar (TSA) is often cited as an example. General-purpose media like TSA allows culturing of several species of microbes, functioning like a buffet (as one of my colleagues always says). Some media can be made and have added components that restrict the growth of certain organisms. For example, a general-purpose media but with the addition of antibiotics will restrict growth of non-resistant cells. A researcher could then place a population of cells, some of which have been transformed with a selectable marker, onto the selective media and allow growth of positive transformants yet restrict growth of non-transformants.

While TSA is generally used for bacterial cultures, other media have been developed that are used for plants. A common media for plant tissue culture is called MS media in honor of the original developers, Toshio Murashige and Folke Skoog.[3] MS media has the required salts, macronutrients, and micronutrients to support many plant species (Table 1). Some media recipes will require only 1/2 MS salts and so the preparer simply adds half the amount called for in the usual recipe.

MS media is a basal media in that only basic requirements for cell growth are met and allows researchers to add other components required for a specific experiment. There are other basal media used in plant tissue culture as well (e.g. Gamborg's B5). The "basal salts" used in plant tissue culture media might often have sugars and added hormones in order to induce the growth of particular organs. Two media of importance in this respect are root initiation media and shoot initiation media.

In many cases, when media is purchased from a company and shipped, the form is a dry powder (but this might vary). When making media the powder needs

to be weighed and hydrated with the appropriate amount of deionized water. The media might require addition of sugars or agar, which are also shipped dry. To prepare media, the amount in grams needed is used in place of water (weight per volume). For example, to make a 500 ml quantity of TSA one might add 20 grams of TSA powder and 7 grams of agar to 473 ml of water.

Name of Chemical	**Formula/Description**
ammonium nitrate	NH_4NO_3
calcium chloride	$CaCl_2$
magnesium sulfate	$MgSO_4$
manganese sulfate	$MgSO_4$
potassium phosphate	KH_2PO_4
ferrous sulfate	$FeSO_4$
zinc sulfate	$ZnSO_4$
potassium nitrate	KNO_3
potassium iodide	KI
cupric sulfate	$CuSO_4$
boric acid	H_3BO_3
cobalt chloride	$CoCl_2$
sodium molybdate	Na_2MoO_4
ethylenediamine tetra-acetic acid	metal chelator

Table 1. Ingredients for MS basal media (pH 5.8). Researchers might also add sugars (e.g. sucrose), agar (to solidify media), shoot hormone (e.g. kinetin), root hormone (e.g. indole acetic acid), amino acids (e.g. glycine), or coenzymes (e.g. niacin).

The ingredients are sometimes mixed in an Erlenmeyer flask, which is a flat-bottomed flask that has a narrow neck. The flask is ready to autoclave after covering the container with aluminum foil or lid that can withstand the autoclave. The autoclave is an oven-like chamber that reaches high pressure and temperatures. The highest temperature most autoclaves reach is 121°C, which kills bacterial endospores that can usually survive harsh environmental conditions. Purging all life from the media by autoclaving is important since any bacteria or fungal spores left in the media might grow, possibly ruining later experiments.

After autoclaving the media is allowed to cool inside a laminar flow hood (Figure 1) where there is negative pressure that aids in keeping the work area sterile. Above the working area within the laminar flow hood, sterile air is forced down. The flow of air is aided by perforations in the back of the hood and lower front part of the hood opening that draw in air.

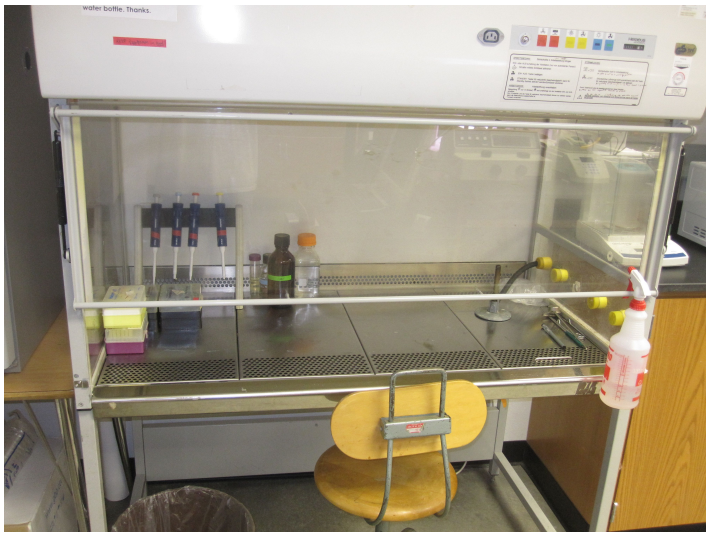


Figure 1. A laminar flow hood where media is prepared or cultures can be manipulated in sterile conditions.

Flow hoods use a high efficiency particulate air (HEPA) filter and are usually checked at least once a year for quality control purposes. The laminar flow hood is the only safe place for carrying out any work that must be kept free from contamination.

Petri dishes are often purchased in bulk and solid media is often prepared as a sleeve (20 plates). This allows one to remove media from the fridge when needed. Pouring plates (i.e., putting molten media into Petri dishes) is the method of removing the Petri dishes from the clear plastic sleeve, making stacks of 4-5 plates, carefully opening each dish, and then pouring molten media inside. Approximately 15-20 ml of media is needed for each plate, which is just more than enough to cover the bottom surface of the plate. The amount of media used in each plate is not measured, rather one simply estimates visually.

The plates are then carefully lifted one at a time from the stack and laid out inside the laminar flow hood in a grid-like fashion. As they are placed onto the surface of the flow hood, the lids are left slightly off from covering the plate completely in order to allow the media to cool faster and let excess moisture escape. During placement and movement of the plates containing molten media one must not splash media onto the inside lid of the plate. If this happens, there will be an increased chance for contamination when working with the plates at a later time.

When the plates have cooled and solidified the lids are gently moved to completely cover the bottom half of the Petri dishes and they are restacked into one column. The sleeve is placed back on from top to bottom and slid down so

that the entire contents can be inverted. One must also appropriately label the sleeve with the preparer's name, the date, and type of media. The media is usually kept refrigerated at 4°C.

The entire process of making media and pouring plates takes approximately 2 hours, although the time is reduced with practice. In industry a different technician other than the one doing the plant biotechnology research might make media. If so, the media labels on the sleeves or even individual plates have a number associated with them that permits tracing the media to the preparer. This is done so that if something goes wrong in the culturing portion of the experiment one can check the quality of the media and possibly determine what might have happened.

Plant hormones and other media components

In some cases, additional media components will need to be added to media after autoclaving. This is largely due to the fact that such molecules might not be heat stable and their chemical bonds might be broken if exposed to a certain heat and pressure (known as the melting point). Therefore, once the media is removed from the autoclave and the media is cool enough to touch (~55°C), additions can be made. Common components added are plant hormones, antibiotics, activated charcoal, and broad-spectrum antifungals, which are also often purchased and shipped in a powder (dry) form.

Hormones and antibiotics vary in their solubility. One will therefore need to know what solvent to use (e.g. water or ethanol) to suspend a particular reagent. Several small tubes of stock solution (solvent with solute) can be made. Often, the best thing to do is filter sterilize (comparable to using a microscopic strainer) the stock solution prior to freezing, ensuring removal of fungal spores or other potential contaminants that might be present. The antibiotic or hormone stock solution is stored at cold temperatures and used as needed.

The addition of plant hormones largely depends on the media type one is making. Plant hormones include auxins, cytokinins, gibberellins, ethylene, abscisic acid, jasmonates, salicylic acid, and brassinolides. They are an important part of *Cannabis* development and biochemistry and play a crucial role in tissue culture. Hormones act as chemical regulators of gene expression to guide development and the morphology (observable shape) of *Cannabis*. Changing the hormones in plant tissue culture media allows for manipulation of organ development in vitro.

Indole-3-acetic acid (IAA) and 2,4-dichlorophenoxyacetic acid (2,4-D) are perhaps the most widely known auxins. The amino acid tryptophan is enzymatically modified to produce indole-3-acetic acid. Auxins are commonly found in developing leaves and seeds and primarily function to control apical dominance. They also play an important role in stimulating flowering and fruit development, and are frequently used to stimulate adventitious roots (growing from the stem) in asexual cuttings (clones). IAA and 2,4-D are used in different degrees in plant tissue culture but are often required in root initiation media.

Similar to auxins, cytokinins vary in their molecular structures but a common one in plants is zeatin. Cytokinins are synthesized in the roots and are moved through the xylem to the shoots to regulate mitosis. Artificial application can induce lateral buds to branch. Cytokinins can also cause a delay in leaf senescence. In plant tissue culture, cytokinins are used primarily for shoot development in shoot initiation media.

Gibberellins were one of the first hormones to be used in plant experiments. They elicit perhaps one of the most obvious effects on a plant. If two plants are grown in separate pots, side by side and gibberellins are applied to one plant, that plant will grow several magnitudes taller than the other. Gibberellic acid is one of the most important and common gibberellins in plants. Gibberellins are produced from mevalonate, a precursor in synthesis of terpenes (a class of plant metabolites). They are found in immature *Cannabis* leaves and seeds. In most species gibberellins help in elongating shoots and regulate some seed enzymes (protein catalysts).

Ethylene (C_2H_4) is a gas, which functions as a plant hormone. The amino acid methionine is the precursor of ethylene. Because ethylene is a small molecule, moving from cell to cell via diffusion is easily accomplished. This hormone gave rise to the old adage that one bad apple can spoil the whole bunch since a rotting fruit emits ethylene, which in turn might ripen other fruits. When tomatoes are picked and shipped, they are green. Just prior to arrival at the grocery store the tomatoes are exposed to ethylene, ripening and reddening the tomato fruits. In plant tissue cultures ethylene has the ability to accumulate and therefore cultures sometimes benefit from venting.

Like gibberellins, abscisic acid is also synthesized from mevalonate. By applying abscisic acid to seeds, they can be kept dormant for shipping, so as not to allow them to mistakenly sprout. Plants manufacture abscisic acid as a messenger molecule in response to a wide array of biotic and abiotic stresses. For example, abscisic acid is an important regulator of stomata. In plant biotechnology, abscisic

acid can induce somatic tissue to form embryonic tissue (somatic embryogenesis) but has also shown to improve resiliency of tissues in vitro.

Although only a basic introduction is provided here, there are entire books written on plant hormones.[4] The important piece of information to remember in cannabis tissue culture is that auxins are used to help initiate root development and cytokinins are used in shoot initiation. Knowing how cannabis tissues are being affected by different media components such as hormones, can be crucial for successful experiments.

In addition to hormones, several antibiotics are used in microbiological experiments and typically only a few are used in plant tissue culture. Kanamycin, which targets the bacterial ribosome, is one such antibiotic. Since plant plastids are essentially cyanobacteria, their ribosomal architecture closely matches that of prokaryotes and is therefore susceptible to protein synthesis inhibition by kanamycin. However, transformed plants can render kanamycin ineffective if they have been transformed with the kanamycin resistance gene (*nptII*). Rifampicin is also used in plant biotechnology to select for Agrobacterium and inhibit *E. coli* that might be carried over from plasmid preparations. Antibiotics are important in other areas of plant biotechnology such as plant cell transformation and so the topic will be revisited.

Bacterial cultures

While making media is a highly useful skill, an equally important skill that any biologist working in a laboratory should have is the ability to obtain and maintain pure cultures. Pure cultures include those of bacterial origin as well as plant origin. Keeping pure cultures can become so important that month-long experiments can be compromised if pure cultures are not maintained. Largely this is due to contamination issues in plant tissue culture but one must also be sure that an *E. coli* or Agrobacterium used for plasmid propagation is indeed the bacteria one thinks, since bacteria of different species growing on a Petri dish can sometimes be difficult to distinguish based solely on colony morphology.

In working with bacteria, streaking for isolation is of central importance. In many cases the investigator might be taking some cells from a frozen stock and plating them. A nichrome loop is used to collect cells, which is sterilized before and after being used. Nichrome is a special metal alloy that rapidly heats and cools, minimizing work time. To generate isolated colonies of bacteria, a small loopful of

bacteria is placed on one edge of a Petri dish using the sterile nichrome loop. Subsequently, the loop is sterilized again and is used to drag a portion of the cells across the media. Two or three more repetitions are performed, streaking only a portion of the cells of the previous streak. The result is a Petri dish with much of its surface utilized in order to gradually spread out the bacteria. After overnight incubation, some of the cells should grow into isolated colonies, which originate from one single cell (theoretically).

Using the sterile loop, one can then pick from one of the isolated colonies and place bacteria onto a new plate for subcultures. This is best done in a laminar flow hood but so much really depends on the research project and individual. Isolated colonies can be grown as larger batches of cells in several ways, one of which is an Erlenmeyer flask containing the appropriate liquid media. Incubation temperature is often kept at a constant temperature, which keeps the cell enzymes most active.

Cells of different species share a pattern of growth known as the growth curve, easily observable in liquid media. The applications of a growth curve extend beyond the microbiology classroom and are widely used in food and industrial microbiology. As cells adjust to a new media, they must synthesize new components for cell metabolism and thus, first there is a lag phase. The lag phase can vary from minutes to days depending on the species and prior growth conditions. Next is the log phase, sometimes called the exponential phase, since the cells are increasing at an exponential rate. The stationary phase follows and occurs when the growth rate is equal to the death rate. Alternatively, the cells in the population might not increase in numbers or decrease due to death. The stationary phase is followed by the final stage, which is the death phase. Metabolites might be building up that become toxic to the cells or the energy reserves of the media have been exhausted.

Microbiologists, which include people interested in manufacturing sought metabolites (e.g. vitamins) for human consumption or insulin production from *E. coli,* have exploited the growth phase by keeping microbes in a stationary growth phase. In such a case, a turbidostat or bioreactor might be used, which are cell culture devices that maintain a desired cell density, allowing removal of waste products and entry of fresh media. Desired metabolites can also be removed and concentrated for use in other applications. Thus, some metabolomic studies frequently incorporate the use of turbidostats or bioreactors.

Plant tissue culture

Plant tissue culture has a fascinating history that can be traced back to the turn of the twentieth century.[5] Culturing plants is similar to culturing microbes in that they both require sterile technique and a cautious hand. Instead of starting with cells to simply streak onto media, culturing plants in vitro requires starting plant tissue or seeds. There is also a core set of items needed for plant tissue culture (Figure 2). In plant tissue culture researchers have used starting plant material that includes hole punches from leaves, sections of stems, pollen, and pieces of roots. Researchers in cannabis tissue culture have found success in using leaves[6,7] and stems[8] and even anthers.[9] Young tissue (a few weeks) appears to be more responsive to in vitro conditions compared to older tissue, likely due to the high rate of mitosis occurring. For this reason, shoot tips and axillary buds are also commonly used.

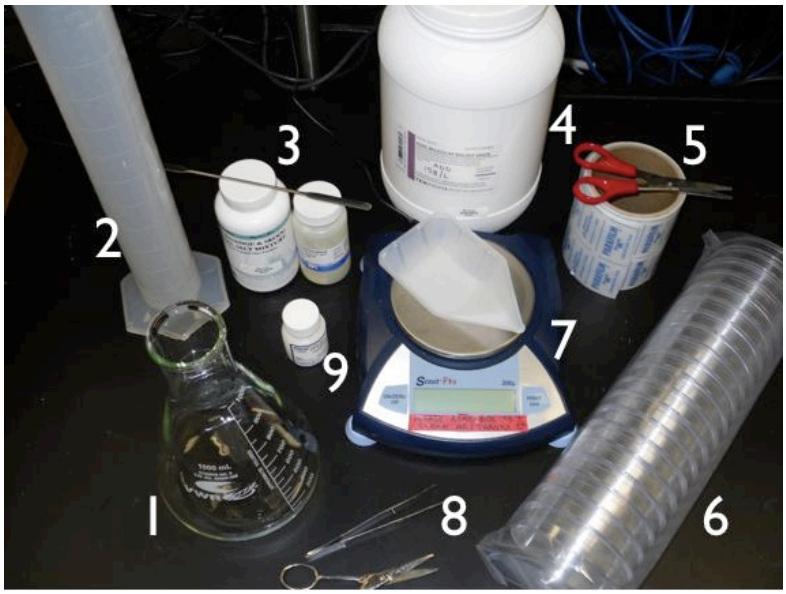


Figure 2. Some of the basic materials needed to begin cannabis tissue culture; 1) Erlenmeyer flask, 2) graduated cylinder, 3) MS and callus media, 4) agar, 5) parafilm, 6) sterile Petri dishes, 7) scale, 8) tweezers and small scissors, 9) antibiotic.

One might choose to remove a short stem from an outdoor or indoor plant but since microbes are ubiquitous, thorough washing is required to help ensure sterility. If even one spore or bacterium is present on the stem surface after the wash, the experiment might be jeopardized. To minimize unwanted contamination, one might desire to start by culturing a cannabis plant in vitro by

seed. Seeds should be washed thoroughly with a 5% bleach solution and rinsed several times. Seedlings grow well on media such as MS agar with 2% sucrose. A Magenta box or larger vessel can be used to allow one to collect plenty of stem tissue (Figure 3). Microbes should be absent from the seed-grown plant and so washing steps can be minimized.

Figure 3. Cannabis plant growing in standard Magenta box (left). Sterilized cannabis seeds can also be germinated and grown in other vessels containing supportive media (right).

Collection of a stem should be carried out with a few simple tools such as a small tweezers and scissors. All tools should be autoclaved in aluminum foil prior to using. Work at this point should be only carried out in the laminar flow hood.

To begin, the stem (or other tissue) is placed in a sterile tube and is ready to be lightly washed with a mild detergent (Tween20) and 70% alcohol solution. Pieces of plant tissue or organs (leaf, stem, etc.) are often referred to as explants after being placed onto Petri dishes. Washing is just as crucial as cutting and trimming the stem. If the wash steps are too long, the plant cells will die. If the washes are too short, microbial contaminants might remain and fungus or bacteria will eventually overrun the growth media. Therefore, a perfect balance must be achieved to successfully wash the stem without killing the plant cells.

Generally, an initial wash with 70% alcohol (e.g., 3 ml water and 7 ml 100% alcohol for a total of 10 ml) is used with 1 µl of Tween20. The detergent is not always necessary, but it does aid in working the alcohol into any grooves on the outside of the stem. The tube is capped and shaken vigorously and allowed to sit at room temperature for 5 minutes. The tube is washed with sterile water and a second wash is implemented in the same fashion as the first wash but without the detergent. A final wash with a 1% bleach solution (1 ml bleach and 99 ml water,

but using only 10 ml of this) is preferred in some tissue culture methods; however when dealing with stems obtained from seedlings grown in vitro, such harsh wash steps might be unnecessary. Sterile water is used for a final series of rinses to ensure that the alcohol and bleach have been completely removed from the container with the stem.

Petri plates with callus initiation media should be removed from the fridge and placed in the hood. A second set of autoclaved tweezers and scissors, an alcohol dish, and a flame are also required. Keeping the tools from touching any part of the inside of the hood, the stem is removed with the tweezers and held steadily over an uncovered Petri dish containing tissue media. Small sections (1-2 mm) of the plant are clipped with the scissors and allowed to fall onto the callus media. Before and after each use, the ends of the scissors and tweezers are intermittently dipped in the alcohol after which the alcohol is burned off. Latex gloves are also used as a precaution to prohibit skin cells or bacteria from falling onto the plate.

When 20-30 pieces of stem have fallen onto the plate, the sterile tweezers can be used to manipulate and move the pieces of cannabis stem. The stem pieces should be placed equidistant from one another and gently pushed down to ensure complete contact with the media (Figure 4). The lid can then be placed back onto the Petri dish. Parafilm, a stretchy plastic film, is wrapped along the edge of the plate. This helps in retaining moisture and keeping the contents sterile. Another plate with sections of young stem or other tissue can then be made.

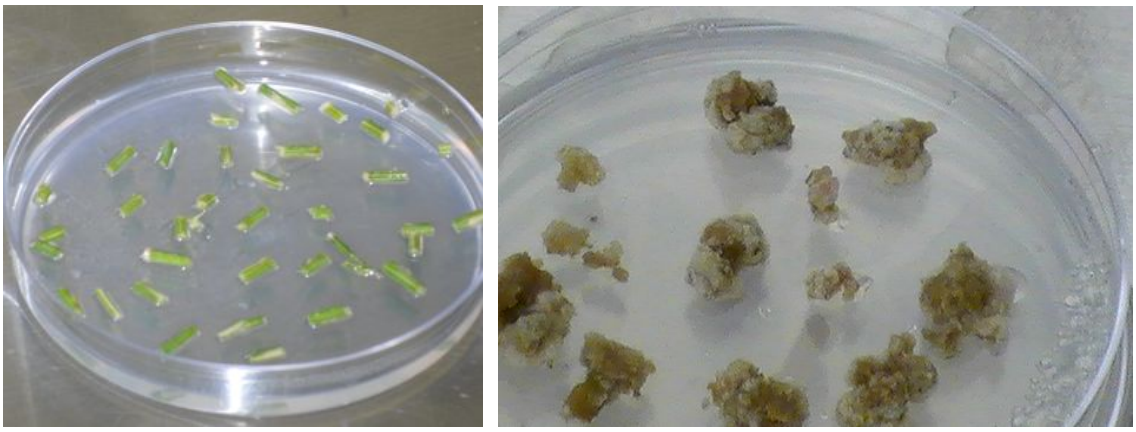


Figure 4. Petri dish with sterilized cannabis stem cuttings (left) and Petri dish with cannabis calli derived from the stem cuttings (right).

The Petri dishes with explants are put away from any disturbance and are usually kept at room temperature (22°C). The plant tissue does not need light since the presence of sucrose in the callus media serves as an energy source. After a few weeks the bits of cannabis stem will slowly start to change into an amorphous aggregate of cells. This is called a tissue callus (Latin for "hard skin") and the cells contain the genetic components of cannabis, but have the distinct quality of being able to develop into any plant organ (totipotency). In fact, callus media is specially balanced to inhibit organogenesis, the developing of roots or shoots. In some cases, calli are transferred every few weeks to new media.

To generate more cannabis calli, one callus can be fragmented into two or more pieces. The calli will continue to grow as long as they have the appropriate media requirements and do not have metabolites accumulating in the media. One callus might be easily broken apart (friable) and so a hundred new calli might be propagated from only a few dozen initial calli (Figure 5). Sometimes calli might be placed in a liquid medium and used in cell suspension culture. Generating more calli for tissue culture by fragmentation of existing calli is akin to cutting clones from large female cannabis plants. However, if propagated for too long, callus cells might undergo changes and result in somaclonal (genetic) variation. In some instances mutations via somaclonal variation can be pursued for further investigation but in general, stable genetic lines are desired.

When the calli have grown into amorphous masses they can be inoculated with a suspension of *Agrobacterium tumefaciens* that carries the appropriate gene of interest. Inoculation can be carried out directly in the Petri dish with the calli or placed in a media with the appropriate Agrobacterium density.

Figure 5. Stacks of Petri dishes with cannabis calli (left). In unfortunate instances some calli will be lost due to contamination from either bacteria or fungi (right). Antifungal agents, therefore, are sometimes added to tissue culture media.

The initial exposure to Agrobacterium is allowed to occur over a period of 1-7 days (varying by plant species) and is called cocultivation. During this time Agrobacterium is provided an opportunity to deliver the gene of interest and a selectable marker to the cannabis cells of the calli. After cocultivation the calli are washed in an antibiotic (e.g. timentin or cefotaxime) solution. The antibiotic will restrict growth of Agrobacterium but does not affect the cells of the calli. The calli are placed on plant growth media with antibiotic (e.g. timentin or cefotaxime) after the wash. Within a few weeks, the calli are transferred to new media containing a selective reagent (e.g. kanamycin) against non-transformed plant cells.

Next, transformed calli are carefully transferred to shoot or root initiation media and will begin to develop shoots or roots, respectively (Figure 7). Upon generating roots, the explants must be transferred to shoot initiation media, which contains IAA or other auxin. The shoot initiation media is typically prepared using a cytokinin such as zeatin. Regenerating plants from calli often incorporate Magenta boxes or test tubes to allow vertical growth of developing plants.

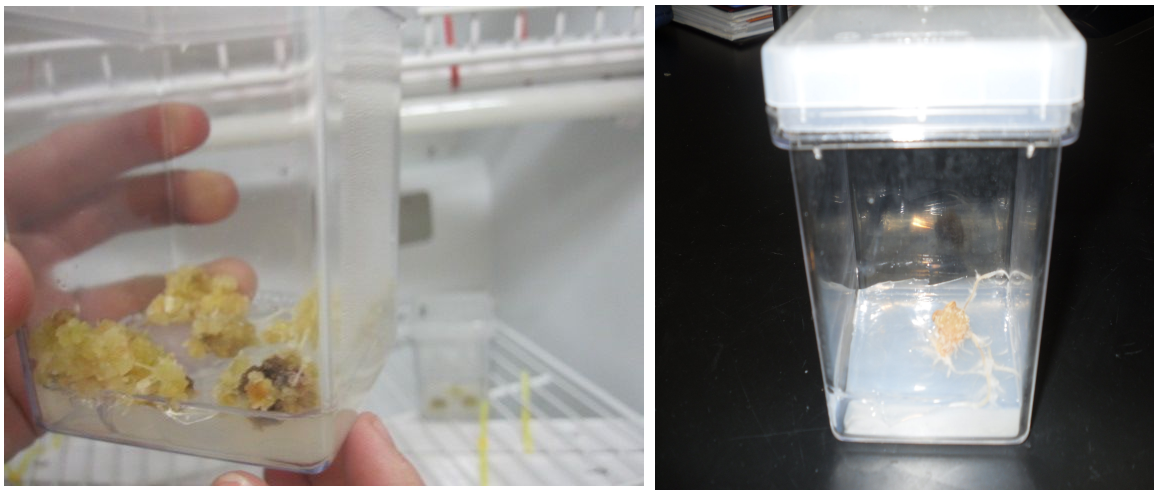


Figure 6. Cannabis calli almost ready for root initiation media (left) and a callus that has developed roots (right), and is ready for shoot initiation media.

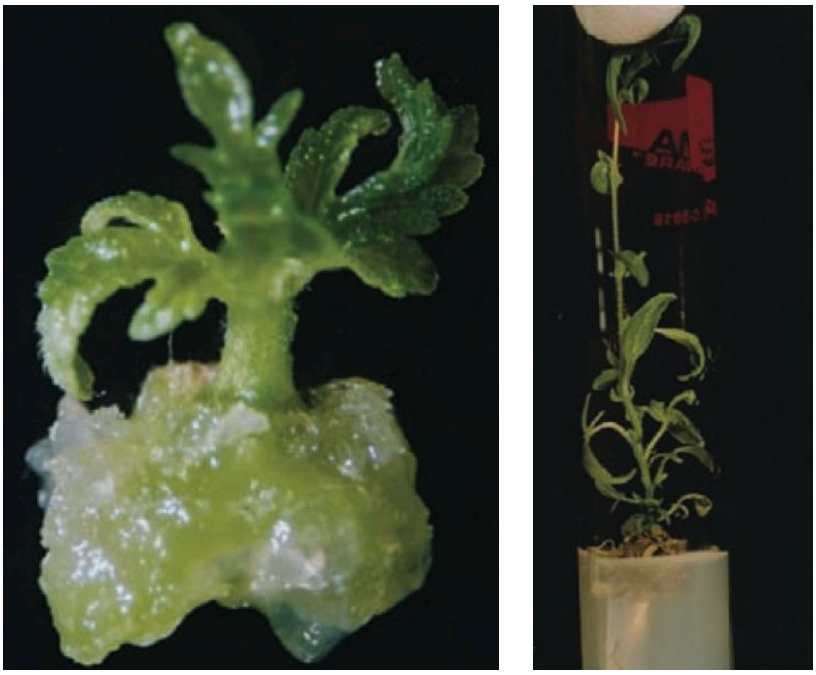


Figure 7. Cannabis callus with developing shoot (left) and a regenerated plant (right) generated by Slusarkiewicz-Jarzina and colleagues[8] who worked to optimize in vitro growth conditions for cannabis.

When shoots and roots have been developed that can begin to sustain the transgenic plant, the lid of the Magenta box (or other culture device) is slightly opened. A slow acclimation to the ambient environment is needed in order to avoid stressing the plant. Within several days the regenerated plant can be moved to a soil environment within a plant incubator and then to a greenhouse.

Cannabis in vitro

Cannabis has a history of being cultured in vitro for at least 40 years. Many of the early studies of growing cannabis in vitro were done in cell suspension cultures, where cells might be placed directly into a flask with liquid media. In many cases, the flask is moved on a "shaker" or agitating device that keeps the media aerated for gases to adequately reach the cells. In 1972 two researchers published a scientific article[10] that reported the growth of cannabis in cell suspension cultures. Their primary interest was in the metabolic products that were produced and they studied those products using chemical analytical techniques. Only five years later, in 1977 three researchers from Japan published a

scientific article[11] entitled, "Biotransformation of cannabinoid precursors and related alcohols by suspension cultures of callus induced from *Cannabis sativa* L."

Another report came in 1983 by three investigators at The Ohio State University.[12] The research team members, Loh, Hartsel, and Robertson, were a collaboration - advancements in science are often through collaborations - that involved the Department of Microbiology and College of Pharmacy. The three researchers sought to optimize the media conditions for cannabis but also wanted to determine if metabolic products could be produced. They first grew seeds on MS media with a B-5 complement of vitamins and harvested the stem sections and leaf discs from four-week-old seedlings. After careful surface sterilization the explants were developed into calli and then placed in liquid MS media with varying combinations and concentrations of three auxins [2,4-dichlorophenoxyacetic acid (2,4-D), 2,4,5-trichlorophenoxyacetic acid (2,4,5-T), and naphthalene acetic acid (NAA)] and three cytokinins [kinetin, 6(gamma, gamma-dimethylallyl-amino)-purine (2iP), and benzylamino purine (BAP)]. They reported that the best media for calli growth was MS media with 2,4,5-T at 3 mg per liter.

While some initial studies focused not only on growing cannabis for optimization of media conditions that was to allow for rapid and sustainable growth of cannabis cells, other studies focused on cannabinoid production. With the rise of plant transformation experiments researchers began to ask if cannabis could be genetically modified. One of the first reports of cannabis transformation was reported by Feeney and Punja in 2003, about ten years ago.[7] Their work, done at Simon Fraser University in British Columbia, experimented with establishing a protocol for in vitro culturing of hemp and its genetic modification using Agrobacterium. They incorporated a balance of 2,4-D and kinetin in their tissue culture media. Additionally, they tested several combinations of salts, vitamins, and hormones. The media resulted in well-developed calli, which would make anyone happy. Unfortunately, they were unable to regrow plantlets from calli. Such inability to regenerate plants had been encountered before in the 1980s.[12] Other reports trickled into the scientific literature on cannabis in vitro during the 1980s. While some luck was encountered in developing roots on cannabis calli, shoots were more difficult to generate.

Interestingly, one of the hormones used by Feeney and Punja while at Simon Fraser University in their attempt to establish an in vitro protocol for cannabis was thidiazuron, a potent cytokinin. They used segments from tissues grown on thidiazuron-impregnated media for four days then subcultured to media

without hormones. In another variation, calli cells were grown in suspension with thidiazuron and after one month were subcultured to media without hormones. In both cases, shoot development was not observed.

However, only a few years ago in 2009 Hemant Lata and her colleagues working at the University of Mississippi, reported that thidiazuron was the key ingredient in generating shoots directly from cannabis axillary bud explants.[13] Although a range of concentrations were tested, the best growth occurred on media with extremely small concentrations (0.5 μM). Amazingly, all cannabis explants were induced to produce shoots. The cost for a 100 milligram bottle of thidiazuron is about 70USD (phytotechlab.com) and is likely to last for several years worth of research. Much time and effort might have been saved through the decades of cannabis in vitro research if such a discovery was made sooner but that is the way of science and experimentation.

Prior to the discovery of thidiazuron's importance in cannabis tissue culture, others were successful in generating shoots from cannabis in vitro. In 2005, Aurelia Slusarkiewicz-Jarzina and two colleagues working at the Institute of Plant Genetics in Poznan, Poland reported shoot regeneration from cannabis calli.[8] They used different combinations of kinetin, NAA, 2,4-D, and dicamba (3,6-dichloro-o-anisic acid) with MS media. After 2-3 weeks in the dark, the explants were moved to light conditions and calli were transferred every 3 weeks to new media. IAA and NAA were used in the rooting media. Calli were generated rather easily in all cultivars. They used internodes, axillary buds, leaves, and petioles. Axillary buds showed a low generation of calli while petioles and young leaves showed the best. Dicamba produced the best calli as well. Amazingly, they tested five different cultivars. Their work was truly a valiant effort. At less than 2.0% of calli regenerating into whole plants (Figure 7), they recognized the additional experimentation required. The authors end their study with saying "further experiments are needed."

Thus, research in Poland continued and three years later, in 2008, Karolina Wielgus and her colleagues published a paper[14] entitled, "Estimation of cannabis tissue culture conditions essential for callus induction and plant regeneration." Their explants were generated from cotyledons, the early "false" leaves of seedlings. In some cases calli could consistently be generated and plant shoot regeneration was observed to be 14%.

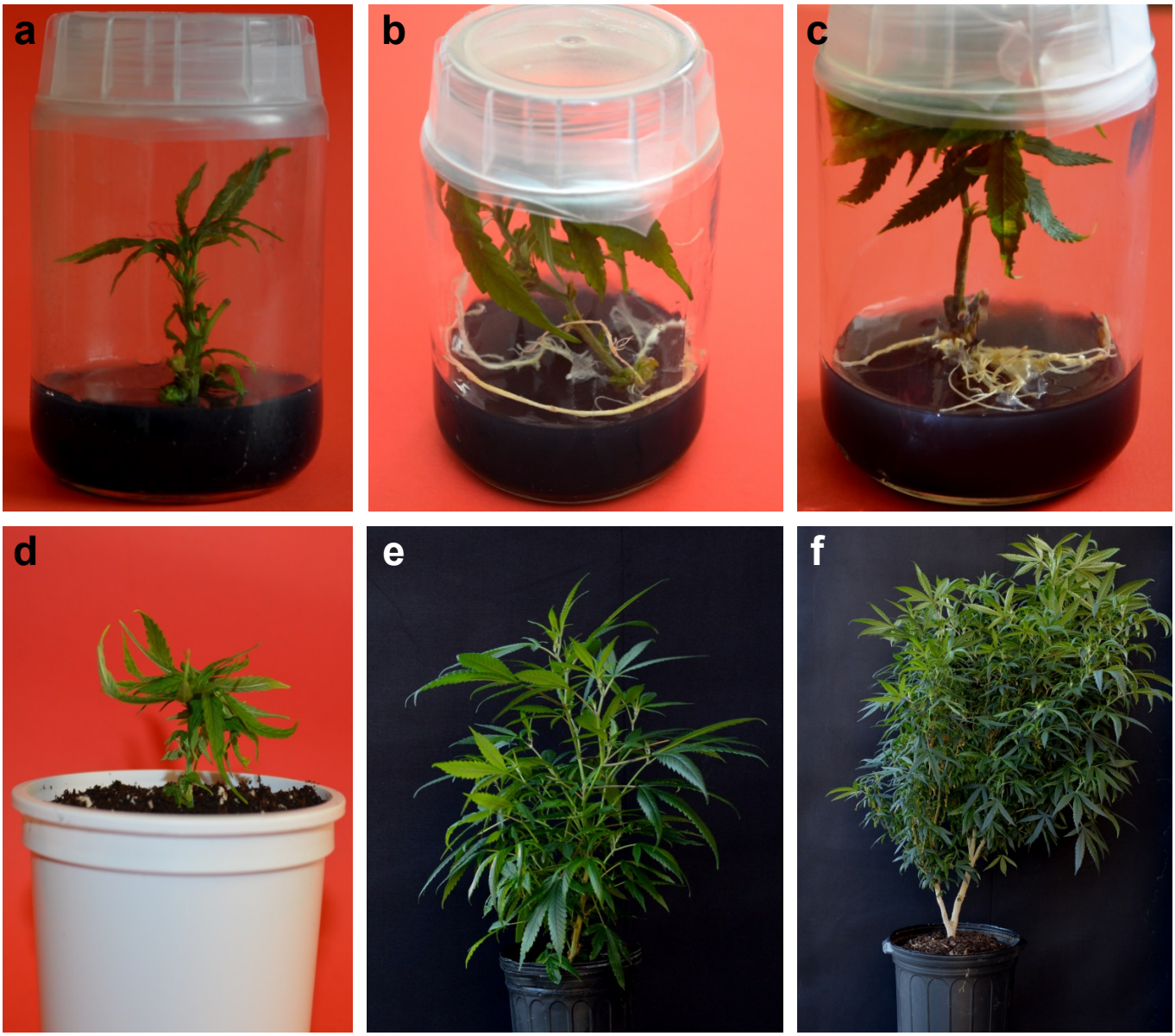
Although Aurelia Slusarkiewicz-Jarzina, Karolina Wielgus, and their colleagues were able to generate shoots from calli, the discovery and optimization by Lata and her colleagues of cannabis tissue culture and the importance of thidiazuron have generated newfound excitement in those working in the area of

cannabis biotechnology. Lata and her colleagues have since published another research paper working with leaf explants from high THC-yielding strains.[15] The explants were coaxed to first morph into calli that were then manipulated to produce several shoots. They had a success rate at regeneration of plants from calli of about 96%. Such results are extremely exciting since they boost the research potentials of producing genetically modified cannabis. Many others have contributed to the study of cannabis (in vitro and otherwise), several of which have recently been discussed by Suman Chandra and his colleagues[16] in a book entitled, "Biotechnology for Medicinal Plants."

As lightly suggested thus far, there are additional purposes for working with cannabis in vitro and calli are not always exposed to Agrobacterium. For example, simply being able to grow cannabis plants in vitro can open up new possibilities in propagation. Such a method for directly generating new plants from existing plants using in vitro methods is called micropropagation, which is a type of tissue culture. Micropropagation requires nearly all of the same tools and techniques as previously discussed for other methods of tissue culture. The main goal of micropropagation is to generate a large number of clones from one plant. Hence, the technique might still require media that encourages roots or shoot formation.

Depending on the research goals, generating clones from micropropagation can be extremely useful. For example, rare strains of cannabis can be rapidly propagated in tightly controlled conditions. Regulating the moisture, nutrients, light, temperature, and CO_2 concentration allows one to fine-tune the ambient environment to produce optimal growing conditions. Additionally, unless present before the micropropagation methods are initiated, pathogen-free plants are nearly guaranteed. Since the culture vessels can be positioned closer together and are sometimes smaller than pots with soil, less space is needed. All of these qualities might lead to lower maintenance costs for facilities that are attempting to reproduce large numbers of plants in a confined space. Thus, micropropagation can be highly economical. Hemant Lata and her colleagues have summarized the method of cannabis micropropagation on the following page. To help understand the process, photographs have also been included.

Micropropagation of *Cannabis sativa* L.



In-vitro propagation of *Cannabis sativa*. (a) Shoot formation on MS+0.5 µM TDZ, (b and c) Rooting on ½ MS medium + 2.5 µM IBA supplemented with 500 mg L^{-1} activated charcoal, (d) Rooted plant under hardening conditions, (e and f) Fully developed micropropagated plants of *C. sativa* at vegetative stage under climate controlled indoor growing conditions (Temperature, 25 ± 3 ^{o}C; Relative humidity, 55 ± 5 %; Photosynthetically active radiation, PAR, ~ 700 ± 24 µmmol $m^{-2}s^{-1}$ at plant canopy level; photoperiod 18 hours). PAR (measured by LI-COR quantum meter, model LI-189) was provided with full spectrum 1000 watts high density discharge, HID, lamps in combination with 1000 watt high pressure sodium bulbs, HPS, Sun Systems, CA.

Micropropagation protocol and photographs by: Hemant Lata, Suman Chandra, Ikhlas A. Khan and Mahmoud A. ElSohly, National Center for Natural Products Research, School of Pharmacy, University of Mississippi, University, MS 38677.

In addition to micropropagation manipulating cannabis in vitro can also find other applications. For instance, sometimes calli can be placed on media that initiates development of embryonic tissue. Since calli are somatic (non-reproductive) cells, the result is a phenomenon known as somatic embryogenesis. Somatic embryos are important because they can be preserved and stored for later research use. An interesting method preserves tissue by using a protective, spherical matrix that surrounds the tissue. The result is synthetic seeds (or synseeds), which look like small capsules containing plant tissue.[17]

Interestingly, researchers are now learning that somatic embryos are not the only tissue that can be used in synseed production. Synthetic seeds can be prepared with calli, shoot buds, shoot tips, or other tissue that can be regenerated. There are also different ways to produce synseeds. While some species are prepared by slowly desiccating the tissue, other species need to be preserved in a more hydrated state. The explant tissue is dropped into a solution containing 5% sodium alginate. After mixing with the sodium alginate, the explants are next placed into a solution of calcium chloride, which results in formation of small, encapsulated pieces of plant tissue, or synthetic seeds (Figure 8).

Figure 8. Cannabis synseeds generated from nodular tissue. Photo courtesy of Joe Oakes and Skunk Pharm Research, LLC.

Like typical seeds, synseeds can be initiated to grow under conducive environments. Although a variant of synthetic seeds have been used in plant biotechnology since the 1970s, newer methods of encapsulation have resulted in increasingly extending the length of time that the tissue is viable. In a recent study

by Hemant Lata and her colleagues[18] cannabis nodal segments were completely enclosed in sodium alginate and regenerated in vitro with no change to the cannabinoid profile, demonstrating synthetic seed technology can be applied to cannabis. Synthetic seed technology has rapidly changed over the last several decades but will likely become more important to cannabis biotechnologists. This is because synthetic seeds offer several advantages that are similar to those of micropropagation systems including increased throughput, smaller space requirements, propagation of genetic lines, preservation of rare genotypes, and can be economical.

Plant tissue culture techniques are very diverse and have a diverse array of applications. A researcher might need to implement several in vitro techniques in order to properly generate a cannabis plant expressing a newly introduced gene. To understand how a gene might become incorporated into plant tissue, let us direct our attention to methods of transgene delivery.

Chapter 5

Gene Delivery

"Crown gall tumors can, in light of our findings, be viewed as a feat of genetic engineering on the part of A. tumefaciens."
-Mary-Dell Chilton and colleagues,[1] 1977

Direct and indirect methods

One can see how a tissue is manipulated in vitro but how gene transfer occurs might not have equal clarity. Gene delivery can occur using two broad techniques that include either direct methods or indirect methods. Each differs in the tools and technology required and thus, there are benefits and drawbacks of each.[2]

Direct gene transfer methods have shown to be quite successful for a large number of species. Directly delivering genes can be done using a propelling force (biolistics) in order to make sure the foreign DNA carried by biologically inert metal particles transverses the cell wall. The device responsible for delivery is called a gene gun and the method incorporating this device is sometimes referred to as biolistics.

In other cases of direct gene transfer methods, such a propelling force is not needed. Plant cells can be gently stripped of their cell walls and grown in suspension cultures. The removal of the cell wall must be done in an isotonic solution since the resulting protoplasts might burst in a hypotonic solution or shrivel in a hypertonic solution. Without the cell wall, the cells are referred to as protoplasts and are placed in a small cuvette with the foreign DNA molecules. Then, the protoplasts are subjected to a short pulse of electricity. The electric

current destabilizes the plasma membrane and allows foreign DNA to transverse to the inside of the cell.

Direct methods of plant transformation have had the benefit of being able to transform monocotyledonous plants (e.g. grasses), which for a long time was difficult to carry out using indirect methods. However, with new strains of Agrobacterium, the ability to successfully transform monocots has increased. While indirect methods also sometimes rely on viral delivery, perhaps the most commonly used indirect method of plant genetic modification is Agrobacterium-mediated transformation.

While one might assume that plant biotechnology is restricted to plants, this is not the case as bacteria are a crucial component in almost every experiment. Knowledge on microbiological processes and understanding every field of microbiology is not necessary, however some background on bacteria are important, since they contribute to plant biotechnology.

There are millions of bacterial species. Only a small percentage has been well characterized thus far, mostly due to the fact that approximately 95% of bacteria do not grow in typical laboratory conditions. Bacteria are also ubiquitous, they grow in nearly every condition or habitat where humans have looked for them including battery tanks, oil wells, benthic zones of the ocean, and even nuclear reactors. There are some exceptions where they cannot grow of course, like molten lava.

In the 1970s two researchers named Carl Woese and George Fox suggested a new taxonomic rank of life, based on ribosomal RNA.[3] The taxonomic rank proposed was above Kingdom and was called Domain. The proposal for a three-domain system was widely accepted in the scientific community and so now life is classified as belonging to either Bacteria, Eukarya, or Archaea. The latter includes extremophiles, or microbes that live in extreme environments like hot springs and acid pools. Unique membrane lipids and lack of a particular cell wall also help characterize the domain Archaea.

Domain Bacteria includes an array of species with different phenotypes including rods, spheres, spirals, and even pleomorphic species. They are mostly single-celled organisms and so lack tissues and differentiated cell types. A key feature, and of interest to the field of medicine, is the presence of a unique cell wall. The cell wall of bacteria is different from those of fungi and plants; fungal cell walls have a carbohydrate known as chitin and plants have cell walls made of mostly cellulose. Bacterial cell walls contain yet another type of polymer known as peptidoglycan. The bacterial cell wall functions to maintain cell integrity under

fluctuating environmental conditions, similar to fungal and plant cell walls.

Bacteria can also be split into two groups based on their cell wall. Those with very thick cell walls are called Gram-positive, while those with a thin layer of peptidoglycan are Gram-negative. Of course, many other differences also exist. After dividing bacteria as either Gram-positive or Gram-negative, they can be further divided into different classes based on their evolutionary relationships.

