INNER WORKINGS

Genomics blazes a trail to improved cannabis cultivation

Elie Dolgin, *Science Writer*

Marijuana legalization continues apace around the globe with governments the world over now recognizing some medical use for cannabis consumption. But that increasing acceptance belies a hidden truth: Researchers still don't really understand the genetic roots of the plant's biochemical bounty.

Yet, over the past few months, large-scale DNA-sequencing efforts have started to chart the genes responsible for the rich spectrum of phytochemicals produced by both drug and hemp varieties of the *Cannabis sativa* plant, offering key insights for research, industry, and policy. "The genome map is a very powerful step forward," says Jonathan Page, chief scientific officer of Aurora Cannabis, one of the largest cannabis companies in the world. It brings the plant into the modern agricultural era, he adds, noting that "the cutting edge is waiting for cannabis."

Legally, the difference between hemp and its more intoxicating cousin is simple: The plant must contain less than 0.3% per dry weight of tetrahydrocannabinol (THC),

the compound that gets you high, to count as a lawful, agricultural commodity—this, according to the latest US Farm Bill passed in December 2018 (1), which delisted hemp as a controlled substance. (Other countries have set the THC limit as high as 1% or as low as 0.2%.) If the level of THC crosses the semi-arbitrary threshold, the plant becomes classified as marijuana, an illicit drug that in the United States, at least, remains subject to a federal ban. Biologically, however, the distinction between the two forms of cannabis is far less clear-cut, which poses a problem for those breeders looking to stay on the right side of the law.

Hemp breeders want to avoid THC and grow plants that predominantly make cannabidiol (CBD), a trendy non-psychotropic substance with calming properties and other purported health benefits. As such, they have a vested interest in understanding the genetic basis of cannabinoid production, as do marijuana growers in countries such as Canada or states such as California—both large markets for the world's multibillion-dollar legal

Biologically, the distinction between the two forms of cannabis is not clear-cut, posing a problem for those marijuana breeders looking to stay on the right side of the law. Image credit: Shutterstock/Cascade Creatives.

Researchers sequenced Purple Kush, pictured here, and compared it with a variety of hemp to identify the genes underlying cannabinoid production. Image credit: Shutterstock/Pablo Trujillo Novoa.

cannabis industry—that are aiming to cultivate new varieties of THC-laden weed for users of the plant.

With genetic markers linked to desirable traits such as THC content, plant breeders can use DNA analyses to screen seedlings for sought-after properties instead of waiting months for the plants to mature into adults. "That's a huge part of being able to rapidly breed," says Page. Others hope to achieve the same end result through direct manipulation of DNA. Pinning down the genetics, though, "was a very tough nut to crack scientifically," Page says.

Distinction with a Difference

Since the mid-1990s, researchers have known that the acid forms of THC and CBD are alternative derivatives of the same cannabinoid forerunner, cannabigerolic acid (CBGA). However, it was unclear whether the enzymes responsible for converting CBGA into either tetrahydrocannabinolic acid or cannabidiolic acid—which are then transformed into THC or CBD upon heating—were encoded by one gene with two variants or by two tightly linked yet distinct genes.

Researchers started tackling the mystery in the early 2000s when a team from The Netherlands cross-fertilized hemp and marijuana plants. They inbred the progeny and analyzed the next generation's cannabinoid profile to find that THC and CBD production seemed to be under the control of a single gene (2). That remained the dominant hypothesis for about a decade until Page, a researcher at the time with the National Research Council of Canada's Plant Biotechnology Institute in Saskatoon, Saskatchewan, Canada, published the first crude map of the cannabis genome in 2011, along with molecular geneticist Tim Hughes and then-postdoc Harm van Bakel from the University of Toronto.

The rudimentary genome sequence, based on a potent variety of marijuana called Purple Kush, contained a functional THC synthase gene as expected. But it also had several nonworking copies of the CBD synthase gene, each with a premature stop signal or some other mutation that rendered the gene inoperative (3). The large number of relevant gene regions seemed to support a multi-locus model of inheritance. Still, it was impossible to know for sure. "The genome was so

fragmented," says van Bakel, now at the Icahn School of Medicine at Mount Sinai in New York City.

