ISB NEWS REPORT
AGRICULTURAL AND ENVIRONMENTAL BIOTECHNOLOGY

January 2016

REGULATORY NEWS

CONTENTS

In Australia: Victory for GE Crop Farmer, Defeat for Gene Patents, and GE Flies on Trial 1

Improving Water-use Efficiency and Drought Tolerance 4

Towards the Pangenome 7

In Australia: Victory for GE Crop Farmer, Defeat for Gene Patents, and GE Flies on Trial

Phill Jones

For the second time, Australian organic farmers failed to convince a court that their neighbor should be held liable for genetically engineered (GE) crop contamination. The neighbor in the case, Michael Baxter, grew conventional crops and Roundup Ready® canola on his farm, Sevenoaks. A road reserve of about 20 meters separates Sevenoaks and Eagle Rest, where Stephen and Susan Marsh grew oats and raised sheep, both certified as organic by the National Association for Sustainable Agriculture Australia (NASAA).

The case began in 2010 when Michael Baxter harvested his GE canola. Canola can be immediately harvested (direct heading) or harvested by swathing. The latter method is achieved by cutting the canola stalks before the canola seed is fully ripe and laying the stalks in rows called "windrows" on top of the cut stubble in the ground. The canola swaths are allowed to dry by exposure to sun and wind before collection. Baxter's contractor swathed GE canola, and while the cut plants were drying, wind blew some of the GE canola onto Eagle Rest land.

The NASAA decided that the airborne swath incursion and the GE canola seeds scattered across the soil of Eagle Rest posed an "unacceptable risk" of "contamination." The group decertified about 70% of Eagle Rest land. During 2012, the Marshes sued Baxter under common law negligence and the tort of private nuisance. They sought $85,000 in damages due to the decertification.

As detailed in the August 2014 issue of the *Information Systems for Biotechnology News Report*, Justice Kenneth Martin of the Western Australia Supreme Court dismissed both of the Marshes' causes of action. The judge decided that no actual contamination occurred, because the Marshes' sheep would not be affected by eating GE canola seeds, and the Marshes' oats cannot cross-breed with GE canola seeds. In addition, the judge found that the swathing harvest method was "entirely orthodox in its implementation." It was the NASAA's "unjustifiable reaction" and erroneous application of group's own standards that led to the Marshes' damages, the judge said.

The Marshes appealed the decision. During September, two of three judges in the Supreme Court of Western Australia dismissed the appeal.

On appeal, the Marshes alleged that Baxter had acted unreasonably by choosing the swathing method to harvest the GE canola. With regard to the negligence claim, the Marshes asserted that the exercise of reasonable care by Baxter required him to use direct harvesting. But the court's majority were unpersuaded that a reasonable person in Baxter's position would have taken the precaution of direct heading for the benefit of the Marshes.

The Marshes also claimed that the presence of the swaths and seeds from GE canola

VirginiaTech
Invent the Future
Virginia Polytechnic Institute and State University

PUBLISHED BY

Information Systems for Biotechnology

Virginia Tech
1900 Kraft Drive, Suite 103
Blacksburg, VA 24061

Tel. 540-231-3747
Fax 540-231-4434

Subscribe to and access archived issues of the ISB News Report online:

www.isb.vt.edu

Editor:
Ruth Irwin
rirwin@vt.edu

ISB welcomes your comments and encourages article submissions. If you have a suitable article relevant to our coverage of agricultural and environmental applications of genetic engineering, please email it to the Editor for consideration.

constituted an unlawful interference with their use and enjoyment of their land, and was a nuisance. The two judges found several reasons why the nuisance claim should be dismissed. One line of reasoning rested on the principle that a "person who puts their land to an abnormally sensitive use cannot thereby unilaterally enlarge their own rights and obtain a higher right to limit the operations of their neighbours than someone who does not put their land to such a use."

The Marsh and Baxter farms are located in a region where organic farming is, at best, an isolated practice, an abnormally sensitive use of the land, the judges said. "The appellants were, of course, entitled to enter into arrangements which had the effect that their land was being put to an abnormally sensitive use," the judges wrote, "but their neighbours did not then fall under an obligation to limit their farming activities on their own land so as not to interfere with that use of the appellants' land." The judges concluded that the decision to swath the GE canola crop had been based on legitimate agricultural considerations, and performed without any expectation of an incursion onto the Marshes' land.

Outside of the courthouse, Baxter told reporters that the dispute should never have gone to trial; the two farmers should have sorted the problem over a couple beers. "He's an organic farmer," Baxter said. "He can't spray. He can't use chemicals. You know he's got red mite. He's got aphids. He's got rust. He's got all the diseases in the world. Does he worry about that? They blow over the fence. I get them all the time. Do I whinge? Do I complain? No, not at all."

Marsh also met with the press. Stressing that the court's decision was not unanimous, Marsh told reporters that he was considering a further appeal to the Australian High Court. Meanwhile, he may be responsible for many hundreds of thousands of dollars in court costs.

The Trials of GE Medflies in the Land Down Under

The Mediterranean fruit fly, *Ceratitis capitata*, targets more than 250 types of fruits and fruiting vegetables. Female Medflies penetrate the skin of a fruit to lay eggs. Larvae feed on the rotting tissues of the fruit. In Western Australia, the fruit flies cost the horticulture industry millions of dollars every year in lost production and costs of controlling the flies. Pesticides offered one method of control. But growers turned to other techniques after organophosphates were banned from Australian orchards.