Agrobacterium (Figure 1) is a Gram-negative rod in the class α-proteobacteria. Agrobacterium is one of millions of microbial species inhabiting the soil environment. Other species of microbes include rhizobia, a general term for microbes that take up residency in plant root nodules. Of two important genera are *Rhizobium* and *Frankia*, both of which are able to fix atmospheric nitrogen and assimilate the nitrogen into organic compounds, some of which are delivered to the plant. In return the plant might supply various carbohydrate sources produced from photosynthesis. Other microbes, including mycorrhyzae (a fungi), might establish even more intimate relationships with plant roots. A particular mycorrhyza called endomycorrhyza enters the root hairs via chemical signaling. The plant might gain an enhanced uptake of nutrients and water, which results in a healthier, more robust plant.

Figure 1. Agrobacterium cells (left) are often kept in an -80°C freezer until they are ready to be used. Broken bark exposes interior tissues to Agrobacterium in soil (right). The swollen portion just above the soil is likely due to crown-gall disease.

Similar chemical signaling occurs with plants and Agrobacterium, although the symbiotic relationship is not mutualistic. To understand the role of Agrobacterium in plant biotechnology, one must first consider its role in the soil

environment. When Agrobacterium is living freely in the soil, the major metabolic processes carried out are saprophytic, meaning the cells live on decaying organic matter. However, plants that have a wounded or open epidermis are susceptible to infection, similar to human skin, which quickly shifts Agrobacterium from a saprophytic lifestyle to a pathogenic one.

The base of the plant closely associated with the upper soil layer is called the crown. The crowns of some plants are highly susceptible to splitting due to uneven rates of growth in their different tissues and thus, are prone to Agrobacterium infection (Figure 1). An Agrobacterium infection does not always occur at the crown, but the crown is a common point of infection. Other species of Agrobacterium also exist. For example, *A. rhizogenes* also occurs in soil and is used to infect plant roots grown in vitro to produce hairy root culture.

One of the first attempts of working with cannabis root infection using *A. rhizogenes* was published in 2006 by Imane Wahaby and his colleagues in Spain.[4] They identified secondary metabolites (choline and atropine) in cannabis roots using a unique chemical analysis approach. A more recent research report published in 2013 by Imane Wahaby, Juan M. Caba, and Francisco Ligero working from the Universidad de Granada in Spain, discusses an optimized protocol for the transformation of cannabis roots using *A. rhizogenes*. Their work studied several variables in transformation efficiency such as plant variety, tissue type, and bacterial strain. They delivered various *A. rhizogenes* strains into different cannabis plants in a variety of locations using a syringe. Part of their work included determining the role of generating hairy roots by infecting plants with transgenic DNA that is important for developing hairy roots (a key phenotypic result of *A. rhizogenes* infection). The authors extended their research efforts by also transforming cannabis calli with *A. tumefaciens*. Among their most important findings are that root cultures, due to their genetic stability, biosynthetic rates, and ease of growing, offer an excellent platform for generating secondary metabolites (e.g. cannabinoids) in vitro. As time progresses, the options of using calli or roots for generating cannabis metabolites will certainly be explored further.

Agrobacterium and the T-DNA

The general infection mechanism in nature begins with Agrobacterium sensing small sugars and plant metabolites that are often leached from a plant wound. Agrobacterium responds by expressing several genes. The genes are

harbored on a special plasmid called a Ti plasmid, also known as a tumor-inducing plasmid. The extrachromosomal plasmid genes encode for proteins that help in transferring bacterial DNA to the plant. The transfer DNA (T-DNA) itself codes for genes responsible for inducing the plant to manufacture plant hormones, or phytohormones. Thus, the T-DNA is indeed transferred to the plant cell nucleus. The hormones include auxins and cytokinins and result in a gall (or tumor) around the infection site. The fact that some genes cause a tumor has led researchers to refer to them as oncogenes. Agrobacterium has been thought to take up residency within the gall caused by oncogene products. Other metabolites are produced in the plant including opines, which can serve as a nitrogen, carbon, and energy source for Agrobacterium.[6]

Plant biotechnologists have taken advantage of Agrobacterium, after years of studying how the T-DNA moves from bacterium to host.[7] The Ti plasmid is a rather large plasmid at more than 200,000 bases and plays a central role in facilitating the DNA transfer. Some of the genes on the Ti plasmid encode for a small molecular syringe that helps deliver the T-DNA. What plant molecular biologists figured out was not only that the T-DNA delivered causes the crown gall, but that parts of the T-DNA could be removed and replaced with other DNA that did not cause tumors. The biological origin of the newly inserted DNA did not matter as much as the placement of the T-DNA on the Ti plasmid. That is to say, the T-DNA always needed to be put in the same position on the Ti plasmid in order to be successfully transferred to the plant. Researchers later learned that the T-DNA on the Ti plasmid is flanked by inverted nucleotide base repeats. Agrobacterium virulence proteins responsible for helping to excise the T-DNA recognize the inverted sequence as cut sites. No matter what DNA sequence was inserted in between the inverted sequences, the T-DNA region was always excised and transferred.

However, more was to be discovered that opened up the field of plant biotechnology to a much larger group of researchers. For practical purposes, the Ti plasmid is rather large to be efficiently manipulated in a plant biotechnology lab. Often, molecular biologists prefer to work with smaller plasmids that have fewer restriction enzyme sites and harbor fewer genes. Easier manipulations result. Researchers wondered if the inverted repeats could be placed onto a second, smaller plasmid, and one other than the Ti plasmid. They termed the second plasmid a binary plasmid. The plasmid also came to be known as a binary vector since it was involved in harboring foreign DNA and would also be present in Agrobacterium with the Ti plasmid.

Although the modified Ti plasmid lacked the inverted repeats and the tumor-causing genes responsible for the phytohormone production, the crucial virulence genes for foreign DNA delivery were retained. This is why the Ti plasmid is now also known as a "helper" plasmid. Subsequently, Agrobacterium strains used in labs are now said to be "disarmed" since the modified Ti plasmid lacks tumor-inducing capabilities.

Disarmed Agrobacterium was then able to be transformed with binary vectors carrying new pieces of DNA. Crucial to selecting transgenic plants that received a copy of the T-DNA was the need to also transform the plant with a selectable marker, a gene whose protein product confers a selective advantage in the presence of a selective agent (e.g. herbicide or antibiotic). Thus, in tandem with a gene of interest on the binary vector is a selectable marker.

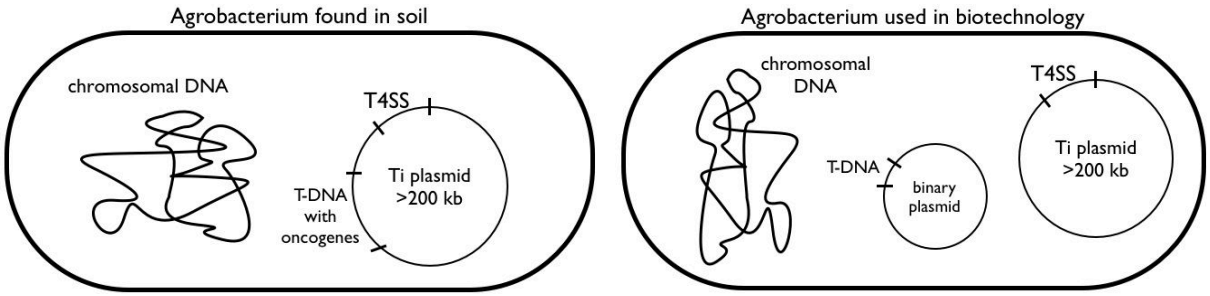


Figure 3. Simplified comparison of Agrobacterium found in soil systems (left) and the laboratory strain (right) that has been modified for plant biotechnology. The lab strain lacks the tumor-causing oncogenes that are found in the wild-type strains.

The resulting Agrobacterium harboring two plasmids (Figure 3) could transform plant cells with foreign DNA. However, one problem was that not all of the transformation events resulted in transfer of the gene of interest. Plants were obviously transformed since they were able to grow in the presence of a selective agent, but upon further investigation the gene of interest was only partially present or not present at all. For some reason the T-DNA was being truncated, and not always at the same position along the sequence of base pairs.

After more research, people realized that the flanking regions of the T-DNA had two different borders, a left border and right border. Further, the DNA transfer occurs with the right border leaving the bacterium first and the left border trailing behind. Thus, if the transfer of DNA from bacteria to plant is interrupted, the trailing segment of DNA moving from bacteria to plant might be interrupted and lost. Since the selectable marker was originally placed near the

right border and the gene of interest was always near the left border, researchers realized the need to switch the position of each.

Research progressed and the gene of interest was repositioned near the right border and the selectable marker was near the left border (Figure 4). This ensured that positive transformants would receive the complete sequence of the transgene first and then the selectable marker would be transferred. Plant cells with the ability to grow on selective media not only had the selectable marker, they contained the complete sequence of the transgene. Thus, the problem of partial DNA transfer and missing segments of transgenes was solved.

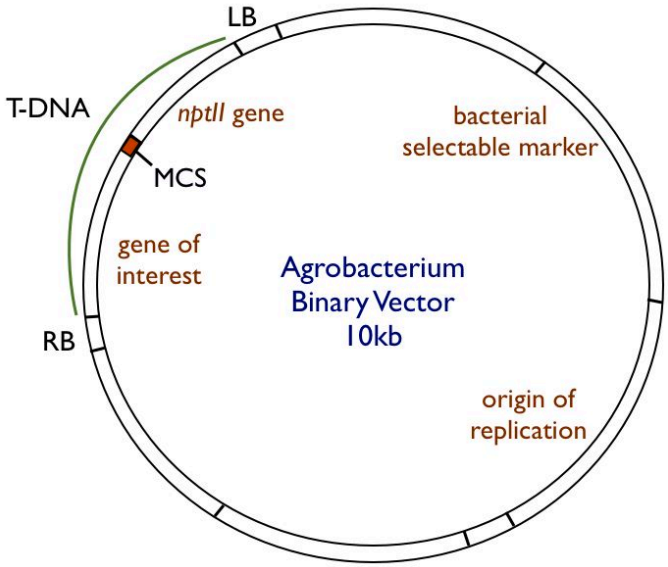


Figure 4. A hypothetical, simplified binary vector showing the T-DNA region flanked by the left border (LB) and right border (RB). The multiple cloning site (MCS) is for the gene of interest. Not shown are the promoters or direction of transcription.

However, as is the case in science, new answers led to new questions. New research into the details about the molecular mechanisms of DNA transfer from Agrobacterium to host plant flourished and much has been elucidated. Stan Gelvin and his colleagues from Purdue University in Indiana, USA has risen as a leader in deciphering how Agrobacterium proteins are responsible for infection.[8] They have helped determine that when a wounded plant releases small sugars or wound-related metabolites, a signaling cascade is induced in Agrobacterium. The cascade begins with an Agrobacterium receptor protein called virA that is activated by molecules released from a wounded plant. VirA then phosphorylates

(adds a phosphate to) a protein called virG located on the cell interior. Interestingly, virG is a transcription factor that is able to turn on several additional virulence genes involved in T-DNA moving from bacterium to plant. Many other virulence genes involved in Agrobacterium gene delivery have also been given alphabetical-related names (e.g. *virB*, *virC*, *virD*, etc.). Thus far, several proteins have been characterized and the function and role of many have been described.[9]

Once the signaling cascade begins and virulence genes are up-regulated, a protein complex is assembled by Agrobacterium that resembles a molecular syringe but is known as a type IV secretion system (T4SS). Some proteins not involved in formation of the T4SS are involved with stabilizing the T-DNA before and after it leaves Agrobacterium. The T-DNA is prone to being diced up by plant nucleases so the protective proteins are crucial to the continued survival of the foreign DNA once inside the plant cell. Eventually, the foreign DNA makes its way to the plant nucleus, where it becomes randomly inserted into the plant genome.

The stochastic (i.e. random) insertion is something that plant biotechnologists are currently working on understanding, since random insertion can lead to variable expression among several transformants. For example, if the transgene is inserted into a relatively quiescent location of the plant genome, the transgene might not be expressed at desired levels. Conversely, production of too much protein from a transgene might not be desirable either. Due to the differences in expression levels of a transgene, several transgenic lines are often generated for a gene of interest.

Regardless, one can now visualize what is occurring during tissue culture. When a callus is formed, the tissue is ready to be infected by Agrobacterium. The process, called cocultivation, gives ample time for Agrobacterium to contact the plant cells and adequately deliver the T-DNA. Some have argued that in order for Agrobacterium infection to occur, acetosyringone dissolved in dimethyl sulphoxide (DMSO) must be added. However, others suggest that acetosyringone, which mimics a plant signaling chemical, does not need be added for infection to occur. That is to say, success of transformation seems to occur with or without addition of acetosyringone. Such discrepancies are not bad since differences in opinion generate further motivation to collect more data, often resulting in a higher confidence to support or refute a position.

As a side note, DMSO is a very dangerous chemical and needs to be used with caution. Although used as solvent for acetosyringone and other chemicals, DMSO allows rapid diffusion of any solute to easily cross cell membranes. In the 1960s some people would infuse DMSO with lysergic acid diethylamide and coat

the door handles of police cars. Hence, this is another good example of the importance of knowing the chemicals and reagents used in lab.

An important detail in Agrobacterium-mediated plant transformation of calli is that the transferred DNA is stably integrated into the host genome. That is to say, the gene will be inherited. Transient expression systems do not incorporate the DNA into the host genome. Instead, they are often concerned with determining protein localization, protein-protein interactions, or producing a desired protein product. These are all areas that are equally important in cannabis biotechnology.

Transient expression might use Agrobacterium, viruses, or biolistic approaches.[10] When using Agrobacterium, the Agrobacterium is applied directly onto leaf tissue and a small vacuum is applied to the leaf in a process called Agroinfiltration. Viruses (e.g. tobacco mosaic virus, potato X-virus) are engineered to express a foreign protein and then applied to leaves as well. One of the largest benefits of transient expression systems is the rapid turn around time to protein production; a transient expression experiment might produce results in only a few weeks. The downside is that the gene of interest is not stably inserted. However, depending on the intentions of the researcher, transient expression systems are very suitable. Thus far, there are many steps in generating a transgenic cannabis line. To help facilitate the understanding of the steps, one might find a flow chart useful (Figure 5).

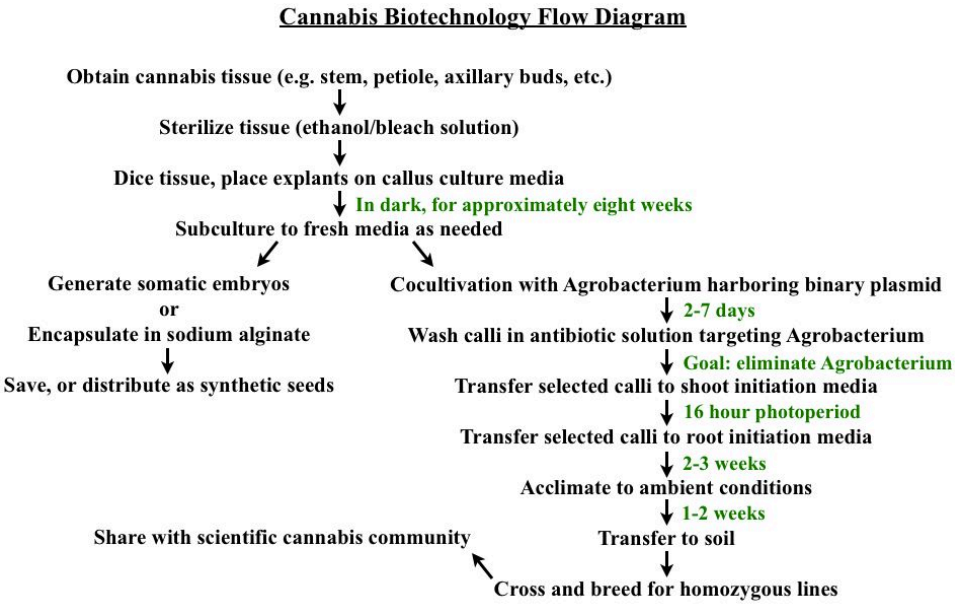


Figure 5. Flow diagram summarizing a few possible research steps when working with cannabis in vitro. While homozygous lines might not be necessary in all cases, downstream scientific assays might require them.

The floral dip method

Of worthwhile mention is a second indirect transfer method of plant genetic modification called the floral dip.[11] In this method, immature seeds (ovules) residing in the ovary of the flower are directly exposed to Agrobacterium carrying a binary vector with a gene of interest. The preparation of Agrobacterium cells involves overnight culture (Figure 6), centrifugation of cells, and resuspending cells to an appropriate density (Figure 7) in a sucrose solution. A small amount of detergent (Tween20) is also added to help the Agrobacterium cells make contact with the ovules. Once Agrobacterium is prepared and the correct density is obtained, the plant is literally dipped into the Agrobacterium suspension.

Figure 6. Overnight culture of two different strains of Agrobacterium, LBA4404 grown in LB broth (left) and EHA105 grown in tryptic soy broth (right). EHA105 does not display the aggregating characteristic like LBA4404.

Figure 7. A spec 20 used to determine cell density. Here the spectrophotometer is set to a wavelength of 420 nanometers (nm).

After dipping, excess Agrobacterium solution is removed by gently shaking the plant. The plant is then placed in a plastic bag or other container to keep humidity high, which is thought to extend the life of Agrobacterium and thus, increase potential ovule infections. After no more than 24 hours, the plants are removed and allowed to recover under normal growing conditions.

The fruits are allowed to ripen and the mature seeds are subsequently harvested, sterilized with a series of ethanol washes, and sown on selective media. If the selectable marker was the *nptII* gene that provides kanamycin resistance, the seeds are plated on MS media with kanamycin. The plates are given a cold treatment (~5°C) in the dark for 48-72 hours, which is called stratification. The subsequent removal from cold treatment and placement under continuous light encourages all seeds to germinate at the same time. While nearly all seeds germinate, only those with the selective marker persist and develop green leaves with long roots (Figure 8).

The main advantage of the floral dip rests with the ability to allow Agrobacterium to deliver foreign DNA directly to the ovules. Such a technique allows one to completely avoid the time consuming and meticulous work that tissue culture requires. Additionally, although the method is called "floral dip" one does not have to completely submerge the flowers. One can also use a spray solution, a syringe, or pipette to apply Agrobacterium cells. Similarly, entire plants do not have to be placed in a plastic bag, only the flowers where Agrobacterium was applied.

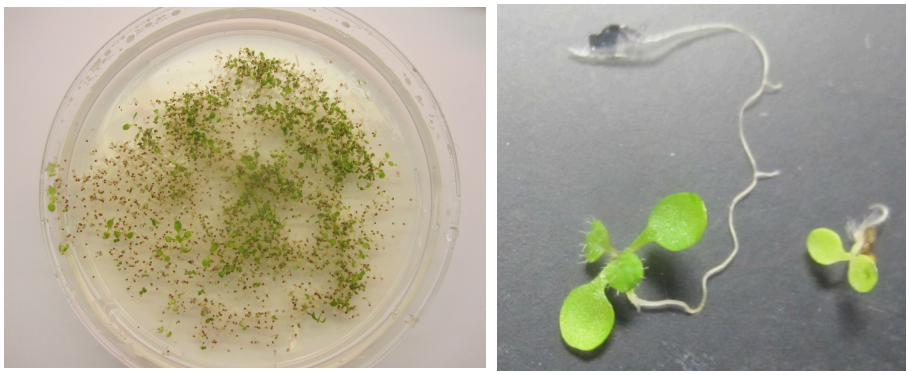


Figure 8. Sterile seeds from the model plant Arabidopsis are plated on selective media (left) and allowed to germinate. Transgenic plants develop true leaves and long roots (right). Semi-solid media is used in germinating seeds, which allows for an easier removal of plant roots from the media.

The model plant *Arabidopsis thaliana* is by far the most commonly used plant in the floral dip, since one of its characteristics is the ability to generate many seeds. While some species have shown stubbornness in taking up foreign DNA via the floral dip, other plant species have been successfully transformed with the floral dip method. At the moment, direct application of Agrobacterium to cannabis flowers as performed in the floral dip has yet to be reported in the scientific literature. We are awaiting such an investigation.

Whether using tissue culture or the floral dip method, a researcher who is successful at generating a transgenic cannabis plant has a few more steps to consider. A newly produced transgenic plant is often not homozygous for a trait. That is, in a diploid individual like cannabis where two copies of each chromosome are present only one copy of the newly inserted gene might be present in the genome (although this needs to be verified). In some cases, both sets of chromosomes of the diploid genome need to carry the gene so that the plant will produce seeds that are all homozygous for the newly introduced trait. For this reason, further generations need to be carried out.

Inheritance of a transgene usually follows Mendelian laws. When considering transgenic plants and their seeds, researchers often use T (for transgenic) instead of F (for filial). For example, a plant originally transformed is the T_0 plant. Seeds from the T_0 plant are T_1 seeds and T_1 seeds develop into T_1 plants. Transgenic cannabis needs to be crossed with itself (if hermaphroditic) or with a transgenic of the opposite sex. Cannabis heterozygosity or homozygosity can also be predicted with a Punnett square but needs to be determined experimentally.

A rather simple method for determining if a plant is homozygous or heterozygous is to plate the seeds on half-strength MS media with a selective agent such as kanamycin. If all seeds germinate and grow, the parent was homozygous, as will be the resulting seeds. The more seeds experimentally verified, the greater the likelihood a person has obtained a homozygous plant. For instance, sewing four seeds on selective media and observing growth in all seedlings will be less assuring than plating several dozen seeds that all germinate and grow normally. Since zygosity will generally resemble the probabilities of a monohybrid (single-gene) cross, seeds need to be grown on selective media each generation in order to ensure one is continually selecting for transgenic plants. The disadvantage is that one must keep a continual collection of seeds.

There are other methods to determine zygosity of transgenic plants besides plating seeds each generation. One method, which is a more technical but faster

approach, requires a variation of the polymerase chain reaction called quantitative PCR.[12] As stated before, many researchers use several independent lines that were generated with the same transgene and investigate them since not all transformation events are equal. Hence, collecting seeds from several transformation events might be important. Proper management of a seed bank when determining homozygosity is of great importance as well but the likelihood is that most serious cannabis growers already maintain excellent records of their seed banks.

Genetic modification by mutation

There is a wide array of techniques to genetically modify cannabis. Two additional methods of genetic manipulation might deserve a brief discussion. The first deals with genetic modification by mutation and the second deals with gene silencing (RNAi). Although not considered "biotechnology" by some, generating mutants has led to the elucidation of many genes, especially in the model plant *Arabidopsis thaliana*.

There are several ways to generate mutant lines of cannabis. In some studies generating mutants is helpful in investigating single-gene function. Although the older methods of generating mutant lines depend on chemicals or radiation to generate random mutation events, interesting phenotypes can be developed and investigated using either approach. An important fact to understand is that mutations must be passed onto progeny via the reproductive cells for them to persist through generations

Mutations in cannabis DNA can be induced by exposing cannabis seeds to ethyl methane sulphonate (EMS) or low levels of radiation. In either approach, seeds are exposed since they harbor copies of the genome that will be propagated as mitosis occurs through development. In other words, the cells of a seed that acquire a new mutation will eventually generate reproductive organs that will likely result in spores for reproduction but that also carry the mutation. EMS and radiation are both mutagens, a broad term for factors that induce mutations in DNA. People who are aware of mutagens, specifically carcinogens, avoid certain foods or environments to minimize DNA mutations in their own somatic cells.

EMS is a chemical mutagen that results in specific types of mutations called transitions. That is, adenine can become guanine or vice versa. Similarly, thymine can become cytosine or vice versa. These mutations are called transitions since the

base becomes another base that has a similar shape. Since the mutations occur at single nucleotide positions they are also known as point mutations. EMS treatment has generated several mutant lines in plant research.[13]

Exposure to radiation also generates mutations in the DNA of seeds but also kills many of the seeds. X-rays are often used, which are at the very far end of the electromagnetic spectrum since they have an incredibly small wavelength. Thus, they are high-frequency, high-energy waves that interact with DNA and other biological macromolecules. Accessibility to an X-ray machine is far less likely than having accessibility to a chemical such as EMS, at least for most cannabis growers.

In either case, exposing seeds to EMS or X-rays, several seeds can be exposed simultaneously and then sown in soil. Selecting mutants relies on looking for phenotypically distinct characters (phenomics). Any mutations that might have occurred might have also resulted in a new protein, resulting in an altered phenotype, or appearance. The new phenotype is selected and propagated. One must appreciate that generating mutant lines using EMS or radiation is random. Several types of mutations can be induced in cannabis DNA and these include the following

point mutations- the change of a single base into another base, point mutations are a broad type mutation

nonsense mutation- change in a base that results in early termination of protein synthesis

missense mutation- a change in base that changes the respective amino acid

silent mutation- a change in one base that does not change the respective amino acid

transition- a base change from adenine to guanine (or vice versa) or cytosine to thymine (or vice versa)

transversion- a base change from a purine (adenine or guanine) to a pyrimidine (thymine or cytosine) or vice versa

deletion- the loss of a nucleotide base, several bases, or pieces of chromosomes

insertion- occurs when a new base (or bases) is inserted within existing bases

frame shift mutation- an insertion or deletion of one base that results in the shifting of the mRNA reading frame, thereby altering codons downstream

In addition to radiation or chemical methods of inducing mutations, there are newer methods that include more specific targeting, known as site-directed mutagenesis or gene targeting. One way to target a specific DNA sequence is to modify existing DNA-binding proteins. One type of such protein is called a zinc finger and is highly specific for the region of DNA that it binds. Zinc fingers are called such because they contain zinc and appear as little fingers that interact with the DNA.

The binding is based on biochemical interactions between the zinc finger amino acids and the DNA topology, which is influenced by the bases present. Genetic modification of the gene for a zinc finger, results in slightly altered amino acids, thereby influencing the zinc finger to interact with a new sequence of DNA. Attached to the zinc finger (or many zinc fingers in tandem) is a restriction enzyme that cuts at a region upstream from where the DNA binding occurs. The two different parts, the zinc fingers and the restriction enzyme, have been artificially joined by linking the genes for each to produce a shared mRNA called a zinc finger endonuclease. The result is a tool generated through biotechnology to further generate products of biotechnology!

RNA interference

An advancement of the last decade or so is the realization that RNA has many more functions than a role in protein synthesis. Researchers working with RNA figured out that genes could be silenced by introducing a complimentary RNA sequence to an mRNA transcript in a technique now known as RNA interference (RNAi) and is really a form of gene silencing. The method relies on genetic transformation of a plant species with an inverted repeat. The inverted

repeat is expressed but hybridizes with itself, leading to an endogenous cell mechanism that dices the double stranded RNA. The small fragments that remain are called small interfering RNAs (siRNA) and hybridize with a target mRNA. The overall result is that the mRNA is destroyed rather than being delivered to the ribosome for translation into a protein. RNAi might play an important role in cannabis biotechnology since if the sequence of a gene is known, the siRNA can usually be obtained.[14]

Interestingly, some plants produce RNA complimentary to RNA material of viruses. Thus, plants evolved to have the RNA silencing method in order to resist viral infection and spread. Generating a genetically modified cannabis plant that expresses siRNA that targets an mRNA encoding a cannabinoid synthesis protein might help engineer new cannabis strains. Similarly, terpene metabolic pathways might also be of interest to silence (chapter 7).

In transgenic studies, an overexpressed gene might result in the production of too much mRNA that might also result in gene silencing. Such instances occur and are known as cosuppression. To understand how transgenes are expressed and how cosuppression might occur, one needs to look deeper into how cannabis genes are expressed. Thus, we are moving to the next area of cannabis biotechnology, transcriptomics.

Chapter 6

Cannabis Transcriptomics and Proteomics

"The idea of the genes' being immortal smelled right, and so on the wall above my desk I taped up a paper sheet saying DNA → RNA → protein."
-James D. Watson, The Double Helix[1]

The central dogma

In the biological sciences there has been an exciting revolution occurring. Specifically, in the field of molecular biology great advancements have been made in the last few decades that are giving humans the ability to alter genes and protein structure with molecular and atomic precision. With more ease than ever before in history, researchers are able to manipulate genes in ways that result in significant modifications of entire organisms. Such ability has huge ramifications for what is possible in the area of plant biotechnology. Perhaps even more exciting is that all of the new technology and discoveries are directly applicable to cannabis.

Information at the molecular level is being produced at an unprecedented rate.[2] Due to the rapid increase in sequence information that can be generated from a project, researchers are sometimes limited to reporting broad overviews of sequence data, unless they are specifically hunting for a gene of interest or sleuthing out a metabolic pathway. Such approaches leave ample room for other researchers to investigate the same set of molecular data. Due to the technological advancements that have led to such large and rapid collections of data, new fields of study have emerged. The high throughput of information has led to the "omics" era, where people are specializing in subdisciplines of molecular biology that were

once filled with only basic knowledge. These subjects include transcriptomics, proteomics, genomics, and metabolomics. Each field is unified by the foundation of introductory-level molecular biology.

Although little "belief" actually exists in the field of science, there are core tenants that have been consistently shown to reoccur. A central theme in the biological sciences is something often referred to as the central dogma. However, the term dogma used here is not referring to belief, as the word is often used to imply. A dogma is simply a set of rules or principles that are set by authorities. A person can get trapped in arguments about religious dogma but scientific dogma is much more objective. The central dogma of biology simply states that DNA is used as a template in order to generate mRNA and the mRNA is then used as a template to generate protein (Figure 1). Whereas the former process is referred to as transcription, the latter is referred to as translation. The central dogma implies that a protein sequence is dependent on the DNA sequence.

Figure 1. The central dogma as usually presented in introductory biology classes.

In higher levels of education, one learns there are several more pieces to the central dogma worth including on a diagrammatic representation (Figure 2). For example, DNA replication occurs with the help of several proteins and can be added to the diagram as a circular loop. Additionally, since DNA also produces tRNA and rRNA, those too can be added. Finally, the protein products of translation function to regulate DNA replication, transcription, and translation and so this can also be illustrated.

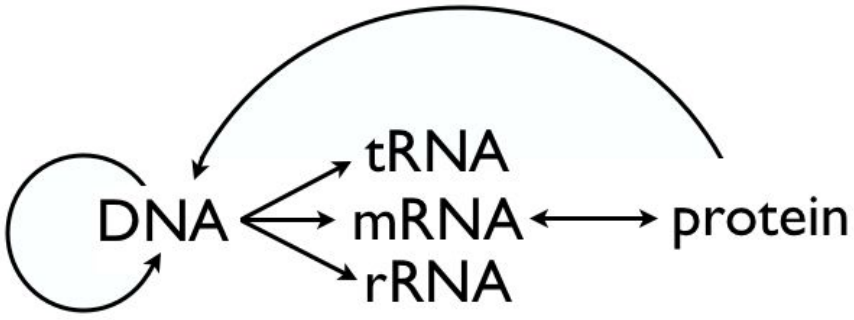


Figure 2. The central dogma as presented to biology students in higher-levels of education. In actuality, all terms can be connected.

The central dogma represents the flow and control of genetic information at the macromolecular level. However, in the cells of different species, there are several variations in their machinery that are involved in the process of making a new protein. Important to cannabis biotechnology are both the understanding of the different enzymes involved in the central dogma and a few basic parts of the gene that code for an mRNA molecule. As with other enzymes, nomenclature for enzymes involved in transcription and translation are given the suffix "–ase." For example, a transferase is an enzyme that transfers one molecule to another, a decarboxylase is an enzyme that removes a carbon, and a hydrolase hydrolyzes. Most of the steps leading from one molecule to another involve an enzyme.

Enzymes need access to DNA but cannabis chromosomes are packed tightly around proteins called histones so that all the genetic material can comfortably squeeze into the nucleus of the cell. A gene becomes available to enzymes that produce mRNA after receiving signals from the nucleus exterior. Some plants might have several thousand genes on a chromosome and also have, in some cases, a dozen or more chromosomes. Further, DNA is double-stranded with four nucleotide bases, which are organized along a side chain of alternating phosphates and deoxyribose sugars. Of the four bases, adenine and thymine always pair together and guanine and cytosine always pair with each other. In RNA, which is often single stranded, thymine is replaced with uracil.

Although DNA is double-stranded (Figure 3), only one side usually codes for mRNA and so enzymes that interact with DNA must be biochemically guided as to which is the coding strand and which is the non-coding strand (sometimes called the template strand).

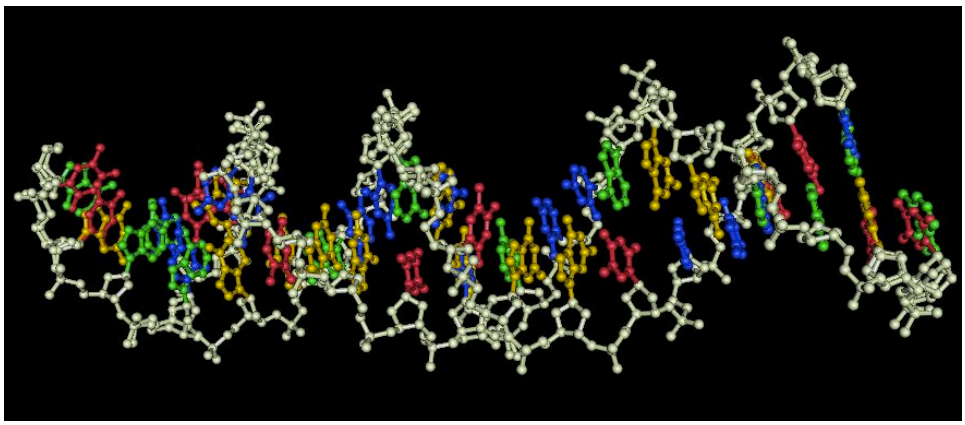


Figure 3. Ball and stick model of deoxyribonucleic acid (DNA) generated by Cn3D (see chapter 9). The sugar-phosphate backbone is white and each of the four bases (cytosine, adenine, thymine, and guanine) are shown in different colors.

The non-coding strand will produce the mRNA. Since the rules of base pairing must be followed, the sequence of mRNA will match the coding strand but have uracil in the place of thymine. Sequences upstream that are just prior to the start site of the gene help guide the transcription machinery to its correct location and orientation.

The direction in which RNA polymerase II (the enzyme that makes mRNA) proceeds can be described by the orientation of the backbone of deoxyribose sugars. In organic chemistry, molecules like carbohydrates might have several carbons associated with them and so chemists and biologists have maintained coherence in the literature by numbering the carbons based on their placement relative to other atoms. For deoxyribose there are two carbons of importance to the molecular biologist, numbered 5′ and 3′ (pronounced "five prime" and "three prime," respectively). RNA polymerase II uses the 3′→5′ strand of DNA to generate an mRNA molecule, which will be read in a 5′→3′ direction. This latter process is important for interaction with the ribosome and producing the correct sequence of amino acids in a protein.

Transcription, making mRNA from DNA, can be broken down into three main stages, initiation, elongation, and termination. Initiation is when small proteins known as transcription factors assist RNA polymerase II in binding to the promoter region to begin mRNA synthesis. Polymerases polymerize, as one might infer from their name, and make longer polymers by gradually including more and more monomers. Different RNA polymerase enzymes catalyze the synthesis of rRNA and tRNA and have been given the names RNA polymerase I and RNA polymerase III, respectively.

The promoter region has a core region, where certain elements or sequences of nucleotides are normally present. Cannabis genes encoding different proteins might share some promoter elements since they share a common biological purpose. Most promoters have a TATA sequence, sometimes called the TATA box. In addition to transcription factors, RNA polymerase II gains access to a promoter region with the help of other proteins (e.g. helicase & topoisomerase), which might work further downstream of the promoter, separating the tightly wound double-stranded DNA.

Elongation, the second stage, occurs as a newly synthesized mRNA strand grows by addition of RNA nucleotides. The enzyme RNA polymerase II slides along the non-coding strand while other enzymes continue to separate the double-stranded DNA downstream of the start site of the gene. Still, other proteins

help stabilize the double-stranded DNA, keeping it separated while RNA polymerase II continues its synthesis of mRNA.

As transcription proceeds, RNA polymerase II will eventually encounter a region on the non-coding strand that destabilizes its association with the DNA molecule. Termination is the term given to the third stage, ending the mRNA synthesis. Termination occurs in slightly different ways in eukaryotes like cannabis compared to bacteria, although several other differences might be pointed out throughout transcription and translation. In eukaryotes, prior to the mRNA leaving the nucleus it is modified on its 5′ and 3′ ends to stay protected from RNA degrading enzymes (RNAses) in the cytosol. The 3′ end has a string of adenine nucleotides added (a poly-A tail) in a process called polyadenylation, which researchers have taken advantage of when separating mRNA from other RNA.

While some mRNA transcripts with a poly-A tail are quickly shuttled out of the nucleus, others might be further acted upon by enzymes prior to leaving the nucleus. This is because some mRNAs have the interesting feature of coding for more than one protein. Some sequence regions within the mRNA transcript are called exons and others that intersperse the exons are known as introns and play an important role in protein diversity (Figure 4). While introns are usually excised out and removed, exons are connected to one another.

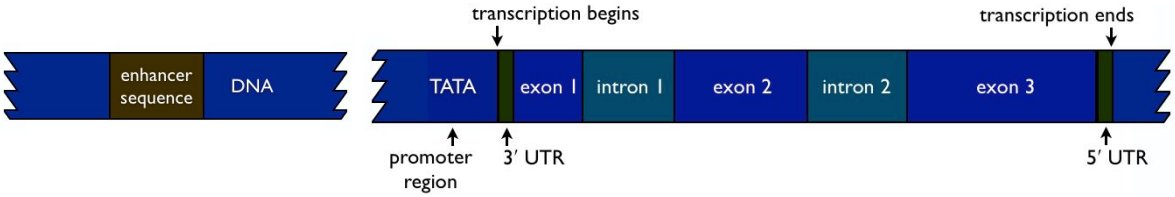


Figure 4. A simplified diagram of a transcriptional region in cannabis showing the promoter upstream of the coding sequence, transcription start site, exons, introns, and untranslated regions (UTR). Exons might differ in length or might be absent in some genes. Several thousand bases can be positioned between the promoter and enhancer. Not shown is the polyadenylation signal sequence.

To generate two different proteins from one cannabis mRNA molecule different exons can be removed and joined together to form mRNA molecules.[3] For example, in an mRNA with five exons, all five exons might be retained or only three exons might be retained while the other two exons are discarded along with the introns. This is just one way nature has found to increase protein diversity while minimizing genome size, thus conserving energy and reducing the amount of nuclear space needed.

There are also other features of a cannabis mRNA transcript worthy of mention. Untranslated regions (UTRs) flank the transcript. Each UTR is referred to as either the 3′ UTR or 5′ UTR. As their name suggests, they are not translated into the protein product but do help in mRNA stability and interactions with the ribosome. In addition, there is a polyadenylation signal sequence within the 3′ UTR that recruits enzymes necessary for adding the string of adenines onto the mRNA. Finally, there are regions on the DNA that in some cases can be thousands of bases away that might help regulate gene expression and these are called enhancers. Proteins interact with enhancers, which impart a positional change on the DNA molecule to allow RNA polymerase II to interact with the promoter. Several other features might be described; however, here we are only concerned with the core features and processes of transcription.

Protein synthesis

As a mature mRNA transcript leaves the nucleus of a cannabis cell, there are two possible destinations. The first is to a free ribosome suspended in the cytoplasm and the second possibility is to a ribosome affiliated with the endoplasmic reticulum, an endomembrane system which functions as a site for post-translational modifications. Cannabis proteins destined to reside within the cell are often manufactured on free ribosomes while proteins to be exported from the cell are synthesized and modified on the rough endoplasmic reticulum.

Translation is the process that cells use to generate protein products from mRNA molecules and, similar to transcription, requires interaction of several different proteins. Some of the associated proteins are needed in small quantities but play a huge role in bringing about changes in cell protein levels. Based on their size, such small proteins might not seem important, but the molecular interactions they help facilitate between an mRNA transcript and their translation machinery is actually very large indeed. These small but important proteins include initiation factors and elongation factors and might be excellent targets of plant biotechnology.

Initiation factors allow for the initial interaction with the mRNA and the small subunit of the ribosome complex. As stated earlier, one might compare the ribosome to a microscopic sewing machine, stitching amino acids together. Upon the interaction of mRNA and the small ribosome subunit, the large subunit is recruited, docks with the small subunit, and results in an mRNA-ribosome

complex. In some cases, without initiation factors, protein synthesis might fail to begin.

While the ends of an mRNA have modifications to enhance its stability in the nucleus, there is another important feature of mRNA central to protein production in cannabis. Each series of three bases is equivalent to one amino acid. The first three nucleotides are known as the start codon, and recruit a transfer RNA (tRNA) carrying methionine, which is almost always the first amino acid of a protein. Codons are highly important in the discussion of protein production, since they represent the amino acids, which in turn comprise a protein. A codon table is often helpful in determining which series of three bases will lead to a particular amino acid (Figure 5).

	U		C		A		G		
U	UUU	Phe (F)	UCU	Ser (S)	UAU	Tyr (Y)	UGU	Cys (C)	U
	UUC		UCC		UAC		UGC		C
	UUA	Leu (L)	UCA		UAA	STOP	UGA	STOP	A
	UUG		UCG		UAG		UGG	Trp (W)	G
C	CUU	Leu (L)	CCU	Pro (P)	CAU	His (H)	CGU	Arg (R)	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	Gln (Q)	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	Ile (I)	ACU	Thr (T)	AAU	Asn (N)	AGU	Ser (S)	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	Lys (K)	AGA	Arg (R)	A
	AUG	START	ACG		AAG		AGG		G
G	GUU	Val (V)	GCU	Ala (A)	GAU	Asp (D)	GGU	Gly (G)	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	Glu (E)	GGA		A
	GUG		GCG		GAG		GGG		G

Figure 5. A codon table. Each three-letter code on the mRNA is associated with an amino acid incorporated into a growing peptide. Shown here are the three letter and single letter abbreviations for each amino acid. Some codons are a start signal (AUG) or stop signals (UAA, UAG, UGA).

Once translation is initiated, elongation factors dock to the mRNA-ribosome complex in order to facilitate the elongation of a growing peptide. A peptide grows by additions of sequential amino acids. However, an amino acid is only incorporated in the presence of the correct mRNA codon. This is largely due to the fact that there is an anticodon on a tRNA that is specific to complementary base pair with the codon on the mRNA. The tRNA molecules

deliver the amino acids, one by one, to an acceptor site on the ribosome. Again, the acceptance is based on the presence of the anticodon and if accepted, the tRNA moves through the ribosome to the next ribosomal position called the peptidyl site, which is where a peptide bond is formed joining two amino acids. From the peptidyl site, the tRNA that has donated its amino acid moves to an exit site and is released from the ribosome.

Each tRNA docks to and enters the ribosome with adjacent tRNA molecules in a somewhat orderly fashion. In other words, there is something akin to a stream of tRNAs carrying amino acids that are entering the ribosome on one end and a stream of tRNAs leaving the opposite end. Subsequently, the peptide grows from the peptidyl site, appearing from another side of the ribosome. The tRNA molecules carrying amino acids eventually cease to enter upon the ribosome when a stop codon is reached, which can be one of three codons (Figure 4) on the mRNA transcript.