Because of technological limitations of early DNA sequencing machines, that initial map was more like a genetic jigsaw containing 136,290 pieces, each a short segment of the genome that presumably fit together. But the researchers didn't know how. That meant, van Bakel says, that "the data weren't conclusive enough to make a hard statement" about how exactly cannabis plants inherit their chemical profiles. The small but budding community of researchers interested in the genetics of cannabis needed a new and improved genome map.

Mystery Solved?

It would take 7 years, a new kind of "long-read" sequencing technology, and a loosening of US federal regulations that allowed researchers to handle cannabis DNA, starting with the 2014 Farm Bill (4). But late last year, Page reunited with Hughes and van Bakel to publish an updated, high-resolution genome build, now with the plant's 820 million or so DNA letters arranged into 10 discrete chromosomes (5). In addition to resequencing Purple Kush, they mapped the genome of a hemp variety called Finola. Plus, they crossed the two types of cannabis to zoom in on the genetic source of cannabinoid biosynthesis.

On chromosome 6, the researchers could clearly chart two unique cannabinoid synthase genes, each separated from the other by around 20 million nucleotides. Only the DNA of the Purple Kush marijuana plant had a working version of the THC synthase gene, though, and only the DNA of the Finola hemp had a working version of the CBD synthase gene.

That would seem to solve the genetic mystery of cannabinoid output in this enigmatic plant. But one "biochemical niggling issue" remains, says Page: If the hemp genome lacks a working copy of the THC synthase gene, then why does the plant still produce low but detectable levels of the euphoria-inducing molecule that can get hemp growers into regulatory trouble?

The answer may lie in the complex nature of chromosome 6, which is chockful of garbled, repetitive DNA derived from viruses—so-called retroelements that retain the ability to copy themselves and jump to other sites in the genome, dragging along other genes in the process. "These kind of elements are just known for accelerating evolution," says Hughes. It's likely, he says, that retroelements helped ancestral synthase genes duplicate and diverge repeatedly throughout the genome, giving rise to a suite of new genes with new functions. Some, however, may have retained their vestigial THC-making ability, albeit only at low levels.

More Genes, More Possibilities

Looking to better understand gene function, Page, Hughes, and their colleagues, as documented in their recent article (5), characterized one of the duplicated genes by inserting its DNA sequence into the genome of a cultured yeast strain. The engineered yeast cells, when fed a diet of CBGA, then spit out the acid form of cannabichromene (CBC), a rare cannabinoid thought to have antiinflammatory effects. To their delight, the researchers had successfully pinpointed the gene

responsible for CBC synthesis. It was 96% identical to the THC synthase gene at the DNA level and 93% identical at the protein level, yet the gene produced a totally different compound, one that's non-psychoactive with potential antiinflammatory properties.

Many more putative cannabinoid-production genes could be analyzed in this way, notes Tim Harkins, a business development advisor to Medicinal Genomics, one of two cannabis companies to post their own genome analyses to preprint servers late last year (6, 7). "We've identified a plethora of new genes that have gone unannotated," he says.

Ultimately, many companies hope to take some of those genes, transfer the DNA into yeast or bacteria growing in large tanks, feed the genetically modified microbes a steady diet of sugar, and derive pure

> "Within three years, none of the plants that we're growing currently will continue to be produced."
> —Jeremy Plumb

tinctures of CBC or any one of the many other obscure cannabinoids with supposed therapeutic properties—a process akin to brewing beer. "There are rare cannabinoids that you just can't get from the plant in any commercial quantity," says Mike Gorenstein, CEO and chairman of Cronos Group, a Canadian cannabis producer.

Last year, Cronos inked a deal with Ginkgo Bioworks, a Boston-based biotech firm, to produce eight cultured cannabinoids in this way for use in the pharmaceutical and nutraceutical markets. And in February, synthetic biology pioneer Jay Keasling of the University of California, Berkeley, became the first researcher to publish a study describing the complete synthesis of cannabinoids from sugars in yeast (8).

Next Wave

A well-characterized genome also opens the door to the genetic engineering of the plant itself, notes Darryl Hudson, cofounder of InPlanta Biotechnology in Lethbridge, Alberta, Canada. Genetically modifications will be the "next wave" of cannabis breeding, he says.

At Canopy Growth in Evergreen, CO, Director of Genetics Research Rob Roscow claims to have developed a method for making THC-free and CBD-free plants via precision gene editing of the relevant genes. And by targeting other gene pathways with his CRISPR-

based approach—a feat described in patent applications (https://patents.google.com/patent/AU2017250794A1/) but not yet peer reviewed for publication in a journal—Roscow hopes to engineer cannabis varieties that are covered shoot to tip in resinous hairs. These "trichomes" are where the plant produces all its valuable chemicals, but they usually only amass on flowers and adjacent leaves. This would make it easier to produce and retrieve those raw ingredients, Roscow says.

Meanwhile, a Toronto-based company called Trait Biosciences is also taking a transgenic approach to stimulate plant-wide cannabinoid production, albeit with a more traditional method for introducing genetic material to cannabis. But rather than induce trichome growth, Chief Scientific Officer Richard Sayre has developed a way to ramp up cannabinoid levels within the leaf tissue itself. Because the plant has its greatest biomass before it begins to flower—when loads of trichomes form but also many leaves get shed—Sayre expects this approach to allow for "a tremendous increase in yield while also cutting harvest time."

Useful Tools

Already, the next-generation genome maps are proving their worth. For one, the raw data, which are freely available online, are helping breeders create new varieties of cannabis with unique chemical profiles, such as plants with elevated levels of CBC. Access to the full genomic code means that "anyone that has bioinformatics skills or molecular biology skills can develop their own in-house marker-assisted selection assays," says Philippe Henry, head of research and development at Flowr, a cannabis company located in Lake Country, British Columbia, Canada.

And beyond cannabinoids, there are myriad other agronomically important traits that could be improved through the same genetics-guided approach. "There's the flower structure. There's disease resistance. There's nutrient uptake and fertility requirements and harvestability," notes Jonathan Vaught, CEO of Front Range Biosciences in Lafayette, CO. "There's still a lot of opportunity to push the research forward."

Because of those new opportunities, Jeremy Plumb, director of production science at Prūf Cultivar in Portland, OR, predicts radical change coming to the cannabis industry. "Within three years," he predicts, "none of the plants that we're growing currently will continue to be produced, and there will be unbelievable new varieties as a result of marker-assisted hybridization and trait-based selection." Plumb believes that we're "at the beginning of an inferno of new cultivars coming forward."

1 Agriculture Improvement Act of 2018, Public Law 115–334, Sect. 7401 (2018).
2 de Meijer EP, et al. (2003) The inheritance of chemical phenotype in *Cannabis sativa* L. *Genetics* 163:335–346.
3 van Bakel H, et al. (2011) The draft genome and transcriptome of *Cannabis sativa*. *Genome Biol* 12:R102.
4 Agricultural Act of 2014, Public Law 113–79, Sect. 7606 (2014).
5 Laverty KU, et al. (2019) A physical and genetic map of *Cannabis sativa* identifies extensive rearrangements at the *THC/CBD acid synthase* loci. *Genome Res* 29:146–156.
6 Grassa CJ, et al. (2018) A complete Cannabis chromosome assembly and adaptive admixture for elevated cannabidiol (CBD) content. *bioRxiv*: 458083.
7 McKernan K, et al. (2018) Cryptocurrencies and Zero Mode Wave guides: An unclouded path to a more contiguous *Cannabis sativa* L. genome assembly. doi:10.17605/OSF.IO/N98GP.
8 Luo X, et al. (2019) Complete biosynthesis of cannabinoids and their unnatural analogues in yeast. *Nature* 567:123–126.

Dolgin