In many cases in molecular biology, examples help immensely in facilitating a deeper understanding. A coding strand of DNA might be represented as follows

```
TGTTCTGCTACTATTGTTGCTATTTCTATTAATACTCATGAAGATAATGCT
```

Although this seems rather meaningless and appears as a scrambled string of four different letters, one must remember that the four letters represent biochemical molecules occupying a three-dimensional space. Since they can be counted in units of three, the codons of the mRNA will be translated based on the codon table and each amino acid can be represented with a single letter.

```
CSATIVAISINTHEDNA
```

The result is a meaningful message. Similarly, in another example, a longer polypeptide can be generated with an equally meaningful polypeptide sequence.

```
CCCACCATCAGCAACACCTGGCACGCCACCAACTGCGAGTGGGCCAGCACCCACGGCCACACCAGCATCAA
CTGCGAGACCGAGTGCCACAACCTGGGCTACATCAGCTGCCACGCCAACGGCATCAACGGCGCCCTGACC
```

Can you determine the amino acid sequence from the above nucleotide sequence using the codon table above (Figure 4)? If not, translating DNA using computer programs will be discussed soon. Here is an example of the coding strand of the DNA sequence for chalcone synthase, an important enzyme in cannabis.

```
ATGGTTACCGTGGAGGAATTTCGCAAGGCACAACGGGCCGAGGGCCCGGCCACCATCATGGCGATCGGCAC
GGCTACACCCGCCAATTGTGTCCTTCAGAGTGAGTATCCGGATTACTACTTCAGGATCACCAATAGTGAAC
ACAAAACTGAGCTCAAAGAAAAGTTCAAGCGCATGTGTGACAAATCCATGATTAGAAAACGTTACATGCAT
TTAACTGAGGAGATTCTCAAAGAAAATCCAAATCTTTGTGCTTATGAGGCACCATCATTGGATGCAAGACA
AGATATGGTTGTTGTTGAAGTACCTAAATTGGGTAAAGAGGCTGCTACAAAGGCCATCAAAGAATGGGGCC
AGCCCAAATCAAAGATAACCCATTTGGTATTTTGTACCACTAGTGGTGTGGACATGCCTGGTGCTGATTAC
CAGCTCACTAAGTTATTGGGCCTTAGACCATCAGTCAAACGTTTGATGATGTATCAACAAGGTTGTTTTGC
TGGTGGAACTGTTCTTCGTTTAGCCAAAGATTTGGCTGAAAATAACAAAGGTGCTCGTGTATTGGTTGTTT
GTTCTGAAATTACTGCTGTTACTTTTCGTGGGCCTAATGATACCCATCTCGATAGTCTTGTGGGCCAGGCC
CTATTTGGTGATGGTTCGGCTGCTTTGATTGTTGGGTCTGACCCAATTCCTGAAGTTGAAAAGCCCATTTT
TGAATTGGTCTCTGCGGCCCAAACTATACTTCCTGATAGTGATGGTGCTATTGATGGTCATTTACGTGAAG
TTGGGCTTACTTTTCATTTGCTTAAAGATGTTCCTGGGCTTATTTCTAAGAATATTGAGAAGAGTTTAAAT
GAGGCTTTTAAGCCTTTGGGTATTTCTGATTGGAACTCACTTTTTTGGATTGCTCATCCTGGTGGCCCAGC
TATTTTGGACCAAGTTGAGTCCAAATTGGCTCTTAAGACTGAGAAATTACGTGCCACTAGACATGTGCTTA
GTGAGTATGGTAATATGTCAAGTGCTTGTGTTTTGTTCATTTTGGATGAGATGAGAAGGAAGTGTGTTGAA
GATGGGCTTAACACCACTGGTGAAGGACTTGAATGGGGTGTCTTGTTTGGATTTGGGCCTGGCCTCACTGT
TGAAACTGTGGTCCTTCACAGTGTGGCTATTTAG
```

The amino acid sequence translated from the above coding sequence results in the following protein (chalcone synthase)

```
MVTVEEFRKAQRAEGPATIMAIGTATPANCVLQSEYPDYYFRITNSEHKTELKEKFKRMCDKSMIRKRYMH
LTEEILKENPNLCAYEAPSLDARQDMVVVEVPKLGKEAATKAIKEWGQPKSKITHLVFCTTSGVDMPGADY
QLTKLLGLRPSVKRLMMYQQGCFAGGTVLRLAKDLAENNKGARVLVVCSEITAVTFRGPNDTHLDSLVGQA
LFGDGSAALIVGSDPIPEVEKPIFELVSAAQTILPDSDGAIDGHLREVGLTFHLLKDVPGLISKNIEKSLN
EAFKPLGISDWNSLFWIAHPGGPAILDQVESKLALKTEKLRATRHVLSEYGNMSSACVLFILDEMRRKCVE
DGLNTTGEGLEWGVLFGFGPGLTVETVVLHSVAI
```

In the example of chalcone synthase above, one might deduce the nucleotide sequence also represents an mRNA transcript. This is apparent from the methionine (M) present at the start of the protein since nearly all mRNAs have the AUG start codon. The ATG is on the DNA coding strand in the chalcone synthase nucleotide sequence. As a side note, the methionine codon can also be found internal to an mRNA transcript but the codon only becomes accessible as the mRNA secondary structure is altered as it progress through the ribosome. Thus, not all methionine codons are equal, only the first is a start codon. Additionally, one should note the presence of the stop codon at the end of the coding sequence, which would be UAG in the mRNA transcript.

In some cases, the string of amino acids will automatically fold on itself and in other cases helper proteins might assist in the folding. The folding is based on physical and biochemical properties of each amino acid side chain. In yet other cases, some proteins require post-translational modifications. Such modifications are crucial if the protein is to function correctly. Modifications might include

additions of acetyl groups, formation of sulfhydryl bonds, addition of phosphates via phosphorylation, or addition of sugar groups in a process called glycosylation. A plant biotechnologist might want to consider the type of modifications a particular protein must undergo in order to properly express that protein in a different host system.

Transcriptomics in biotechnology

Biotechnology projects often begin with first knowing what transcripts are found in a cell at a given moment. A major goal of research projects is to collect, analyze, and group (i.e. cluster) transcripts into categories based on what molecular function or biological process their protein products might perform. Additionally, transcripts can be categorized based on the cellular component to which the protein products belong.

Due to the dramatic lowering in cost of DNA sequencing, research has generated copious amounts of data. For example, the cost to sequence one base in the early 1990s was about 1USD. For the same price, a person can now obtain the sequence for 1,000,000 bases. Thus, with some sequencing technologies, the entire transcriptome (all the mRNA) of a cell is analyzed in a single study at a very affordable price. Although some transcripts can be more abundant than others, this is simply a reflection of either the research methods or the actual events occurring in the cell. In either case, resulting sequence data is analyzed with the help of computers (see chapter 9). Prior to discussing how sequence data is analyzed, a person should further understand some of the methods used to collect transcripts for analysis.

Large collections of transcripts are known as libraries. The libraries are often called a complimentary DNA (cDNA) library if the whole transcripts are represented. This differs from an expressed sequence tag (EST) library, which is when only part of the transcript sequence has been obtained. Researchers might be limited in using an EST library if wanting to perform investigations such as whole gene cloning. However, generating full-length cDNAs might not be necessary for identification of a transcript.

Transcriptomics has developed into a field that has made possible the investigation of not just an organism's transcriptome, but also the transcriptomes of individual organs, tissues, and cell types. Prior to discussing research projects involving the cannabis transcriptome, let's discuss library construction.

Library construction

There are different methods for constructing a cDNA or EST library. Each varies in the amount of time required, starting material characteristics (total RNA vs. mRNA) and quantity needed, as well as the vector choice. These factors affect the time and effort needed for library construction in addition to downstream analysis and applications.

First one must gather the plant material of interest. For example, an investigator might not be interested in the transcripts of an entire cannabis plant, rather, they might solely be interested in the transcripts of a flower. Regardless of the tissue or organ of interest, the plant material must be immediately snap frozen in liquid nitrogen. Since mRNA can be detected within a few minutes of environmental stimuli, rapid, or snap freezing provides a relatively good representation of the current transcriptional state. The frozen plant material is then ground while in liquid nitrogen using a mortar and pestle.

Once the liquid nitrogen has evaporated, the resulting powder is suspended in a mixture of phenol:chloroform:isoamyl alcohol. There is also a small amount of guanidinium isothiocyanate, which acts as a chaotropic agent, inhibiting enzymes that might degrade RNA. As an important note, the mixture is pH specific in order to encourage separation of RNA from DNA into different layers (Figure 6).

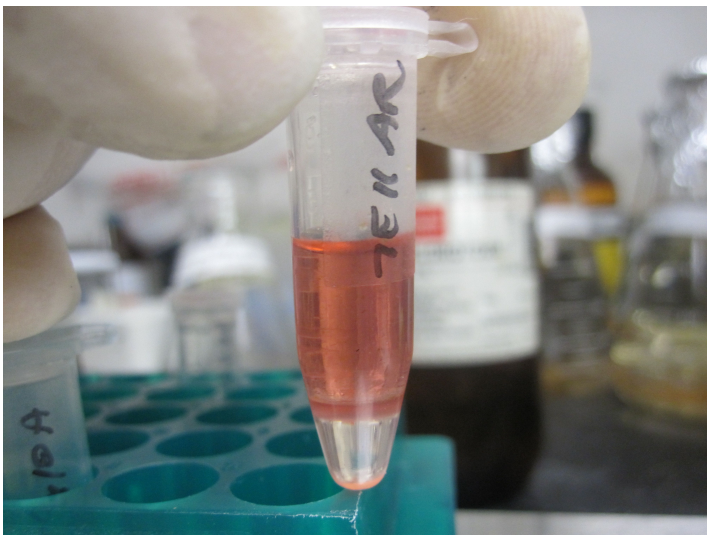


Figure 6. RNA extraction showing layers of phenol solution above (aqueous) and clear (organic) layer below. Oxidized phenol appears red in color.

After a series of cold centrifugation steps and transfer of the aqueous layer, the total RNA is precipitated out with ethanol. The pellet, which consists of total RNA, is resuspended in DPEC-treated (diethyl pyrocarbonate) water and is ready for mRNA isolation. The mRNA only consists of about 1-5% of the total RNA. Getting this small amount of mRNA separated from the rest of the RNA (tRNA, rRNA, etc.) required a bit of ingenuity from original researchers.

They found that the mRNA could be separated by mixing the total RNA in a microcentrifuge tube with tiny paramagnetic beads coated with short chains of thymine repeats (oligo-dTs). Since the mRNA has a series of adenines as a polyA tail, the mRNA hydrogen bonds with the thymines coated on the paramagnetic beads. With added warmth, secondary structures of the RNA relax and allow additional hydrogen bonding to occur. The microcentrifuge tube is placed on a small magnetic stand, which captures the magnetic beads and thus, captures the mRNA. The binding solution is also a high-salt solution that encourages hydrogen bonding of the polyA tail and oligo-dT beads.

A series of washes slowly decreases the salt content and keeps the T-A hydrogen bonds intact. A final washing solution (that is preheated) removes the mRNA, which is placed in an RNase-free tube along with a strong solution of ammonium acetate and 1 μl of glycogen. The ammonium acetate is used to precipitate out the mRNA upon centrifugation, which takes place after an overnight freeze at -20˚C or 1 hour freeze at -80˚C. After isolating the mRNA pellet and rehydrating, the quantity obtained should (hopefully!) give a total 2-4 μg. Importantly, for this method the concentration must be high for the next steps.

The mRNA is then combined with oligo-dT primers, deoxynucleotide triphosphates (dNTPs), a first strand buffer (Tris, Mg2+) and MMLV (Maloney Murine Leukemia Virus). MMLV is a reverse transcriptase that catalyzes the addition of dNTPs to the 3´ end of the oligo-dTs and was originally obtained from a mouse (family muridae) leukemia virus. The reaction with MMLV is allowed to proceed in the presence of DTT, (dithiothreitol) which is based off of a threose (three carbon sugar) that functions as a reducing agent to ensure protein sulfhydryl groups are protonated and not interlinking, thus avoiding disulfide bonds. After 1 hour at 37˚C the reaction is stopped and MMLV is denatured by heating the microcentrifuge tube.

At this point the first strand has been synthesized. To make a double-stranded DNA molecule, a second strand must be synthesized as well. Second strand synthesis uses the reverse transcribed mRNA (the cDNA) from first strand

synthesis as a template. Second strand synthesis relies on the activities of part of DNA polymerase I, which is also called the Klenow fragment. The Klenow fragment has 5′⟶3′ polymerase activity but lacks exonuclease activity. In addition to polymerase I, dNTPs, DTT, and a buffer for polymerase I to work properly are added. Finally, an enzyme called RNase H is added, which cleaves the RNA portion of cDNA-mRNA hybrids. This removes the remaining mRNA and allows complete (or theoretically complete) synthesis of a double-stranded cDNA molecule. After allowing the reaction to proceed, a phenol:chloroform extraction is used to separate the proteins from the nucleic acid and the aqueous layer is moved into a fresh tube. The proteins stay behind in the organic layer.

To generate more meaningful sequence data, longer cDNAs are often desired. Therefore, longer cDNA molecules need to be selected. This can be done using electroelution. This is the process of running the cDNA on an agarose gel and cutting out a region (500bp-5kb). The gel is melted but contains cDNA. Another method such as column chromatography can also be used and also separates molecules based on size. Once a researcher has obtained the larger cDNA molecules, they can be precipitated using 5M ammonium acetate. The cDNA is ready to be inserted into a vector and transformed into *E. coli* cells.

Subsequently, the *E. coli* cells are plated onto media with a selective component that hinders growth of nontransformed cells. Individual colonies result after overnight growth, and these colonies are picked for sequencing reactions that result in sequence data. There are many more steps required in analyzing a cDNA library such as trimming reads, overlapping sequences to form contigs, and scaffolding to orient sequences to their positions on chromosomes. However, these steps require bioinformatics programs and so will be considered as we proceed. For the moment, one can realize the various steps required to obtain transcriptomics data can become somewhat laborious and attention to detail is an absolute necessity.

Alternate methods of generating transcriptomic data

Transcriptomics began to expand as a field when researchers learned they could select coding sequences from non-coding sequences When researchers first learned it was possible to only study transcripts and not an entire genome, more and more research was dedicated to isolating mRNA and obtaining the sequence

data. As time progressed, the methods used to acquire mRNA and make cDNA improved.

One advancement that was made was the use of PCR in generating cDNA molecules. The method, abbreviated as the SMART™ method,[4] uses a "*s*witching *m*echanism *a*t 5´ end of *R*NA *t*ranscript." The method has had some revision for improvement and now there is a "SMARTer" method. The SMART method was invented by researchers working at Clonetech in Palo Alto, California. The latest technology uses a thermocylcer to amplify full-length cDNA transcripts from total RNA. Primers that anneal to the polyA tail are added along with reverse transcriptase. In most reactions, the 5´ end of genes would be underrepresented but the technology incorporated in Clonetech's SMARTer approach, is able to enrich for 5´ ends.

The technology has two major advantages. The first is that only a few nanograms of total RNA are required. This is important since obtaining microgram amounts of RNA can sometimes be challenging. Since PCR is incorporated, researchers will worry less about working with RNA and tedious transfer of microliter amounts and focus more on downstream work such as data analysis. A second advantage is that full-length transcripts are generated. More known bases of a transcript help identify and characterize the function of the resulting protein.

A more recent concept called RNA-seq has been designed to only sequence partial RNA molecules but still give the researcher enough sequence length to determine the role of the mRNA after comparison with full-length sequences in databases. Whereas full-length cDNAs generated in traditional library construction might be more than a thousand bases, RNA-seq reads are only a few dozen to a few hundred bases long. Cloning and plating *E. coli* is not required; however, isolated mRNA molecules are still used. The mRNA is fragmented, a special primer is annealed to the fragments, and cDNA is generated. Small sequences that flank the cDNA called adaptors are then annealed. The adaptors function in downstream sequencing reactions. The result is that more than 100,000 (in some cases millions) of transcripts can be sequenced, increasing the throughput to bring RNA-seq to the front of race in transcriptome technology.

Binary libraries

Although RNA-seq generates copious amounts of sequence data, there are some disadvantages that might otherwise exist. For example, methods that require cloning cDNA into vectors results in *E. coli* colonies that harbor the cDNA insert. After selecting for transformed *E. coli* cells by selecting colonies on media with an antibiotic, colonies are grown in small indentations, called wells, on a rectangular plate (Chapter 7, Figure 3). The plate is inserted into a sequencing machine and a series of reactions signals to the computer what base is present. Although the rectangular sequencing plate can be discarded, often a sequence technician will freeze the plate in case the sequence data is unsatisfactory. Thus, the sequencing plate can act as a back-up. This is lacking in the RNA-seq approach. Upon request, researchers who generate libraries can obtain sequencing plates and use them to choose a cloned gene of interest. Thus, an advantage of generating a library using *E. coli* is that if the sequencing plate is obtained, a person can further manipulate the clones directly from the plate.

Working late in the lab one night in graduate school I had an idea. *What if cDNA could be directly cloned into a plant expression vector?* If possible, a plant cDNA cloned into a binary vector might result in several thousand genes that could then easily be used in several thousand plant transformation studies. That night after going home, more thought was dedicated to the idea. Upon asking the primary investigator of the lab the next day, he warned that direct cloning into large vectors are typically discouraged if there are downstream applications such as sequencing, since large vectors can hinder the sequencing reaction. However, after much protest, he finally agreed. The method was a success. After obtaining the sequence data, the sequencing plates that harbored my frozen *E. coli* colonies in wells were also obtained. The sequence data was mined for genes of interest. Those genes were correlated with a coordinate position on the sequencing plate.

Since each *E. coli* in each well carried a different vector-insert construct, I was able to choose a gene of interest based on the sequence data and locate its position on the sequencing plate. Following up on the idea, the *E. coli* was grown overnight and a plasmid miniprep was performed. The binary vector was then transformed into Agrobacterium and those cells were in turn used to transform Arabidopsis plants. For this project more than a dozen transgenic lines each harboring a unique gene of interest were generated. Such a large number of transgenic lines were only able to be produced since the gene isolation and cloning steps could be omitted. The result was high-throughput plant transformation.

Since library construction used a binary vector, I call this the "binary library" approach to plant transformation.

There are several applications for a binary library. Suppose a person wanted to transform a plant such as cannabis. A person would first need to clone a gene into a plant expression vector (sometimes a lengthy process in itself), before even attempting to infect plant tissue. However, if a binary library were available thousands of genes of interest could be immediately ready to use. Binary libraries could not only be used to generate thousands of varieties of transgenic cannabis, but might also provide experience in plant biotechnology for undergraduates in college and possibly high-schools. Certainly, pre-cloned genes in a plant expression vector would simplify the process of plant biotechnology for many people.

However, there might be one additional step that might streamline a business attempting to generate several lines of transgenic plants using the binary library approach. After the cDNA is cloned into the expression vector, the mixture of vector-insert constructs are typically delivered to *E. coli* using electroporation or heat shock. The *E. coli* harboring the constructs are then plated and individual colonies are used in sequencing reactions. Perhaps an alternate approach, one that has not been apparently reported in the literature, is instead directly transforming Agrobacterium with the vector-insert constructs, then performing the sequencing step. A person would still be able to select vector-insert constructs from a position on a sequencing plate but the resulting colony would be an agrobacterium colony not *E. coli*. Thus, the requirement of plasmid mini-prep and transformation of Agrobacterium would be avoided.

Proteomics

Transcriptomics is directly related to proteomics, the field that studies proteins of an organism. This includes knowing the proteome, all of the proteins that might be present in an organism. Similar to transcriptomics, there are proteomic studies that seek to determine only the proteomes of specific organs, tissues, or cells. Also, similar to transcriptomics, the field of proteomics is rapidly growing with advancements in technology.

Knowing how proteins might function in a cannabis cell is largely based off of three-dimensional shape. The three-dimensional shape is largely based off the sequence of amino acids, dictated by the order of codons on an mRNA transcript. Amino acids have both an amino group and a carboxylic acid group. These are

joined to a central carbon, the alpha-carbon, which is also bonded with a hydrogen. The fourth bond in the alpha-carbon is with a side group, which varies depending on the amino acid. As previously mentioned, the side group of each of the approximately 20 amino acids imparts different physicochemical properties on a protein. For example, glycine has a hydrogen for its side group while phenylalanine has a large aromatic (circular ring) side group. The characteristics of side groups can be small and nonpolar, hydrophobic, polar, or carry positive or negative charges. The difference results in different interactions between each side group, thus influencing the overall shape of the protein.

As a newly-formed protein is synthesized, the molecular forces between amino acid side groups are already helping dictate the folding pattern. However, in some cases, nascent proteins are helped into the functional and correctly folded positions with the assistance of other proteins. In either case, a protein can best be described using a series of levels, a type of hierarchy.

The primary structure of a protein is simply the sequence of amino acids. Often the primary sequence is represented by a series of one letter abbreviations (Table 4). The secondary structure includes basic conformations that include alpha-helices, which resemble cylinders and beta-sheets, which resemble sheets of paper. In addition to the several variations on these two themes, several other secondary structures exist but by and large alpha-helices and beta-sheets are the most common. Tertiary structures include the interactions of secondary structures with one another. The last hierarchical level, or quaternary structure, describes the interaction among two or more tertiary structures.

Some cannabis proteins, like enzymes, function only with the help of cofactors, atoms that help facilitate transfer of electrons to rearrange bonds in other molecules. Coenzymes are technically larger than cofactors, and while they still assist with redistributing electrons in bonds, they are molecules themselves. Examples include vitamins but technically coenzymes are any nonprotein molecule that is required for proper enzyme function. In addition to electrons, coenzymes might also function to carry smaller molecules such as amines ($-NH_2$) or methyl ($-CH_3$) groups in transfer reactions.

There are several tools available to study the cannabis proteome. One method, which is often introduced in cellular biology undergraduate courses, is two-dimensional gel electrophoresis. If a researcher were trying to understand the protein differences between two plants that produced either a high or low level of cannabinoids, they might consider generating a 2-dimensional (2-D) gel. To do this, protein extraction from two plants (experimental and control) is usually

performed using an acetone solution and then the precipitate is weighed. The equally weighted samples are loaded into a column where the proteins are separated by a charge. The separated proteins in the column are separated again on a flat gel. However, this second dimension is where the proteins are separated based on size (molecular weight) by applying a voltage, similar to the process of separating PCR products. A protein-binding dye (often Coomassie blue but this can vary) is applied to identify the position of each protein in the gel. The result is a 2-D gel that allows one to visualize proteins from the original sample separated by their charge and size.

Tri Raharjo and his colleagues took a proteomics approach to study all the proteins in leaves, flowers, and glands of cannabis using 2-D gel electrophoresis.[5] One of their objectives was to determine differences in proteins among tissues regarding cannabinoid synthesis, and also attempt to characterize an enzyme involved in the cannabinoid pathway (olivetolic acid synthase, which we will see has been somewhat elusive). In some cases, greater protein quantity was produced in the flowers compared to the leaves although higher protein diversity was seen in the leaves. Some of the spots on the gel were further characterized but were not thought to directly play a role in the cannabinoid pathway.

In many labs 2-D gels are not just compared with the researcher's unaided eye but are digitally contrasted with the assistance of computer technology. Differences in protein spots might show an absence of, or differences in, levels of particular proteins. In some cases, a one dimensional (1-D) gel can be generated where proteins are separated much like DNA fragments are separated in gel electrophoresis. This provides a very quick overview when comparing two tissues or cells. Researchers can also go a step further by excising protein spots from 2-D or 1-D gels and analyzing the amino acid sequence to determine the protein primary structure. Comparing the sequence to a database might then result in identifying the protein.

If a researcher had a protein they were interested in isolating, the DNA-coding sequence could be cloned into a bacterial (or yeast) expression vector. Following the coding sequence is sometimes a sequence coding for a string of histidine amino acids. The expressed protein will then have a series of histidine amino acids attached. After a certain period of growth the bacterial (or yeast) cells are centrifuged and lysed, freeing the protein product. In some cases, several rounds of growing and centrifugation are repeated, which might increase the amount of protein product. The lysate is put through a small column that is prepared with a filter specifically designed to bond with the repeating sequences of

histidine residues attached to the protein. As the sample runs through the column, the protein of interest is retained and all other proteins travel through and out of the column. The column is then washed with a special solution that causes release of the protein of interest from the column.

Proteins of interest can be characterized by determining their sequence or shape. For determining sequence, many labs will obtain a purified protein sample and then subjugate the sample to various digestive enzymes. The enzymes will yield different sized proteins and the amino acid sequence can be determined. Although possible, some researchers do not rely on sequencing the amino acids of a protein to determine the sequence of amino acids. Instead, the cDNA or genomic DNA (gDNA) is obtained and sequenced. The protein sequence is inferred using relatively straightforward bioinformatics programs that translate the DNA sequence in silico.

To determine the atomic detail of protein structure, techniques such as X-ray crystallography or nuclear magnetic resonance (NMR) might be used. In either case, a protein of interest must first be isolated, purified and crystallized. In X-ray crystallography, the small protein crystal is positioned between an X-ray machine and detector. The X-rays are short enough wavelengths that they interact with the atoms in the protein crystal. The wavelengths that are scattered onto a detector are recorded and positional information of electrons is relayed to a computer. The coordinates of the electrons, or an electron density map, is given either in a spreadsheet or it might be interpreted visually with one of several protein viewing software programs. Thus, X-ray crystallography ultimately provides the ability to view structures of proteins in three dimensional space.

There are tens of thousands of proteins available through the Protein Data Bank,[6] a large database for data files containing information from experimentally-derived protein structures. The files are free to download and view using protein visualization software. Although more will be discussed about the software, for the moment the important point is that proteomic studies find value in studying atomic interactions and viewing proteins can be of large importance.

Proteins of interest in cannabis biotechnology

Cannabis proteins carry out crucial functions in cellular metabolic processes, regulation of gene expression, and cell communication. While many readers of cannabis literature have traditionally been interested in enzymes that

carry out metabolic processes involving the production of cannabinoids, interest in other secondary metabolites like terpenes is increasing. This is good news since cannabis has several additional proteins involved in secondary metabolism worth studying.

Growing interest might be partially attributed to data generated by X-ray crystallography or NMR that is manipulated by imaging software. Although many proteins have been characterized in Arabidopsis, the evolutionary likelihood is that in highly conserved processes, cannabis has very similar proteins to Arabidopsis and so one can expect the protein structures to be very similar.

Some of the best characterized proteins of plants are those that are involved in photosynthesis. Light has been shown to bring about shifts in enzymatic activity of proteins involved in photosynthesis, often due to reduction-oxidation (often abbreviated as redox) reactions. Redox reactions are involved in transfer of electrons. Molecules undergoing reduction are those that accept electrons while oxidation involves the loss of electrons. Two amino acids, cysteine and methionine, are commonly involved in redox reactions since they have a sulfur group that can be oxidized and reduced. Upon oxidation of and subsequent bonding between two sulfur groups, large changes in protein conformation might occur.

In an interesting study by Gayathri Gopalan and colleagues,[7] X-ray crystallography was used to determine the positions of cysteine residues in a protein in the thylakoid lumen called FKBP13, which is a type of binding protein. In the dark, proteins involved in photosynthesis have disulfide bonds (the oxidized state) but these bonds are reduced in the presence of light, bringing about conformational changes and thus, activating the enzymes. The posttranslational modification allows for the continual presence of the proteins that can be rendered inactive or active, depending on the environmental stimuli (e.g. visible light). The work done by Gopalan and colleagues suggest that the active form of at least some proteins involved in photosynthesis such as, FKBP13, are functional when reduced.

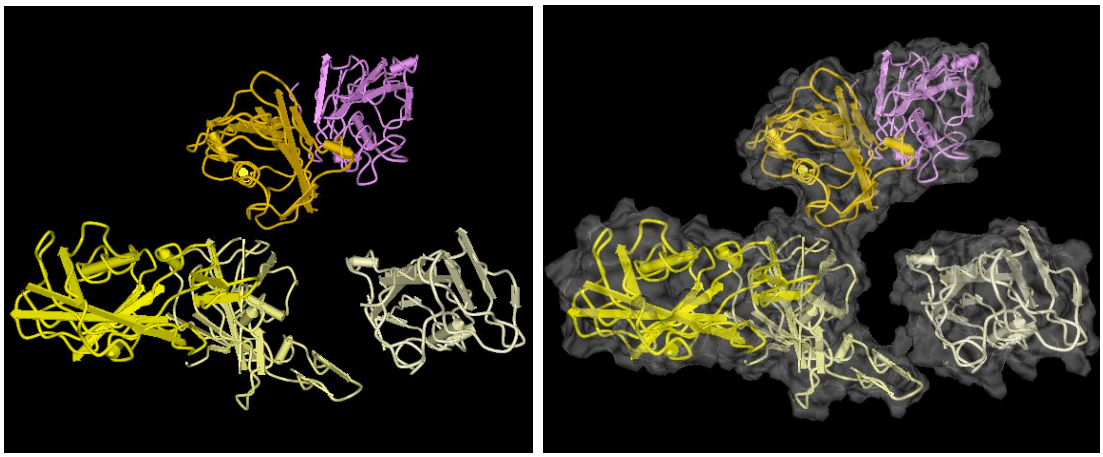


Figure 7. Structure of FKBP13 generated by UGENE,[8] a free, cross-platform and multi-purpose software package. The beta-sheets and alpha-helices are emphasized with flat arrows and with tubes, respectively. The image on the right shows the solvent-excluded surface, which helps to visualize the molecular surface.

In addition to interesting plant proteins involved in photosynthesis, plant proteins are also important in cell communication. As discussed previously, plant hormones are the signaling molecules plants use to elicit changes in gene expression. In some cases, hormones might bind to protein receptors on the cell exterior and a signaling cascade is generated. The signal is eventually transduced to the nucleus, which results in an alteration of gene expression.

One interesting group of hormones is called the brassinosteroids. Although first discovered in the mustard family (Brassicaceae), they are found throughout the plant kingdom, primarily in young tissues (i.e. shoot tips). Their main function is in regulating development and growth, especially in response to light. The molecular underpinnings of the initial signal transduction was previously hypothesized; however, crystallography data by Julia Santiago and colleagues[9] at the Max Planck society in Germany has helped to characterize the interaction on a molecular level.

Specifically, they determined the role of an extracellular domain, a leucine rich repeat, in initially interacting with brassinolide. Upon interacting, the protein's conformational shape changes and the signaling mechanism is initiated. The mechanism begins with the brassinolide receptor known as BRI1, which upon interaction with brassinolide, results in a conformational change that in turn influences BRI1 to interact with a co-receptor known as SERK1. The interaction between receptor and co-receptor results in phosphorylation of one another, which sets off a signaling cascade to the cell interior. Such a complex series of reactions is common in the molecular world of biology and determining how proteins participate in a signaling cascade is better understood through direct visualization.

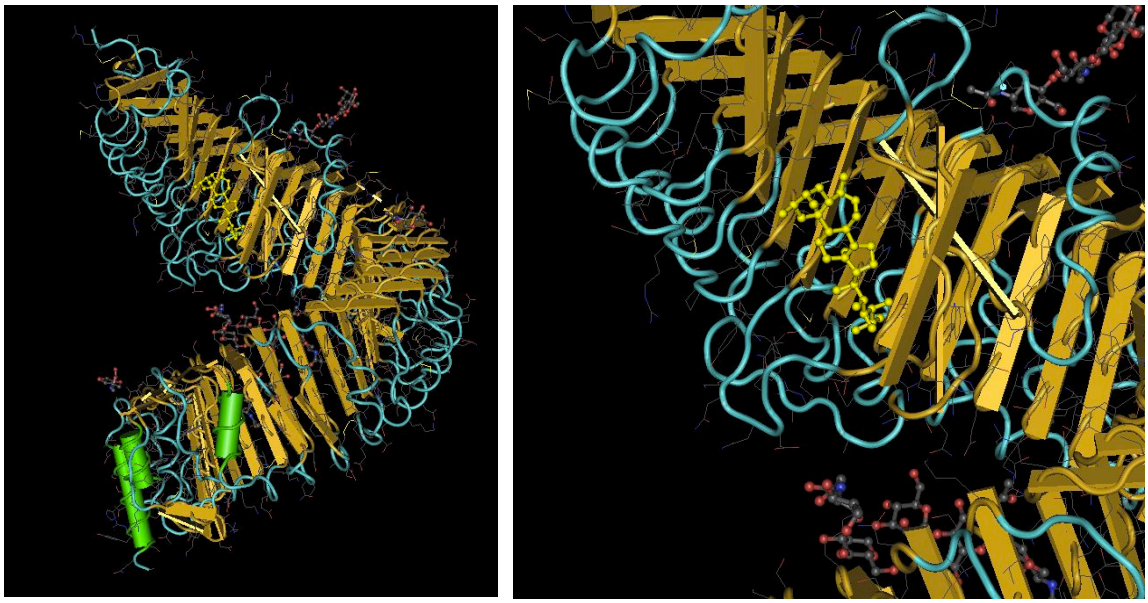


Figure 8. The ectodomain with the leucine rich repeat is shown here in complex with brassinolide, which is highlighted in yellow. Using Cn3D,[10] an imaging software freely available from NCBI, distant (left) and close-up (right) views are shown.

Regulation of gene expression in plants is also becoming better understood through X-ray crystallography. Some domains of cannabis proteins associate with DNA and assist in transcription and thus, participate in gene expression. Some of the important proteins that interact with DNA include transcription factors, often small proteins that can result in drastic changes in gene expression. Specific DNA-binding domains of transcription factors are the main interest, and the DNA molecule is often included in the resulting crystallography data. That is to say, the DNA is crystallized with the protein.

In a study by Cécile Hamès and colleagues[11] the LFY (LEAFY) transcription factor responsible for switching on several other genes in flower development was characterized. The authors suggest that knowing the structure of LFY (Figure 9) and comparing it to other proteins involved in gene regulation might help increase our understanding about how flowers evolved. Certainly anyone who looks at such a beautiful interaction between protein and DNA would surely recognize the intimate molecular interaction between proteins and DNA. Individual atomic interactions can also be viewed, helping bring what once was limited to the imagination and hypotheses, into an amazing visual reality.

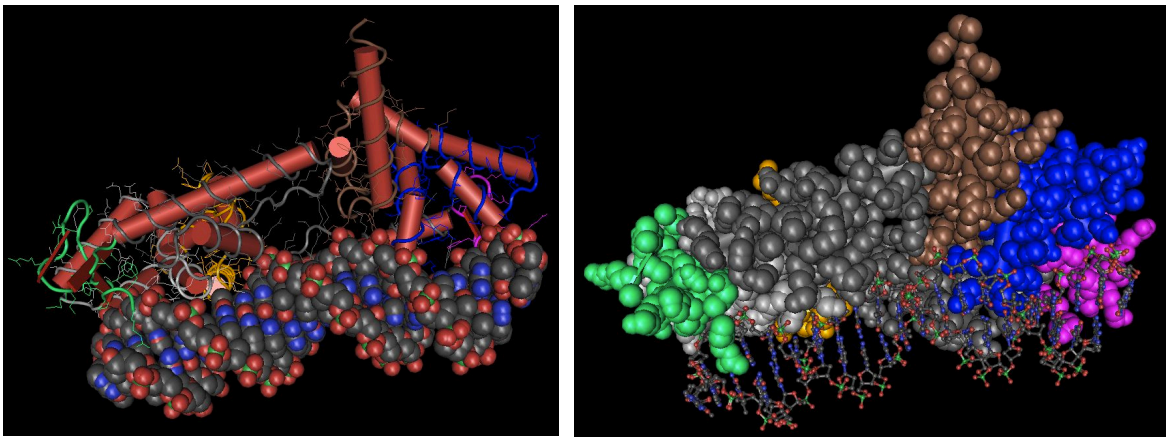


Figure 9. DNA in complex with the LFY transcription factor generated with Cn3D.[10] The DNA helix is represented as a "space-filling" model (left) and a "ball and stick" (right). Different views can help better visualize the molecular interactions.

Almost any protein can be characterized at the atomic level if the crystal can be obtained. Many bacterial proteins are of interest to investigators since bacteria can contribute to important processes of ecosystems. A small percentage of microbes also cause disease and unfortunately, this is where some bacteria have acquired their negative reputation. To advance knowledge on the molecular mechanism of disease-causing bacteria, the structure for several virulence proteins have been determined.

The bacterial type four secretion system, as you might recall, is important in transferring DNA from Agrobacterium to the plant cell. To better characterize the type four secretion system, Vidya Chandran and colleagues[12] generated the crystallographic structure of the outer membrane complex. The giant macromolecular complex reveals a dense assortment of proteins interacting with one another (Figure 10). Looking at the structure one can better understand how foreign DNA is able to be physically transferred from the Agrobacterium cell to the plant cell, since the pore that the DNA travels through is clearly seen.

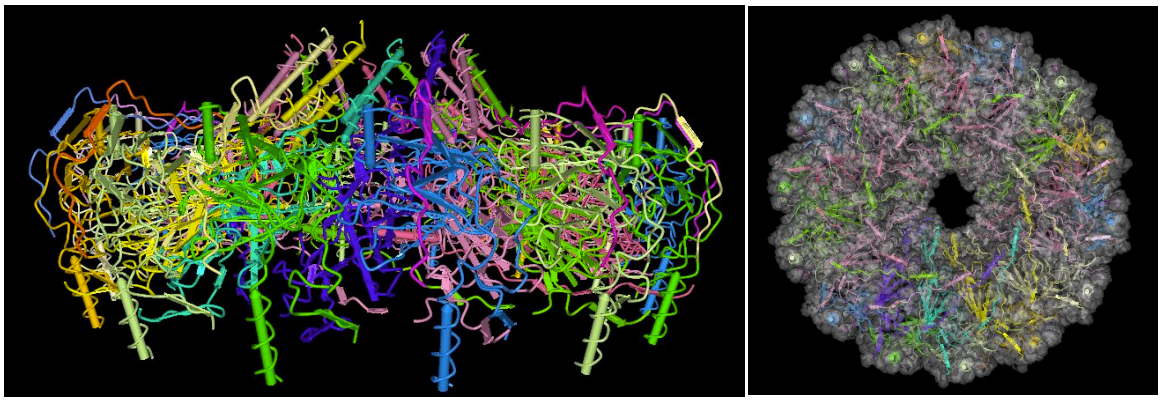


Figure 10. The macromolecular complex of the bacterial type four secretion system (PDB 3JQ0), which is the route for Agrobacterium T-DNA transfer. Represented is the side view (left) and view of the outer membrane pore (right) generated using UGENE.

Limited structural information has been generated for cannabis proteins. However, several human proteins have been characterized and some of the work relates to cannabinoid metabolism. Since cannabinoids bind to particular receptors, determining the structure of the receptors or the proteins involved in the metabolism of cannabinoids might give insight into how cannabinoids interact with cellular proteins.

In some cases, synthetic analogs of cannabis metabolites are used in studies. One such study incorporated ajulemic acid. The importance of ajulemic acid in cannabinoid metabolism is that ajulemic acid is a synthetic analog of THC-11-oic acid, a product of tetrahydrocannabinol metabolism that elicits anti-inflammatory and other positive medicinal effects. Thus, determining how ajulemic acid interacts with protein receptors might lend insight into some of the therapeutic effects of THC-11-oic acid or its analog.

Andre Ambrosio and colleagues[13] generated data for a protein receptor known as the peroxisome proliferator-activated receptor γ (PPARγ). Their work focused on the domain that binds with small molecules such as ajulemic acid. They determined the structure of PPARγ is changed upon binding with ajulemic acid, which leads to a signaling cascade to allow transcription of genes involved with the immune system. Although previous work suggested cannabinoid analogs bind cannabinoid receptors (CB1 and CB2), the data generated provided structural evidence that another receptor can be a target. Ambrosio and colleagues also proposed that the structural data might be a good starting point for drug design aimed at treating maladies such as type 2 diabetes.

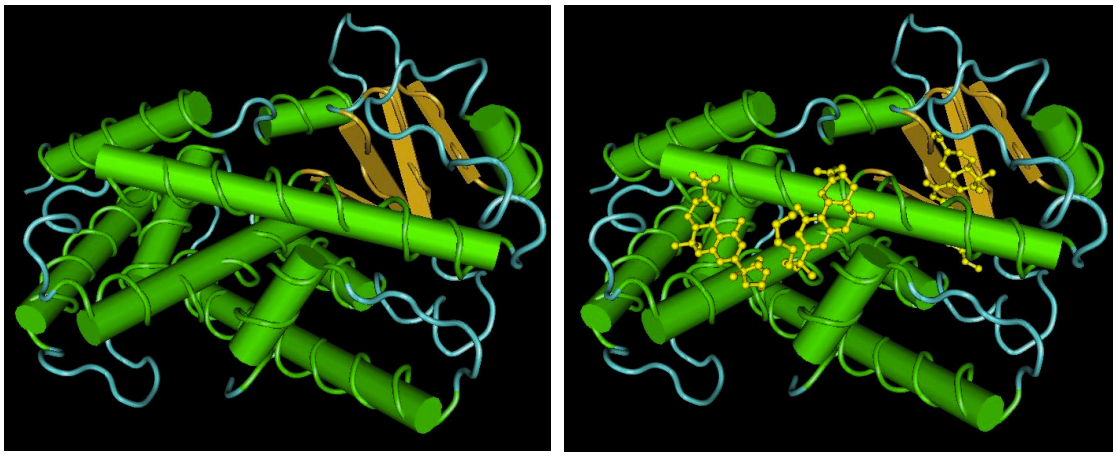


Figure 11. Protein structure of a region of PPARγ without (left) and with ajulemic acid (right) generated by Cn3D. The image on the right has three molecules of ajulemic acid, which are highlighted in yellow.

After looking at only a few different protein structures, one appreciates the diversity of proteins. Additionally, proteins that participate in a variety of cell activities including metabolism, signaling, and gene regulation are shaped based on their function. At other times, determining protein structure using crystallography adds to known genetic data, further strengthening hypothesized functions.

Proteomics is often of direct value to plant biotechnology. For example, a researcher interested in overexpression of a transcription factor in a transgenic plant might first want to study the protein architecture and understand its molecular interaction with DNA. Mutations within a DNA-binding protein might also lead to binding new regions of DNA. Thus, zinc finger endonucleases are often designed with the help of visualization software. Obviously proteomics is highly intertwined with the field of bioinformatics. Indeed, almost all areas of molecular biology are now deeply dependent on computers.

In protein studies, one begins to appreciate the relationship between DNA and protein. While transcriptomics lends insight into gene expression patterns, protein studies help determine the transcripts that are being translated. Although maintained by proteins, the genome is the major player in storing the genetic material that will become the transcriptome and eventually protein. Hence, there is a need to understand the regulation of sequences within the genome and genome organization.

Chapter 7

Cannabis Genomics

"Understanding the cannabis genome in its totality - the composition, structure, regulation, and evolution - is like knowing the secret to the universe."
-Source Unknown[1]

Introduction to Genomics

Genomics is the study of an organism's entire genome. In recent years, a genome has taken on a more complicated definition but most can agree that there are two major components. The first is the complete set of genes that guide an organism through its life and the second is the associated mechanisms to help regulate the genes.

There are several fascinating areas within genomics including functional genomics and structural genomics. Functional genomics investigates the role of the gene products (e.g. proteins). Structural genomics seeks to understand the genome architecture, which is the physical arrangement of chromosomes. While functional and structural genomics are considered here in relation to cannabis, there are several other interesting areas of genomics that include personal genomics, comparative genomics, and metagenomics.

Thus far, the fact that genes encode products (protein, RNA, etc.) that help determine cell processes has been made clear. Importantly, cannabis genetic material can be described in a hierarchical fashion with the genome being made up of chromosomes, which are in turn made up of chromatin. Chromatin is simply the genetic material (nucleotide bases) in association with histones. Related to this is the fact that there can be several modifications to the nucleotide bases besides just the sequence of the bases.

Modification to the DNA that does not include changing the sequence of bases is known as epigenetics[2] and has important consequences. As an example, some areas of the cannabis genome have bases that might be coated with methyl groups, a single carbon with three hydrogens (-CH_3), which inhibits transcription from initiating. Thus, methylation of DNA bases is a form of gene regulation in that methylated genes are not expressed.

Another example of epigenetic regulation of genes is seen in histones, the proteins that eukaryotic DNA is sometimes tightly wrapped around. Histone proteins have a positively charged lysine residue that attracts the negatively charged backbone of DNA. Acetylase enzymes are able to add a negatively charged acetyl group to the lysine residue, thereby freeing the DNA from its histone association. The subsequent disassociation results in accessibility of DNA by RNA polymerase to begin transcription of mRNA. Similar to methylases, histone acetylases (and de-acetylases) can be up- or down-regulated in a cannabis cell, influencing accessibility of DNA by transcriptional machinery.

The DNA sequence, of course, is also important and even a single base change can have drastic effects on the protein product. In the case of enzymes, a difference in one base might result in a change in enzyme activity. When comparing gene sequences, single bases that are different between two species are referred to as single nucleotide polymorphisms (SNPs or "snips"). SNPs are used extensively in genomic studies, such as helping to understand variation in drug metabolism among humans.[3]

Other mutations might result in the complete halting of translation (nonsense mutations) or result in deletion of one or two bases, which causes shifting of the reading frame of the mRNA molecule. Whether mutations result in meaningful alterations to protein products has been extensively discussed in evolutionary theory. In general, most mutations are not harmful to the organism while a few mutations might occasionally be beneficial.

Finally, genomics has not only seen rapid advancements in the way we view gene expression patterns and understand mutations, but also in the structure of the chromosomes themselves. We now know that genes are not distributed equally along chromosomes. In eukaryotes, chromosome tips usually have repeating elements (telomeres) and areas of the middle (centromeres) that are important in mitosis and meiosis. Other areas between the telomere and centromeres can have non-coding DNA, which has become an important player in understanding genome function and evolution. Interestingly, less than 2% of the human genome codes for proteins. Perhaps what is more interesting is that large portions of DNA

in animals and plants are non-coding. That is not to say non-coding DNA is unimportant, we have just been slow in understanding its role.

Gene dosage and polyploidy

Some of the most interesting non-coding regions of DNA are those that are flanked by inverted repeats. Sometimes, these pieces of DNA can excise themselves out of one location of the genome and insert themselves into another location. Since they change place such sequences have become known as transposable elements or "transposons." The popular literature refers to transposons as "jumping genes" since they appear to literally jump from place to place around the genome.

Transposons are incredibly important to genome architecture and gene expression. For example, if two transposons flank a gene and both transposons are induced to change place (in some cases due to stress signals), a gene can be reinserted into another part of the genome or the gene might be lost. Much larger pieces of DNA can be rearranged due to the effects of transposons. In other cases transposons will propagate themselves by first duplicating, excising themselves out of a chromosome, and then inserting themselves into a new location, leaving behind a copy of the original transposon. In yet other instances, transposons might lead to gene duplication, where the fate of the duplicated gene is determined over time. However, whole-genome duplications or species divergence can also result in genes changing function.

Regardless of the mechanism, gene duplication has provided several excellent examples of molecular evolution. A duplicated gene might evolve to take on entirely new functions in a process called neofunctionalization. In some instances the genes might become quiescent and become what are known as pseudogenes, although not all pseudogenes originally result from gene duplication. In other cases, a duplicated gene might evolve to catalyze a new reaction in a metabolic pathway through a process called subfunctionalization. Several examples of gene duplication have been suggested to play important roles in the evolution of secondary metabolite production.[4] Interestingly, the genes involved in secondary metabolite production have also been shown to be clustered closely together in genomes.[5]

A possible result of gene duplication might be an increase in gene dosage, or the amount of protein product that a cell produces. If the particular protein is

responsible for a step in a metabolic pathway, more metabolic product might result. Organism-wide effects of gene dosage are easily seen in many polyploid crops, whose entire genomes have been duplicated.[6] Polyploid plants often have larger flowers, bigger fruits, or show an increase in overall size.

Many organisms are diploid and contain two sets of chromosomes. For example, cannabis has 20 total chromosomes in its somatic (non-reproductive) cells.[7] Sexual reproduction in cannabis requires one set of 10 chromosomes derived from the male flower and the other set of 10 chromosomes derived from the female flower. Upon pollination, the cannabis zygote will have 20 total chromosomes with 2 copies of each chromosome. The shorthand notation in writing is 2n = 20. The 2 represents the ploidy level (the number of copies of each chromosome), the n is the haploid number (the number of chromosomes received from each parent), and the 20 is the total number of chromosomes. Compared to fungi, animals, and protists, the plant kingdom appears to have the most members that are polyploid (multiple copies of chromosomes are present).

Shorthand for a polyploid individual will include a base number, which is the haploid number of chromosomes for the ancestor of a group of plants under study. For example, bread wheat has 42 total chromosomes and is hexaploid. Therefore the shorthand is 2n = 6x = 42. Here the base number (x) is 7, which is the haploid number of the ancestral lineage of wheat. In other words, bread wheat has seven different chromosomes and six copies of each chromosome. However, during meiosis, the genetic material is (under normal circumstances) split evenly into spores, each containing 21 total chromosomes. The spores will eventually fuse with one another to regenerate a plant with 42 total chromosomes. A tetraploid cannabis plant might be presented as 2n = 4x = 40. In this case the haploid state (n) would have 20 chromosomes and the tetraploid (4) condition would be derived from an ancestor with a haploid number of 10 (x).

Although hexaploid or tetraploid conditions are presented here, polyploid individuals might range in having three or more full sets of chromosomes. Several crop plants now consumed or used by societies are polyploids. Potatoes (*Solanum tuberosum*) are tetraploid and many apples (*Malus sylvestris*) are triploid. One could easily provide an endless stream of example polyploid crops that are important to humans. Perhaps a more important piece of the genomics puzzle is that polyploidy events have naturally happened in the majority of flowering plant lineages.[8]

To understand how polyploidy might happen one must know some of the basics of meiosis, the process that results in the generation of the reproductive cells. Plants do not produce gametes like animals, instead they produce spores. Spores

and gametes are unique in that they contain half of the genetic material of the parent, or half the genome. Somatic (non-reproductive) cells such as those in the leaves, stem, or root contain a complete genome. Different types of somatic cells, regardless of where they are found in the plant, contain the same genetic information. The difference in cell types is a result of different genes being expressed. In plant reproductive organs, spores are generated that have a reduced chromosome number (Figure 1). Upon fertilization in cannabis, the genetic material of the spores joins (10 + 10) to reconstitute two sets of chromosomes (20).

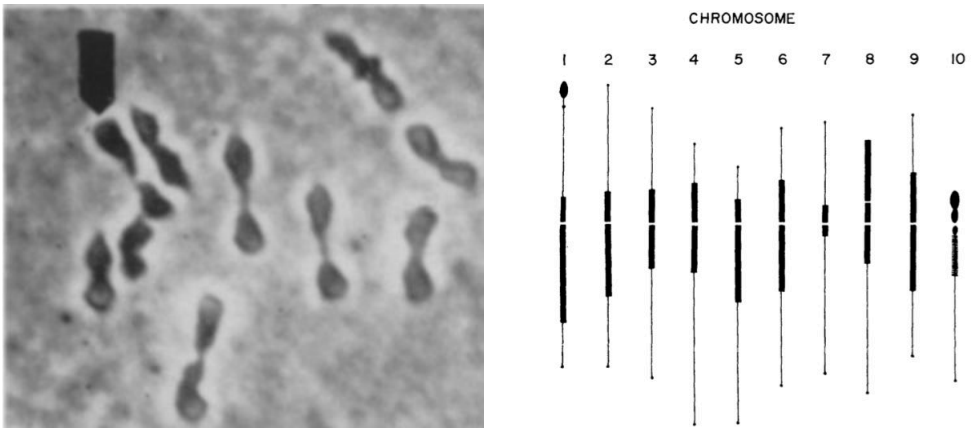


Figure 1. Photograph taken by P. L. Redfearn, Jr. (left) of chromosomes from pollen cells of a hemp variety. Drawing of cannabis chromosomes (right) from a haploid cell produced by G. W. Reinert. Both images were published in a study by Margaret Y. Menzel in 1963.[7]

Specialized plant cells generate spores by going through meiosis, which is similar to mitosis of somatic cells except that meiosis results in four genetically distinct daughter cells instead of two genetically identical daughter cells. Before meiosis, the genetic material is duplicated and a linear chromosome will appear as "X" shaped, commonly depicted in textbooks. The identical strands of the X-shaped chromosome are called sister chromatids and remain attached until the second major stage of meiosis. To determine the number of chromosomes in a cell at any point, one simply needs to count the number of centromeres, the center point of the "X."

Meiosis is initiated when chromosomes carrying the same genes (homologous chromosomes) match up, sometimes with portions of their arms crossing over and resulting in a complete exchange of those regions. The homologous chromosomes align near the center of the cell and their centers (centromeres) are connected to protein polymers called spindle fibers that function

to pull the homologous chromosomes away from one another and to opposite poles of the cell. The spindle fibers connect to an attachment site on the chromosome centromere and form a complex called the kinetochore. As the fibers depolymerize at the poles, they pull the chromosomes away from one another. After the homologous chromosomes are equally divided and have reached the opposite poles of the cell, the cell pinches into two and the result is two cells with half the genetic material of the original cell.

However, another event immediately occurs in the two new daughter cells. The chromosomes align in the center of each cell again, the spindle fibers attach at the centromere, and the sister chromatids are pulled apart to the opposite poles. The two cells then split again and the result is four cells, each with half the number of chromosomes as the original starting cell. No matter what the starting number of chromosomes, the goal (from the cell's perspective, if the cell had a perspective) of meiosis is to reduce the number of chromosomes by half.

The spindle fibers that pull the chromosomes apart in meiosis are polymers of a smaller protein. Polymerization that leads to spindle fibers is sometimes disrupted, which results in unequal distribution of chromosomes. That is to say, if formation of spindle fibers is disrupted, some spores might not contain any chromosomes and others might retain the original full sets of chromosomes. If a spore that has not undergone reduction in chromosome number lands on a stigma and fertilization occurs, the result might be a seed containing an extra set of chromosomes, a triploid individual. In other situations, an unreduced spore might fertilize an unreduced egg, which results in tetraploidy. Tetraploids will generate spores that are haploid, relative to themselves, but that are diploid relative to the ancestor. Subsequent fertilization events might lead to persistence of the polyploid or lead to new ploidy levels. Important to cannabis genomics is the fact that the chromosomes that are pulled apart by the spindle fibers can be disrupted with chemicals such as oryzalin or colchicine (Figure 2).

Figure 2. Colchicine is a chemical easily obtained but is highly mutagenic, sometimes leading to errors in meiosis and mitosis.

Colchicine was being used to induce polyploidy even before the structure of DNA was published; however, the exact mode of action of colchicine was not determined until years after its wide use in biology.[9] Inducing polyploidy by applying colchicine can be done in one of two ways. Direct application of colchicine to leaves or apical regions that form flowers might result in production of spores that contain an unreduced number of chromosomes. Spores can then be collected and dusted onto stigmatic surfaces.

Alternatively, seeds can be soaked in a colchicine solution and then germinated, resulting in many of the somatic cells having unreduced chromosome sets. Since the somatic cells carry out mitosis and form more cells, polyploid cells will simply generate more polyploid cells. Entire organisms or organs (i.e. reproductive organs) of the organism will also be polyploid and thus, all spores generated will lead to polyploid progeny. Many cannabis polyploids have been reportedly generated with colchicine that have resulted in increased flower size.

In 2013 Tian Bin and colleagues[10] generated cannabis polyploids using colchicine. Some seeds were immersed in a 0.20% colchicine solution for 24 hours while another trial used a 0.15% solution for 36 hours. After soaking the seeds they were sown and allowed to germinate. The germinated plants showed differences in their tissues regarding their ploidy levels. Such plants that have tissues that vary in their genetic content are called chimeras. Cells that were polyploid had an increased size compared to normal, diploid cells. The research team of Tian Bin,

who is from China, had similar goals of others who attempted to generate polyploid lines with enhanced metabolic profiles.

One must be careful when handling colchicine, since contact with skin can result in mutated epidermal cells. Even a low percentage working solution that is sometimes used in generating polyploids, might cause mitotic problems in those improperly handling the solution. Still, colchicine is routinely but carefully used in the floral industry in generating bigger blooms, which are often more attractive to consumers.

Many animals have trouble surviving when they are polyploid due to confusion from unbalanced gene dosage. However, plants can cope with extra sets of chromosomes by inactivating either single chromosomes or a set of chromosomes by hypermethylation of bases or hypoacetylation of histones. Due to the common place of the phenomenon, polyploidy has been studied extensively in plants and two main types of polyploidy have been described. Autopolyploid is when the added set of chromosomes results from fertilization of a species with itself (potatoes). Allopolyploids are generated through hybridization of two different plant species that are genetically different (apples). Although more could easily be said on polyploidy in plants, for the moment we must shift our attention to technologies that have been used to study the genome.

Genomic technologies

Biologists have been characterizing genomes for decades by first generating a karyotype, a picture of an organism's complete set of chromosomes organized as homologous pairs of chromosomes. In organisms that sexually reproduce, the sex chromosomes follow the autosomes (chromosomes not involved in sexual determination) in the arrangement. Cells are kept in a mitotic state and are photographed with the help of a microscope. To determine homologous chromosomes, nucleotide base-specific dyes are used that bind either AT- or GC-rich areas, for example. The banding patterns are similar for chromosomes that carry related alleles.

The technique of using dyes to characterize regions of chromosomes has been referred to as chromosome painting but is more appropriately known as fluorescence in situ hybridization (FISH). The technique is based on the fact that areas of non-homologous chromosomes are known to consist of different base composition. The regions with similarities can be compared, depending on the

fluorescent dyes used. In this way, chromosomes can be compared by visual inspection through a microscope, sometimes a fluorescent microscope if using FISH. As in other uses of fluorescent technology, a sample is bombarded with short wavelength light. Interaction with the different fluorophores will result in emission of lower energy, longer wavelength light in the visible spectrum and might appear as different colors. To generate a karyotype, a picture is taken of the stained chromosomes and later arranged together in pairs (using software to cut and paste), based partly on size, usually from largest to smallest.

While genome structure has been revealed by chromosome staining (and now whole-genome sequencing), newer tools have allowed investigations into functional genomics in order to determine the functions of different genes. One of the most powerful tools that researchers began using in the mid-1990s (and has been used in cannabis studies) is the microarray. While useful in quantifying and qualifying transcripts, many consider microarrays part of the genomics "toolbox" since one can study expression patterns of genomes. The technique allows for whole genome analysis in the sense that all transcripts of a cell type, specific tissue, or organism can be studied simultaneously.

Microarrays require glass slides similar to those used for microscope slides by students in a biology lab. However the special glass slides used in microarrays have arrays of spots that are actually clusters of single-stranded DNA. Each spot corresponds to a gene that an organism might express. In other words, the sequences on the microarray slide are complimentary to transcripts that might be expressed in a specific organism. Using a series of steps, samples are hybridized to the slide. Often, comparisons between a non-treatment (control) and treatment group are studied using microarrays.

While a control would be an organism growing under typical conditions, the treatment sample might be subjected to heat stress or exposed to a particular chemical, for example. The treatment might be anything the researcher applies, but will hopefully result in an altered expression of different (or some similar) transcripts. Total mRNA collected from the non-treatment and treatment samples are reverse transcribed to make single-stranded cDNA. Incorporated into the cDNA are nucleotides that are labeled with cyanine dyes, which have a fluorescent characteristic. The treatment group incorporates a green fluorescing dye (e.g. Cy3) and a red fluorescing dye (e.g. Cy5) is used for the control sample.

Both of the samples of the single-stranded cDNA are then hybridized onto the glass slide harboring the arrays of complimentary single-stranded DNA. After hybridization and a series of wash steps, the microarrays are scanned using a

special microarray scanner that aims a beam of light at the hybridized array and reports the fluorescence, which is recorded by a computer and placed into a data spreadsheet. Green fluorescence suggests a gene was expressed after treatment while red fluorescence suggests the gene was not expressed. A yellow spot indicates equal expression between both treatment and control (Figure 3).

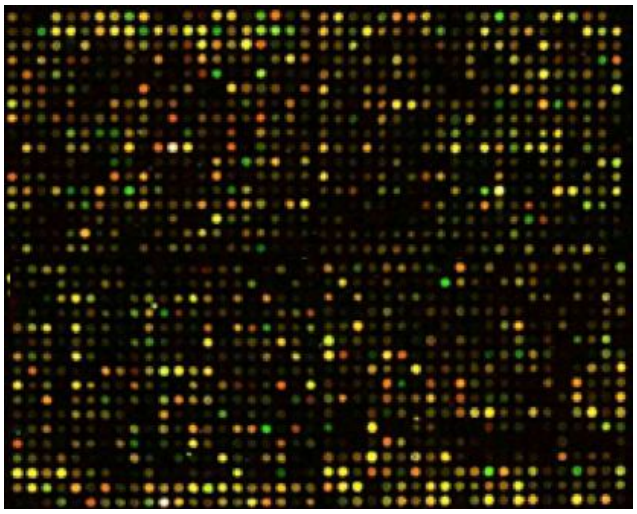


Figure 3. Partial microarray after being scanned with colors representing genes either up-regulated (green), down-regulated (red), or no change in expression (yellow).

Performing microarray experiments is labor intensive and certain labs often specialize in performing microarrays. Like some other techniques of molecular biology, researchers might opt to send RNA samples for another lab to analyze. Statistical design is important as is quality control measures. For example, in a procedure called dye-swap, the mRNA samples are converted to cDNA using different dyes; red-fluorescing dyes are used for treatment samples and green-fluorescing dyes are used for control samples. Expression profiles should closely match the original microarray experiment, except that the colors should be inversely related. Several other important variables should be controlled and tested for as well.[11]

Only a few people have used microarray technology to understand cannabis; however, the emphasis has been on hemp fiber and cell wall development.[12,13] Others have used microarrays to help understand cannabis-related phenomena in humans and other animals. To try to improve knowledge on how delta-9-THC and cannabidiol influence the inflammatory pathway, researchers in Israel collaborated with North American researchers.[14] They used microarrays to

compare human cells that were exposed or not to delta-9-THC and cannabidiol. Their results confirm previous studies that suggest cannabinoids modulate the inflammatory and immune response in humans. Importantly, they were able to correlate several genes involved in responding to delta-9-THC and cannabidiol.

Microarrays are not available for all organisms, since the genetic profiles of the organism under study must be known in advance. Having the sequence data is needed since the spots on the glass slide are physically printed by a machine that requires computer input of sequence data. If sequence data is lacking, the glass microarray slide cannot be prepared. To study the genome of organisms that lack genomic data, researchers have resorted to a technique known as heterologous hybridization. In this process, a microarray for an organism that is closely related to the organism under study might be used. For example, an Arabidopsis microarray slide might be used to hybridize cannabis cDNA. However, heterologous hybridization has been argued to lead to erroneous results. Heterologous hybridization is also becoming less of a necessity since genomic and expression sequence data is becoming more widely available due to the decrease in cost of generating sequence data.

DNA sequencing

Genome sequencing began in the 1970s with viruses called bacteriophages that infect bacteria. In the late 1980s the human genome project began and was "finished" 11 years later. Thus far, more than 50 plant genomes have been sequenced[15] and many of those genomes have been studied for crop improvement.[16] The first plant to be sequenced was the model organism *Arabidopsis thaliana* in 2000 and new genome sequence data is often contrasted with Arabidopsis. An Agrobacterium species has also had its genome sequenced.[17] Since genome sequencing began, researchers have realized that genomes share several features at the macromolecular and molecular level.

One of the first sequencing methods in widespread use was a method known as Sanger sequencing. While earlier methods required radioactive labeling and laborious hours reading individual bases, Sanger sequencing incorporated fluorescent dyes that were much safer. Base calling also became computer automated. The Sanger method has changed over the years from requiring DNA to travel through large slabs of polyarcylamide gels to traveling through small capillary tubes.

The process of Sanger sequencing exploits the chemistry of DNA, specifically the deoxyribose portion of the molecule. DNA polymerase elongates DNA in the 5′⟶3′ direction, where the dangling 3′ OH group attaches to the 5′ carbon of the incoming nucleotide. In a method similar to PCR, DNA is amplified except a small amount of dideoxynucleotide triphosphates are added into the PCR reaction mixture. Incorporation of dideoxynucleotides terminate DNA elongation since they lack a 3′ OH group that DNA polymerase needs in order to attach another nucleotide. The dideoxynucleotides are also covalently bonded to a fluorophore. A different fluorophore is used for each nucleotide base.

The result after the DNA amplification is varying lengths of DNA that are tagged at the 3′ end with one of four fluorophores, depending on the final base that was incorporated. The varying lengths of DNA are pulled through a small capillary tube and interact with a laser beam at the end of the capillary. The wavelength (color) that is reflected back to a detector is dependent upon the fluorophore, which indicates the identity of the specific nucleotide base. Since smaller pieces move through the capillary faster than larger pieces, there is a size-dependent relationship between the called base and the position along the chain of DNA, resulting in a sequence of nucleotides that reflects the order of bases in a DNA sequence. Base calls are also accompanied by scores that reflect the degree of confidence that a base is indeed the base called. Confidence is also indicated by the height of the peak in the resulting chromatogram, which is the file containing the called bases that can be viewed (Figure 3).

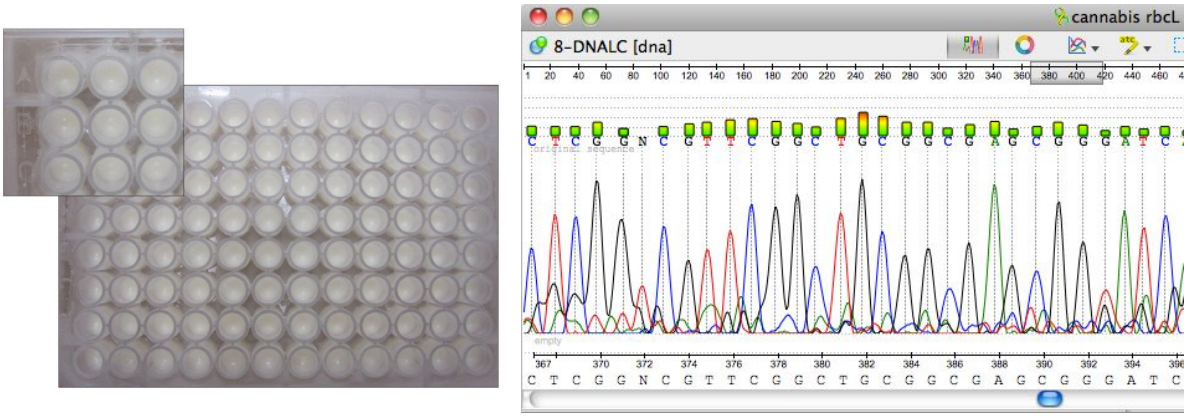


Figure 3. A 96-well plate used in automated Sanger sequencing. Each well has a unique piece of DNA in a vector, which is PCR amplified and pulled through a capillary. A chromatogram (right) shows a different color for each of the four bases as they traveled through the capillary.

Newer methods of DNA sequencing have not completely replaced Sanger sequencing but have resulted in higher throughput, or the ability to generate more information. Many people often refer to the newer methods as "Next-Gen" or next generation sequencing technologies. While they vary in their approaches and length of DNA sequences generated, they all exploit the biochemistry of the DNA molecule. Sanger sequencing and Next-Gen sequencing share the same basic procedures of needing to purify DNA, and fragmentation of DNA into smaller pieces. However, the main difference between the two is in how the DNA is amplified and how sequence information is generated.

One common next generation sequencing method (Illumina) relies on first shearing and fragmenting DNA into small strands. The DNA fragments are joined at both ends with smaller pieces of DNA called adapters. The adapters are then ligated so they flank both ends of all the pieces of sheared DNA. The short (~100-300 bp) pieces of adapter-DNA-adapter molecules are separated into single strands and allowed to hybridize with a short chain of single-stranded DNA attached to a glass slide. The hybridization occurs through the adapters and bases on the glass slide and leaves the genomic pieces of single-stranded DNA attached to the slide. The adapter on the free end of the DNA serves as a primer region for DNA polymerase to attach and begin synthesizing a second strand. Instead of incorporating chain-terminating dideoxynucleotides as in Sanger sequencing, nucleotides are added stepwise, followed by a wash step. After each addition of a nucleotide, a laser is used to help determine the incorporated base, which is subsequently recorded by a computer. The result is that many more pieces of DNA can be "read" in a much more time efficient manner.

An exciting sequencing technology recently invented relies on a novel method of sequencing and is the size of a small USB drive.[18] In fact, the mini DNA sequencer can attach to a computer USB port. The technology might allow anyone to generate their own sequence data from any organism they would like. The sequencing device has sensors that detect the electrical charge of each base as the DNA molecule passes through a nanopore. The nanopore is actually derived from natural bacterial proteins, similar to the type four secretion system in Agrobacterium. The bases are recorded and stored on the computer as the DNA travels through the nanopore. According to the company who came up with the idea, Oxford Nanopore, the USB sequencer can sequence a small bacterial genome in minutes. A human genome should be able to be sequenced in under an hour and for less than 1,000USD. Such an advancement would certainly allow sequencing of a cannabis strain well under an hour and facilitate several

comparative studies to better understand genetics or transcriptomes of any particular strains.

In most genome sequencing projects, sequence reads must be joined together (assembled). Researchers often rely on comparing sequences using bioinformatics tools in order to find overlapping regions on two apparently different strands of DNA. The order of assembly of different sequences is therefore heavily reliant on shared sequence overlaps between two sequence reads. Two sequences are thought to be adjacent to each other (continuous) if they share a particular amount of overlap. If the sharing is above the threshhold, the sequences can be merged (in silico) into one contiguous sequence. The goal is to continue generating contiguous sequences (called contigs), which eventually results in assembly of the sequences to generate an entire chromosome, mostly with the help of a process called scaffolding. Contig assembly is also performed in bioinformatics analysis of EST or cDNA libraries to generate longer transcripts.

Genomes that are sequenced also need to be annotated. Genome annotation is the process of determining the role of each of the coding and non-coding regions (or transcripts in the case of expression libraries). There are several steps that can be time consuming but the work is streamlined with the use of bioinformatics. Determination of what a transcript or gene codes for sometimes relies on supportive evidence from to known sequences. There are several bioinformatics programs that help almost anyone who has an introductory knowledge of gene organization annotate genes.

Competition for obtaining the sequence of an organism can sometimes lead to propelling techniques or technologies forward. In the human genome sequencing project, two very different approaches in sequencing genomes were used. The first was whole-genome shotgun sequencing and the second was a map-based approach. The first method, led by a genius named Craig J. Venter,[19] relied upon randomly fragmenting the DNA, sequencing the small parts, and then piecing the parts back together. The second method, led by another genius named Francis S. Collins,[20] relied on first generating a map of a chromosome and then determining where each piece of sequenced DNA fit onto the chromosome based on the map. Mapping first is more time consuming but argued to be more accurate, while the shotgun approach is faster but not as accurate. More could be said here, but the topic of the human genome project has been beat to death in several other books and journal articles.

The first draft of the *C. sativa* genome

Anticipated by several researchers seeking to better understand the cannabis plant was the publication, "The draft genome and transcriptome of *Cannabis sativa*," by Harm van Bakel and colleagues[21] from Canada. Their work was possible with the help of Medicinal Genomics, LLC. Not only did they seek to generate a draft genome, they also used the sequence data to investigate the genes involved in cannabinoid biosynthesis, analyze the expression patterns of different organs and different flowering stages of the plants, and provide further evidence for the biochemical difference between medical cannabis and hemp. The results of their studies, which greatly expanded the molecular information available on cannabis, also allowed a genetic basis for hemp breeders who aim to improve oil and fiber production in different hemp varieties.

The cannabis genome was estimated to be 534 Mb (million bases) with approximately 30,000 genes. By comparison, the human genome has about three billion bases and 21,000 genes, and the Arabidopsis genome has about 26,000 genes. In the main part of the study, genomic data from Purple Kush and the hemp strain known as Finola was generated and compared. Two different Next-Gen sequencing technologies were used (Illumina and 454) that generated variation in length of the reads, which helped in the assembly (joining contigs) and scaffolding (lining along chromosomes) process. The assembly used 40,000 transcripts, in which 83% of these were also found to be present in other plant species.

Important to the process of science, the genomes were made publically available in "The Cannabis Genome Browser" where all genes involved in the biosynthesis of cannabinoid can be viewed (genome.ccbr.utoronto.ca). Genomic data for organelles was also generated, but was removed since the primary interest was plant genomic DNA. The size of the mitochondrial genome was estimated to be 450 kb (thousand bases) and the plastidial genome 150 kb.

The authors also compared the expression patterns of Finola and Purple Kush flowers and found that while both apparently contained the genes for synthesizing cannabinoids, expression levels were nonexistent in Finola. This led the researchers to conclude some of the biosynthesis genes involved in cannabinoid production have become quiescent (as pseudogenes) in Finola, and might help explain the genetic difference between medical cannabis and hemp. Interestingly, Finola did not show genetic expression of delta-9-tetrahydrocannabinolic acid synthase; however, this enzyme was expressed in Purple Kush. This finding is not

too surprising, since most cannabis consumers seek plants with high delta-9-tetrahydrocannabinolic acid and low cannabidiolic acid. Hemp cultivators typically grow plants with high levels of cannabidiolic acid and low THCA levels. Overall, genes involved in cannabinoid biosynthesis were much higher in Purple Kush. Not surprisingly, delta-9- tetrahydrocannabinolic acid synthase was found in the genome and transcriptome of Purple Kush.

In addition to investigating the transcriptome of Purple Kush flowers, expression patterns of the leaves, stem, and root were also explored. These tissues showed apparent similarity in the biochemical pathways required to carry out photosynthesis. However, the flowers showed the highest expression of genes involved in terpenoid and cannabinoid pathways. The different flower stages of Purple Kush all showed a constant presence of tetrahydrocannabinolic acid synthase; however, cannabidiolic acid synthase showed a complete absence. Some genes involved in the cannabinoid biosynthetic pathway for Purple Kush showed a 15-fold increase over Finola. Transcription factors involved in forming trichomes were also found in higher abundance in Purple Kush.

The authors also searched the transcriptomic data of Purple Kush to investigate the presence of the enzyme responsible for synthesizing cannabichromenic acid. The enzyme is hypothesized to be an oxidocyclase, essential in formation of cannabichromenic acid. The authors identified 23 candidates, of which the top ones are currently under biochemical study.

Of interest to hemp breeders, the authors investigated the relationship between cannabis known for its high cannabinoid content and hemp strains used for seed oil and fiber production. The authors obtained additional sequence information from the hemp strain USO-31 and Chemdawg. Finola is a short plant, produces high levels of seed oil, and is dioecious (male and female flowers are on two different plants). USO-31, which is taller, produces lower amounts of oil, and is monoecious (male and female flowers are on the same plant). Using Purple Kush as a reference genome, USO-31 showed the least amount of heterozygosity or genetic variation in alleles, presumably since USO-31 is a monoecious variety that likely self-pollinates. Using single nucleotide variations, the authors justified the recognition of hemp and cannabis as following two different evolutionary trajectories.

The first draft of the cannabis genome, in association with transcriptome analysis, has provided insight into genetic and biochemical differences between cannabis that has been selected for its therapeutic effects and hemp varieties. The next phase of characterizing the cannabis genome will require complete genome

annotation. The transcriptome data from flowers, stems, leaves, and roots has also allowed for comparison of biochemical synthesis between these different organs. The availability of the first draft of the cannabis genome will surely serve as a reference for future transcriptomic investigations and offers an excellent starting point for cannabis biotechnology research.

Emerging areas within genomics

The process of obtaining whole genome sequence data has exploded and is expected to continue to improve in speed. Thus far, sequencing technologies have generated a deluge of sequence data, which has opened up new areas of study and offered new approaches to evolutionary studies, medicine, and agriculture. Metagenomics, comparative genomics, pharmacogenomics, and phytochemical genomics are only a few of the fields that have been allowed to evolve due to the increase in sequence data.

Metagenomics generates sequence data without culturing organisms. An environmental sample is first obtained and the total DNA is extracted. A researcher is not typically concerned with what species are present, rather identifying the taxa is often a major goal. However, metabolic pathways can also be elucidated based on the presence of particular enzyme-encoding genes, which might provide insight into the metabolic processes of an ecosystem. Metagenomics data can also be useful in understanding the microbes that might be present in or on the leaves[22] of a plant or soil of a plant[23] but no metagenomic data specifically regarding cannabis has been yet acquired.

In the field of comparative genomics, one might be interested in how genomes have changed over time. As its name implies, comparisons of genomes are made between taxa to determine if movement of parts of chromosomes (translocations) or entire chromosomes have been lost or gained (anueploidy). Syntenic regions, or areas that have the same sequence, might therefore be retained by two diverged taxa but have migrated to different parts of the genome. The results of comparative genomics often provide deeper insight into genome architecture and genome evolution. Whole genome comparisons between cannabis and related species is presently awaiting investigation.

A subdiscipline of genomics that has changed the way medicine is studied is pharmacogenomics, which helps to analyze an individual's genetic profile. By knowing if an individual has certain point mutations (SNPs) in their genome,

doctors are able to better serve patients. However, not all genes are studied equally. Instead only genes encoding drug-metabolizing enzymes, drug receptors, or genes that might predispose one to a particular disease are analyzed. Pharmacogenomics helps people working in medicine, therefore, to better predict the response of a patient to a drug, tailor a particular drug prescription to a patient, or estimate the likelihood of developing a particular disease. Cannabis patients might someday wish to have their genomes analyzed to know which strain of cannabis would best suit their medical needs.

A more recent area in genomics called phytochemical genomics[24] seeks to understand the metabolic pathways in plants that generate important plant secondary metabolites. Although the desire to determine such pathways has a deep history, genomic technologies and increased throughput are allowing a firmer grasp on all of the genes involved in a pathway. There is also the potential to generate new, genome-wide knowledge on enzymes that might be involved in more than one pathway, understanding the evolutionary divergence of various genes involved in metabolic pathways, and the influence of such pathways on other genes in the genome. Again, technological advancements are making phytochemical genomics a developing reality.

With so many areas emerging from sequencing technologies, the field of cannabis genomics seems only to be waiting on more people to join in the effort. Might we be approaching a day when cannabis growers can subjugate their favorite strain of cannabis to an at-home sequencing machine or better understand their own genome? The possibility seems extremely likely. If so, there would be multiple opportunities and research avenues to pursue. More information and knowledge on the unique sequence combinations that give rise to a distinct cannabis strain might lead to a deeper understanding and more meaningful relationship with the cannabis plant.

Chapter 8

Engineering Cannabinoids

"Cannabis is produced in virtually every country of the world, making it the most widely illicitly produced and consumed drug plant."
- United Nations World Drug Report, 2013

Metabolomics

The science of metabolomics seeks to characterize the small molecules (metabolites) in a cell. Metabolomics also seeks to understand all anabolic and catabolic processes of a cell or tissue. Anabolic reactions that build up cell constituents and catabolic reactions that break down substances are the two overarching processes of cannabis metabolism. Thousands of different products and reactants might occur in a cell at any given moment. Although diverse in their structure and chemistry, only certain metabolites are of primary interest when discussing cannabinoids. However, due to the nature of cross-talk between metabolic pathways, seemingly unimportant metabolites might play a significant role in cannabinoid biosynthesis.

Metabolites, whether found in animal or plant systems, can be classified as either primary or secondary metabolites. Primary metabolites are necessary requirements of the organism and are macromolecules that include proteins, nucleic acids, carbohydrates, and lipids. Secondary metabolites are not necessarily needed by a cell or organism to carryout essential functions; however, secondary metabolites might provide an evolutionary advantage.

Secondary metabolites, like primary metabolites, often contain chemical groups that contribute to their overall biochemistry, thus they are known as functional groups. Examples of functional groups and types of macromolecules they can be found in include hydroxyl groups (-OH) in sugars and alcohols, amino groups ($-NH_3$) and carboxylic acid

(-COOH) groups in amino acids (hence the name), and phosphate groups (-PO_4) in nucleic acids and adenosine triphosphate. Sulfhydryl groups (-SH) might also be found and are perhaps most familiar to people who have sliced onions for too long.

Metabolomics has taken huge leaps in recent years and in several areas of biology, including human biology.[1] Although knowing many of the biochemical reactions that can occur might be intellectually satisfying, memorizing every single biochemical step that occurs in a cell is unnecessary and is likely to be impossible (without the aid of a computer chip implant). In regards to cannabis biotechnology, one might want to primarily concern themselves with the products and reactants that lead to cannabinoids and related therapeutic compounds. That said, the major cannabinoid metabolites include pathways that are also important in primary metabolism.

Many secondary metabolites have always been important to human society due to their medicinal properties, mind-altering effects, and ability to add flavors to otherwise bland foods. The molecular structures of many isolated plant metabolites have been characterized with the advancements of chemical analysis tools such as chromatography, mass spectroscopy, and nuclear magnetic resonance imaging.[2] Several types of chromatography (e.g. high-performance, liquid, etc.) can be used in studying cannabis metabolites.

Many school science labs use thin layer chromatography (TLC) to separate a metabolite mixture to determine the compounds present. TLC offers an inexpensive, reliable, and relatively fast method for metabolomic analysis. To perform TLC analysis, a sample is prepared by crushing the plant tissue and adding it to a solvent. A small drop of solution containing the sample is spotted onto a chromatography plate and allowed to dry. The plate is placed inside a jar containing a few milliliters of a solvent. Since only the bottom portion of the plate is contacting the solvent, the spotted samples are pulled up the chromatography matrix as the solution migrates via wicking. The different metabolites are carried with the solvent and separated. The distance each metabolite migrates is compared to previously characterized metabolites on a control (known) chromatography sample. Growers or consumers of cannabis have the option of testing the metabolomic profile of their cannabis, since at-home kits are available through the Internet (thctestkits.com; cannalyticssupply.com).

Metabolites of interest in cannabinoid profiles are mostly secondary metabolites, which are generally divided into three major classes that include alkaloids, phenolics, and terpenes. Alkaloids are defined as bitter-tasting, nitrogenous compounds. Well-known alkaloids include atropine, caffeine,

psilocybin, strychnine, quinine, and cocaine. Many alkaloids have been isolated from the Solanaceae family, also known as the deadly nightshade family. Some of the earliest reports of alkaloids come from the Egyptians. Cleopatra, the famous Egyptian queen, was said to have put belladonna extracts into her eyes, which dilated her pupils and supposedly increased her attractiveness. Alkaloid biosynthesis has been well-documented in plants from the Solanaceae but can also be found in fungi such as ergot.[3]

Phenolics are another class of secondary metabolites that include tannins, vanilla, nutmeg, capsaicin (the spicy hot molecule in peppers), and anthocyanins (plant pigments). All phenolic compounds contain the phenolic ring, a six-carbon ring (benzene) with a hydroxyl group (-OH). Lignin, a huge polymer of phenolic rings found in cell walls, is the most common phenolic compound among plants.

Terpene production

Terpenes are the third and largest class of secondary metabolites that provide a wide array of functions in plants. This is not surprising since there are more than 30,000 described terpenes from plants.[4] Examples of terpenes include the tail portion of the chlorophyll molecule, which is composed of a diterpene called phytol. The fragrances of citrus fruits are due to the presence of various types of terpenes.

The basic enzymatic pathways leading to terpene metabolites incorporate carbon molecules based on multiples of fives. Therefore, a nomenclature system has emerged that follows this pattern (Table 1). Plants can anabolically produce terpenes and might later recycle the parts by breaking them down catabolically. Often, large terpene compounds can be broken down and released in the form of 2-methylbutane (isoprene), show in Figure 1. Some mountains (e.g. Blue Ridge Mountains of North America) appear blue because of the high levels of isoprene released from the trees.

Terpene type	Formula	Molecular weight	Isoprene units	Example
Hemiterpene	**C_5H_{16}**	**76.2**	**0.5**	**2-methylbutane (isoprene)**
Monoterpene	**$C_{10}H_{16}$**	**136.2**	**1.0**	**geranyl diphosphate**
Sesquiterpene	**$C_{15}H_{24}$**	**204.4**	**1.5**	**farnesol**
Diterpene	**$C_{20}H_{32}$**	**272.5**	**2.0**	**phytol**
Sesterterpene	**$C_{25}H_{40}$**	**340.6**	**2.5**	**leucosceptrine**
Triterpene	**$C_{30}H_{48}$**	**408.7**	**3.0**	**squalene, THC**
Tetraterpene	**$C_{40}H_{64}$**	**544.9**	**4.0**	**carotenoids**
Polyterpene	**$C_\infty H_\infty$**	**100,000-1,000,000**	**1,500-15,000**	**natural rubber**

Table 1. Introduction to terpene types and formulas with estimated molecular weights and number of isoprene units. Although only one is given, several examples could be provided for each terpene type.

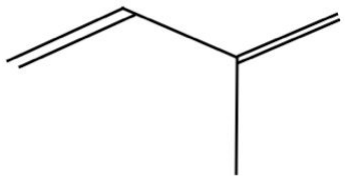


Figure 1. A single isoprene molecule. Terpenes with larger molecular weights are constructed by cells using phosphorylated isoprene derivatives.

The five carbon units for building terpenes consist of the phosphorylated starting materials isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). These can be joined in either "tail to tail" or "head to tail" reactions. In cannabis, there are two pathways, which lead to production of terpenes. The first is known as the mevalonate (or MVA) pathway. The mevalonate pathway occurs in the cell cytoplasm and leads to sesquiterpenes and triterpenes. The second pathway is deoxyxylulose (or DXP or non-MVA) pathway and occurs in the plastid. The latter pathway is required for synthesis of monoterpenes and diterpenes, which are used in cannabinoid biosynthesis. One of the most important

larger terpenes is geranyl diphoshpate, which is needed for delta-9-tetrahydrocannabinolic acid biosynthesis.

The mevalonate pathway first relies on a thiolase to catalyze the synthesis of acetylacetyl-CoA by fusing two acetyl-CoA molecules. HMG-CoA synthase synthesizes acetylacetyl-CoA with a third acetyl-CoA to produce 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA). A final reaction catalyzed by HMG-CoA reductase uses 2 NADPH to reduce HMG-CoA to the six-carbon molecule mevalonate.

The high-energy molecule adenosine triphosphate (ATP) is required for the next three reactions, which ultimately lead to isopentenyl diphosphate. These reactions involve MVA kinase, MVAP kinase, and MVAPP decarboxylase, and proceed with mevalonate, mevalonic acid 5-phosphate (MVAP), mevalonic acid 5-diphosphate (MVAPP), and isopentenyl diphosphate (IPP), respectively. The result is one molecule of isopentenyl diphosphate that can be used in constructing larger terpenoid molecules

The plastidial pathway is initiated with the joining of a pyruvate molecule to a glyceraldehyde 3-phosphate (which can be obtained in photosynthesis) molecule facilitated by the enzyme DOXP synthase. This forms 1-deoxy-D-xylulose-5-phosphate (DOXP). This is reduced by the enzyme DOXP reductoisomerase (DOXP-R) to form 2-C-methyl-D-erythritol 4-phosphate (MEP). A cytidine triphosphate is then incorporated to form 4-(cytidine-5-diphoshpo)-2-C-methyl-D-erythritol (CDP-ME) via the enzyme CDP-ME synthase.

An ATP is used to add a phosphate to form 4-diphosphocytidyl-2C-methyl-D-erythritol 2-phosphate (CDP-ME-2P). The enzyme that catalyzes this reaction is CDP-ME kinase. This product is then cyclized to form 2-C-methyl-D-erythritol 2,4-cyclodiphosphate (CDP-ME diphosphate) via CDP-ME diphosphate synthase. After removing a water molecule, (E)-4-hydroxy-3-mehtylbut-2-enyl diphosphate (HMBPP) is formed via HMBPP synthase. The final step removes an additional water molecule while simultaneously reducing (E)-4-hydroxy-3-mehtylbut-2-enyl diphosphate to yield isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP).

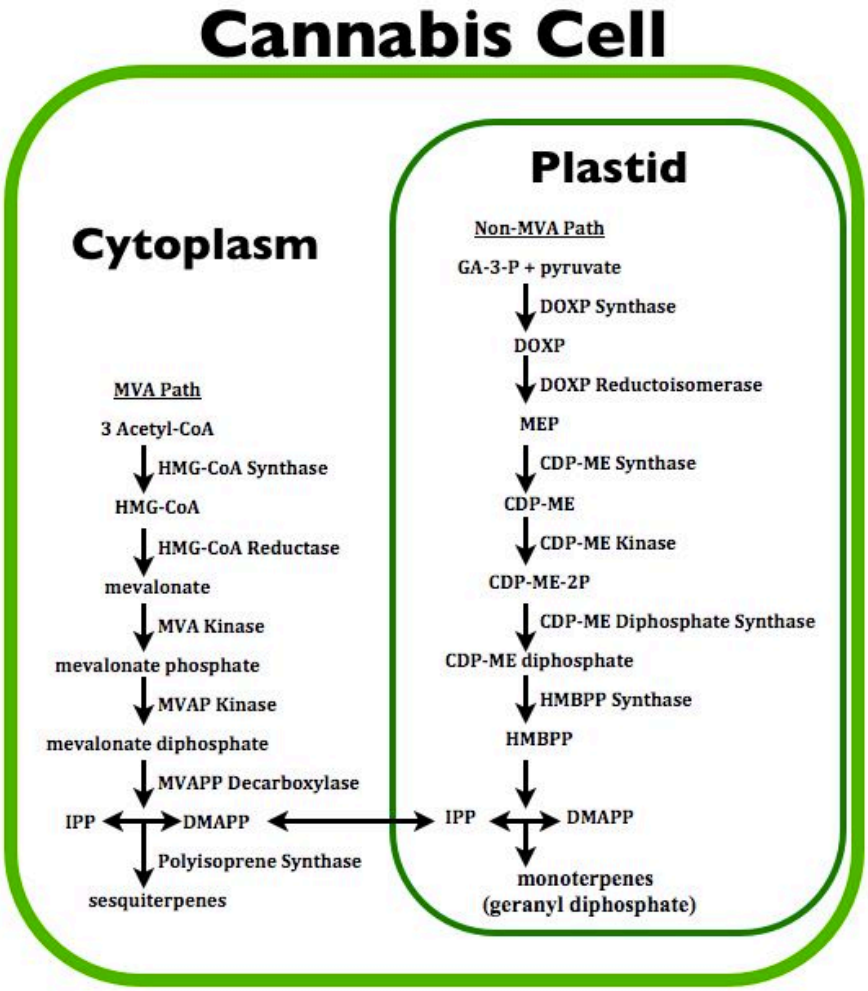


Figure 2. Outline of the cytosolic and plastidial terpene pathway. Each enzyme might be a potential target of cannabis biotechnology.

Since they are phosphorylated, the IPP and DMAPP can be used in the so-called "head to head" or "tail to tail" combinations to build terpenes. DMAPP can also be produced from IPP by the enzyme isopentenyl-diphosphate isomerase (IPP isomerase). Dimethylallyl transferase uses either IPP or DMAPP to form geranyl diphosphate or farnesyl diphosphate via polyisoprene synthase. Geranyl diphosphate and farnesyl diphosphate are monoterpenes and sesquiterpenes, respectively. Geranyl diphosphate is used in the synthesis of several important cannabinoids. Finally, one should recognize that there can be cross-talk and exchange of terpene products between the cytosolic and plastidial pathways (Figure 3).

Other terpenes produced by cannabis are not directly associated with cannabinoid production. Rather, they impart flavor or aroma to the plant. Limonene is a terpene that has an extremely fruity smell. By its name one would

guess (correctly) that the smell is similar to a lemon. In fact, lemons are loaded with limonene. Pinene is also found in cannabis and was named in honor of pine trees since they have an abundance, which gives the distinct pine smell. Another terpene in cannabis is linalool, which has a rather sweet smell. Beyond pleasant smells, limonene, pinene, and linalool have either anti-inflammatory or soothing effects. There are more than a hundred types of terpenes in cannabis and include other well-characterized types such as terpineol, myrcene, menthol, and cymene.

Many biochemical reactions taking place within plant cells to generate terpenes (or other metabolites) are not carried out in continuous sequential steps. Although biochemical pathways occur when precursor molecules are available, reaction rates can only proceed as fast as products or previous reactions are made or accumulate to sufficient concentrations. Enzymes are often suspended within an intracellular matrix or might be found clustered together via attachment to a cellular membrane, so that reactants are more likely to interact with the correct enzyme. In this way, a complex interaction between enzymes and their substrate is played out where an enzyme may only be produced on demand via other methods of cell signaling networks. In other instances, increasing the concentration of a substrate might cause an increase of the products.

In general, enzymatic reactions taking place within the cannabis cell all occur very rapidly and depend heavily on the temperature and concentration of reactants and enzymes. However, the terpene pathway is one of many plant biosynthetic pathways. Therefore one should not be too surprised to find that terpene biosynthetic pathways overlap with other plant pathways, including plant hormone synthesis. For example, gibberellins and auxins (plant hormones) are both formed starting with a molecule of mevalonate derived from the MVA pathway.

Besides being important to the plant, there are increasing numbers of reports suggesting terpenes might be directly responsible for the therapeutic effects of cannabis, since terpenes interact with proteins of human cells. A researcher might find genetically modifying the terpene profile of cannabis entertaining in addition to being therapeutic. Cannabis might be engineered to have high levels of limonene and could be consumed during warm weather. Something akin to having a cannabis lemonade. During the Christmas season, other cannabis lines could be consumed. Personally, cinnamon nutmeg cannabis in front of a warm fire sounds enjoyable.

A note on the polyketide pathway

Before proceeding to the cannabinoid pathway, readers will best be served to know that formation of several important and predominant cannabinoids (tetrahydrocannabinolic acid, cannabidiolic acid, and cannabichromenic acid) are synthesized from a starting molecule known as cannabigerolic acid.[5] However, cannabigerolic acid is, importantly, produced from terpene pathway-derived molecules. The first, already introduced, is geranyl diphosphate and is also known as geranyl pyrophosphate (GPP).

The second important molecule needed for cannabigerolic acid synthesis is olivetolic acid, a metabolite derived in part from a polyketide pathway. Olivetolic acid is characterized as being an alkylresorcinol, presumably derived from acetate. A polyketide compound literally means the compound has many ketides where a ketide is, for simplistic purposes, two carbons. Polyketide synthases add two carbons at a time, where the two carbons are often derived from small molecules also found in the terpenoid pathway.

The pathway leading to olivetolic acid is thought to be synthesized from three molecules of malonyl-CoA and one molecule of hexanoyl-CoA. Malonyl-CoA can be derived from malonic acid (from glucose) and hexanoyl-CoA can be derived from a hexane (from fatty acids). Formation of hexanoyl-CoA begins with a fatty acid that is desaturated and then acted upon by a lipoxygenase followed by further action from a hydroperoxide lyase to give hexanoate. Acyl-activating enzyme, which uses one acetyl-CoA molecule, finally produces the hexanoyl-CoA.[6] The hexanoyl-Co-A is acted upon in a series of reactions by polyketide synthase enzymes that combine it with molecules of malonyl-Co-A resulting in the olivetolic acid precursor.

To form olivetolic acid, olivetolic acid cyclase circularizes the chain structure of the polyketide, making a ring structure with a tail. The characterization of olivetolic acid cyclase has an interesting history. As an important intermediate in cannabinoid synthesis, understanding its production is key to understanding how cannabinoids are synthesized; however, parts of this pathway have been rather elusive.

Early research attempted to pin down an olivetol synthase enzyme, a polyketide synthase, responsible for generating olivetol as its product. In 2004 Tri Raharjo and his colleagues[7] worked to try to identify the enzyme. Although they had found possible candidates, their isolates could only form olivetol, not olivetolic acid. In a separate experiment,[8] they cloned and expressed a polyketide synthase

in *E. coli*. Subsequent addition of hexanoyl-CoA and malonyl-CoA into the *E. coli* suspension were hoped to generate olivetolic acid or olivetol but neither was detected. More research on the pathway was needed.

In 2009 work led by Taura and his colleagues,[9] focused on cloning another polyketide synthase. Their approach relied on cloning a polyketide synthase that was expressed in *E. coli*. However, their next step was isolation of the polyketide synthase, then subsequent experimentation with enzymatic assays using various substrates as starting compounds. The assays suggested synthesis of olivetol was occurring but not olivetolic acid. They also found extracts of cannabis failed to generate olivetolic acid from starting compounds. The results led them to hypothesize that a second protein might be required with olivetol synthase or that other temporal effects (cannabis developmental stages) might be playing a part in generating olivetolic acid.

In 2012 Steve J. Gagne and his colleagues from the University of Saskatchewan in Canada sought to study the polyketide pathway in cannabis as well.[10] They began their research by first searching through cannabis transcriptome (EST) data generated by David Marks and his colleagues.[11] Since the EST library was derived from the mRNA transcripts in cannabis flowers, such a place to begin a search was thought to offer a good chance to identify the enzyme involved in olivetolic acid synthesis. Gagne and his team used criteria based off structural or sequence similarity to other polyketide synthases to identify candidates. Their search resulted in three possible enzymes, one of which was of primary interest due to its similarity with a bacterial polyketide cyclase. Subsequent cloning and expression of one of their candidates in yeast cells resulted in olivetolic acid production.

They have since renamed the olivetol synthase/polyketide enzyme sought by earlier researchers as tetraketide synthase, which is more reflective of its function. They further suggest that tetraketide synthase might generate olivetol as a substrate for olivetolic acid cyclase within the cannabis plant. Finally, their report encourages the exploration of culturing microbes for production of cannabinoids, since they had shown this to be possible with yeast. More research is needed to determine other possible intermediate steps and to characterize the additional enzymes that might be involved in the entire polyketide pathway in cannabis. For the moment, cannabis biotechnologists have a few interesting enzymes responsible in olivetolic acid production to consider in their research.

The cannabinoid pathway

Similar to the terpene and polyketide pathways, the cannabinoid pathway consists of different enzymatic steps and has intermediate molecules. After learning about the terpene and polyketide pathways, one can much more easily understand the origins of the predominant cannabinoids. Additionally, there are fewer enzymatic steps involved (once geranyl diphosphate and olivetolic acid are generated) to reach the cannabinoid end products.

Interestingly, tetrahydrocannabinolic acid can be heated and decarboxylated to become chemically modified into a more psychoactive form, which is delta-9-tetrahydrocannabinol (THC). (The "delta-9-" refers to the location of a double bond.) Since the tetrahydrocannabinolic acid component of cannabis is the precursor of THC, tetrahydrocannabinolic acid (THCA) formation and accumulation within the plant directly influences the amount of THC present in cannabis when consumed. THC has been recognized as a major component of cannabis smoke and for this reason, the chemical structure was first determined in the 1930s and within a decade was chemically synthesized.[12] The heating (by an oven or lighter) causes a decarboxylation reaction, or a loss of a carbon group that is on the TCHA molecule, thereby converting it to the more psychoactive THC molecule (Figure 3).

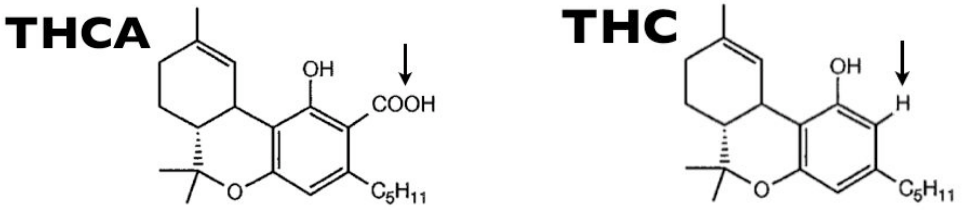


Figure 3. Molecular structure of delta-9-tetrahydrocannabinolic acid (THCA) and delta-9-tetrahydrocannibinol (THC). The arrow indicates the carbon lost upon heating.

Synthesis of tetrahydrocannabinolic acid begins when a molecule of geranyl diphosphate is joined to one molecule of olivetolic acid, which yields cannabigerolic acid (CBGA). The enzyme performing this reaction is called geranylpyrophosphate:olivetolate geranyltransferase (GOT) and is a type of prenyltransferase. The resulting product, cannabigerolic acid (CBA), can be modified in several ways. Of importance to many cannabis consumers is the production of cannabichromenic acid (CBCA), tetrahydrocannabinolic acid, and cannabidiolic acid (CBDA). All require cannabigerolic acid as a starting compound.[5]

If cannabigerolic acid is acted upon by cannabidiolic acid synthase, the result is cannabidiolic acid. Similarly, if cannabigerolic acid is acted upon by cannabichromenic acid synthase, cannabichromenic acid is formed. Another final product after cannabigerolic acid formation might be tetrahydrocannabinolic acid by way of tetrahydrocannabinolic acid synthase. Each of these enzymes are potential targets of cannabis biotechnology and up-regulating their production might increase different cannabinoids in a cannabis plant.

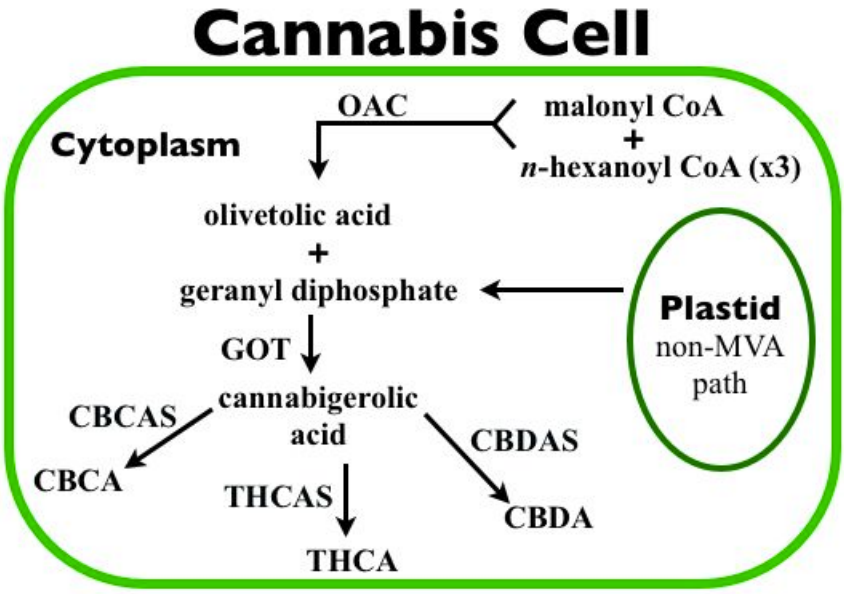


Figure 4. Cannabis cell depicting olivetolic acid cyclase (OAC) generating olivetolic acid. Geranyl diphosphate and olivetolic acid combine to yield cannabigerolic acid, which in turn yields tetrahydrocannabinolic acid (THCA).

Although not shown here (Figure 4), many more compounds besides CBCA, THCA, and CBDA can be formed from cannabigerolic acid. Additionally, much more can be said of the cannabinoid pathway since so many cannabinoids have been identified. Nearly all the cannabinoids have been found to have a medicinal application; cannabinol has sedative and antibiotic properties; tetrahydrocannabivarin elicits a euphoric effect and reduces pain; and delta-8-THC combats nausea.

Clearly cannabinoids do not merely induce altered states of consciousness; they are recognized for their non-psychotropic effects as well. Angelo A. Izzo and his colleagues, who are all members of the Endocannabinoid Research Group in Italy, have produced an excellent review article[13] with the assistance of Raphael

Mechoulam (a person who will receive a bit more discussion momentarily). Their article discusses the pharmacological activities and therapeutic uses of phytocannabinoids in beautiful detail. Such an outline emphasizes the fact that, while many might be interested in generating genetically enhanced cannabis for recreational use, a much-needed application of the biotechnology of cannabis will be for therapeutic use to treat an array of maladies.

Engineering cannabinoids

Becoming familiar with both the terpene pathway and the cannabinoid pathway allows one to understand key enzymes that are important for cannabis metabolic processes. This is crucial to providing ideas for genetic modification of any biosynthesis system to produce desired cannabinoids. For example, in order to increase the concentration of the psychoactive components of cannabis, an increase in geranyldiphosphate or desaturated fatty acids might be important. The gene coding for the protein that synthesizes geranyldiphosphate or a particular desaturase therefore might need to be overexpressed.

Since the final reactions in the biochemical pathway of cannabinoid biosynthesis have also been elucidated, selecting genes that produce enzymes with a role in the late stages of cannabinoid production is also possible. For example, constitutive expression (via the CaMV35S) of cannabidiolic acid synthase or cannabichromenic acid synthase might result in accumulation of cannabidiolic acid or cannabichromenic acid, respectively. Up-regulation of geranylpyrophosphate:olivetolate geranyltransferase (GOT) might result in the accumulation of more cannabigerolic acid, the starting compound for all three of the above-mentioned cannabinoids.

Generating engineered cannabis with high levels of THCA synthase might result in a plant with an increased THCA content. Subsequent consumption of the plant might produce a heightened euphoric feeling, since that is the main effect imparted by THC. One should not limit themselves to cannabinoids. Choosing any gene that codes for any enzyme within the terpene pathway might produce a similar increase. Using biotechnology, metabolites could be expressed in cannabis that are not normally found. Perhaps generating a plant that has endocannabinoids, cannabinoids found within the human body, might be a project worth pursuing.

The important aspect to remember from the complex metabolic pathways of cannabinoid synthesis is that transferring any of these genes with a distinct promoter for manipulating gene expression levels is possible with tools of biotechnology. Although cannabis lines with unique metabolic profiles have been produced via breeding, biotechnology might offer an even more accurate method of delivering a more tailored metabolic profile.

To further help cannabis biotechnologists in their research, the raw sequence data for each enzyme from each pathway is available through the genome draft, thereby providing a starting point for generating novel cannabis lines that over express genes of key steps. Similarly, novel genetic lines could be established through gene silencing of a particular step in the pathway. A reduction in one cannabinoid might result in accumulation of a different cannabinoid that might be more active in certain patients. Such an approach might produce a novel line that is tailored to a particular patient's need. For example, combinations of cannabinoids in vitro have shown promising results in slowing the growth of leukemia cells.[14] Of additional interest is the potential of cannabis terpenes acting in synergy with cannabinoids.[15]

Some molecular structures can elicit similar results as cannabinoids but are not synthesized by cannabis. Chemistry often uses the term "synthetic" to refer to chemical reactions taking place in vitro, which is often the process of addition or removal of chemicals and solvents into a reaction vessel. Biochemists and biologists, on the other hand, use the term "biosynthetic" (and its derivatives) to describe reactions happening in a cell. For example, phytocannabinoids are biosynthesized in cannabis plants.

There are reports of people obtaining and using "synthetic marijuana" such as K2 or spice. Cannabinoid analogs that are found in mixtures like K2 or spice only have molecular structures that mimic the effects of cannabinoids, they are not the same structure chemically. Sometimes the chemicals are more specifically referred to as functional analogs, since they mimic the effects of another drug. Synthetic cannabinoid analogs such as JW200 have found uses in the medical arena.[15] There are other functional analogs of cannabinoids in nature and there are terms used to distinguish them as well. For example, cannabinoids from cannabis are phytocannabinoids and cannabinoids that can be found within the human body (i.e. endogenously) are called endocannabinoids. For the purposes of this book, cannabinoids are implied to be phytocannabinoids unless otherwise specified.

Synthesis of cannabinoids in vitro has a history of more than four decades.[16]

In the 1960s Raphael Mechoulam and his colleagues from Israel reported in vitro synthesis of tetrahydrocannabinolic acid, cannabigerolic acid, and cannabidiolic acid.[17] Mechoulam is a leader in the field of cannabis research and has given interviews discussing his findings, trying to educate the public on the importance of cannabinoids. After his publications and research gained attention, others joined in the effort and in later years a string of publications followed, describing how to produce several other cannabinoids in vitro.[18]

As biotechnology, cloning, and heterologous expression systems became more widespread, even more people began synthesizing cannabinoids in vitro. In 2004 Supaart Sirikantaramas and his colleagues from Japan published a scientific paper[19] describing how they cloned the gene for THCA synthase. They further analyzed the in vitro activity of THCA synthase and determined its activity is at least partly dependant on the presence of the coenzyme FAD. After cloning the gene into a binary vector, tobacco root explants were infected with Agrobacterium carrying the vector-insert construct. Resulting hairy roots were placed into liquid culture media. Interestingly, upon analysis of the culture media, the root cells were able to transform cannabigerolic acid into THCA.

In the same study, Sirikantaramas and his colleagues used genetically modified bacterial and insect cells to produce THCA synthase. Bacterial cells produced an insoluble protein so were not pursued further. The insect cells were able to produce THCA synthase; however, the enzyme was not accumulating within the cells and was being secreted instead. As a final note in their publication, the authors cite Raphael Mechoulam and his colleagues' work on in vitro synthesis of cannabigerolic acid, emphasizing that the precursor metabolites for THCA, are inexpensive and easy to produce. Thus, the authors return to the idea that in vitro production of cannabinoids might have a therapeutic application.

In 2009, as part of a joint effort between researchers in the Netherlands and the Czech Republic, Isvett Josefina Flores-Sanchez and colleagues reported on the results of cannabis cell suspension cultures exposed to a variety of different factors.[20] They subjugated the cultures to more than a dozen different abiotic and biotic compounds such as jasmonic acid, fungi, plant cell wall components, and ultra-violet light. They monitored the expression of THCA synthase gene for more than a month and found there was no expression. However, several precursor metabolites were detected, leading them to hypothesize that expression of THCA synthase must be dependent upon tissue type or developmental period. Based on their results, the authors encourage analysis of the promoter regions controlling cannabinoid genes.

Many of those working to understand cannabinoids meet every year at a symposium hosted by the International Cannabinoid Research Society (ICRS). The ICRS (icrs.co) is an organization whose goal is to facilitate discussion among researchers working on all areas of cannabinoids. Currently, there are more than 500 members of the ICRS across the globe, all working together in the interest of science to understand synthetic cannabinoids, endocannabinoids, and phytocannabinoids.

A part of the ICRS's research focuses on understanding molecular structures of cannabinoids. Metabolites that are made in cells have a distinct chirality, or handedness. That is to say, if molecules could be placed in front of a small mirror, one would see two images. The first would be the original molecule and the second would be the molecule's reflection. Since the arrangement of the atoms differs, they are called left and right-handed images. In chemistry the two molecules are recognized as being distinct due to the fact that each usually has different biochemical properties. They also fall into a distinct class of isomers called enantiomers (mirror images).

Biochemists are well aware of chirality, since most enzymes often only recognize one handedness or the other. In other cases, enzymes or cell receptor proteins weakly interact with one isomer and bind strongly with the other isomer. Most enzymes also only synthesize one form of isomer. For example, amino acids are almost all of the left-handed type. Exceptions exist, especially in bacteria. Sugars are often of the right-handed form. However, chemical synthesis reactions in non-cell (in vitro) systems often produce an even mixture of right-handed and left-handed molecules and is called a racemic mixture. In some instances, a particular percentage of one handedness is produced over the other, varying by the chemical being synthesized.

Thus, endocannabinoids in humans also show handedness. We have subsequently evolved cell receptors (proteins on our cells' plasma membranes) for these internally produced (endogenous) molecules to bind and cause a cascade of biochemical reactions. Lumír Hanuš, a long-time colleague of Raphael Mechoulam in Israel, has provided an extensive, in-depth review entitled "The Pharmacological and Therapeutic secrets of plant and Brain (Endo) Cannabinoids".[21] I reference the paper here because the subject is one that deserves much attention but is beyond the scope of our present goal.

Cannabinoids binding with cell receptors ultimately provide the euphoric feeling, or high, after smoking. However, the binding of phytocannabinoids to our cells' receptors is actually due to cross-reactivity. Endocannabinoids and

phytocannabinoids might bind the same receptors but they have a noticeably different molecular structure. In some cases cannabinoids bind to slightly different cell receptors, eliciting different cell signaling cascades.

For this reason, some companies have taken the approach of using only one or two cannabinoid metabolites and packaging them as a medicine. In other words, pure THC, CBD, CBC, or CBG are produced and isolated. David Watson had the idea of single cannabinoid cannabis varieties almost 20 years ago. He now helps run HortaPharm B.V. in Amsterdam, which is known for breeding plants with single-cannabinoid profiles. Such profiles are highly sought in regards to the therapeutic application of cannabis.

A question that many might be asking is, "Why should anyone rely on a company to generate cannabis when you can grow your own plant?" Certainly, not everyone relies on companies to supply their cannabis needs. However, there are so many strains of cannabis each with a specific metabolic profile (chemotypes), that one might have difficulty in finding a distinct cannabis strain wanted. Different companies who are each able to produce several strains of cannabis are able to consistently deliver reliable products to consumers.

To enhance current methods, cell culture systems might offer a novel way to quickly scale up production of desired metabolites like cannabinoids. In some cases starting compounds might be needed to feed into a cell culture system and the cellular machinery then leads to an end product. Various substrates such as hexanoate, geranyl diphosphate, and olivetolic acid are already produced and sold commercially (sigmaaldrich.com). The substrates can then be added into systems to boost production of a desired product. The factor of importance is more likely to circle around what type of cell culture system to manipulate into producing cannabinoids. Using genetically modified cells in a suspension culture might increase cannabinoid production compared to non-genetically modified systems. This is the idea behind molecular pharming, which is the use of genetically modified plants or other organisms to produce chemicals (or proteins) with pharmaceutical importance.[22]

Bacterial systems (e.g. *E. coli* or cyanobacteria) are often easier to culture and are considered easier to genetically modify compared to eukaryotes such as cannabis. They also have a high ratio of surface area to volume, which increases the rates of metabolic exchange between the media and the cells. Some studies seek to identify particular microbial strains by screening for mutants that might have increased rates of metabolite production, similar to cannabis breeders selecting for plants that might have increased metabolites. Other studies

overexpress particular genes in a pathway in order to increase distinct secondary metabolites (Figure 5).

However, smaller eukaryotic organisms that are less complex than cannabis are often genetically modified. *Saccharomyces* (an ascomycete fungi) and *Chlamydomonas* (a green algae) are great examples. One might foresee the use of bioreactors, similar to those used in large-scale microbial fermentations or other areas of industrial microbiology, in order to mass produce cannabinoids. Upon choosing a host organism, different but interconnected bioreactors each containing a genetically modified host might allow several stages of production of each metabolite and necessary subsequent reactions. Each bioreactor would simply need to be sequenced in the same order as the pathway proceeds in the cell.

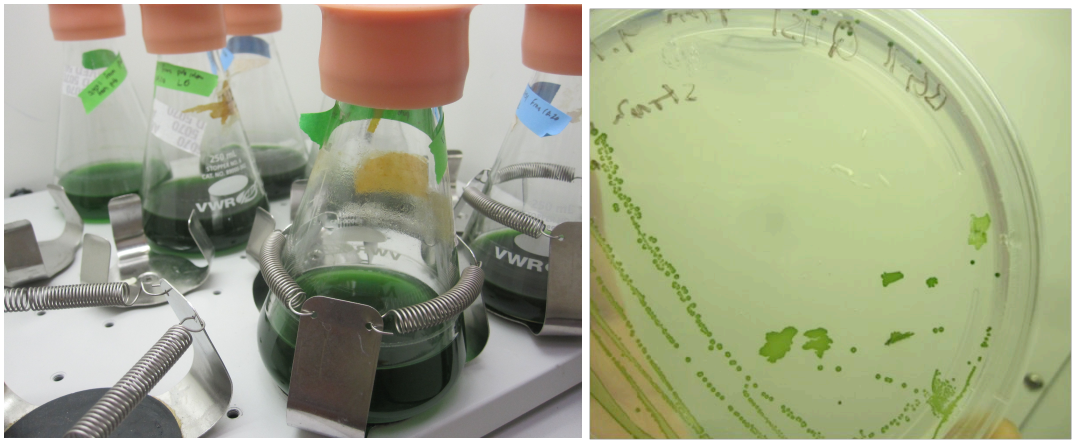


Figure 5. Algae or cyanobacteria can be genetically manipulated to increase or decrease a particular secondary metabolite. Here, *Synechocystis* (a cyanobacterium) has been genetically modified to overexpress beta-caryophyllene synthase resulting in higher levels of beta-caryophyllene, an important terpene found in cannabis.

Others have proposed plant tissue culture might be a good way to produce cannabinoids for mass production. Several benefits in reliability and stability might be offered in plant tissue culture systems. W. H.-T. Loh and colleagues had predicted such benefits in 1983 by stating, "The production of secondary metabolites and the biotransformation of precursor compounds by plant tissue cultures may lead to both the commercial synthesis of valuable substances and an elucidation of their biosynthetic pathways."[23]

Some have argued using citrus plants might be an optimal starting point for developing cannabinoid-infused plants. This idea was apparently stimulated by an Internet hoax, claiming a university professor from Florida had attempted such an experiment. The hoax continued that, prior to his discovery by law enforcement

and subsequent detainment, transgenic orange seeds were already mailed to several recipients but whose identity could not be determined due to mail privacy laws. The generation of a transgenic citrus plant could not be confirmed and was later revealed as a hoax through another web site. Many such rumors can be found on the Internet but in the realm of science rumors have a finite life span.

Of course, direct modification of cannabis should also be considered and several different enzymes might be important to consider for transferring back into cannabis albeit with higher expression levels. For example, if tetrahydrocannabinolic acid synthase is preceded by the CaMV35S promoter it will be constitutively expressed. New genetic lines of cannabis that are loaded with tetrahydrocannabinolic acid might result. A cannabis plant overexpressing cannabidiolic acid synthase might result in more cannabidiol. Since cannabidiol has anti-inflammatory effects, the resulting cannabis smoke might be less irritating when consumed. In either case, the cannabis plant should be considered the ultimate system for genetic modification, since it so often seems that cannabis consumers are not stimulated by any single cannabinoid, but a unique blend of several dozen cannabis metabolites.

Still, finding the right combination of metabolites might be more easily obtained through biotechnological means. There are attributes of cannabis that might also be desirable to completely remove such as the phytometabolites that induce sleepiness (such as cannabidiol) or those that increase heart rate. While traditional breeding might be able to obtain cannabis lines that approach desirable strains, a faster and more accurate approach might be via biotechnology. The metabolic pathways thus far discussed hopefully serve as a starting ground for those interested in exploring genetic engineering possibilities.

How to engineer genes

There are several possible problems with using bacterial cell suspension cultures to produce cannabinoids, which might arise in one of any stage of metabolite production. For example, a person must consider the fact that cannabis genes contain non-coding intervening sequences (introns) between the coding sequences (exons). Many eukaryotes differentially express a gene, altering the combination of exons. However, bacterial species lack introns in their coding sequences and therefore lack the cellular machinery to splice genes and remove the

introns. Any eukaryotic gene that is cloned and expressed in a bacterial system will need to have introns removed prior to cloning.

Promoters, not just genes, are of huge importance to plant biotechnologists. Plants and other eukaryotes harbor a variety of promoter sequences that are positioned upstream of coding regions, which function in regulating gene expression. Selecting the correct promoter is needed to help guide the RNA polymerase machinery to begin making mRNA, which ultimately produces protein.

Gene expression depends upon the presence of the proper promoter region upstream of the newly introduced gene, but not all promoter regions are recognized equally in all organisms. For example, bacteria have promoters that only they predominantly recognize but which also can differ between species (e.g. *E. coli* and *Agrobacterium*). Similarly, the same can be said of two different plant species. Often, the degree of difference in promoter regions of the same gene is reflected in their evolutionary divergence, with more diverged species showing greater differences in promoters being recognized.

There are some promoters that are found upstream of developmental genes, only permitting expression of such genes according to an organism's developmental progression. There are also promoters for particular organs as well, which is well-recognized among plant biotechnologists working on edible vaccines. For example, carrots are often considered as a potential system to express antigenic proteins that might be able to stimulate a human immune response. Using carrot to deliver a vaccine might be beneficial for developing countries where basic medical supplies are often lacking. Plant biotechnologists, in this case, use a root-specific promoter so that the plant primarily accumulates the antigens in the consumed part of the plant. Such a method helps control the location of the protein product but also limits the energy input required by the transgenic organism. If a cannabis biotechnologist desired a protein such as a cannabinoid synthesizing enzyme to be produced in other parts of a cannabis plant besides the flowers, a leaf specific promoter could be used. Trimmings produced from a harvested plant would be much more valuable to growers.

Plant expression vectors are obtained with promoters as part of the vector. That is to say, promoters are already packaged into a vector. A researcher simply obtains a vector and inserts their gene of interest. Promoters are better understood than they were a decade ago and the whole promoter region of a native gene is not needed. Instead, only short, specific sequences in plant promoters have shown to be of high importance in directing the RNA polymerase machinery. Thus,

researchers now rely on synthetic promoters, which are promoters that are still specific to a tissue or developmental pathway, but that are much shorter in length. Synthetic promoters might also be much stronger in regards to stimulating gene expression.[24]

Generating maximum amounts of metabolic products using in vitro cell cultures or plant transformation methods both rely on first determining rate limiting steps of reactions. In a cell, reactions can only proceed as fast as a product accumulates. In other words, the concentration of a substrate is a necessary factor when considering the rate of progression of a metabolic pathway. A low concentration of the starting reactants results in little production in the ending product, regardless if the end product is three enzymatic steps later or eleven enzymatic steps later. In other words, a metabolic pathway will progress unhindered until the rate limiting step is met. All subsequent reactions will be slowed down accordingly since the rate-limiting step is similar to a bottleneck. To help engineer cannabis metabolites, the rate-limiting steps of a pathway need to be determined. By some estimates, olivetolic acid has shown to be produced in the least amount. The gene encoding olivetolic acid cyclase might, therefore, need to be ligated to a constitutive promoter and used in transgenic experiments.

Cells also are able to limit the amount of end product accumulation by either directly or indirectly sensing the amount accumulated. The mechanism, known as feedback inhibition, helps control the metabolic pathway and limits unnecessary accumulation of a particular metabolite. In some instances, a metabolic product can serve as the signaling molecule that turns the pathway off. For example, at a certain concentration the product might bind with an upstream enzyme that halts early steps in the pathway. Alterations of particular enzymes might need to be examined to avoid metabolites that serve as inhibitors so that they can no longer inhibit the pathway.

A consideration when expressing a bacterial gene in plants is that the end of the transcribed region will need to be attached with several adenines called a polyA tail. In eukaryotes the polyA tail interacts with proteins that protect the mRNA from degradation once in the cell cytoplasm. Although the polyA tail is added to an mRNA before the transcript has exited the nucleus, there must be a polyA recognition site on the mRNA. Since the recognition site is a specific series of bases, some transgenes are more likely to be incorrectly polyadenylated if the transgene contains a series of bases within its coding region that are similar to the polyA coding sequences. This is one property of the transgene, therefore, that might need to be investigated. Studies have also shown that longer polyA tails

result in a longer mRNA life span. One might therefore consider engineering transcripts involved in the cannabinoid pathway to have longer polyA tails.

Some enzymes require transport after they are translated on the ribosome. Many mRNA transcripts therefore have another region that is not essential for protein function but is essential in directing the transport of the newly synthesized protein. The region codes for a signal sequence, a sequence of amino acids that serves as a message to a carrier protein. The apparent ability of cells to direct newly synthesized proteins to a particular location is purely based on biochemical interactions between the newly synthesized protein and the carrier protein. Transport by a carrier protein might be to a different region of the organism or to an organelle.

Plant biotechnologists (and molecular biologists) also realize that not all codons in a coding region are equally used by different species. In many introductory biology classes DNA is discussed as being redundant. The redundancy of the code is apparent when one looks at a codon table. Leucine, for example, can be incorporated into a growing peptide chain if either CUU, CUC, CUA, or CUG is present on the mRNA transcript. However, on average Arabidopsis mRNA has CUG 10% of the time while *E. coli* has CUG 51% of the time (kazusa.or.jp/codon). Codon usage is important since each codon requires its own tRNA to carry the amino acid to the ribosome. Thus, not accounting for codon usage in different species might lead to insufficient levels of protein expression. Accounting for and altering codons to maximize protein production is called codon optimization.

Other considerations to facilitate the expression of newly introduced genes into cannabis include determining post-transcriptional modifications of the translated product. Post-transcriptional modifications such as attachment of sugar groups (glycosylation) or attachment of phosphates (phosphorylation) are often important in cell communication and enzyme regulation, respectively. To complicate matters, not all newly produced proteins spontaneously fold into their proper, functional state. Some require helper proteins to direct the folding in order to generate a functional protein.

Thus, one can readily see that there are several potential hurdles in generating a useful genetically modified cannabis plant. However, knowing what to avoid and where to focus attention should be viewed not as research hurdles but rather as starting points to generate novel cannabis lines. Moreover, considering the several species of transgenic plants that have been generated and successfully marketed demonstrates the feasibility of cannabis biotechnology.

Engineering metabolic pathways

Transgenic manipulations incorporating one or two genes are actually relatively straightforward. Inserting multiple genes, sometimes called gene stacking, has become improved over the last ten years. Currently, multiple gene transfer is becoming more and more common as advancements in knowledge are made.

To generate a transgenic plant with more than one gene, researchers previously performed an initial transformation starting with one gene. The resulting plant was then grown into an adult and bred for multiple generations. Only then could the transgenic plant be regrown in vitro for a second transformation in a process known as sequential transformation. Crossing two different transgenic lines has also produced plants with two transgenes, granted homozygosity is established in each transgenic line. Another method, called cotransformation, is able to deliver two different genes that are oriented next to each other on a DNA molecule. In other cases, using special high-capacity vectors, up to 200,000 base pairs of DNA have been transferred to plants. Indirect transformation methods (e.g. Agrobacterium-mediated transformation) are required in cotransformation; however, there are various vector choices, which affect the size of the transferred material.

One of the factors limiting DNA length is the physical inability of Agrobacterium to remain attached for a long enough time so that the delivery conduit via the type four secretion system (i.e. the molecular syringe) is not broken. Conjugation between two bacterial cells, which uses a similar connection system, also sometimes results in incomplete DNA transfer. If too large a piece of DNA is transferred, the stability of the newly transferred DNA is also a factor. Longer DNA sequences often show less stability and higher rates of degradation once in the plant cell. Long pieces of transferred DNA are also prone to being "kicked out" of the plant genome.

Thinking outside the box is important in pushing a new science forward. Recently, plant biotechnologists have implemented a technique called chromosome engineering that has been used in mammalian models. The technique allows one to deliver several genes at once, by mimicking the characters of natural chromosomes. The technology incorporates minichromosomes, so-called due to their capacity to carry several genes while also having features similar to natural chromosomes.[25] Early research is suggesting that the size of minichromosomes can likely increase to contain hundreds of genes, and thus the

structures are also referred to as plant artificial chromosomes. The newly delivered genetic material might be for producing different enzymes or other protein products (e.g. cell structures). Alternatively, a suite of genes on a minichromosome might encode for an entire metabolic pathway.

There are important factors to consider when implementing artificial chromosomes. Similar to *Agrobacterium*-mediated transformation, there must be a selectable marker so that transformed plants can be selected from a pool of possible transformants. However, there also must be a telomere sequence and a special site that will facilitate DNA recombination events. Linear chromosomes have several features that were previously discussed. Two important regions are the centromere where spindle fibers attach to pull sister chromatids apart (forming the kinetochore) and the telomeres.

The telomeres, as you will recall, are at the ends of the chromosomes. They act like the ends of shoestrings, keeping the double-stranded DNA together after each replication event. In plant artificial chromosomes telomeres are needed to act as chromosomal termini as well. Mini-chromosomes cannot be too small, since for unknown reasons they are lost in meiotic events once below a certain size. Conversely, mini-chromosomes that are too big have resulted in poor transfer across generations. Much work remains to be done to understand the factors of chromosome size and its relation to meiotic and mitotic stability.

In addition to the telomere region, some plant artificial chromosomes have sites that allow for downstream manipulations via a process known as site-specific recombination, similar to the Gateway vectors. This allows downstream manipulation with a known insertion site for the transferred DNA. Such an option avoids random insertion into the plant genome. Subsequently, this avoids the potential of disrupting the expression of other genes where the foreign DNA might be inserted.

Engineering plant artificial chromosomes occurs via one of two ways. The first is a top-down approach and relies on manipulation of chromosomes present in the cell, but the pieces are delivered intact. The bottom-up approach to chromosome engineering relies on combining the features of chromosomes, like centromeres (but not always telomeres), and introducing these into a cell so they can be assembled by the cell itself. The chromosomes targeted are not part of the major component of the necessary genomic DNA. Many karyotypes are presented with only A type chromosomes. In actuality, some species might have lingering bits of DNA that might act as their own chromosome, called B chromosomes. Chromosome engineering often incorporates modified B chromosomes, since they

apparently seem to be more amenable to the technique. For the moment, the early results of mini-chromosomes are exciting news to plant biotechnologists.

The ability to incorporate several genes without using Agrobacterium or biolistic methods has placed plant artificial chromosomes ahead in the race to genetically modify plants, although much still needs to be learned. Still, the plant biotechnology literature is progressing fast, with phrases such as "genome editing" and "gene design" becoming more common place. Plant artificial chromosomes are providing an exciting new reality of what plant biotechnology is capable of generating, especially in the realm of plant secondary metabolites. As research in the area of multiple-gene transfer continues, the complete cannabinoid pathway will undoubtedly have the potential to be transferred to many plant species.

Chapter 9

Bioinformatics in Cannabis Biotechnology

"The programs are designed for use in planning experiments, locating experimental errors, determining all possible sequences consistent with the experimental data and testing the reliability of the final answer...The possibility of accurate systematic manipulation and presentation of data makes attractive a much wider range of experimental techniques than are now used."
-Margaret Dayhoff, 1965[1]

Computers and biology

Bioinformatics is a hybrid discipline that manipulates biological information with computer programs. Much of the information currently being manipulated using bioinformatics is macromolecules (DNA, RNA, protein). Computers are used to store, search and retrieve, work with, and redistribute data. With the aid of computers, discoveries can be made that help researchers understand cellular and molecular processes of life. Since rapid advancements are being made in sequencing technology, computers are now indispensable.

Bioinformatics has a rich history that has resulted in the ability to compare nucleotide sequences, determine structural properties through comparison and predictive algorithms, and hypothesize protein interactions and functions. Prediction of promoter regions, genes, or regulatory elements is often useful when characterizing (annotating) genomes. Importantly, although computers do the work, humans dictate how the work is done through rules outlined in programs (software). The applications of bioinformatics are very broad and include storing data, performing sequence alignments, analysis of transcriptomic and genomic

data, constructing phylogenetic trees, predicting subcellular location of proteins, and drug design.

Bioinformatics is also made possible by sharing data and exchanging program code, which has largely been driven by the advent of the Internet over the last two decades. Perhaps this is one of the most important features of science, the sharing of data. In general, the process of science usually proceeds more rapidly with community efforts instead of one person attempting to make leaps and bounds without help and oversight by colleagues. Today, bioinformaticians and molecular biologists are able to share sequence information by submitting data to a database, which can be retrieved by another researcher. Databases that store nucleotide sequences and amino acid sequences are open for all researchers to use and access, thereby facilitating joint efforts on similar projects.

Several different databases house different types of molecular data. One of the largest bioinformatics databases in the United States is known as GenBank (ncbi.nlm.nih.gov/genbank). Established in the 1980s, GenBank is part of the National Center for Biological Information (ncbi.nlm.nih.gov). In the interest of science, data is equally shared between GenBank, the DNA Databank of Japan (ddbj.nig.ac.jp), and the European Molecular Biology Laboratory (embl.org). All databases share information on a daily basis so that they all contain the same information. The databases allow users to search for sequences or search submitted sequences against a database to determine a "best match" for a submitted sequence using various algorithms. Several thousand cannabis sequences have been submitted to GenBank and are available to the public.

Although results are sometimes rapidly generated in bioinformatics programs, good researchers should understand how the major components of the algorithms function. Database searching is seemingly straightforward; however, many hidden components are working as the search is being performed. Since the search usually happens very rapidly, a database user might not realize the steps that the program is performing.

Within the NCBI database, there are several details of the search options and results. If a user is searching to retrieve a sequence within a database one of the most common sequence formats is known as "fasta" format, a flat file format with a description line preceded with a ">" symbol, as shown in Figure 1. A nucleotide or amino acid sequence follows the description line. The sequence can range from a few dozen bases to several thousand bases.

```
>gi|449061817|sp|A6P6V9.1|CBDAS_CANSA RecName: Full=Cannabidiolic acid
synthase; AltName: Full=CBDA synthase.; Flags: Precursor
MKCSTFSFWFVCKIIFFFFSFNIQTSIANPRENFLKCFSQYIPNNATNLKLVYTQNNPLYMSVLNSTIHN
LRFTSDTTPKPLVIVTPSHVSHIQGTILCSKKVGLQIRTRSGGHDSEGMSYISQVPFVIVDLRNMRSIKI
DVHSQTAWVEAGATLGEVYYWVNEKNENLSLAAGYCPTVCAGGHFGGGGYGPLMRNYGLAADNIIDAHLV
NVHGKVLDRKSMGEDLFWALRGGGAESFGIIVAWKIRLVAVPKSTMFSVKKIMEIHELVKLVNKWQNIAY
KYDKDLLLMTHFITRNITDNQGKNKTAIHTYFSSVFLGGVDSLVDLMNKSFPELGIKKTDCRQLSWIDTI
IFYSGVVNYDTDNFNKEILLDRSAGQNGAFKIKLDYVKKPIPESVFVQILEKLYEEDIGAGMYALYPYGG
IMDEISESAIPFPHRAGILYELWYICSWEKQEDNEKHLNWIRNIYNFMTPYVSKNPRLAYLNYRDLDIGI
NDPKNPNNYTQARIWGEKYFGKNFDRLVKVKTLVDPNNFFRNEQSIPPLPRHRH
```

Figure 1. The flat file format for the amino acid sequence for cannabidiolic acid (CBDA) synthase retrieved from NCBI. Each letter represents an amino acid.

The cannabidiolic acid synthase sequence (Figure 1) was obtained from NCBI, but the database also contains 12,907 cannabis ESTs.[2] Thus, one can see that databases have become a central repository for researchers who have worked to generate data. Perhaps such ability seems odd, that we are now able to look at the string of nucleotides and characterize the functional role they might play. For a researcher to be able to recognize that the sequence file actually has meaning in the cell, brings them one step closer towards realizing the significance of bioinformatics.

Although searching databases might help obtain sequence files, databases can also serve more specialized roles. Some researchers have developed databases that focus on more specialized research such as trichomes, secondary metabolites, and siRNA. Generating smaller, more specific databases helps facilitate research goals and centralize data that is used by groups of researchers that share a common research interest.

Important to the cannabis community is the TrichOME database[3] (planttrichome.org), put together by an international collaboration of researchers from the USA, China, and France. The goal of the TrichOME database is to unify data generated from trichome studies, which have generated ESTs and unigenes, microarray data, and metabolic information acquired by mass spectrometry and assemble the data in a comprehensible, searchable form. The database also includes information for genes and proteins related to trichomes. A person can also search the literature that has been published relating to trichome. As of January 11, 2014 the databases housed 12,575 ESTs and 8,396 unigenes (unique genes) for cannabis.

Some of the nice features of the TrichOME database are the ability to mine literature published regarding important enzymes and genes in trichome development and their secreted compounds. The TrichOME database also

integrates metabolic databases and protein databases involved with trichome development and secondary metabolite synthesis.

Another important database worth mentioning that focuses on plant secondary metabolites is the PlantCyc database[4] (plantcyc.org). Metabolic databases include organisms in addition to plants. A good example of such a database is the BRaunschweig ENzyme DAtabase, or BRENDA[5] (brenda-enzymes.org). Readers are encouraged to explore these databases with cannabis-related search terms. After all, exploration is often the first step in learning.

Aligning sequences

The first step in many bioinformatics analyses is sequence alignment. The fundamentals of sequence alignment are therefore important to understand. There are two main types of sequence alignments; pairwise sequence alignment involves two sequences where one can either attempt to have local regions of sequences aligned (local alignment) or entire lengths of sequences aligned (global alignment). Comparing three or more sequences is called multiple sequence alignment.

Pairwise sequence alignment is useful when comparing nucleic acids and proteins to determine the amount of similarity or identity between them. In nucleic acids similarity and identity have the same meaning. In proteins similarity and identity have different meanings. Identity means that the amino acid (residue) from one sequence is identical (the same) as the residue in the aligned residue. However, similarity suggests the two aligned residues have the same physicochemical properties. For example, leucine and leucine are identical. However leucine and isoleucine are not identical but share similarity since they are both hydrophobic. One might encounter other amino acids that have similarity such as phenylalanine and tryptophan, which are both large, aromatic, and nonpolar.

An easy way to visualize the relationship of two sequences (or to compare a sequence to itself) is by generating a dot matrix, which is a grid containing a pattern that reflects the relationship of the sequences (Figure 2).

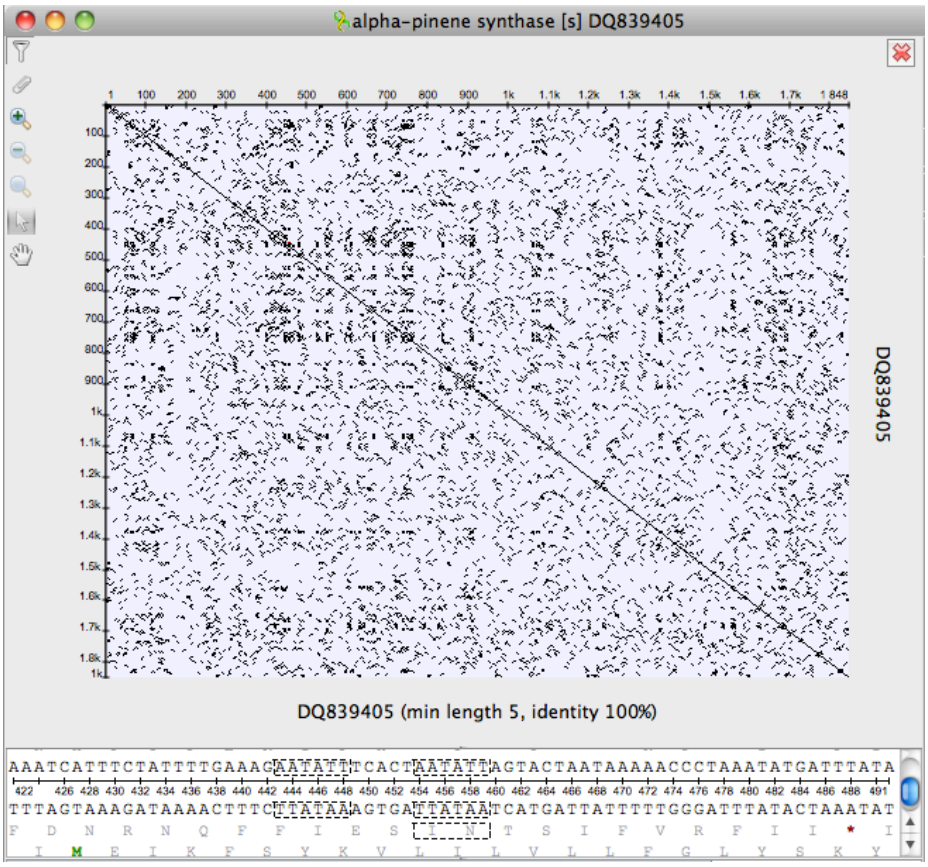


Figure 2. Dot matrix of the nucleotide sequence of alpha-pinene from cannabis generated with UGENE.[6] The gene was compared against itself with a minimum string of five nucleotides with 100% identity. Several inverted repeats are apparent from position 400 to approximately 750.

To generate a dot matrix (dot plot), one sequence is placed on the top axis, while a second sequence is placed on the adjacent (or side) axis. Any identical match is indicated with a dot. Over the course of the entire comparison a matrix is generated, which gives a good overall view of the sequence comparison. Several patterns might be observed. A diagonal line is observed if the sequences are the same, while gaps in the diagonal line might indicate gaps present in one of the two sequences.

In addition, parallel lines might be observed if there are repeats in the sequence and inversions are seen as perpendicular slashes. While providing a nice overview, increasing the length of the residue string (the "window size") will result in a better viewing resolution since much of the noise might be reduced. However, one of the drawbacks of dot plots is that the results are not quantifiable.

To quantify the relationship of two sequences, optimal mathematical alignment can be done using exhaustive methods. These methods employ a

mathematical technique known as dynamic programming. This seeks to find the most mathematically correct alignment based on a score, regardless of the biological reality. Dynamic programming might seem difficult to follow; however if you are lost at any point in this section, the Sequence Alignment Teacher[7] (melolab.org/sat) is an excellent, free program to help understand the process of scoring and filling in a matrix.

The dynamic programming method relies on establishing a matrix using a match value, a mismatch penalty, and a gap penalty. As an example suppose the following two nucleotide sequences need to be optimally aligned

TACATGT
TAAGATC

The matrix used in dynamic programming is set up with two sequences on different axes (similar to the dot matrix) and then the side values are set according to multiples of the gap cost. The procedure begins in the upper-left corner (if the sequences are set accordingly) and one row of the matrix is filled in at a time. For each diagonal, two sequences are implied to be aligned, even though they might be a mismatch. The resulting score is placed in the box. Using the match value and penalty scores above, a match is desirable but a mismatch might be more optimal since a mismatch will give a higher score than incurring a gap.

For this example we can use a match score of 2, a mismatch score of -3, and a gap penalty of -4. A horizontal or vertical move through the matrix is considered a gap (-4). A diagnol move might be either a match (2) or a mismatch (-3) but is considered to be an alignment. A completed matrix using these values is shown in Figure 3. Remember, we are looking for the optimal score.

-	-	T	A	C	A	T	G	T
-	0	-4	-8	-12	-16	-20	-24	-28
T	-4	2	-2	-6	-10	-14	-18	-22
A	-8	-2	4	0	-4	-8	-12	-16
A	-12	-6	0	1	2	-2	-6	-10
G	-16	-10	-4	-3	-2	-1	0	-4
A	-20	-14	-8	-7	-1	-5	-4	-3
T	-24	-18	-12	-11	-5	1	-3	-2
C	-28	-22	-16	-10	-9	-3	-2	-6

Figure 3. Dynamic programming for two hypothetical nucleotide sequences generated with SAT.[7] The traceback path (discussed below) is highlighted in yellow and is one optimal mathematical solution to the alignment.

The aligned sequences would then appear as

T-ACATGT
TAAGAT-C

The optimal score for this alignment is -6; however, the same optimal score might be obtained by generating other possible alignments. Additionally, (not to confuse matters more) other scoring factors can be manipulated. For example, a high gap penalty and lower mismatch penalty might be more likely to allow for incorporation of a mismatch with fewer gaps in the alignment.

The optimal alignment between the entire length of two sequences requires a traceback that begins where the two sequences end, at the bottom right of the matrix. The traceback allows a person (or the computer) to retrace the steps to find the optimal score while recording the alignment. The use of gap penalties is important in global alignment (aligning the entire length of sequences) but is not used in local alignment (aligning sections of sequences, while disregarding others).

Since gap costs do not exist in local alignment, the local alignment will give indication of where conserved domains are potentially located. Local alignment, therefore, might initiate the matrix in the top corner but insertion of a gap is set to 0. In this way the matrix of a local alignment gives the opportunity for the choosing of ending the alignment without incurring a penalty cost. Similarly,

starting the alignment can begin again without penalty. This will be reflected in the output of the final alignment (one will observe gaps).

The scoring matrix used in the DNA alignment above is

A	2			
G	-3	2		
C	-3	-3	2	
T	-3	-3	-3	2
	A	G	C	T

For proteins, the two most commonly used scoring matrices (also called similarity matrices) are the PAM and BLOSUM matrices (although others such as JTT exist). The percent accepted mutation (PAM) matrix comes in PAM1 and goes up to the PAM250. In general the larger the PAM number used the less similar the two sequences are assumed to be. The PAM matrices were developed by Margaret Dayhoff in the 1970's and were based off of a distinct set of proteins. These protein sequences were related according to mutational frequencies. All of the PAM matrices are based off of the PAM1 values. More divergent matrices are used (PAM250) for more divergent sequences (millions of years).

In contrast the BLOSUM matrices are based on observed blocks of amino acid similarities. These are conserved blocks within 2000 proteins of only a few dozen protein families. The blocks substitution matrices range from BLOSUM1 to BLOSUM100 and are different than PAM matrices. They are used to find homology between blocks of residues, not infer evolutionary relationships. A BLOSUM1 and PAM250 both use values that suggest either non-homologous or distantly related proteins.

To give an example of scoring in protein alignment, while a valine might readily be substituted for an alanine, the likelihood for a protein to substitute a glycine for a tryptophan is unlikely, and therefore this is reflected in the scoring matrix (Figure 4). Comparatively, a match of tryptophan with another tryptophan is given a large value.

A	2																		
R	-2	6																	
B	0	0	2																
D	0	-1	2	4															
C	-2	-4	-4	-5	12														
Q	0	1	1	2	-5	4													
E	0	-1	1	3	-5	2	4												
G	1	-3	0	1	-3	-1	0	5											
H	-1	2	2	1	-3	3	1	-2	6										
I	-1	-2	-2	-2	-2	-2	-2	-3	-2	5									
L	-2	-3	-3	-4	-6	-2	-3	-4	-2	2	6								
K	-1	3	1	0	-5	1	0	-2	0	-2	-3	5							
M	-1	0	-2	-3	-5	-1	-2	-3	-2	2	4	0	6						
F	-3	-4	-3	-6	-4	-5	-5	-5	-2	1	2	-5	0	9					
P	1	0	0	-1	-3	0	-1	0	0	-2	-3	-1	-2	-5	6				
S	1	0	1	0	0	-1	0	1	-1	-1	-3	0	-2	-3	1	2			
T	1	-1	0	0	-2	-1	0	0	-1	0	-2	0	-1	-3	0	1	3		
W	-6	2	-4	-7	-8	-5	-7	-7	-3	-5	-2	-3	-4	0	-6	-2	-5	17	
Y	-3	-4	-2	-4	0	-4	-4	-5	0	-1	-1	-4	-2	7	-5	-3	-3	0	10
V	0	-2	-2	-2	-2	-2	-2	-1	-2	4	2	-2	2	-1	-1	-1	0	-6	-2	4
	A	R	B	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V

Figure 4. Scoring matrix for PAM250 generated by SAT.[7] Substituting a valine (V) for an alanine (A) results in a 0 and is a better score compared to substituting glycine (G) for tryptophan (W), which is -7.

Global alignment uses the Needleman and Wunsch algorithm and preceded the local alignment algorithm established later by Smith and Waterman. Some very fundamental differences exist between the two algorithms and the result is either a much more global or localized alignment. Regardless if performing a global or local alignment, dynamic programming often serves as the basis for understanding how alignments are computationally optimized. However, one must be cautious about interpreting such alignments, since the optimal mathematical alignment might not make the most sense biologically. Finally, other methods of alignment use similar techniques of comparing bases or amino acid residues but attempt to match short strings of sequences. Remember, alignments are important to the cannabis biotechnologist because comparisons between cannabis genes and proteins to those of other plants can reveal several pieces of useful information.

Bioinformatics is often combined with research methods that generate large amounts of data. Sequence reads can number in the tens of (sometimes hundreds of) thousands, which demands the use of computers for analysis. Procedures for handling large amounts of transcriptomics data might include vector trimming and trimming of poor-quality base calls, comparing sequences against a database to find the closest matches, annotation to help characterize the function of cannabis protein products, and generating graphical output to summarize the data.

Raw cannabis sequence data might need poor-quality bases removed. These bases are seen by observing an 'N' instead of an A, T, C, or G. In some cases, the partial sequence of the cloning vector might also be detected. Ambiguous base calls and vector contamination should be removed prior to database searching. Many software programs facilitate a method known as batch processing, which allows the researcher to trim (or otherwise manipulate) unwanted bases in all sequences simultaneously. Parameters are set by the user to determine how many Ns should be removed or to enter the vector sequence that needs to be removed. Longer sequences of a set of sequence reads tend to generate better results in database searching and so some researchers will omit any sequence reads with less than a few hundred bases, for example.

Cleaned sequences can then be exported as a batch file, which is a single file containing the sequences arranged in a designated format (e.g. fasta). Several different programs are available for users to import a batch file. One of the best programs, which allows users batch blasting is Blast2GO,[8] a user-friendly, platform-independent application. Users can compare the sequences against the NCBI non-redundant database and specify the type of BLAST algorithm to use. The publication[9] that first described the basic local alignment search tool (BLAST) was the most highly cited paper in the 1990s, which indicates BLAST usability. For analysis of protein products, BLASTx can be used, which translates the nucleotides in all six frames and then searches the result of each translation. The hits are recorded as Blast2GO continues to run.

Eventually, all sequences with a best match from BLASTx hits are available to view within the program. To help characterize the protein products, BLASTx results are then searched against the gene ontology[10] database, also as a batch process. The gene ontology (GO) results are annotations based off of previously described proteins that have been, in some cases, manually inspected and usually accompany a published article. The GO database and GO project aim to use a set of controlled vocabulary to better integrate databases in order for those working on different organisms (e.g. fly, worm, plant) to avoid using different words for the same gene. Finally, the output can be given in many ways such as number of BLASTx hits, GO hits, GO categories (which can overlap), and graphical output.

Multiple sequence alignment

Researchers sometimes need to compare more than two sequences and in this case a method for multiple sequence alignment is required. Multiple sequence alignment can use exhaustive methods, similar to the DP method just explained; however, a working space of the number of dimensions is needed for the number of sequence comparisons. This demands copious amounts of computing power and is beyond the current processing speed of many computers.

Instead exploratory or heuristic approaches are used that might not find the optimal mathematical alignment but will still give reliable results. Using just sequence data an ab initio (from the beginning) comparison can be made. These comparisons are based on primary protein structure and use values established in the 1970's, established from a set of proteins available at the time.[11] There are newer values based on the GOR (named after three researchers) method, which uses a larger protein data set and can incorporate residues that flank the residues of interest.

A common method in multiple sequence alignment is to use what is called the sum of pairs (SP) method, where scores from each pairwise alignment are compared. Several multiple sequence alignment programs are currently available that incorporate SP. Two that have been in use for many years are ClustalW and DAlign. In any software performing multiple sequence alignment, sequence type is important to consider. A nucleotide residue has the chance to mutate into 3 bases while proteins are more specific in their matches. In other words, 25% of alignment can be caused by random chance with DNA, where as with proteins only 5% is attributed to random chance.

Software programs might incorporate progressive or iterative methods of alignment. The progressive alignment is also called a "greedy" method since any mistakes in alignment are transferred downstream as the multiple sequence alignment progresses. Iterative alignment allows for establishment of a tentative alignment, which can be gradually changed to optimize newer sequences as they are compared. In this way an original phylogenetic tree, for example, that is constructed can be reconfigured based on new information.

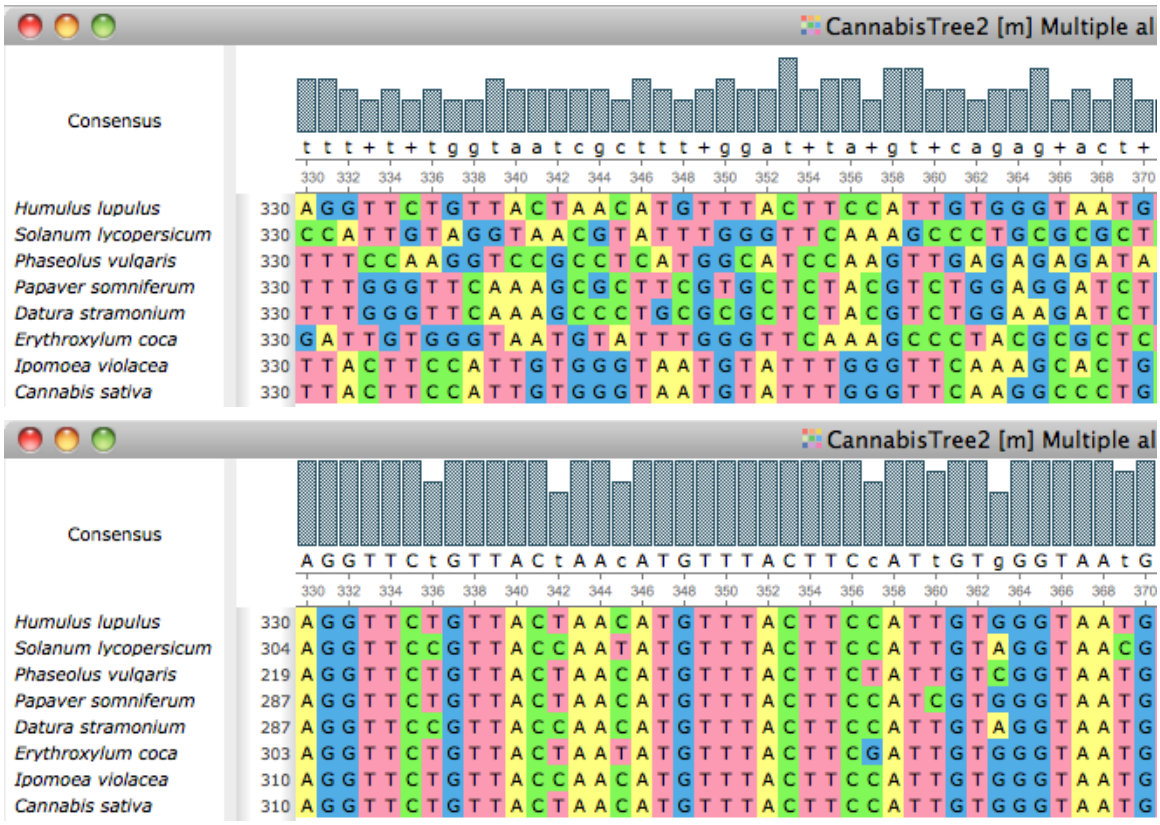


Figure 5. Multiple sequence alignment generated with UGENE.[6] Sequences here are for the gene encoding the large subunit of rubisco (*rbcL*) acquired from the NCBI database (above). After alignment (below), the consensus increases as shown by the vertical bars.

Applications for multiple sequence alignment of a set of sequences could be for downstream structural analysis (secondary and tertiary prediction), functional analysis (similar sequences could have similar function), inferring if genes are orthologs (the same gene found in different species), paralogs (two genes from a common ancestor that can be found in the same organism) or xenologs (a gene that has crossed species). Multiple sequence alignment is also necessary when generating phylogenetic trees (Figure 5). Beyond these comparative reasons, a researcher might also be interested in how a cannabis protein might change yet retain original functionality or have evolved to have a new functional role.

Database searching, revisited

A short note on database searching is necessary, since an introduction to sequence alignment methods and algorithms has been provided. A common algorithm, the BLAST algorithm, might provide relevant hits (results) that are

closely related to the query sequence but one should know how to interpret the results. Statistical significance of alignment to the BLAST database is provided and often summarized and reported as an expected likelihood value (evalue or Evalue). A smaller evalue indicates a more likely match. Various equations to determine evalues exist but a simplified one is

$$\text{Evalue} = m \times n \times P$$

The equation above[12] can be interpreted to mean that an evalue is equal to the number of residues in the database (m) times the number of residues in the query sequence (n) times the P value derived from the likelihood of finding the sequence to match with a random sequence. This latter value is determined from an extreme distribution curve generated by comparing a sequence against another sequence repetitively (>1000x) and reshuffling one of the sequences each time they are compared. A lower evalue indicates higher similarity.

Since a database changes over time (sequences are added) an evalue might increase, which might later render a particular hit unfound in subsequent searches. Stated another way, as a database or query sequence length grows the evalue changes. To help standardize the evalue and obtain a constant resulting value, a raw score that is translated as a standard score called the bit score (S´) is determined.

A point should be stressed here; since a translated nucleotide query will give more accurate results, the blastx algorithm is often used. In addition, different search algorithms have been developed that include BLAST and FASTA (FAST Alignment). Each is slightly different but they are also similar in some ways. FASTA gives one optimum output while BLAST provides multiple hits, with varying degrees of significance. FASTA also gives a z value, which can be described as the standard deviation of the hit occurring by chance. Each search algorithm starts by scanning the query sequence for short words (called ktups in FASTA) of usually 2-3 amino acids (or 6-9 nucleic acids). These are compared against the database and the sequences with the most matching words are retained for further analysis. In FASTA, a tup1 score is given and ten other sequences are also scored, giving a tup*n* value. The final score is an opt score and results in a single best hit. Additionally, FASTA is often used for proteins and BLAST is often used for nucleic acids. More might be said about databases here but perhaps energy would be best spent by encouraging readers to explore the NCBI database using cannabis-related terms.

Molecular phylogenetics

Molecular data is also rising in its importance for comparing the evolutionary relationships of organisms such as cannabis. Traditional methods of classifying organisms and determining a monophyletic group (an ancestor and all its descendants) depended on macromolecular structures such as the number and arrangement of floral whorls (sepals, anthers, etc.), leaf morphology, stem anatomy, and sometimes microstructures (phytoliths). However, problems might arise in determining monophyly with using plant phenotypic characters since some species show variation depending on temporal or environmental conditions. Such variation is known as phenotypic plasticity and is observed throughout the plant kingdom. For example, some plants such as arrowhead (*Sagittaria latifolia*) display heterophylly depending on the abundancy of water.

Since splitting a species that displays phenotypic variation into two different species is inappropriate according to the biological species concept, one might find molecular data more reliable. Molecular data has helped resolve confusion and disagreements among seemingly ambiguous relationships of several plants and also help establish time of divergences.[13] One of the main advantages in molecular data is that one can obtain thousands of characters (nucleotide bases or amino acid residues) very rapidly whereas traditional methods can be laborious and time consuming, even for one or two characters. Measuring or recording a single feature, such as seed length or leaf width, can take several weeks for a collection of specimens.

One of the most user-friendly programs available for constructing molecular phylogenies is the Molecular and Evolutionary Genetic Analysis (MEGA) software.[14] For our purposes here, constructing a phylogenetic tree to convey evolutionary relationships for cannabis and other organisms can be decomposed into five steps. The first step in constructing a phylogenetic tree using molecular data is to decide which type of data to use. (We will use the partial DNA-coding region for the large subunit of rubisco from the alignment in Figure 5.) For analyses of higher taxonomic ranks (e.g. domain, kingdom) slowly evolving nucleotide sequences are often used. The famous example, which helped establish the three domains of life as the highest rank, is the ribosomal RNA genes.

For closely related organisms, nucleotide sequences are preferred since they typically evolve at a faster rate compared to proteins. In some studies noncoding regions might be used. Similar to the nucleus, there are some noncoding sequences in regions of mitochondrial DNA. For cannabis and other plants, part of the

nuclear noncoding DNA called the internal transcribed spacer regions might be used. These regions are found within and among several nuclear ribosomal genes.

DNA also allows for detecting synonymous substitutions, when a change in base does not cause a change to the amino acid. Similarly, one can detect nonsynonymous substitutions, a change in base that does change the amino acid. Comparing rates of synonymous and nonsynonymous substitutions helps predict the type of evolutionary selection process at work.

Several types of genes, including analogs and homologs might also be compared. Two analogous genes are similar to the spines of a euphorb (not a true cactus) or spines of a cactus. While the spine of a cactus is a modified leaf, the spine of a euphorb is a modified stem. The structures are not analogous and the genetic mechanisms that control the development of both structures are not genetically related. Importantly, homologs do share a similar ancestral sequence. Homologs can be orthologs (found in two different organisms) or paralogs (sequences in the same organism that have evolved separate functions).

The second step in generating a phylogenetic tree is to perform a multiple sequence alignment. This step (like many others) is actually performed by the computer but the researcher needs to understand what the computer is doing since an incorrect alignment produces an incorrect tree. The purpose of aligning sequences in phylogenetic analysis is actually the process of stipulating the positions of individual amino acids or nucleotide bases where they are assumed to be related. There are several problems in alignment that include a position with multiple substitutions over time, mutations that have reverted to a previous state, and mutations of two different bases evolve into the same base.

Figure 6. Unrooted tree of a variety of plants, generated based on the alignment of the *rbcL* (large subunit of rubisco) gene using MEGA.[14] As expected, cannabis and hops are closely related. The approach for tree construction used maximum likelihood and incorporated the Kimura 2-parameter model. Bootstrap values are provided for each branch.

A researcher might not know what mutations have occurred in a sequence. After all, the researcher was not there through history to watch the sequence change. Therefore to determine the probabilities of each base evolving into a new base or the chances that two different bases might have evolved into the same base, models of evolution have been developed. Choosing a model of evolution is often considered to be the third step in constructing a phylogenetic tree using molecular data. Some models will use an algorithm that states all bases have an equal chance of mutating into another base (Jukes-Cantor) while other models incorporate mutation rates of the physicochemical properties of another base (Kimura 2-parameter). There are limitations to many of the current models of evolution but there are also many advantages.

Once the sequence alignment is performed, the fourth step is to determine the tree-building method. Distance-based methods assign a single score based on the difference of two sequences. The sequences with the best scores are paired together in a process called neighbor-joining.

Parsimony, maximum likelihood, and Bayesian analysis are all examples of character-based methods. There are differences in each method of analysis but the unifying theme is that they all consider each base in a nucleotide or residue of a protein as an individual character. Thus, several trees are constructed for each character. The best tree is chosen based on similarities among branches of all the tress produced, although more than one tree might also be obtained as the best tree. Character-based methods have shown superiority in their performance and as computers have improved in their speed, character-based methods are becoming more predominant.

The final step in constructing a phylogenetic tree is to assess tree reliability. One method, called bootstrap analysis, is a type of resampling technique that uses the original data to generate a new alignment and construct new trees from that alignment. The trees are compared to the first set of trees, which will either match or not match. If the trees match then they are recorded as such. In bootstrap analysis, new alignments and new trees are made for 1000 replicates and the frequency of each branch that was supported is given as a percent. The numbers above a branch in a phylogenetic tree often represent the bootstrap value where a higher value suggests stronger support for the relationship.

Ranks are used in biology to help organize organisms into categories, where several levels of categories are used (Table 1). Latin names are provided for all species and are referred to as a binomial, consisting of a genus (generic epithet) and a species (specific epithet). Sometimes people will also specify the authority,

the person who was first to describe the species. The authority's name is usually abbreviated. For example, since Linnaeus (L.) was first person to describe cannabis (at least under his binomial system), one might see the species name as *Cannabis sativa* L. The two-word name is always italicized or underlined in order to honor the Latin. When abbreviated, the name should appear as *C. sativa*. Exceptions exist for species that have taken on their generic epithet as their common name. Hence, Agrobacterium and Arabidopsis are not always in italics.

Rank	Example Organisms		
Domain	**Eukarya**	**Eukarya**	**Eubacteria**
Kingdom	**Plantae**	**Animalia**	**Bacteria**
Division/Phylum	**Eudicots**	**Chordata**	**Proteobacteria**
Class	**Rosids**	**Mammalia**	**α-proteobacteria**
Order	**Rosales**	**Primates**	**Rhizobiales**
Family	**Cannabaceae**	**Homindae**	**Rhizobiaceae**
Genus	***Cannabis***	***Homo***	***Agrobacterium***
Species	***C. sativa***	***H. sapiens***	***A. tumefaciens***

Table 1. Taxonomic ranks for two eukaryotes (cannabis and humans) and one prokaryote (Agrobacterium). Unranked groups are shown in orange.

One of the ranks commonly taught in Plant Systematics courses is family. Cannabis is in the family known as Cannabaceae. All plant families end with the suffix "aceae" which is derived from Latin that can be loosely translated as "from the nature of." The Cannabaceae is a widely distributed family and can be found in temperate and tropical ecosystems. The family consists of 11 genera with about 80 species.[15]

Researchers form China, the Netherlands, and Iran have recently collaborated in using chloroplast DNA to generate a molecular phylogenetic analysis of cannabaceae.[16] Molecular methods have suggested that hackberries (*Celtis* spp.), which include several common trees, are placed within the cannabaceae. Another member of the cannabaceae includes hops, *Humulus lupulinus*. Since hackberries and hops are closely related to cannabis one might hypothesize that their metabolic profiles are similar. Hops is known to synthesize several interesting secondary metabolites (e.g. terpenes) that closely resemble or match those found in cannabis.[17]

Understanding relationships of cannabis to other plants in the plant kingdom is crucial to appreciating the natural history but can also lead to new discoveries of metabolic pathways. Cannabaceae members that have not yet been investigated for their metabolic profiles could offer a source for new therapeutic metabolites. Closely related species might also harbor genes with shared ancestry to cannabis genes that might have useful features to the cannabis biotechnologist.

Protein databases

While many different databases exist that have nucleotide sequences, other databases contain information on proteins. In addition to sequence information, protein databases also have structural information as well. Often attributed to being the founder of bioinformatics, Margaret Dayhoff was the first to collect and implement the use of her own protein database to study the relationship of proteins. Not many years after Margaret Dayhoff's protein database was established, the Brookhaven National Lab established the Protein Data Bank[18] (PDB), which has become an important protein database (rcsb.org/pdb). The PDB stores more than 96,000 structures and allows users to deposit and download structural data.

Bioinformatics analysis of a protein might incorporate structural data for evolutionary or functional comparisons. Protein databases also might contain hierarchical information on a protein structure. For example, this might entail using CATH (class, architecture, topology, homology) or SCOP (structural comparison of proteins). CATH (cathdb.info) and SCOP (scop.bic.nus.edu.sg) are similar in that they both use a hierarchical classification scheme. For example, CATH[19] organizes proteins by

C- class - the highest rank, includes all alpha helices, all beta strands, neither alpha or beta strands or other secondary structure features

A- architecture - the folding of secondary structures on themselves

T- topology - includes how folded secondary structures interact

H- homology - homology of the sequence to determine evolutionary relationships

There are several specialized protein databases, many of which are integrated with transcriptome databases. In other cases, structural databases contain links to additional protein or related databases. As examples, provided here are three important databases with some of the main goals of the database described:

The Gene Ontology Database[10] (geneontology.org) provides a set of rules and controlled vocabulary (i.e. terms) for describing gene products (not just proteins). Three large ontologies group gene products based on the biological process, cellular component, or molecular function. Hundreds of thousands of gene products are categorized in each ontology with more than 700,000 gene products. To help characterize a gene product, terms are arranged in a hierarchical manner called parent and child terms, but can also be interrelated across ranks. The goal is to unify the vocabulary used to describe a gene product that might occur in different species in order to improve standardization in bioinformatics and across biologists.

InterPro[20] (ebi.ac.uk/interpro) is a protein database that classifies proteins and also integrates several other databases into protein searches. InterPro seeks to characterize proteins based on algorithms that help predict domains and other sites within the protein that might be of structural importance such as active sites. Information for cannabinoid receptors can be found through InterPro using their search feature.

The KEGG Pathway Database[21] (genome.jp/kegg/pathway.html) stands for the Kyoto Encyclopedia of Genes and Genomes. The database provides excellent map-based views of metabolic pathways, including secondary metabolites. The database also provides information relating to the central dogma, drug development, and several other pieces of information that relate to molecular interactions and biochemical reactions. Maps for endocannabinoids can be found using the search function.

Two additional databases of huge importance are the Molecular Modeling Database[22] (MMDB) or the Protein Data Bank[18] (PDB). Files containing information for protein structures can be obtained from either of these web sites. Since some people prefer to study protein structure by manipulating a protein in three-dimensional space, several software programs are available to read the

MMDB or PDB files. Cn3D[23] (pronounced "see in 3-D") is a software program that is easy to use and has several options for viewing proteins. The file type recognized is .cn3. Users can view the sequence of different domains and select individual amino acid residues or nucleotide bases. Another great bioinformatics program is UGENE,[6] which has several additional features besides being able to view proteins. Some of the tools can be used for sequence alignments, translating DNA, viewing chromatograms, assembling sequence reads, constructing phylogenies, and analyzing potential restriction digest sites for a sequence. Perhaps one of the best features of both Cn3D and UGENE is that they are freely available to anyone.

Although one might hope for information to be available for cannabis proteins, the selection at the moment is rather slim. However, in 2012 the crystal structure for THCA synthase (THCAS) was determined by Yoshinari Shoyama and colleagues.[24] As discussed previously, THCAS is important for converting cannabigerolic acid into THCA, which is then decarboxylated upon heating and is recognized as one of the more "psychoactive" components (THC) of cannabis. Shoyama and his group used X-ray crystallography to help characterize the two domains of the enzyme (Figure 7 & 8). The cyclooxidation reaction that THCA synthase carries out requires the presence of the coenzyme flavin adenine dinucleotide (FAD). Shoyama and his team determined that FAD attaches to a histidine at amino acid position 114 and a cysteine at position 176. The active site was also characterized as involving a tyrosine at position 484; however, the researchers recognized that other amino acids might also interact with cannabigerolic acid.

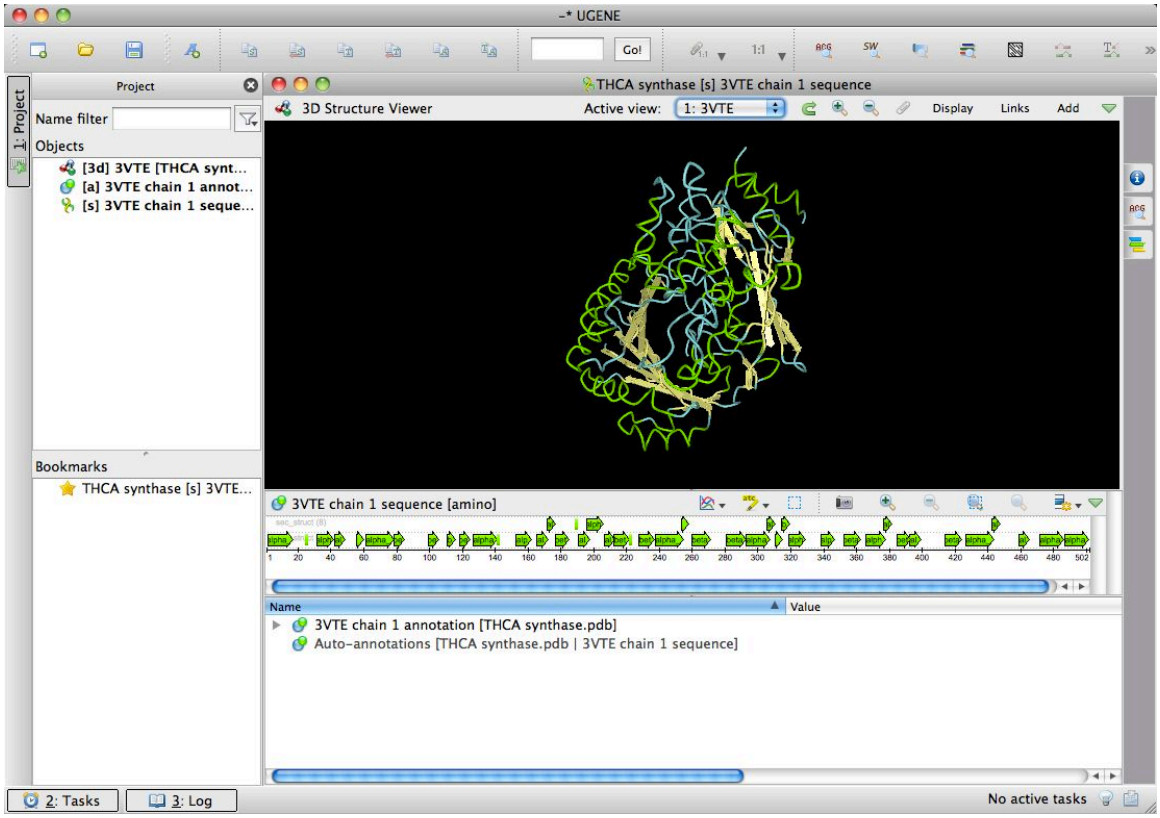


Figure 7. Screen shot of UGENE,[6] a freely available multi-purpose bioinformatics tool. Shown is THCA synthase, secondary structures as indicated by green arrows, and additional detail about the protein that can be viewed.

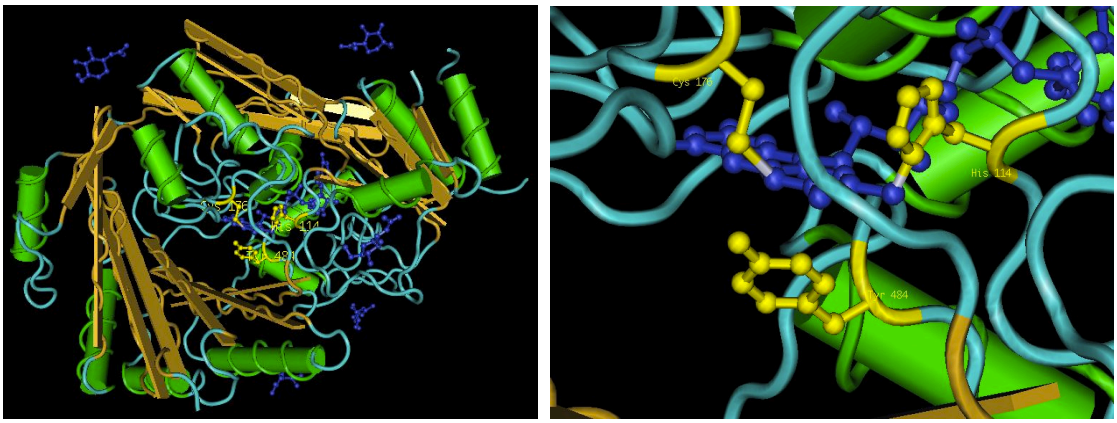


Figure 8. THCA synthase using Cn3D[23] based off work by Shoyama and colleagues.[23] The FAD coenzyme within the active site (center, left) is highlighted in dark blue. Smaller heterogens are also shown in dark blue. A closer view (right) shows FAD attached to histidine and cysteine. The tyrosine (Tyr 484) that interacts with cannabigerolic acid is highlighted in yellow.

If little structural information existed, a researcher might also use a "threading" technique to help determine how a protein might behave. More information can be gleaned through homology modeling techniques such as scoring the C-alpha backbone distances, residue positions, considering the lowest energy, or solvation tendencies of a residue. There are also methods to predict how a primary protein structure might fold. For example, one might use a hydrophobicity plot (isoleucine and leucine are highly hydrophobic) based on the Kyte-Doolittle, a hydrophilicity plot (residues such as lysine and arginine are hydrophilic) based on the Hopps-Woods scale (for looking for antigenic regions), or determine transmembrane domains (TMpred or HDhtm). Several tools might be important when considering cis- or transgenic expression of a protein in the cytoplasm of a cannabis cell and so several modeling approaches might be used.

An important application of in silico protein manipulation is drug design and docking prediction. Some software programs provide atomic-level detail and changing of protein side groups at distinct positions in the protein backbone. Altering amino acids might produce closer bonds or help determine how a drug might interact with an enzyme active site. Drug design methods aim to increase the speed and accuracy (thereby reducing the workload) for finding drug targets. This is important since the average time to bring a drug to market takes nearly 13 years and almost two billion dollars. The Swiss Institute for Bioinformatics provides a web site called Click2Drugs[25] (click2drug.org/directory_Docking.html), which has 781 links to software useful in drug design such as homology modeling, binding site prediction, docking, and ligand design. As with any software, one needs only to explore to begin the process of self-education.

Synthetic and systems biology

The rise in sequencing throughput has led to the need for manipulating large amounts of data. Bioinformatics has helped organize and dissect large amounts of information into meaningful, comprehendible results. Thus, computers are now indispensable in molecular biology. Bioinformatics has also helped stimulate new areas of research that have the potential to take cannabis biotechnology one step further than simply generating transgenic plants. Synthetic biology and systems biology are two examples of new fields that are set to revolutionize how biologists visualize, implement, and interpret research. Perhaps their capabilities will be applied to cannabis in the near future.

Synthetic biology includes artificial synthesis of genes. Synthesized genes generated by a machine can be tailored made to account for codon usage among species. As previously described, one of the problems that can occur in transgenics is that there is a difference in codon usage. That is, multiple codons might code for the same amino acid, but not all codons are used in equal frequency among species. Therefore, if a person desired to express a human gene in cannabis, the codons might need to be modified via codon optimization. This process might increase the amount of protein being produced.

Codon optimization is brought up again here because gene synthesis techniques form a major part of synthetic biology. Gene synthesis is increasingly utilized by industry and other research entities. Synthesized pieces of DNA that are up to 3,000 bases can be generated. Joining smaller pieces has resulted in DNA molecules that are 10,000 bases. One of the more impressive results of synthetic biology is the de novo synthesis of an entire bacterial genome. However, such an experiment has not been performed for a plant. Perhaps cannabis will be the first.

Synthetic biology has been used for more than a decade to alter cellular processes but is becoming increasingly common for manipulating plant metabolism.[26] However, more researchers seem to prefer simply to order a desired gene that has been synthesized over traditional PCR amplification in order to save time and guarantee that the gene will be obtained.

Several groups are now participating together in a global effort to generate "parts" that can be used in protein engineering (parts.igem.org). Essentially, researchers have streamlined the process of cloning individual genes (although traditional cloning methods are still routinely performed in many labs). Since biological pathways, often called networks in systems biology, consist of several parts (e.g. promoters, coding sequence, terminators), researchers have made parts available to add to a network being constructed. Users can select from an array of different parts in a catalog. The building of cloning constructs and linking of genes is a reflection of the methods of cloning and determining complimentary regions that can be ligated together. However, synthetic biology streamlines the effort.

Synthetic biology uses molecular toolboxes, where the so-called parts can be found. GoldenBraid[27] (gbcloning.org) is a cloning strategy that allows different parts to be pieced together and interchanged among researchers. The method also incorporates a binary vector system that is compatible with all parts, which makes the cloning steps simple and uniform. The GoldenBraid system facilitates research questions by providing the engineering tools for investigating protein-protein

interactions, expression assays, gene silencing, and multigene cloning. Building insert-vector constructs is made much easier using synthetic biology approaches.

While synthetic biology has generated excitement in the plant biotechnology community, systems biology is also generating much-deserved attention. One of the other goals of modern day molecular biology is to understand a cell from a whole systems approach, which provides an understanding of how each gene or signaling event is related. Thus, systems biology seeks to analyze the complete network of a cell. A network can consist of several genes and/or enzymatic modules working together that result in a cellular process or attribute.

Systems biology also seeks to understand entire metabolic pathways and their relationships. This includes metabolic pathways that are important to cell function, cell networks, and communication systems. Knowledge of the pathways gives a global view of metabolic processes. Thus, multiple layers of the metabolic networks that are intertwined are simultaneously analyzed. Although using systems biology to help determine metabolic circuitry is rather new, the process would not be possible without the aid of computers and advanced bioinformatics tools.

Researchers working in systems biology often use methods of disruption of cell processes to study the effects on downstream events and reprogram cellular pathways. That is to say, information of the molecular components of the cell system is gleaned by studying the fluctuations of cell physiology and chemistry by altering one or more factors of the cell itself or the cell environment. Thus, systems biology is a large scale approach to attempt to understand genome-wide signals and provide detailed descriptions of those processes. Similar to synthetic biology, several software programs are available to help in systems biology research. OptFlux[28] is perhaps one of the best freely available programs for those interested in metabolic engineering.

In summary, several new areas within biology are leading towards generating novel plants. Synthetic biology is allowing researchers to develop new pathways or perturb existing pathways. Complimentary to synthetic biology, systems biology provides the opportunity to explore the effects of perturbing a network to study cell-wide influences. Integrating cannabis into these scientific fields will certainly lead to a better understanding of the cannabis cell and might lead to the production of superior cannabis plants. Although there are several existing and emerging tools to generate new lines of cannabis, perhaps some of the potential outcomes of genetically enhanced cannabis should be considered.

Chapter 10

Addressing Questions from Cannabis Consumers

"We mustn't be afraid of inventing anything...Everything there is in us exists in nature. After all, we're part of nature. If it resembles nature, that's fine. If it doesn't, what of it?"
Françoise Gilot, Life with Picasso, 1964

The science of cannabis biotechnology

Lack of understanding has led to many harsh campaigns against cannabis but our understanding has greatly improved over the years. Most people now realize that cannabis is safe. Companies are rushing to develop new strains of "green gold" while also delivering better medicines to replace manufactured pharmaceuticals. We now know that the physiological influence of cannabis can, at least in part, be attributed to cannabinoids. By sheer chance, they are able to bind with receptor proteins on the cell exterior, which sets off a signaling cascade on the inside of the cell. One might suggest that the chance plants could make such products is slim but considering the diversity of the plant kingdom, such a chance of cross-reactivity is rather high.

Many people that have brought scientific principles to the general public have consumed cannabis, either openly or secretly. Popularizing science is not always easy among the daily rush of people. Carl Sagan, a populizer of science and a leader in the field of astrobiology, was thought to have used cannabis. He helped several people realize other habitable planets might surround our "pale blue dot" called Earth. Other people including painters, dentists, writers, philosophers, parents, researchers, clergy members, college students, actors, grandparents,

politicians, lawyers, mountain climbers, musicians, garbage collectors, doctors, surfers, farmers, chefs, and health care workers also use cannabis. Not to worry, the scientific data is in. Cannabis is safer for recreational and therapeutic uses compared to other drugs. But is cannabis biotechnology safe?

A broad look shows that people have questioned the safety of plant biotechnology. Questioning the safety is important since plant molecular biology has not only allowed for generating transgenic plants but transgenic plants are also playing an increasingly important role in society. Recent advancements in techniques and tools have only sped up the process of bringing the idea of transgenic cannabis to reality. In addition to the potential advancements in cannabis biotechnology, several other plants that have been modified through molecular biological techniques are playing a larger role in the food and fiber supply.[1] According to the International Service for the Acquisition of Agri-Biotech Applications (isaaa.org), more than 15 million farmers in about 30 countries (most of which are developing countries) have adopted biotech crops.[2] Since the first biotech crops were adopted in 1996, biotech crops have increased in popularity and in 2012 170 million hectares were planted. With such large numbers of farmers supporting biotechnology and dedicating agricultural area to biotech crops, an increase in public curiosity and awareness is not surprising.

Although corporations do make money from biotechnology-derived products, most plant biotechnologists are not profit-driven, nor are they careless or negligent in their work. They are dedicated scientists and are well aware of the opportunities their skill set has to offer. A large part of being a scientist is not much more than meeting repeated failures while realizing anything is possible with the human mind, dedication, and imagination. A successful scientist might not always make discoveries that change the world, but is still able to incite interest in nature among their colleagues and those that will listen. One should consider that a cannabis biotechnologist would likely have the best interest of humanity in mind when genetically modifying a plant and also be in tune with nature. Thus, they are likely to be aware of the need for human safety and sympathetic to the concerns of the natural environment.

Plant biotechnologists are also well-trained individuals. They work in a field that requires reproducibility of results, which is what makes science as a whole so useful, its predictability. Of course scientists who generate precise genetic modifications have differing positions on their research products. Not only do opinions differ among scientists, but also across time scientists are likely to change

their attitudes and feelings on a subject. However, many plant biotechnologists hope their work simply results in an increase in choice for consumers.

Biologists are very much aware of what the genetic code does and is capable of doing. Although there are more peculiarities arising with time, such as the function of RNA on gene expression, DNA as a macromolecule has been characterized to a level of impressive comprehension. Genetic engineering is also well understood, a fact that is apparent when one considers the amount of technology invested, applied intellectual energy, and number of organisms that have been modified over the last two decades.

However, just because humans have a technology does not mean that the technology must be used. One job of scientists is to educate the public on the importance and relevance of a technology so that while not everyone directly uses the technology, the core principles are understood. If particular questions on a new technology are not well explained, the questions morph into unwarranted fears. Thus, several papers have been published regarding the concerns of plant biotechnology safety.[3] Proponents have argued that most of the worries surrounding plant biotechnology are due to poor information and misconceptions.[4]

Yet others are so opposed to genetically modified crops that they have formed campaigns, sometimes going so far as burning test plots of transgenic crops in the middle of the night. In one such demonstration against genetically modified crops, 40 tons of sugar beets were burned in the northwest USA.[5] The arsonists were quick in their work and fled before being found. Although burning of genetically modified crops is extreme, such a tactic against a plot of transgenic cannabis might not allow arsonists to sneak off in such a hurry!

One of the leading arguments for the need of agricultural biotechnology is the rising human population and the increased demand for food. Plant biotechnology might offer a way to feed the nine billion humans that are expected to inhabit the planet in the next few decades.[6] Still, worries and concerns exist. Thus, the major concerns of cannabis biotechnology must be addressed.

Ahh, ahh, allergies

Transgenes that confer some advantage or trait in a plant can, in some cases, be consumed in the food crop in which they are placed. People have argued that merely by eating a transgenic crop they could be harmed, for example, by

developing allergies to the protein product of the introduced gene. An interesting case involved the accidental incorporation of a genetically modified maize variety that made its way into the human food supply.[7] The particular protein of concern is known as the Bt protein. To some insects such as the European corn borer (*Ostrinia nubilalis*), the protein is a toxin.

The Bt protein is derived from the bacterium *Bacillus thuringiensis*, which protects itself by making the protein. The protein binds to a cell receptor on the insect gut and inhibits the insect from carrying out digestive processes. Therefore, maize plants with the gene for Bt protein are less likely to be fed on by corn borer and other insect larvae. Thus, Bt maize (called StarLink) reduces pesticide and petrochemical use. (Maize, *Zea mays*, is sometimes called corn in certain parts of the world).

Somehow, along the circuitous path of food production, Bt maize became mixed with non-Bt maize. Such a probability seems inevitable considering Bt maize looks identical to maize that does not express the Bt gene. Although the Bt maize should have been destined to cattle feedlots, the mixed types of maize were used to make taco shells and other products for human consumption.

In the year 2000 the story about StarLink corn was made public and millions of taco shells had to be recalled from stores. The concern was that since certain taco shells were made with corn containing Bt protein, people who ate the Bt protein might have their digestive systems somehow compromised. The company (Aventis) behind Bt corn tried to allay fears by explaining that the Bt toxin is specific for a protein receptor only found in the corn borer gut. Molecular biologists from Aventis further explained that humans lack the receptor targeted by the Bt protein, which was the precise reason why the Bt protein was chosen. Nonetheless, the finding prompted public concern and recalls in taco shells that cost millions.

While the case above pertains to maize, a similar situation might happen with transgenic cannabis. How are people to know if their seed supply contains foreign genes or not? What proteins or other macromolecules might a person be consuming without their knowledge? More importantly, could cannabinoids that are expressed in another plant somehow become mixed into the food supply and cause unintended harm? Could genetically modified maize saturated with cannabinoids become mixed with unmodified maize?

Many cannabis consumers might not be too concerned with cannabis metabolites entering the food supply, some might actually welcome such a situation. However, the same might not be true of people who are allergic to

cannabis. There are a small percentage of people who generate an immune response from exposure to cannabis smoke or the plant itself. In an interesting study, two workers at a forensic lab showed an allergic reaction to cannabis.[8] Upon further analysis researchers determined the cannabis allergy was an IgE type, similar to the response seen in people suffering from hay fever.

Unintentional ingestion of cannabis might lead to an immune response and might require medical attention for some people. Certainly such a scenario would be largely avoidable with cannabis. If people are allergic to cannabis they simply avoid consumption. However if cannabis metabolites are produced in food crops, unintentional ingestion might occur. Considering the fact that maize carrying the Bt toxin has previously entered the human food supply, a scenario where cannabis metabolites engineered into food crops and later enter the food supply is not unreasonable.

An obvious solution is to carefully keep track of genetically modified cannabis or food crops with cannabis metabolites. A situation such as the StarLink scenario of 2000 is now less likely to happen since agricultural producers are more aware of plant biotechnology. That is to say, most people who grow crops are now aware of how two seeds might appear the same yet carry introduced DNA. Cannabis growers are also becoming more aware of the characteristics of plant biotechnology.

However, cannabis biotechnology might help those who are allergic to cannabis consume the plant without developing an immune response. Determining the causative agent of the immune response might lead to the engineering of hypoallergenic cannabis, knockout strains that lack allergy-inducing properties. Pets like cats and dogs are already being genetically modified to decrease the amount of dander they produce, thereby allowing those allergic to cats or dogs to own them as pets. Perhaps we are reaching a point where those with cannabis allergies will demand genetically engineered varieties that they are able to consume.

Genetic pollution

Genetic pollution is the opinion that suggests an introduced trait into a plant could, given enough time, find its way into natural (wild) populations. The trait might then be propagated in wild populations and cause unintended results. While some have argued that genes do not pollute the environment any more than

whole genomes can pollute the environment, others contend that an escaped gene might damage entire ecosystems.

In 1990, an article was published in the popular journal *Nature* that, without a doubt, showed major environmental destruction.[9] The Bt protein was to make its first major debut on scientific news headlines around the world. Top-level researchers showed that when monarch caterpillars consumed pollen from a maize plant producing Bt protein, the monarch butterflies died. Although the study was performed in a lab setting, the researchers hypothesized that the pollen from maize plants in the field could be dispersed by wind onto adjacent milkweed plants. Since monarch larvae feed on milkweeds, they would likely be consuming maize pollen loaded with Bt.

Alarm was sent through the plant biotechnology and ecology circles that were now faced with considering such horrible, unintended effects of transgenic maize. As panic quickly spread, other researchers began criticizing the original study. Some people realized that the monarch butterflies in the study were fed amounts of pollen that would never occur in nature. Moreover, the amount of pollen fed to the monarchs in the original study would need to be systematically collected and nearly force-fed to captive caterpillars in order to obtain the same results. That is to say, the researchers, not the pollen, killed the monarch caterpillars in the original study. As more scientific articles were published news headlines related the findings suggesting that the Bt protein in pollen was only present in miniscule amounts and therefore would, under normal circumstances, pose no threat to larve.[10] However, the story of the original publication seemed to have pushed some peoples' fears beyond recovery.

Those opposed to transgenic plants continue to use the monarch butterfly article as part of their arsenal against the development and release of transgenic plants. Some opponents of genetically modified crops rightly contend that, while the monarch study was a poorly designed scientific experiment, there is still a concern for environmental pollution. Might a newly introduced gene have unintended effects that are ecosystem-wide? Or could an insect-repelling protein be incorporated into the cannabis genome as a protective measure (Figure 1)? We now know that gene transfer happens all the time in nature. Genes are being transferred between eukaryotes and between bacteria, even between bacteria and plant in the case of Agrobacterium. As with all areas of life, a probability value can be associated with any event. Minimizing the probability is one of the goals of plant biotechnology.

Figure 1. Might cannabis plants be genetically engineered to have increased resistance to abiotic and biotic stress? If so, insect damage as shown here might become rare.

One of the major concerns with transgenic plants is the incorporation of the selectable marker. Any trait that is transferred to a plant is associated with a selectable marker such as an antibiotic resistance gene. Some are concerned that the antibiotic resistance gene might somehow spread into the environment and increase the potential for the evolution of antibiotic-resistant bacteria. However, systems have been developed that do not require the use of an antibiotic resistance gene. In some cases, the selectable marker can be for synthesis of a particular sugar or nutrient that is deficient in the plant tissue culture media. Only transgenic plants that are able to grow on the nutrient-deficient media therefore have the gene of interest.

Systems also exist to completely remove the selectable marker from the plant.[11] Recombinase systems are sometimes used in this approach. After selecting for transgenic plants, a recombinase enzyme is induced to snip out the selectable marker gene. The enzyme cuts at recognition sites that flank the selectable marker gene. Subsequently, the genome loses the DNA region between the recognized flanking sites.

Other methods of generating plants with less foreign DNA have been proposed, one of which simply discards the use of a selectable marker or the reliance on a recombinase system. Such an approach requires regenerating all plants after a potential transformation period and then later selecting for only

transgenic plants using genetic detection methods. The work is much more laborious, especially in transgenic systems with low transformation rates.

Evolution of super weed

Some plants are engineered to be resistant to herbicides. An herbicide resistance trait could be spread to a second species that is closely related to the transgenic species. The spread, via pollen, might result in a hybrid weed that is resistant to the pesticide. For example, if a gene for herbicide resistance were put into sunflowers (often grown for their seed oil) the pollen would have the capacity to carry the transgene into wild sunflowers or close relatives. Resulting seeds from the pollen might also carry the herbicide resistance gene, as would the young plants that germinated from the seeds. Subsequent crops that were sprayed with the herbicide would still be resistant but so might the surrounding populations of weeds.

An example of herbicide resistance that has developed in plants is glyphosate (Roundup). Glyphosate inhibits aromatic amino acid synthesis in plants and is one of the best-studied herbicides. Many crops are currently grown that are resistant to glyphosate, allowing farmers to spray fields without harming the crop plants. Although glyphosate resistance has been found in some weed species such as horseweed (*Conyza canadensis*), this has been attributed to the fact that horseweed developed resistance via natural selection, not from gene transfer.[12] Such resistance is simply a product of evolutionary selection, where plants with slightly different structural properties of the target enzyme of glyphosate might have an advantage over other plants. Some plants also export herbicides and other toxins out of their cells via membrane transport proteins and such a method for glyphosate has also been shown to occur in plants.[13]

Techniques to limit the spread of herbicide resistance genes via pollen has been studied. In this respect, chloroplasts have become good candidates for transgenes due to a variety of reasons. Since the chloroplast is maternally inherited in most angiosperms, pollen lacks the chloroplasts. Therefore, pollen from a transgenic plant is absent of any transgene. Thus, the potential of genetic pollution is completely avoided. Plants with engineered chloroplasts also often produce high levels of protein product due to the number of chloroplasts present. Additionally, since pollen and seeds are two primary vectors for gene flow, cannabis plants that lack the ability to develop mature anthers or produce pollen can be engineered.

One potential technology that has not been explored in the scientific literature is generating a transgenic line that is only compatible with itself. Many plants rely on a particular combination of S alleles for successful fertilization. The S alleles encode surface recognition proteins on pollen and the stigmatic surface that triggers the pollen to germinate and grow a pollen tube, allowing cell nuclei to fuse (fertilization). Preventing germination or pollen tube formation would generate a reproductive barrier for plants and limit the spread of transgenes. Perhaps a distinct combination of S alleles for transgenic lines could be generated so that only transgenic pollen from a male can pollinate a transgenic female. Perhaps pollen or seed production should be eliminated altogether.

Targeting the genes responsible for seed production has already been investigated. An interesting technology has been proposed is called Genetic Use Restriction Technologies (GURTs), where seeds can be bought but the resulting adult plant lacks the capacity to develop viable seeds. In other words, a cannabis plant with GURT would make seeds but the seeds would not germinate. GURTs rely on three genes to produce a toxin, a recombinase enzyme, and a small protein known as a repressor. The toxin that kills the embryo of the seed from the adult plant is only produced when the recombinase cuts out sequences in the toxin promoter. The recombinase is produced when tetracycline (an antibiotic) is applied. Any seeds incorporating the GURTs can be treated with tetracycline, sold, and subsequently planted.

Even if pollen or seed production is eliminated, a chance still exists that transgenic plants can enter the environment. For example, if a cannabis plant that lacks the ability to form seeds or pollen is genetically modified to be a hundred times more vigorous in its growth, an escaped plant might literally take over ecosystems, perhaps even the world. Obviously, laws might be needed so that growers of genetically modified cannabis must control vigorous plants or face harsh penalties. Then again, who would complain about cannabis plants growing over the entire earth?

Decreased choice in variety

When transgenic cannabis seed comes to market, growers will have many more choices of strains to grow. Genetic modification of cannabis will inevitably lead to an increase in variety since more genetic traits are being introduced into the population. Generating more genetic variety results in an increase in

phenotypic variety. Genetic variety decreases based on a decrease in choice on what varieties to cultivate. Thus, stifling cannabis biotechnology and hindering the research efforts to generate genetically modified cannabis leads to a decrease in varieties. Varieties arise from the breeding of new traits, regardless if they are derived with current biotechnology or from old methods of artificial selection.

Almost all gardeners love variety, not just cannabis growers. Everyone seems to prefer a particular strain, different from another. People, once aware of the potentials of cannabis biotechnology, might even begin to demand novel strains. Those who prefer not to try cannabis with enhanced traits, have the choice of deciding. But how should a cannabis consumer know if a particular cannabis variety has been modified through biotechnology or not? In general, biotechnology companies might need to communicate with growers who in turn might be responsible in relaying that their strains have either been modified in the lab or not. At the same time, companies might prefer to not sell genetically modified seeds and simply sell their products directly to consumers.

Growers might choose to cultivate genetically modified strains of cannabis themselves, perhaps when modified cannabis plants show more vigorous growth compared to non-transgenic varieties. Generating a more vigorous plant is a major goal of plant biotechnology. Currently, growers often prefer to cultivate more vigorous strains. Since biotechnology might lead to more vigorous strains that outperform non-transgenic strains, a person would likely rather grow a genetically modified strain. Whether they are indoor cultivars or outdoor cultivars, most growers prefer a more vigorous plant.

Modification of one gene does not lead us towards making a cannabis plant with unknown properties since we have technology to study the entire transcriptome. With cannabis biotechnology, we are generating plants that are better understood, biochemically and genetically. With more precise tools to modify cannabis, we have the ability to generate strains that pose less harm and more benefit.

Perhaps generating genetically modified cannabis is less of a concern than the patenting of new plant varieties or cannabis genes. How would genetically engineered cannabis be regulated? Regulatory processes are already in place for genetically modified crop plants like corn and cotton. Genetically modified cannabis might follow similar guidelines in regulation.

Might we begin to see cannabis genes patented? A gene can only be patented after someone shows and adequately describes the application of the gene.[14] People or companies cannot simply provide a sequence of DNA and expect

a patent. Further, the US Patent Office clearly states that discoveries are not patentable but inventions are. Therefore, sequencing a genome and mining it for "valuable" genes is no more worthy of patenting than a plant biotech company sequencing a cannabis gene.

At the present time, applications for cannabis gene patents can easily be found on through a patent search on the Internet. For example, a patent for an "aromatic prenyltransferase from cannabis" was submitted in August 2010. This is the gene that encodes for production of cannabigerolic acid, the substrate for cannabinoid-producing enzymes (TCHA synthase). Additionally, some forms of metabolites have been patented after being extracted and sold as a medicine. For example, GW pharmaceuticals (GWPharma) patented a combination of cannabidiol and THC called Sativex. They have also patented a delivery method for cannabinoids and a cannabinoid-free cannabis plant developed through artificial selection (selective breeding) methods. Several other patents surrounding cannabis have been submitted and some of them have been granted.

Although cannabis metabolites, extraction processes, and drug delivery methods have been patented, should we expect to see a rise of patent applications for genetically engineered cannabis? Companies who dedicate time and effort to developing a novel plant also invest financial resources. Therefore, some companies have argued that they should collect on their investments and patenting is a logical means to do so. However, patenting any cannabis plant with the expectation that growers will honor the patent is unlikely, given the characteristic underground growing operations that has surrounded the cannabis culture.

Still, some people have argued that using cannabis genes in biotechnology might lead to patent rights and control by biotechnology companies like Monsanto. I too have wondered if biotech companies like Monsanto might be interested in cannabis biotechnology. After contacting one of their company representatives the response was that Monsanto is not involved or interested in cannabis. However, there might be cannabis companies that are interested in generating genetically enhanced cannabis. Leading cannabis companies might want to directly communicate their position on genetically enhanced cannabis with their clientele. At this point, one can only wonder if a company is currently attempting to generate genetically engineered lines of cannabis and are keeping quiet. We will have to wait and see.

Genetic engineering successes

Genetic engineering has been helping humans since the first gene was cloned into an *E. coli* cell. Insulin, which was once collected from the pancreas of pigs and cows, yielded limited quantities and was processed to inconsistent purities. After cloning of the insulting gene into an expression vector and transforming *E. coli*, nearly limitless supplies and consistently high purity have become available. One simply needs to ask a diabetic needing insulin shots if they think genetic engineering has been a success.

However, resistance from some societies to adopt genetic modification technology can be severe and in some cases, causes unnecessary harm to themselves. For example, golden rice was a new rice strain developed containing extra beta-carotene.[15] In some countries, lack of beta-carotene hinders proper eyesight development and actually contributes to blindness. Since many poor people live on rice and also lack a healthy mix of vegetables, researchers thought that by generating rice with beta-carotene the rates of blindness could be dramatically decreased. Although golden rice was developed over ten years ago, a few vocal opponents of genetic modification have halted the distribution of golden rice. Much of their opposition is generated by emotion and a rational understanding of any technology is difficult to provide when people are in a heightened emotional state.

Might a better-informed public result in the growing of certain transgenic crops such as golden rice? If golden rice is continued to be held back from those who would benefit, perhaps a different crop could be developed to harbor the beta-carotene. Perhaps we will soon see the engineering of a golden cannabis variety. Other essential vitamins could also be produced in cannabis to increase the health and benefit of cannabis consumers.[16]

Interesting transgenic plants have been generated that have been met with immediate awe. In 1986, tobacco was genetically modified to express the gene responsible for the glow in fireflies.[17] The photo of "glowing tobacco" can easily be found on the Internet and in most introductory biology textbooks. But the plant was not glowing with any apparent strength. The first published picture that has become famous was taken as a time-lapse photo (24 hours exposure). The plant also had to be watered with a luciferin (the luciferase substrate) solution to produce the resulting picture. The protein luciferase was responsible for the glow but the protein could not be made in sufficient quantity as to give a continuous, obvious glow to the plant. The goal was not only a proof of concept but was done to

investigate potential reporter genes for plant biotechnology. Still, people came to realize the new science of plant biotechnology after seeing a glowing tobacco plant.

Of course, there are ethical concerns against generating transgenic crops. But the ethical concerns should address the growing world population and its ever-demanding supply of food and proper nutrition. In some areas of the world the environment is suffering where plant biotechnology could help. For example, transgenic plants could be engineered to require less nutrient (nitrates and phosphates) input. Currently, many bodies of water on the earth are accumulating fertilizer run-off and leading to massive algal blooms. As they die, the algae are consumed by bacteria that deplete the water of dissolved oxygen, which result in large "dead zones." Aquatic life cannot persist. Planting genetically modified crops that required fewer fertilizers would decrease nutrients required for traditional crop plants. Aquatic systems would see less algal blooms and thus, more aquatic life.

Even so, for the first time in history, humans have the ability to introduce single, novel traits into a species using molecular methods and ever-increasing precision. We often worry that other people around the world are unable to access sufficient quantities of food. The potential is here to begin generating plants that are able to withstand the harshest environments, driest conditions, grow in nutrient-poor soils, tolerate high levels of salts, remove metals from polluted soils, increase crop yield or even, in the case of cannabis, improve a seemingly perfect plant.

Certainly there are many concerns over transgenic cannabis. Whether cannabinoid profiles should arise from breeders or biotechnologists might be an ongoing discussion for the next several decades. However, in the former case, less precision is available and more time is required. In the meantime, plant biotechnology is moving forward and will not wait for laws to be passed or ethical debates to be settled.

Additional considerations

Whether genetically modified cannabis will result in good or bad effects on society is of the highest importance. Perhaps the question is most important to people ready to dedicate time, perhaps their lives, to the field of cannabis biotechnology. Perhaps pausing, even if momentarily, is beneficial in allowing time to reflect on some potential questions, even if they cannot be immediately answered.

- There are people who view cannabis to have spiritual-inducing effects. In fact, some cultures depend on cannabis as a component of their religious traditions. Might consuming more cannabinoid-saturated herb be seen as an obligation rather than a choice?

- Cannabis awareness is leading to an increased demand of cannabis. The world is slowly becoming transformed with billions of people demanding a steady supply of cannabis. When cannabis becomes too popular, will global supplies rely on biotechnology to increase production?

- Perhaps another important consideration is whether or not cannabis biotechnologists or their products might hurt another person. Are there ways in which genetically modified cannabis might hurt a person who does not otherwise wish to consume cannabis?

- Cannabis has been cultivated for thousands of years. Thus, people have been genetically modifying cannabis equally as long. However, at the present moment we are able to transfer a gene from another organism into cannabis. For example, an arctic fish gene expressed in cannabis might allow cannabis to tolerate cold climates. If so, cannabis consumers living in the arctic could grow their own medicine. Would a cannabis consumer living in cold climates appreciate such technology and whole-heartedly desire such a unique cannabis strain?

- Evolution on a molecular level is random. Organisms are constantly evolving. Meanwhile, plant biotechnologists have been told they are changing the course of evolution of a species when inserting new genes. That is what makes the science of plant biotechnology paradoxical. Are cannabis biotechnologists really changing the evolutionary direction of cannabis if there is not a direction to begin with?

- Breeding new varieties and artificially selecting for traits has changed plants. Might there have been anti-breeders ten thousand years ago suggesting early human civilization not start cross-pollinating different varieties of apples? Will humans look back thousands of years from now and be surprised there were those who wanted to stop cannabis biotechnology?

- Consider that evil cannabis biotechnologists might stealthily insert cannabinoid pathways into a nation's food supply. If uptight, stressed individuals were to become stoned eating their mashed potatoes one evening, would their neighbors mind? Perhaps in this case, would the actions of an evil cannabis biotechnologist be justified or even desirable? Might our global society be changed for the better if all irate, type A personalities ceased to exist?

- Some people argue we should not play God and tamper with nature. Since humans are a product of nature, can people hold the view that genetic engineering is unnatural?

- With the implementation of cannabis biotechnology, might we inadvertently become a cannabis-infused planet?

Some of these considerations might be easily addressed, others might need expert panels to debate, and others simply produce a good laugh. Admittedly, there are likely to be dozens of more questions in addition to the ones above. Asking questions is good, since discussion helps to realize in what direction the science of cannabis biotechnology might proceed. Considering the changes that have already occurred in the field of plant biotechnology, questions that address *how* humanity should proceed seem to be more realistic questions than *if* humanity should proceed.

There is no doubt that more and more people are becoming familiar with plant biotechnology each passing day. We must remember that young scientific fields are almost always met with concerns initially but with time the worries fade and the foreseen potentials become reality. One way to proceed in the field of cannabis biotechnology is with caution. The likelihood is that the more people involved in and aware of the details of cannabis biotechnology, the more humanity is likely to proceed cautiously.

Chapter 11

The New Age of Cannabis

"We have yet to develop means for satisfactorily involving the public in determining directions for plant biotechnology research."
–Lawrence Busch, MSU, in US-EC Task Force on Biotechnology Research[1]

Directions for cannabis biotechnology

New lines of cannabis are waiting to be engineered. These include cannabis varieties with tendrils to help them climb, larger flowers with increased amounts of metabolites, and leaves, stems, or roots saturated with cannabinoids. Varieties might also be engineered that have rhizomes or adventitious roots, tubers, completely lack the ability to produce seeds or pollen, contain attractive pigments, or impart fruit flavors. No longer will we just have the cannabis taste with an occasional hint of blueberry. Cannabis consumers might soon have the experience of strawberry, orange, banana, lemon, cherry, or grape. Perhaps even chocolate cannabis might soon be available, something that would make Charley Bucket blush. Metabolic pathways that produce cannabinoids or terpenes might also be transferred to roses, citrus trees, amaranth, ailanthus, bananas, or common houseplants and the cannabis plant might be something of the past.

In spite of all the exciting possibilities in cannabis biotechnology, not all agree that genetic modification should be done on cannabis. Such opinion is certain to exist, especially in an open and democratic society. Educated citizens often know that while directions can be taken by societies, individuals make choices. What path should we take as a society?

Perhaps most people in a society would agree that whether a person chooses to participate in a technology should be left to the individual. While improvements

made to cannabis through biotechnological means are available, one does not necessarily need to actively participate in the technology. Even though some might accept improving cannabis through biotechnology, opponents might want to know how and why the change is occurring. Since transgenic cannabis might lead to increased choices, the choices made by consumers might then influence cannabis producers in which directions to proceed.

Research projects to increase choice

Considering the diversity of life on earth, there are millions of different traits one could insert into cannabis. Traits range from very obvious phenotypic characters to molecular changes in biochemical pathways and altering secondary metabolite profiles. Perhaps initially, cannabis biotechnology will begin with small, single-gene traits being altered or inserted but the number of genes simultaneously manipulated will undoubtedly increase in time with technological advancements. Such an evolution in capacity is already visible in other areas of plant biotechnology. Even looking at different cultivars among houseplants and garden flowers, one can see that several traits are popular to florists and gardeners.

Coleus plants have some of the most interesting foliage patterns among the commonly seen decorative plants (Figure 1). They have beautiful reds, blues, and yellows along different regions of their rather large, marbled leaves. Similarly, hostas have shown their aesthetic value among horticulturists and consumers wanting to introduce foliage patterns into gardens. While some of the white, non-photosynthetic tissue needs energetic and nutritional support from the other, more photosynthetically productive tissues, hostas breakup the colors of flowers and abundance of green so often present in gardens. Many other plants have so-called variegated leaves with alternating patterns of colors as well. Variegated cannabis might not only attract cannabis consumers, but also find a firm place in grandma's garden. Such a hardy plant with unforgettable fragrance would certainly be an attractive sight.

Figure 1. Plants have an array of pigments and patterns on foliage, all of which might be of use to increase the aesthetics of cannabis. Shown here clockwise starting from upper left *Coleus*, striped inch plant (*Tradescantia zebrina*), and polka dot plant (*Hypoestes phyllostachya*).

As times are changing, so are the attitudes towards cannabis. However, the dwindling few that still scrutinize the captivating powers of cannabis are still quick to judge. In some areas of the world, cannabis is still illegal and cannabis cultivation results in harsh penalties. Perhaps the time is not only desired but also truly needed to generate an "in cognito" cannabis plant. Such a plant would look as its own species typically does, but have the added-value traits of cannabis metabolic pathways. Deciding on a good surrogate plant that can adequately produce the medicine in a harvestable form, is easy to genetically modify, adjusts itself well to a variety of weather conditions, and grows with some rapidity will be necessary. Perhaps a species of mint (*Mentha*) would be ideal, considering that tissue culture and transformation have been established for some time.[2] Perhaps the floral dip method might be an easier approach to transforming mint. Many mints would likely have the starting metabolites in adequate supply as starter terpenoid compounds to feed the cannabinoid pathway. Mint is also extremely easy to vegetatively propagate, and often becomes a weed in gardens.

Many such characteristics are presently found in several species of weeds besides mints. A herbaceous weed that might be a good candidate is Palmer's amaranth (*Amaranthus palmeri*), which grows in poor soil, is a herbaceous annual, and is a common weed found in gardens and city streets. Cannabinoid-producing amaranth would therefore easily blend in with other plants. A passerby, landlord, or visiting relative would never know or suspect the amaranth growing harbored the same healing potential as a healthy, female cannabis plant.

The idea of using weeds as carriers of cannabis metabolic pathways might sound strange. If something is a weed, the implication is that the plant is simply out of place. Through the eyes of a water conservationist, dandelions are no more weeds than the fescue grass in lawns. Weeds (e.g. amaranth) that harbor cannabis

metabolic pathways might no longer be considered weeds, since they would be highly desired by their growers.

Perhaps the engineering of a weedy, woody perennial would be better than a weedy, herbaceous annual. Still, the species chosen would preferably have weed-like traits so that growers would not need to worry about meticulously working towards keeping the specimen alive. Instead they could dedicate their time to the enjoyment of simply watching the growth and preparing for harvest. A great candidate tree might be ailanthus (*Ailanthus altissimum*). This tree is perhaps one of the weediest tree species known to botanists. Upon finding an adult, one only needs to look at the ground to see the dozens of saplings sprouting from the previous year's fruit. Prophetically, *A. altissimum* is already commonly known as the "tree of heaven."

One might scoff at the idea that cannabis pathways can be inserted into another plant. Perhaps starting with plants that have similar metabolic profiles would be a better starting point. But technology is poised to surprise. Apparent hindrances to generating the cannabis surrogate can be overcome with a little ingenuity and dedicated lab time. For example, although not all species have outlined protocols for plant tissue culture, the optimal culture conditions can be determined through experimental methods. Thus, experimentation will be a deciding factor in which plants will be the cannabis surrogate.

Any plant that harbors the metabolic pathways to produce the products desired by the cannabis consumer will need a distinct metabolic profile, the correct blend and balance of terpenoids and cannabinoids that act synergistically to produce the desired euphoria or pain relief. Producing cannabis plants with such tailored profiles is currently happening in the circles of cannabis growers. Integrating computers, molecular models, and biotechnology would provide consumers with an even more tailored product, specialized in the precise components that results in a desired effect. Slight modifications of enzymatic pathways, including silenced genes, would also likely produce slight variations in metabolic profiles. Each cannabis consumer would essentially be able to obtain his or her personalized medicine.

In thinking farther ahead, cannabis might no longer be limited in the therapeutic effects that are currently provided. Beneficial plant secondary metabolites, antibodies, and even vaccines might be better produced through cannabis. A major beneficial plant secondary metabolite is resveratrol, commonly found in grapes and is often promoted as being beneficial by wine drinkers (and of course grape and wine producers). Resveratrol has in fact been found to decrease

blood pressure and slow production of free radicals.[3] An extensive body of information exists on the enzymatic pathway for resveratrol production as well. Perhaps a cannabis biotechnologist might want to insert the metabolic requirements for resveratrol into cannabis plants. The result might be a cannabis plant that would decrease blood pressure and slow production of free radicals.

Other beneficial traits could be inserted into cannabis besides pathways for plant secondary metabolites. To extend longevity for cannabis consumers, a strain of cannabis expressing a sirtuin synthase gene could be generated. Sirtuins are a group of proteins that have been found to increase longevity in almost all animal species. They are expressed under conditions of starvation and decreased rates of cell division. In some cases, sirtuins act on the telomere regions of chromosomes, regions shown to be important in lifespan. Perhaps a cannabis plant that is harboring cells loaded with sirtuin synthase could be orally ingested, and taken up by resident cells of the digestive tract, and delivered throughout the body. Consumers of sirtuin-loaded cannabis might be able to stay medicated and increase their life span.

Vaccines are often weakened states of a pathogen and are introduced to the immune system as a form of "education" occurring on a molecular and cellular level. Proteins are a major factor in stimulating the immune response in vaccines. Edible vaccines are currently being tested on mice but perhaps one could incorporate a vaccine into the cannabis plant. In order to avoid destroying the vaccine, the genetically engineered cannabis plant would likely have to be eaten, not smoked. Edible vaccines are already becoming a reality and have some benefits over traditional vaccines.[4]

Beyond increasing the health benefits of cannabis consumption and looking past the potentials of cannabis to deliver medicine, there might be undeveloped technologies that allow rapid transformation of a plant species. Perhaps focusing on a few metabolic products of cannabis could be used in closely related species that made cannabinoid intermediates. In this respect, a person might want to develop a plant virus with the capacity to infect several plant species. The virus could also be engineered to carry the gene for cannabigerolic acid synthase and/or other enzymes. Application of the virus to plants might facilitate the delivery and the cannabis genetic material might be stably integrated in the host genome. Viruses already have much to offer in the way of gene delivery and are commonly used in many different areas of biotechnology.[5]

Human retroviruses deliver genetic material to host cells and the genetic material is then incorporated into the host genome. Incorporation of viral DNA is

what occurs with infections of human papilloma virus, herpes virus, and human immunodeficiency virus. Biotechnologists have found they can simply replace some of the viral DNA with genes they hand pick. Now, in the new age of cannabis, perhaps the time has come to consider using retroviruses in cannabinoid gene delivery to patients that need to stay constantly medicated.

One might want to consider inserting the suite of cannabinoid metabolic pathways directly into those who are in need. Of particular interest to target with cannabinoid metabolic pathways are those cells that give rise to new cells such as hematopoietic stem cells, which are responsible for producing new blood cells. Although red blood cells lack nuclei and therefore would be unable to express the cannabinoid synthase genes, white blood cells continually circulate in predictable numbers throughout a healthy person's blood stream. Delivery of foreign DNA to stem cells of adults (and children) is the basis behind gene therapy techniques.

Genetically modifying humans to increase endocannabinoid expression or express phytocannabinoid pathways might sound unappealing, but perhaps examples should be considered. First, there are some parents who might frown on the behavior of their teenagers, a restless age group always getting into trouble. Gene therapy for pre-teens predicted to show problematic behavior could begin upon diagnosis. When the children reach their early teen years a dramatic change would occur. They would be calm, almost sedate but focused. Controlling expression of the cannabinoid synthase genes would likely rely on a developmental-associated promoter site. The diet could be supplemented with cannabinoid pathway precursors. They would not drive erratically, there would be no difference in appetite, and they might go to bed at a decent time.

Other people might be interested in performing genetic modifications on humans so they are able to endogenously produce cannabinoids. Surely the U.S. Department of Homeland Security would be interested in injecting an engineered retrovirus with a cannabinoid metabolic pathway into terrorists. Although, perhaps instead of a developmental promoter, a testosterone promoter could be incorporated. The suite of cannabis genes would be up-regulated in response to rapid testosterone increases and the would-be terrorist might immediately get high and relax. The war on terrorism could easily be won using these extreme but perhaps necessary applications of cannabis biotechnology.

Edibles like never before

"I eat pot, and I like it," is something often overheard in cannabis dispensaries. One aspect that has undoubtedly changed with cannabis over the last several years is edibles, foods infused with cannabis or cannabis extracts. With the smorgasbord of candies, drinks, ice creams, cakes, cookies, and even meats, one has to wonder what cannabis biotechnology holds "in store" for our food supply (Figure 2).

Figure 2. Cannabis consumers have seen a sharp rise in the diversity of edible medicines available, including drinks and candy bars.

To save time in the process of growing cannabis for edibles and then adding the cannabis to the ground flour when baking, perhaps the time is approaching when consumers will prefer to have cannabinoids produced directly in the grain where the flour is derived (Figure 3).

Figure 3. Wheat field in Kansas, USA (left) and harvesting of wheat (right). Might large-scale production of genetically modified crops that produce cannabinoids save time and decrease the effort required in baking cannabis-infused foods?

Importantly, some of the most widely cultivated crops have also already been genetically modified. That is to say, the transformation systems are already in place for several crop plants. Prominent examples are soybean and maize, which by many estimates are already the most widely grown transgenic crops in the world.

Time might be saved and effort conserved if maize were engineered to harbor an extra chromosome that carried a suite of cannabis metabolic pathways. Flour made from cannabis maize could be sold directly in supermarkets or be sold in foods like cakes and cookies. Maize might be a crop that is consumed by too many people to be considered a good surrogate for producing cannabis-induced states of mind (let's not forget the StarLink incident). On the other hand, considering the numerous food items that are made from maize, perhaps such widespread infusion of food with cannabinoids would not be all bad. Certain parts of the world might eat their way to peace.

Much to the dismay of helicopter parents, cannabis edibles include candy and are one of the easiest edibles to conceal. Cannabis candy also usually lacks the distinct smell that many cannabis edibles have. The main ingredients of cannabis candy are sugar, cannabis extract, and flavoring agents. To simplify the cannabis candy production, perhaps engineering sugarcane that manufactures select cannabis metabolites is needed.

Some people who consume cannabis are also very health conscientious, and are hesitant to eat candy but instead chew gum. People have realized the decrease in cavities resulting from chewing gum are not from the gum itself but from stimulation of salivary glands, which continually clears microbial growth. Gum is a carbohydrate extracted from several plants, a main one being the chicle tree (*Manilkara*). Deriving gum from a genetically modified chicle tree loaded with cannabinoids might result in cannabis chewing gum that would be able to help fight cavities and deliver medicine. As another feature the gum could have a slow release property so the gum chewer could have a prolonged medication period over the course of several hours, perhaps even throughout the day. (As a side note, most "gum" used in chewing gum today is synthetic.)

Cows too could be genetically engineered to produce cannabinoids right in their milk. Goats have already been considered as genetically engineered producers for spider silk. Such an idea was developed so that high levels of the protein, which has several useful applications, could be produced. Transgenic goats were successfully made and the silk protein was targeted to be produced in their milk. Although levels of actual production were low[6] research in this area

continues. All dairy products from cannabis cows would be produced the same and taste identical but they would contain cannabis metabolites. Yogurt would still be a health food but people would likely eat two servings instead of one. Cheeses such as Swiss, Roquefort, mozzarella, and cheddar made from cannabis cow milk would be best served on hemp crackers. Sticks of cannabis butter at the store would need to be labeled as "better butter." Cannabis milk might need to be artificially colored green but could be sold on the grocery shelves right between white milk and chocolate milk.

In cannabis dispensaries, edibles are not limited to sweets. More and more cannabis is appearing in foods considered non-snack. This might lead to grocery stores that eventually sell cannabis-loaded foods derived from transgenic plants with cannabis compounds (Figure 4). Imagine going to the super market and everything in the frozen food section has various amounts of cannabis. Wrapped foods like a pizza, a type of pot pocket, might vary based on cannabinoid profiles rather than traditional pizza toppings. Spinach would be the ultimate health food and would contain several B vitamins as well as medicine for the cannabis patient. Would you prefer the cannabinol apple or the cannabidivarin banana? Do you need a bag of whole-wheat flour enriched with THC or a cannabichromene orange?

Figure 4. Might grocery stores begin to sell genetically enhanced fruit and vegetables infused with cannabinoids and terpenoids produced directly in the plant?

Perhaps there are those who will ask the purpose of generating transgenic plants that contain cannabis metabolites to harvest and use in food production. Such a position is also likely to point out that cannabis foods are already available. There might not really be a need for such a drastic approach since cannabis is traditionally grown, harvested, then put in with food as it is cooked. However, one

needs to realize the scale of the comparison. A dozen zonky bars or a handful of cannabis candies pale in comparison to millions of hectares of cannabis sugarcane and cannabis wheat. One needs to compare a pint of cannabis ice cream to dozens of milk trucks loaded with cannabis cow milk at a dairy. A six pack of cannabis drinks and bag of cannabis jerky are mere snacks compared to having half of the global maize supply infused with cannabinoids.

If we truly are to begin growing crop plants with cannabis metabolites a concern might be dosing. How can one properly dose themselves without becoming too medicated? Perhaps the fact should be pointed out that some people choose edibles over other methods of cannabis consumption for this very reason. They also enjoy the variety in effects that edibles offer. Such a variety is part of their medical needs. With edibles, cannabis consumers have more choice.

Thus, we return to the words of Lawrence Busch,[1] a leader in the field of analyzing the relationship between plant biotechnology and society. The importance is not if the propositions of cannabis biotechnology are engineered. Rather, since biotechnology tools are here, the importance is to develop a means to involve the public in the discussion of the direction of biotechnology. By realizing the potential products of cannabis biotechnology, one realizes the amount of discussion that should be occurring.

The biotechnology of *Erythroxylum cocao*?

Exploring the potentials of cannabis biotechnology is exciting. There are dozens of potential research projects screaming for attention. Many such projects require dedication of time in the lab but also demand a strong educational background of the researcher and imagination is indispensable. We must therefore open ourselves to other members of the plant kingdom and ask if there are other plants besides *Cannabis sativa* worthy of biotechnological applications. Several come to mind and include the more well-known plants that have helped shape our human global society. They include chocolate, cocaine, coffee, poppy, and datura.

Chocolate (*Theobroma cacao*) is a member of the Sterculiaceae. The family also includes the genus *Cola*, from which the cola nut was used as an ingredient in the successfully delicious Coca-Cola soft drink. A component of chocolate includes caffeine but the actual pleasurable chemical often attributed to chocolate is given to theobromine,[7] literally translating to "God molecule." To speed up tinkering with the chocolate genes, the genome has been published.[8] The chocolate scent is

also given off by about a half-dozen other plants (e.g. *Berlandiera lyrata*). Might cannabis biotechnology be able to produce a cannabis plant with chocolate flavor? How about a cacao plant with cannabinoids?

Cocaine (*Erythroxylum cacao*) is in the Erythroxylaceae. Might biotechnology be used to capture the genetic pathways leading to the production of cocaine and transfer them to another plant? Perhaps given the huge money making industry that cocaine has, such genetic modifications have already happened. One might wonder if the global decrease in tobacco consumption is related to the fact that tobacco is serving as a surrogate plant for cocaine biosynthesis. Tobacco is easily genetically modified and since other alkaloids are already present, inserting cocaine metabolic pathways are certainly feasible. Although scientific literature of the in vitro culturing or genetic transformation of cocaine is not (at least publically) available, sequence data for genes encoding various enzymes involved in alkaloid synthesis in Erythroxylaceae have been published and are available through NCBI.[9] To the pleasure of some, several hurdles would be overcome by growing cocaine-producing tobacco directly in a country instead of the drug being smuggled across borders.

Coffee (*Coffea arabica*) is in the Rubiaceae. Caffeine is the main component sought by consumers of coffee, usually resulting in good discussion and a stimulated pulse. Several genetic resources of coffee are available through the International Coffee Genome Network (coffeegenome.org). Coffee was engineered more than a decade ago that had reduced levels of caffeine by 70% through RNAi.[10] Might coffee be engineered to have cannabinoids? How about cannabis lines generated that have caffeine? Such a feat would certainly modify the "wake and bake" philosophy of some.

Poppy (*Papaver somniferum*) belongs to the Papaveraceae, the poppy family. Poppy was first genetically modified using Agrobacterium more than a decade ago.[11] Metabolites of the poppy plant induce pleasure and relieve pain. To collect the resin that contains the metabolites, the unripe capsules are usually split and the resin slowly exudes over a few days. However small quantities of opiates can be found in leaves of closely related species like *Argemone polyanthemos* (Figure 5).

Figure 5. The leaves of the prickly poppy (*Argemone polyanthemos*) are known to contain interesting alkaloids. Might the genes for the alkaloid pathway be engineered into cannabis?

For those interested, the opiate pathway has been elucidated and described with great detail.[12] Could alkaloids naturally found in the Papaveraceae be produced in cannabis? Would cannabis consumers want to be sedated with opiates and cannabinoids too, perhaps for dual medication?

Datura (*Datura*) belongs to the Solanaceae, the deadly nightshade family. The solanaceae provide many different alkaloids, one of the other major groups of plant secondary metabolites. Several pathways for alkaloid production are currently being elucidated.[13] Many datura species have an extremely potent hallucinogenic effect when consumed. However, some of the hallucinogen-inducing metabolites such as scopolamine[14] can be deadly in uncontrolled amounts but can also have medicinal qualities when used appropriately. Scopolamine inhibits the function of involuntary muscles such as the diaphragm muscles, therefore overdosing on a datura extract is not good for someone who wants to keep breathing. However, so many alkaloids are produced in this family that the sol genomics network (solgenomics.net) has been established.[15]

Each of the above considerations can be decomposed into their metabolic parts. Expressing other metabolite-synthesizing genes in cannabis or expressing cannabinoid pathways in these well-known plants might be desirable. Surely looking at the genetic and metabolic composition of each species leads one to have a steady stream of questions. However, each plant is also greater then the sum of its parts. While the system is dissectible the entire organism is composed of such interlocking complexity, one wonders if incorporation of select genes would merely result in partial products. Answers to these questions might be best provided

through experimentation. Hence, at the present moment, we need plant biotechnologists to perform these experiments.

Cannabis biotechnology for everyone

Cannabis biotechnology is not a field that is restricted to scientists with advanced degrees and expensive labs. One might quickly decide such processes of obtaining a particular gene, growing bacteria, tissue culturing, delivering genes, and selecting for plants costs tens of thousands of dollars and many years of research effort. However, the field of cannabis biotechnology is open to anyone who has a drive or motivation to improve and genetically enhance the most fascinating plant on earth.

As of the writing of this book, cannabis has become legal to purchase in more parts of North America for recreational use. Additionally, Uruguay has become the first country in the world to completely legalize cannabis. In time, more countries will legalize cannabis and anyone who enjoys herbs in their garden will likely plant cannabis seeds. Given a little more time, more people will attempt to culture cannabis in vitro. Legalization efforts, whether intended or not, are likely to increase those interested in cannabis biotechnology. Thus, more people will begin to study what might be needed to genetically modify cannabis.

While the tools to clone a gene might not seem readily available, with the advent of gene synthesis companies, the reality is that plant biotechnology might be more achievable to those who understand the methods. Having a postal address and a credit card is enough for ordering a several molecular biology reagents. DNA is extremely stable and could be stored safely in a home refrigerator or freezer. The same can be said of several different bacteria, which are easily obtainable from hundreds of labs around the world. While many protocols suggest storing bacteria at -80°C, the reality is that there have been several times where bacteria have shown subsequent recovery after being stored in a glycerol solution at -20°C.

Plant tissue culture is not an expensive technology if one can afford Petri dishes, culture media, and synthetic hormones. Antibiotics can get expensive but not to the degree they are completely unaffordable. Manipulating bacteria and making media could rely on a home-built sterile area, something similar to a laminar flow hood. Microbiology was a science long before laminar flow hoods existed. Isolating bacteria and growing them without protection from the ambient

air was once common. One simply took more care in their motions, used Bunsen burners much more often, and allowed less time for their culture vessels to be open.

Tissue culture, like culturing bacteria, would also require sterile technique and an adequate workspace. Making media would not be a problem, since most people have access to a pressure cooker or a good-working stove. There are also companies that sell in-home tissue culture kits (planttc.com; kitchenculturekit.com). However, companies are now becoming more involved in cannabis tissue culture. Skunk Pharm Research, LLC (skunkpharmresearch.com) is one such company.

Gene delivery might prove a bit more challenging than tissue culture. Two methods could streamline gene delivery for at-home cannabis biotechnologists. The first is the ability to order a binary vector with a gene of interest already inserted. One would simply order the vector-insert construct in Agrobacterium and be ready to transform their calli. The second is avoiding tissue culture all together and seeking to work instead directly with cannabis flowers. The floral dip is an approach that could be tested on cannabis and the resulting seeds could be sown on selective media. If possible, genetically modified cannabis could be a matter of shopping for genes and simple lab work. A simple model for many businesses might incorporate several technologies discussed thus far (Figure 6).

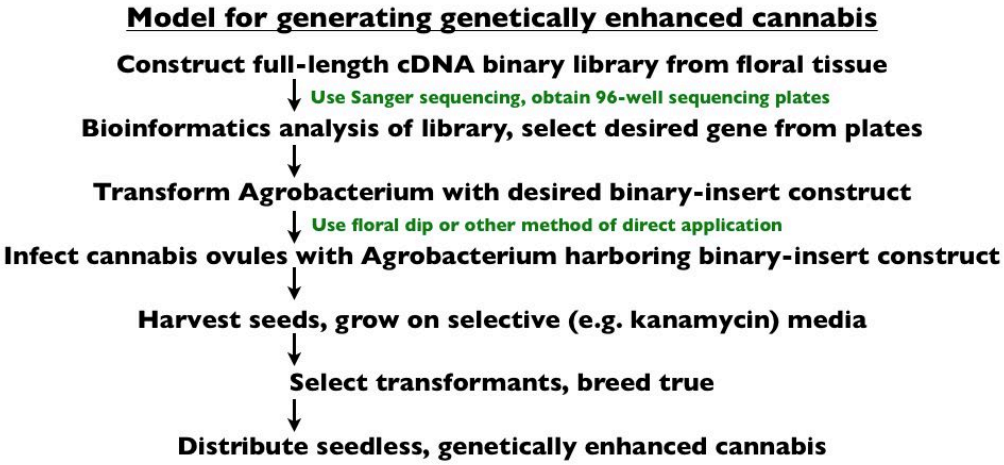


Figure 6. Outline for developing genetically enhanced medical cannabis for distribution. Details of most steps are discussed in this book.

Competition between businesses will likely drive production of novel cannabis strains. There are thousands of interesting metabolic pathways and genes in other plants and thousands of genes in cannabis that can be used in genetic modification studies. With so many options available, there is more than enough research for several companies to pursue.

Is cannabis biotechnology going to become available to the every day gardener? The United States Department of Agriculture (USDA) states on a web site,[16] "K-12 teachers and students learn and use the laboratory techniques used in crop improvement, such as DNA extraction, gel electrophoresis, PCR, and microarray technology." The USDA information is linked to the American Phytopathological Society (APS), where some members are directly involved with making the educational leap to the youth by supplying several web pages with information and activities that include plant tissue culture, PCR, and genetic modification of bacteria.[17] Other researchers have published studies that have tried to analyze methods of teaching agrobiotechnology to high school students in hopes to improve teachers' confidence in teaching biotechnology. Some approaches to teaching biotechnology to young kids have even incorporated comics.[18]

Thus, students are increasingly being taught the skills of biotechnology. Like the energetic sun for a growing flower, today's students of biotechnology might be very important to the future success of cannabis. Students should be encouraged to learn all facets of plant biotechnology. They should never be told they are "obscure" or that their work is irrelevant. Critics who give such negative comments often fall victim to their own hubris, eventually giving talks where only four or five people might attend. We must stay positive about the potentials of students, for someday they will be the instigators and innovators for the new age of cannabis.

Considering the increased rates of transgenic crops being planted and influential education regarding plant biotechnology, the science is obviously becoming more common in peoples' lives. With such early ages of exposure, more and more young people might be likely to directly participate and contribute to the field of plant biotechnology. Which of these young students will continue on and genetically enhance cannabis in a way that will benefit humanity?

A final toke

The botany of *Cannabis sativa* has stimulated endless interest in improving its cultivation. The plant has many features that have served humanity well. The nutritional benefits of cannabis seeds, to humans and other animals, are now widely realized. The durability of its fibers has influenced societal interactions. Perhaps more influential than cannabis being used as a food or fiber is the realization of the unique blend of secondary metabolites. As mysterious as they are, scientific methods are slowly unraveling the activities of phytocannabinoids on the human mind and body.

One cannot help imagine what biotechnology, a truly multidisciplinary science, has in store for cannabis. Anyone with even a slight interest in the subject of cannabis will certainly want to understand the lab tools and techniques that are being used to alter its genetic composition. Although seemingly complex, a step-by-step look at the core set of molecular biology tools helps one understand DNA isolation, amplification, and cloning.

Perhaps most important to plant biotechnology has been the ability to grow plants in vitro using techniques of tissue culture. One might almost be under the impression that plants were somehow created with capabilities to assist in manipulating plant tissue in vitro. However, clearly from an evolutionary view, cannabis is merely being manipulated by human scientific ingenuity. Many people have also worked hard in teasing out the details of plant infection by Agrobacterium. Deciphering the molecular details of gene delivery and stable transformation has made plant biotechnology a reality.

Other technology has improved our understanding of the inner workings of the cannabis cell. Transcriptomics has helped determine the expression patterns of developing plants and specific transcriptomes of plant organs and cell types. Methods for analyzing mRNA transcripts are improving, resulting in increased throughput that gives greater depth and comprehension of gene expression. Technologies are also allowing humans to characterize cannabis proteins with molecular precision, providing researchers with a new outlook on cellular functions of enzymes or structural roles of proteins involved in photosynthesis.

Genomics is helping us understand chromosome architecture, genome evolution, and the function of thousands of genes. The first draft of the cannabis genome has provided a broad yet focused look at cannabinoid pathways. Genomics is also helping us determine the relationship of metabolic pathways to one another. The result is a blueprint for analyzing and engineering cannabinoids

and related secondary metabolites. Metabolic pathways are only becoming better characterized as time progresses. Sites of metabolite synthesis are known, as are the enzymatic steps of hundreds of intermediates. With such knowledge, perhaps we are inevitably heading into an era where individuals can determine their cannabinoid combination preference, tailored to fit their pharmaceutical needs.

Almost every area of cannabis biotechnology is assisted by the tools of bioinformatics. Bioinformatics is a powerful subject to learn since data mining and manipulation can be done by nearly anyone with a computer. Computers are becoming inseparable from molecular biology. The acquisition of molecular data and storage in databases gives researchers the ability to search for sequenced genes, proteins, and pathways. Algorithms for searching databases, aligning two or more sequences, and phylogenetic construction are providing ample room for comparing evolutionary relationships of cannabis to its botanical relatives. Synthetic and systems biology are poised to make the process of generating highly-personalized medicine much easier and much better understood.

To adequately understand the potentials that cannabis biotechnology holds for the future, one must appreciate the evolutionary chances that led to the cannabis plant we know today. Additionally, there has never been a time in history when such an interest in cannabis has been so high or when cannabis research has been so advanced. More and more academic, government, and private labs are performing scientific experiments with cannabis.

Thus, we live in an exciting age, where cannabis biotechnology could potentially change humanity in ways never before imagined. There remain several questions for future research, a need for public understanding of genetically modified cannabis, and increased discussion about the potentials of cannabis biotechnology. Cannabis is clearly undergoing a radical transformation and there are many difficult decisions and possible paths ahead. Undoubtedly, the paths chosen will be decided by those who embrace the biotechnology of *Cannabis sativa*.

References

Chapter 1

1. Brookes G, Barfoot P. Key environmental impacts of global genetically modified (GM) crop use 1996–2011. GM Crops and Food: Biotechnology in Agriculture and the Food Chain. 2013 Apr 1;4(2):109–19.

2. Herman RA, Garcia-Alonso M, Layton R, Raybould A. Bringing policy relevance and scientific discipline to environmental risk assessment for genetically modified crops. Trends in Biotechnology. 2013 Sep;31(9):493–6.

3. Holst-Jensen A, Bertheau Y, de Loose M, Grohmann L, Hamels S, Hougs L, et al. Detecting un-authorized genetically modified organisms (GMOs) and derived materials. Biotechnology Advances. 2012 Nov;30(6):1318–35.

4. Bruce TJA. GM as a route for delivery of sustainable crop protection. J Exp Bot. 2012 Jan 1;63(2):537–41.

5. Jacobsen S-E, Sørensen M, Pedersen SM, Weiner J. Feeding the world: genetically modified crops versus agricultural biodiversity. Agron Sustain Dev. 2013 Oct 1;33(4):651–62.

6. Chandra S, Lata H, Khan IA, ElSohly MA. The Role of Biotechnology in *Cannabis sativa* Propagation for the Production of Phytocannabinoids. In: Chandra S, LATA H, Varma A, editors. Biotechnology for Medicinal Plants [Internet]. Springer Berlin Heidelberg; 2013 [cited 2013 Dec 3]. p. 123–48.

7. Clarke RC, Merlin MD. Cannabis: Evolution and Ethnobotany. University of California Press; 2013. 464 p.

8. Schultes RE, Klein WM, Plowman T, and Lockwood TE. Cannabis: an example of taxonomic neglect. In Rubin V. Cannabis and Culture. Walter de Gruyter; 1975.

9. Russo EB, Jiang H-E, Li X, Sutton A, Carboni A, Bianco F del, et al. Phytochemical and genetic analyses of ancient cannabis from Central Asia. J Exp Bot. 2008 Nov 1;59(15):4171–82.

10. Honnay O, Jacquemyn H, Aerts R. Crop wild relatives: more common ground for breeders and ecologists. Frontiers in Ecology and the Environment. 2012 Apr 1;10(3):121–121.

11. Gaudeul M, Rouhan G. A plea for modern botanical collections to include DNA-friendly material. Trends in Plant Science. 2013 Apr;18(4):184–5.

12. Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature. 1953 Apr 25;171(4356):737–8.

13. Van Lijsebettens M, Angenon G. Thirty years of transgenic research in plants. Int J Dev Biol. 2013;57(6-7-8):445–7.

14. De la Cruz F, Davies J. Horizontal gene transfer and the origin of species: lessons from bacteria. Trends in Microbiology. 2000 Mar 1;8(3):128–33.

15. Mohamed S, Syed BA. Commercial prospects for genomic sequencing technologies. Nat Rev Drug Discov. 2013 May;12(5):341–2.

16. Bakel H van, Stout JM, Cote AG, Tallon CM, Sharpe AG, Hughes TR, et al. The draft genome and transcriptome of *Cannabis sativa*. Genome Biology. 2011 Oct 20;12(10):R102.

17. El-Alfy AT, Ivey K, Robinson K, Ahmed S, Radwan M, Slade D, et al. Antidepressant-like effect of delta-9-tetrahydrocannabinol and other cannabinoids isolated from *Cannabis sativa* L. Pharmacol Biochem Behav. 2010 Jun;95(4):434–42.

Chapter 2

1. Schultes RE. Random thoughts an queries on the botany of cannabis. In: The Botany and Chemistry of Cannabis. J & A Churchill, London,1970.

2. Lane N, Martin WF, Raven JA, Allen JF. Energy, genes and evolution: introduction to an evolutionary synthesis. Phil Trans R Soc B. 2013 Jul 19;368(1622):20120253.

3. Dzieciol AJ, Mann S. Designs for life: protocell models in the laboratory. Chem Soc Rev. 2011 Dec 5;41(1):79–85.

4. Miller SL. Production of Some Organic Compounds under Possible Primitive Earth Conditions1. J Am Chem Soc. 1955 May 1;77(9):2351–61.

5. Deamer DW. Role of amphiphilic compounds in the evolution of membrane

structure on the early earth. Origins Life Evol Biosphere. 1986 Mar 1;17(1):3–25.

6. McPartland JM, Nicholson J. Using parasite databases to identify potential nontarget hosts of biological control organisms. New Zealand Journal of Botany. 2003;41(4):699–706.

7. Robson PRH, Smith H. Fundamental and biotechnological applications of phytochrome transgenes. Plant, Cell & Environment. 1997;20(6):831–9.

8. Bruley C, Dupierris V, Salvi D, Rolland N, Ferro M. AT_CHLORO: A Chloroplast Protein Database Dedicated to Sub-Plastidial Localization. Front Plant Sci [Internet]. 2012 Sep 11 [cited 2013 Nov 28];3. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3438710/

9. Berman J, Zhu C, Pérez-Massot E, Arjó G, Zorrilla-López U, Masip G, et al. Can the world afford to ignore biotechnology solutions that address food insecurity? Plant Mol Biol. 2013 Sep 1;83(1-2):5–19.

10. Ruan C-J, Shao H-B, Teixeira da Silva JA. A critical review on the improvement of photosynthetic carbon assimilation in C $_3$ plants using genetic engineering. Critical Reviews in Biotechnology. 2012 Mar;32(1):1–21.

11. Peterhansel C, Krause K, Braun H-P, Espie GS, Fernie AR, Hanson DT, et al. Engineering photorespiration: current state and future possibilities. Plant Biology. 2013;15(4):754–8.

12. Howell SH. Molecular Genetics of Plant Development. Cambridge University Press; 1998.

13. Hallahan DL, Callow JA, Gray JC. Plant Trichomes. Elsevier; 2000.

14. Anderson RC. Ecology and management of the Prairie Division. 2012 [cited 2013 Nov 28]. p. 175–201. Available from: http://www.treesearch.fs.fed.us/pubs/41098

15. Zhao Y, Huang LH, Peng Y, Peng ZZ, Liu XZ, Kuang FC, et al. Trichome expression of iaaM transgene influences their development and elongation in tobacco. Russ J Plant Physiol. 2013 Nov 1;60(6):839–44.

16. Christensen S, Weigel D. Plant development: The making of a leaf. Current Biology. 1998 Sep 10;8(18):R643–R645.

17. Rehman MSU, Rashid N, Saif A, Mahmood T, Han J-I. Potential of bioenergy production from industrial hemp (*Cannabis sativa*): Pakistan perspective. Renewable

and Sustainable Energy Reviews. 2013 Feb;18:154–64.

18. Friis EM, Endress PK. Origin and Evolution of Angiosperm Flowers. In: J.A. Callow, editor. Advances in Botanical Research [Internet]. Academic Press; 1990 [cited 2013 Nov 28]. p. 99–162. Available from: http://www.sciencedirect.com/science/article/pii/S0065229608601330

19. Bowman JL, Smyth DR, Meyerowitz EM. The ABC model of flower development: then and now. Development. 2012 Nov 15;139(22):4095–8.

Chapter 3

1. Ragan GA. Innovative recycling options for biomedical research facilities. Chem Health Saf. 2007;14(6):17–20.

2. Murray MG, Thompson WF. Rapid isolation of high molecular weight plant DNA. Nucl Acids Res. 1980 Oct 10;8(19):4321–6.

3. Quattrocchio FM, Spelt C, Koes R. Transgenes and protein localization: myths and legends. Trends in Plant Science. 2013 Sep;18(9):473–6.

4. Goldberg RB, Beals TP, Sanders PM. Anther development: basic principles and practical applications. Plant Cell. 1993 Oct;5(10):1217–29.

5. Pance A. Tailoring the Models of Transcription. Int J Mol Sci. 2013 Apr 8;14(4):7583–97.

6. Barnes WM. Variable patterns of expression of luciferase in transgenic tobacco leaves. PNAS. 1990 Dec 1;87(23):9183–7.

7. Chen Q-J, Zhou H-M, Chen J, Wang X-C. A Gateway-based platform for multigene plant transformation. Plant Mol Biol. 2006 Dec 1;62(6):927–36.

Chapter 4

1. Haberlandt G., c. M. k. Akad. Culturversuche mit isolierten Pflanzenzellen. Vorgelegt in der Sitzung am 6. Februar 1902.

2. Hitchens AP, Leikind MC. The Introduction of Agar-agar into Bacteriology. J Bacteriol. 1939 May;37(5):485–93.

3. Murashige T, Skoog F. A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. Physiologia Plantarum. 1962;15(3):473–97.

4. Galun E. Phytohormones and Patterning: The Role of Hormones in Plant Architecture. World Scientific; 2010.

5. Thorpe TA. History of plant tissue culture. Mol Biotechnol. 2007 Oct;37(2):169–80.

6 Lata H, Chandra S, Khan IA, Elsohly MA. High frequency plant regeneration from leaf derived callus of high Δ9-tetrahydrocannabinol yielding *Cannabis sativa* L. Planta Med. 2010 Oct;76(14):1629–33.

7. Feeney M, Punja ZK. Tissue culture and Agrobacterium-mediated transformation of hemp (*Cannabis sativa* L.). In Vitro Cell Dev Biol -Plant. 2003 Nov 1;39(6):578–85.

8. Slusarkiewicz-Jarzina A, Ponitka A, Kaczmarek Z. Influence of cultivar, explant source and plant growth regulator on callus induction and plant regeneration of *Cannabis satavia* L. Acta Biologica Cracoviensia 2005; 47(2): 145-151.

9. Luwanska, A., and K. Wielgus. "Anther culture and callus induction in *Cannabis sativa* L." New Biotechnology 25 (2009): S302.

10. Veliky IA, Genest K. Growth and metabolites of *Cannabis sativa* cell suspension cultures. Lloydia. 1972 Dec;35(4):450–6.

11. Itokawa H, Takeya K, Mihashi S. Biotransformation of Cannabinoid Precursors and Related Alcohols by Suspension Cultures of Callus induced from *Cannabis sativa* L. Chemical & pharmaceutical bulletin. 1977 Aug;25(8):1941–6.

12. Loh WH-T, Hartsel SC, Robertson LW. Tissue Culture of *Cannabis sativa* L. and in vitro Biotransformation of Phenolics. Zeitschrift für Pflanzenphysiologie. 1983 Sep;111(5):395–400.

13. Lata H, Chandra S, Khan I, ElSohly MA. Thidiazuron-induced high-frequency direct shoot organogenesis of *Cannabis sativa* L. In Vitro CellDevBiol-Plant. 2009 Feb 1;45(1):12–9.

14. Wielgus K, Luwanska A, Lassocinski W, Kaczmarek Z. Estimation of *Cannabis sativa* L. Tissue Culture Conditions Essential for Callus Induction and Plant Regeneration. Journal of Natural Fibers. 2008;5(3):199–207.

15. Lata H, Chandra S, Khan IA, Elsohly MA. High frequency plant regeneration from leaf derived callus of high Δ9-tetrahydrocannabinol yielding *Cannabis sativa* L. Planta

Med. 2010 Oct;76(14):1629–33.

16. Chandra S, Lata H, Khan IA, ElSohly MA. The Role of Biotechnology in *Cannabis sativa* Propagation for the Production of Phytocannabinoids. In: Biotechnology for Medicinal Plants. Springer Berlin Heidelberg; 2013. p. 123–48.

17. Ara, Hussain, Uma Jaiswal, and V. S. Jaiswal. "Synthetic seed: prospects and limitations." *Curr Sci* 78.12 (2000): 1438-1444.

18. Lata H, Chandra S, Khan IA, ElSohly MA. Propagation through alginate encapsulation of axillary buds of *Cannabis sativa* L. — an important medicinal plant. Physiol Mol Biol Plants. 2009 Jan 1;15(1):79–86.

Chapter 5

1. Chilton M-D, Drummond MH, Merlo DJ, Sciaky D, Montoya AL, Gordon MP, et al. Stable incorporation of plasmid DNA into higher plant cells: the molecular basis of crown gall tumorigenesis. Cell. 1977 Jun;11(2):263–71.

2. Hooykaas PJ. Plant Transformation. eLS. John Wiley & Sons, Ltd; 2001. http://onlinelibrary.wiley.com/doi/10.1002/9780470015902.a0003070.pub2/abstract

3. Woese CR, Fox GE. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. PNAS. 1977 Nov 1;74(11):5088–90.

4. Wahby I, Arráez-Román D, Segura-Carretero A, Ligero F, Caba JM, Fernández-Gutiérrez A. Analysis of choline and atropine in hairy root cultures of *Cannabis sativa* L. by capillary electrophoresis-electrospray mass spectrometry. Electrophoresis. 2006 Jun;27(11):2208–15.

5. Wahby I, Caba JM, Ligero F. Agrobacterium infection of hemp (*Cannabis sativa* L.): establishment of hairy root cultures. Journal of Plant Interactions. 2013;8(4):312–20.

6. Pitzschke A, Hirt H. New insights into an old story: Agrobacterium-induced tumour formation in plants by plant transformation. EMBO J. 2010 Mar 17;29(6):1021–32.

7. Venturi V, Fuqua C. Chemical Signaling Between Plants and Plant-Pathogenic Bacteria. Annual Review of Phytopathology. 2013;51(1):17–37.

8. Gelvin SB. Traversing the cell: Agrobacterium T-DNA's journey to the host genome. Agrobacterium. 2012;3:52.

9. Gelvin SB. Plant Proteins Involved in Agrobacterium-Mediated Genetic Transformation. Annual Review of Phytopathology. 2010;48(1):45–68.

10. Hefferon KL. Plant virus expression vectors set the stage as production platforms for biopharmaceutical proteins. Virology. 2012 Nov 10;433(1):1–6.

11. Clough SJ, Bent AF. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. Plant J. 1998 Dec;16(6):735–43.

12. Mieog JC, Howitt CA, Ral J-P. Fast-tracking development of homozygous transgenic cereal lines using a simple and highly flexible real-time PCR assay. BMC Plant Biology. 2013 Apr 30;13(1):71.

13. Martín B, Ramiro M, Martínez-Zapater JM, Alonso-Blanco C. A high-density collection of EMS-induced mutations for TILLING in Landsberg erecta genetic background of *Arabidopsis*. BMC Plant Biology. 2009 Dec 14;9(1):147.

14. Li F, Orban R, Baker B. SoMART: a web server for plant miRNA, tasiRNA and target gene analysis. The Plant Journal. 2012;70(5):891–901.

Chapter 6

1. Watson JD. The Double Helix: A Personal Account of the Discovery of the Structure of DNA. Scribner; 2011. 361 p.

2. Baginsky S, Hennig L, Zimmermann P, Gruissem W. Gene Expression Analysis, Proteomics, and Network Discovery. Plant Physiol. 2010 Feb 1;152(2):402–10.

3. Simpson CG, Manthri S, Raczynska KD, Kalyna M, Lewandowska D, Kusenda B, et al. Regulation of plant gene expression by alternative splicing. Biochem Soc Trans. 2010 Apr;38(2):667–71.

4. Zhu YY, Machleder EM, Chenchik A, Li R, Siebert PD. Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. BioTechniques. 2001 Apr;30(4):892–7.

5. Raharjo TJ, Widjaja I, Roytrakul S, Verpoorte R. Comparative proteomics of *Cannabis sativa* plant tissues. J Biomol Tech. 2004 Jun;15(2):97–106.

6. Rose PW, Bi C, Bluhm WF, Christie CH, Dimitropoulos D, Dutta S, et al. The RCSB Protein Data Bank: new resources for research and education. Nucleic Acids Research. 2012 Nov 27;41(D1):D475–D482.

7. Gopalan G, He Z, Balmer Y, Romano P, Gupta R, Heroux A, et al. Structural analysis uncovers a role for redox in regulating FKBP13, an immunophilin of the chloroplast thylakoid lumen. Proc Natl Acad Sci U S A. 2004 Sep 21;101(38):13945–50.

8. Okonechnikov K, Golosova O, Fursov M. Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics. 2012, *doi: 10.1093/bioinformatics/bts091*.

9 Santiago J, Henzler C, Hothorn M. Molecular Mechanism for Plant Steroid Receptor Activation by Somatic Embryogenesis Co-Receptor Kinases. Science. 2013 Aug 23;341(6148):889–92.

10. Wang Y, Geer LY, Chappey C, Kans JA, Bryant SH. Cn3D: sequence and structure views for Entrez. Trends Biochem Sci. 2000 Jun;25(6):300–2.

11. Hamès C, Ptchelkine D, Grimm C, Thevenon E, Moyroud E, Gerard F, et al. Structural Basis for Leafy Floral Switch Function and Similarity with Helix-Turn-Helix Proteins. Embo J. 2008 Jul 16;27:2628–null.

12. Chandran V, Fronzes R, Duquerroy S, Cronin N, Navaza J, Waksman G. Structure of the outer membrane complex of a type IV secretion system. Nature. 2009 Dec 24;462(7276):1011–5.

13. Ambrosio ALB, Dias SMG, Polikarpov I, Zurier RB, Burstein SH, Garratt RC. Ajulemic Acid, a Synthetic Nonpsychoactive Cannabinoid Acid, Bound to the Ligand Binding Domain of the Human Peroxisome Proliferator-activated Receptor γ. J Biol Chem. 2007 Jun 22;282(25):18625–33.

Chapter 7

1. The source for this quote is stated to be unknown but Nate at GMBudz claims to have been the person who said it. I can neither confirm nor deny so the source remains a mystery.

2. Kinoshita T, Jacobsen SE. Opening the Door to Epigenetics in PCP. Plant Cell Physiol. 2012 May 1;53(5):763–5.

3. Ahmadi KR, Weale ME, Xue ZY, Soranzo N, Yarnall DP, Briley JD, et al. A single-nucleotide polymorphism tagging set for human drug metabolism and transport. Nat Genet. 2005 Jan;37(1):84–9.

4. Kliebenstein DJ, Lambrix VM, Reichelt M, Gershenzon J, Mitchell-Olds T. Gene Duplication in the Diversification of Secondary Metabolism. Plant Cell. 2001 Mar;13(3):681–94.

5. Nützmann H-W, Osbourn A. Gene clustering in plant specialized metabolism. Current Opinion in Biotechnology. 2014 Apr;26:91–9.

6. Chen ZJ. Molecular mechanisms of polyploidy and hybrid vigor. Trends Plant Sci. 2010 Feb;15(2):57.

7. Menzel MY. Meiotic Chromosomes of Monoecious Kentucky Hemp (*Cannabis sativa*). Bulletin of the Torrey Botanical Club. 1964 May;91(3):193.

8. Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH. The frequency of polyploid speciation in vascular plants. PNAS. 2009 Aug 18;106(33):13875–9.

9. Eigsti OJ. A Cytological Study of Colchicine Effects in the Induction of Polyploidy in Plants. Proc Natl Acad Sci U S A. 1938 Feb;24(2):56–63.

10. Tian Bin, Guo HongYan, Yang Ming, Xin PeiYao. Preliminary study on polyploid inducement of *Cannabis sativa*. Acta Agriculturae Jiangxi. 2013;25(6):51–4.

11 Brazma A. Minimum Information About a Microarray Experiment (MIAME); Successes, Failures, Challenges. The Scientific World Journal. 2009;9:420–3.

12 Van den Broeck HC, Maliepaard C, Ebskamp MJM, Toonen MAJ, Koops AJ. Differential expression of genes involved in C1 metabolism and lignin biosynthesis in wooden core and bast tissues of fibre hemp (*Cannabis sativa* L.). Plant Science. 2008 Feb;174(2):205–20.

13 De Pauw MA, Vidmar JJ, Collins J, Bennett RA, Deyholos MK. Microarray analysis of bast fibre producing tissues of *Cannabis sativa* identifies transcripts associated with conserved and specialised processes of secondary wall development. Funct Plant Biol. 2007;34(8):737–49.

14. Juknat A, Pietr M, Kozela E, Rimmerman N, Levy R, Gao F, et al. Microarray and Pathway Analysis Reveal Distinct Mechanisms Underlying Cannabinoid-Mediated Modulation of LPS-Induced Activation of BV-2 Microglial Cells. PLoS ONE. 2013 Apr 24;8(4):e61462.

15 Michael TP, Jackson S. The First 50 Plant Genomes. The Plant Genome. 2013;6(2):0.

16. Edwards D, Batley J. Plant genome sequencing: applications for crop improvement. Plant Biotechnology Journal. 2010;8(1):2–9.

17. Trimble WL, Phung LT, Meyer F, Gilbert JA, Silver S. Draft Genome Sequence of *Agrobacterium albertimagni* Strain AOL15. J Bacteriol. 2012 Dec 15;194(24):6986–7.

18. USB stick can sequence DNA in seconds - 17 February 2012 - New Scientist [Internet]. [cited 2013 Dec 27]. Available from: http://www.newscientist.com/article/dn21495-usb-stick-can-sequence-dna-in-seconds.html.

19. Venter JC. A Life Decoded: My Genome, My Life. Viking; 2007. 424 p.

20. Collins FS. The language of life: DNA and the revolution in personalized medicine. New York: Harper; 2010.

21. Van Bakel H, Stout JM, Cote AG, Tallon CM, Sharpe AG, Hughes TR, et al. The draft genome and transcriptome of *Cannabis sativa*. Genome Biol. 2011;12(10):R102.

22. Müller T, Ruppel S. Progress in cultivation-independent phyllosphere microbiology. FEMS Microbiology Ecology. 2013;n/a–n/a.

23. Berendsen RL, Pieterse CMJ, Bakker PAHM. The rhizosphere microbiome and plant health. Trends in Plant Science. 2012 Aug;17(8):478–86.

24. Saito K. Phytochemical genomics--a new trend. Curr Opin Plant Biol. 2013 Jun;16(3):373–80.

Chapter 8

1. Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, et al. HMDB: the Human Metabolome Database. Nucl Acids Res. 2007 Jan 1;35(suppl 1):D521–D526.

2. Allwood JW, Goodacre R. An introduction to liquid chromatography–mass spectrometry instrumentation applied in plant metabolomic analyses. Phytochemical Analysis. 2010;21(1):33–47.

3. Wallwey C, Li S-M. Ergot alkaloids: structure diversity, biosynthetic gene clusters and functional proof of biosynthetic genes. Nat Prod Rep. 2011 Mar 1;28(3):496–510.

4. Degenhardt J, Köllner TG, Gershenzon J. Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. Phytochemistry. 2009 Oct;70(15–16):1621–37.

5. Sirikantaramas S, Taura F, Morimoto S, Shoyama Y. Recent advances in *Cannabis sativa* research: biosynthetic studies and its potential in biotechnology. Curr Pharm Biotechnol. 2007 Aug;8(4):237–43.

6. Stout JM, Boubakir Z, Ambrose SJ, Purves RW, Page JE. The hexanoyl-CoA precursor for cannabinoid biosynthesis is formed by an acyl-activating enzyme in *Cannabis sativa* trichomes. Plant J. 2012 Aug;71(3):353–65.

7. Raharjo TJ, Chang W-T, Choi YH, Peltenburg-Looman AMG, Verpoorte R. Olivetol as product of a polyketide synthase in *Cannabis sativa* L. Plant Science. 2004 Feb;166(2):381–5.

8. Raharjo TJ, Chang W-T, Verberne MC, Peltenburg-Looman AMG, Linthorst HJM, Verpoorte R. Cloning and over-expression of a cDNA encoding a polyketide synthase from *Cannabis sativa*. Plant Physiology and Biochemistry. 2004 Apr;42(4):291–7.

9. Taura F, Tanaka S, Taguchi C, Fukamizu T, Tanaka H, Shoyama Y, et al. Characterization of olivetol synthase, a polyketide synthase putatively involved in cannabinoid biosynthetic pathway. FEBS Letters. 2009 Jun 18;583(12):2061–6.

10. Gagne SJ, Stout JM, Liu E, Boubakir Z, Clark SM, Page JE. Identification of olivetolic acid cyclase from *Cannabis sativa* reveals a unique catalytic route to plant polyketides. PNAS. 2012 Jul 31;109(31):12811–6.

11. Marks MD, Tian L, Wenger JP, Omburo SN, Soto-Fuentes W, He J, et al. Identification of candidate genes affecting Delta9-tetrahydrocannabinol biosynthesis in *Cannabis sativa*. J Exp Bot. 2009;60(13):3715–26.

12. Pertwee RG. Cannabinoid pharmacology: the first 66 years. Br J Pharmacol. 2006 Jan;147(Suppl 1):S163–S171.

13. Izzo AA, Borrelli F, Capasso R, Di Marzo V, Mechoulam R. Non-psychotropic plant cannabinoids: new therapeutic opportunities from an ancient herb. Trends Pharmacol Sci. 2009 Oct;30(10):515–27.

14. Scott KA, Shah S, Dalgleish AG, Liu WM. Enhancing the activity of cannabidiol and other cannabinoids in vitro through modifications to drug combinations and treatment schedules. Anticancer Res. 2013 Oct 1;33(10):4373–80.

15. Russo EB. Taming THC: potential cannabis synergy and phytocannabinoid-terpenoid entourage effects. Br J Pharmacol. 2011 Aug;163(7):1344–64.

16. Mechoulam R, Hanuš L. A historical overview of chemical research on cannabinoids. Chemistry and Physics of Lipids. 2000 Nov;108(1–2):1–13.

17. Mechoulam R, Ben-Zvi Z. Carboxylation of resorcinols with methylmagnesium carbonate. Synthesis of cannabinoid acids. J Chem Soc D. 1969 Jan 1;(7):343–4.

18. Shoyama Y, Yagi M, Nishioka I, Yamauchi T. Biosynthesis of cannabinoid acids. Phytochemistry. 1975 Oct;14(10):2189–92.

19. Sirikantaramas S, Morimoto S, Shoyama Y, Ishikawa Y, Wada Y, Shoyama Y, et al. The gene controlling marijuana psychoactivity: molecular cloning and heterologous expression of Δ1-tetrahydrocannabinolic acid synthase from *Cannabis sativa* L. J Biol Chem. 2004 Sep 17;279(38):39767–74.

20. Flores-Sanchez IJ, Peč J, Fei J, Choi YH, Dušek J, Verpoorte R. Elicitation studies in cell suspension cultures of *Cannabis sativa* L. Journal of Biotechnology. 2009 Aug 20;143(2):157–68.

21. Hanuš LO. Pharmacological and therapeutic secrets of plant and brain (endo)cannabinoids. Medicinal Research Reviews. 2009;29(2):213–71.

22. Obembe OO, Popoola JO, Leelavathi S, Reddy SV. Advances in plant molecular farming. Biotechnology Advances. 2011 Mar;29(2):210–22.

23. Loh WH-T, Hartsel SC, Robertson LW. Tissue culture of *Cannabis sativa* L. and in vitro biotransformation of phenolics. Zeitschrift für Pflanzenphysiologie. 1983 Sep;111(5):395–400.

24. Ranjan R, Patro S, Pradhan B, Kumar A, Maiti IB, Dey N. Development and functional analysis of novel genetic promoters using DNA shuffling, hybridization and a combination Thereof. Zhang Z, editor. PLoS ONE. 2012 Mar 14;7(3):e31931.

25. Houben A, Mette MF, Teo CH, Lermontova I, Schubert I. Engineered plant minichromosomes. Int J Dev Biol. 2013;57(6-8):651–7.

Chapter 9

1. Dayhoff MO. Computer aids to protein sequence determination. Journal of Theoretical Biology. 1965 Jan;8(1):97–112.

2. NCBI dbEST database search December 25, 2013.

3. Dai X, Wang G, Yang DS, Tang Y, Broun P, Marks MD, et al. TrichOME: A Comparative Omics Database for Plant Trichomes. Plant Physiol. 2010 Jan;152(1):44–54.

4 Plant Metabolic Network (PMN), http://www.plantcyc.org/tools/tools_overview.faces on www.plantcyc.org, January 1, 2014.

5. Schomburg I, Chang A, Placzek S, Söhngen C, Rother M, Lang M, et al. BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. Nucleic Acids Res. 2013 Jan;41(Database issue):D764–772.

6. Okonechnikov K, Golosova O, Fursov M. Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics. 2012 Feb 24;091.

7. Ibarra IL, Melo F. Interactive software tool to comprehend the calculation of optimal sequence alignments with dynamic programming. Bioinformatics. 2010 Jul 1;26(13):1664–5.

8. Conesa A, G&#xF6, Tz S. Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics. International Journal of Plant Genomics [Internet]. 2008 Apr 30 [cited 2013 Dec 18];2008.

9. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. Journal of Molecular Biology. 1990 Oct 5;215(3):403–10.

10. The Gene Ontology Consortium. The Gene Ontology: enhancements for 2011. Nucleic Acids Research. 2011 Nov 18;40(D1):D559–D564.

11. Chou PY, Fasman GD. Prediction of protein conformation. Biochemistry. 1974 Jan 1;13(2):222–45.

12. Xiong J. Essential Bioinformatics. Cambridge University Press; 2006. 360 p. The equation was taken from this book, which is perhaps one of the best bioinformatics books.

13. Smith SA, Beaulieu JM, Stamatakis A, Donoghue MJ. Understanding angiosperm diversification using small and large phylogenetic trees. Am J Bot. 2011 Mar 1;98(3):404–14.

14. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol Biol Evol. 2013 Dec 1;30(12):2725–9.

15. Judd WS. Plant systematics: a phylogenetic approach. Sinauer Associates, Incorporated; 2008. 636 p.

16. Yang M-Q, van Velzen R, Bakker FT, Sattarian A, Li D-Z, Yi T-S. Molecular phylogenetics and character evolution of Cannabaceae. Taxon. 2013 Jun 17;62(3):473–85.

17. Page JE, Nagel J. Chapter eight Biosynthesis of terpenophenolic metabolites in hop and cannabis. In: John T. Romeo, editor. Recent Advances in Phytochemistry. Elsevier; 2006. p. 179–210.

18. Rose PW, Bi C, Bluhm WF, Christie CH, Dimitropoulos D, Dutta S, et al. The RCSB Protein Data Bank: new resources for research and education. Nucleic Acids Research. 2012 Nov 27;41(D1):D475–D482.

19. Sillitoe I, Cuff AL, Dessailly BH, Dawson NL, Furnham N, Lee D, et al. New functional families (FunFams) in CATH to improve the mapping of conserved functional sites to 3D structures. Nucleic Acids Res. 2013 Jan;41(D1):D490–D498.

20. Hunter S, Jones P, Mitchell A, Apweiler R, Attwood TK, Bateman A, et al. InterPro in 2011: new developments in the family and domain prediction database. Nucleic Acids Research. 2011 Nov 16;40(D1):D306–D312.

21. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. Nucl Acids Res. 2013 Nov 7;1076.

22. Madej T, Addess KJ, Fong JH, Geer LY, Geer RC, Lanczycki CJ, et al. MMDB: 3Dstructures and macromolecular interactions. Nucleic Acids Research. 2011 Dec 1;40(D1):D461–D464.

23. Wang Y, Geer LY, Chappey C, Kans JA, Bryant SH. Cn3D: sequence and structure views for Entrez. Trends Biochem Sci. 2000 Jun;25(6):300–2.

24. Shoyama Y, Tamada T, Kurihara K, Takeuchi A, Taura F, Arai S, et al. Structure

and function of delta-1-tetrahydrocannabinolic acid (THCA) synthase, the enzyme controlling the psychoactivity of *Cannabis sativa*. JMolBiol. 2012 May 28.

25. Last accessed 1-1-14. http://www.click2drug.org/directory_Docking.html

26. Facchini PJ, Bohlmann J, Covello PS, De Luca V, Mahadevan R, Page JE, et al. Synthetic biosystems for the production of high-value plant metabolites. Trends in Biotechnology. 2012 Mar;30(3):127–31.

27. Sarrion-Perdigones A, Vazquez-Vilar M, Palací J, Castelijns B, Forment J, Ziarsolo P, et al. GoldenBraid2.0: A comprehensive DNA assembly framework for Plant Synthetic Biology. Plant Physiol. 2013 May 13;pp.113.217661.

28. Rocha I, Maia P, Evangelista P, Vilaça P, Soares S, Pinto JP, et al. OptFlux: an open-source software platform for in silico metabolic engineering. BMC Systems Biology. 2010 Apr 19;4(1):45.

Chapter 10

1. James C. A global overview of biotech (GM) crops: Adoption, impact and future prospects. GM Crops. 1(1):8–12.

2 http://www.isaaa.org/resources/publications/briefs/43/pressrelease/default.asp

3. Nicolia A, Manzo A, Veronesi F, Rosellini D. An overview of the last 10 years of genetically engineered crop safety research. Crit Rev Biotechnol. 2013 Sep 16;

4. Fedoroff NV. Will common sense prevail? Trends in Genetics. 2013 Apr 1;29(4):188–9.

5. GMO crops vandalized in Oregon [Internet]. Biology Fortified, Inc. [cited 2013 Dec 31]. Available from: http://www.biofortified.org/2013/06/gmo-crops-vandalized-in-oregon/

6. Godfray HCJ, Beddington JR, Crute IR, Haddad L, Lawrence D, Muir JF, et al. Food Security: The Challenge of Feeding 9 Billion People. Science. 2010 Feb 12;327(5967):812–8.

7. Bakshi A. Potential Adverse Health Effects of Genetically Modified Crops. Journal of Toxicology and Environmental Health, Part B. 2003;6(3):211–26.

8. Herzinger T, Schopf P, Przybilla B, Rueff F: IgE-mediated hypersensitivity reactions to cannabis in laboratory personnel. Int Arch Allergy Immunol 2011;156:423–426.

9. Losey JE, Rayor LS, Carter ME. Transgenic pollen harms monarch larvae. Nature. 1999 May 20;399(6733):214–214.

10. Gatehouse AM., Ferry N, Raemaekers RJ. The case of the monarch butterfly: a verdict is returned. Trends in Genetics. 2002 May 1;18(5):249–51.

11. Ow DW. GM maize from site-specific recombination technology, what next? Current Opinion in Biotechnology. 2007 Apr;18(2):115–20.

12. Powles SODSB. Glyphosate-Resistant Crops and Weeds: Now and in the Future [Internet]. 2010 [cited 2013 Dec 19]. Available from: http://www.agbioforum.org/v12n34/v12n34a10-duke.htm

13. Powles SB, Preston C. Evolved Glyphosate Resistance in Plants: Biochemical and Genetic Basis of Resistance. Weed Technology. 2006 Apr;20(2):282–9.

14. Admin. U.S. Supreme Court Strikes Down Human Gene Patents [Internet]. Science/AAAS | News. 2013 [cited 2013 Dec 19]. Available from: http://news.sciencemag.org/people-events/2013/06/u.s.-supreme-court-strikes-down-human-gene-patents

15. Potrykus I. Golden Rice and Beyond. Plant Physiol. 2001 Mar 1;125(3):1157–61.

16. Beyer P. Golden Rice and "Golden" crops for human nutrition. New Biotechnology. 2010 Nov 30;27(5):478–81.

17. Ow DW, Wet JRD, Helinski DR, Howell SH, Wood KV, Deluca M. Transient and Stable Expression of the Firefly Luciferase Gene in Plant Cells and Transgenic Plants. Science. 1986 Nov 14;234(4778):856–9.

Chapter 11

1. Busch, L. Department of Sociology Institute for Food and Agricultural Standards Michigan State University. Can Plant Biotechnology Become a Truly Social Science? US-EC Task Force on Biotechnology Research "Future of Plant Biotechnology" June 21- 22, 2005 In US-EC Task Force on Biotechnology Abstracts.

2. Niu X, Lin K, Hasegawa PM, Bressan RA, Weller SC. Transgenic peppermint (Mentha×piperita L.) plants obtained by cocultivation with *Agrobacterium tumefaciens*. Plant Cell Reports. 1998 Jan 1;17(3):165–71.

3. Kiselev KV. Perspectives for production and application of resveratrol. Appl Microbiol Biotechnol. 2011 Apr 1;90(2):417–25.

4. Gunn, Karyn Scissum, et al. "Using transgenic plants as bioreactors to produce edible vaccines." Journal of Biotech Research *[ISSN: 1944-3285]* 4 (2012): 92-99.

5. Komarova TV, Baschieri S, Donini M, Marusic C, Benvenuto E, Dorokhov YL. Transient expression systems for plant-derived biopharmaceuticals. Expert Rev Vaccines. 2010 Aug;9(8):859–76.

6. Tokareva O, Michalczechen-Lacerda VA, Rech EL, Kaplan DL. Recombinant DNA production of spider silk proteins. Microbial Biotechnology. 2013;6(6):651–63.

7. Smit HJ. Theobromine and the Pharmacology of Cocoa. Methylxanthines [Internet]. Springer Berlin Heidelberg; 2011 [cited 2013 Dec 1]. p. 201–34. Available from: http://link.springer.com/chapter/10.1007/978-3-642-13443-2_7

8. Argout X, Salse J, Aury J-M, Guiltinan MJ, Droc G, Gouzy J, et al. The genome of *Theobroma cacao*. Nat Genet. 2011 Feb;43(2):101–8.

9. Jirschitzka J, Schmidt GW, Reichelt M, Schneider B, Gershenzon J, D'Auria JC. Plant tropane alkaloid biosynthesis evolved independently in the Solanaceae and Erythroxylaceae. PNAS. 2012 Jun 26;109(26):10304–9.

10. Ogita S, Uefuji H, Yamaguchi Y, Koizumi N, Sano H. RNA interference: Producing decaffeinated coffee plants. Nature. 2003 Jun 19;423(6942):823–823.

11. Park SU, Facchini PJ. Agrobacterium rhizogenes-mediated transformation of opium poppy, *Papaver somniferum* L., and California poppy, *Eschscholzia californica* Cham., root cultures. J Exp Bot. 2000 Jun;51(347):1005–16.

12. Page JE. Silencing nature's narcotics: metabolic engineering of the opium poppy. Trends Biotechnol. 2005 Jul;23(7):331–3.

13. Glenn WS, Runguphan W, O'Connor SE. Recent progress in the metabolic engineering of alkaloids in plant systems. Current Opinion in Biotechnology. 2013 Apr;24(2):354–65.

14. Palazón J, Navarro-Ocaña A, Hernandez-Vazquez L, Mirjalili MH. Application of

Metabolic Engineering to the Production of Scopolamine. Molecules. 2008 Aug 18;13(8):1722–42.

15. Mueller LA, Solow TH, Taylor N, Skwarecki B, Buels R, Binns J, et al. The SOL Genomics Network. A Comparative Resource for Solanaceae Biology and Beyond. Plant Physiol. 2005 Jul 1;138(3):1310–7.

16. Retrived from http://www.nifa.usda.gov/nea/plants/sri/pbgg_sri_outreach.html on Jan. 1, 2014.

17. Stephens. Classroom Activities in Plant Biotechnology. The Plant Health Instructor [Internet]. 2008 [cited 2013 Jan 11]; Available from: http://www.apsnet.org/edcenter/K-12/TeachersGuide/PlantBiotechnology/Pages/default.aspx

18. Rota G, Izquierdo J. "Comics" as a tool for teaching biotechnology in primary schools. Electronic Journal of Biotechnology. 2003 Aug;6(2):85–9.

Index

About the Author

Dr. Zwenger completed his PhD in 2011 and immediately acquired a teaching position at a small liberal arts university. After teaching for two years and at the age of 33, he resigned from his position and left academia. He is currently studying a rare species of *Psychotria* in Central America.

the end