The current popular method for controlling Medflies is the Sterile Insect Technique. Male flies are bred and treated with radiation to render them sterile. Then, sterile male flies are released in fruit growing areas that are plagued with Medflies. Female flies that mate with the sterile males fail to produce offspring, leading to a decline in the Medfly population.

The technique has a serious drawback: Irradiated male flies can be weaker and have shorter lives than wild-type males. Because sterile male flies are poor competitors, sterile males must be released in large numbers to increase the likelihood that female flies mate with sterile males, rather than with their wild-type counterparts. A reduction in Medfly population requires a series of releases of sterile males.

Using genetic engineering, United Kingdom-based Oxitec Ltd. developed

an improvement of the Sterile Insect Technique that eliminates the weakening effects of radiation. GE male flies carry two new genes: a "self-limiting gene" and a marker gene. The self-limiting gene prevents female offspring of a GE male from developing to adulthood and laying eggs. The number of females in a population declines with repeated releases of GE males. The marker gene produces a fluorescent protein that enables researchers to track GE male flies and their offspring.

On November 24, the Western Australia Department of Agriculture and Food announced that its researchers will perform an indoor assessment of Oxitec's GE flies in 2016. Mating studies will take place in a tent enclosed within a glasshouse. The GE flies were imported as eggs during October and raised in a laboratory environment. The Department's Horticulture Director David Windsor explained that the "trials will determine whether the Oxitec fruit flies can be reared successfully and cost-effectively and if the males are compatible mates with the pest females." In addition to verifying compatibility between GE males and wild-type females, researchers will decide if GE male flies can successfully compete with wild-type males and irradiated, sterile males. The test results will determine the role of the GE Medflies in new, commercial pest management programs.

The case creates uncertainty about the degree to which a naturally-occurring molecule must be altered before it can be deemed worthy of patent protection.

Australian High Court Pulls the Plug on DNA Patent

Myriad's DNA patent claims continue to inspire new law. Despite years of precedent, the US Supreme Court decided that "genes and the information they encode are not patent eligible . . . simply because they have been isolated from the surrounding genetic material" (*Association for Molecular Pathology v. USPTO* (2013)). For a while, Australian courts treated Myriad's DNA patent claims with kindness in the case, *D'Arcy v. Myriad Genetics, Inc.* Justice John Nicholas of the Federal Court of Australia concluded that a valid patent may be granted for a claim that covers naturally-occurring nucleic acid that has been isolated from human cells. The plaintiff appealed the decision. The five judges of the Federal Court of Australia, however, unanimously decided to dismiss the appeal in 2014. The court not only agreed with Justice Nicholas, but also criticized the US Supreme Court. "Myriad's claim," the judges wrote, "properly considered is not, as the US Supreme Court considered, concerned '*primarily with the information contained in the genetic sequence [rather than] with the specific chemical composition of a particular molecule.*'" The Australian plaintiff appealed again. This time, she won.

During October, the Australian High Court allowed the appeal and ruled that Myriad's DNA claims are revoked. The claims concern nucleotide sequences of the BRCA1 gene linked to susceptibility for breast cancer. The three claims on appeal describe various BRCA1 gene mutants. For example, claim 1 covered an "isolated nucleic acid coding for a mutant or polymorphic BRCA1 polypeptide, said nucleic acid containing in comparison to the BRCA1 polypeptide encoding sequence set forth in SEQ.ID No:1 one or more mutations or polymorphisms selected from the mutations set forth in Tables 12, 12A and 14 and the polymorphisms set forth in Tables 18 and 19."

In the court's view, the claims had nothing to do with chemical, structural, or functional differences between isolated nucleic acids and nucleic acids within cells. "[T]he information stored in the sequence of nucleotides coding for the mutated or polymorphic BRCA1 polypeptide is the same information as that contained in the DNA of the person from which the nucleic acid was isolated," the court said. "It is the existence of that information which is an essential element of the invention as claimed. The product is the medium in which that information resides."

The court went on to address the patentability of cDNA molecules and decided that the same principle applies to cDNA. That is, a cDNA molecule "replicates a naturally occurring sequence of exons."

Like the US Myriad decision, the Australian decision overturns decades of accepted practice about the patentability of DNA molecules. The case creates uncertainty about the degree to which a naturally-occurring molecule must be altered before it can be deemed worthy of patent protection.

Selected Sources

Creese, Chris, "Australia Trials Oxitec's Environmentally Friendly Medfly for Protection of Fruit Crops," *Oxitec* Website (November 24, 2015). Available at: http://www.oxitec.com.

D'Arcy v Myriad Genetics Inc. [2015] HCA 35.

Gregg, Claire and Martin O'Brien, "News from Abroad: Isolated Nucleic Acids Not Patentable in Australia," *Patent Docs* Website (October 7, 2015). Available at: http://www.patentdocs.org.

Marsh v Baxter [2015] WASCA 169.

"Mating Trial to Examine New Control Option for Medfly," *Western Australia Department of Agriculture and Food* Website (November 24, 2015). Available at: https://www.agric.wa.gov.au.

"Organic Farmer Loses WA Court Appeal Against GMO Crop Neighbour," *Australian Food News* Website (September 7, 2015). Available at: http://ausfoodnews.com.au.

Palmer, Kirrily, "Fruit Fly Trial in Western Australia," *Western Australia Department of Agriculture and Food* Website (November 24, 2015). Available at: https://www.agric.wa.gov.au.

Weber, David and Tyne McConnon, "Organic Farmer Steve Marsh Loses GM Appeal for Compensation from Neighbour Michael Baxter," *ABC* Website (September 3, 2015). Available at: http://www.abc.net.au.

Phill Jones
Biotech-Writer.com
PhillJones@nasw.org

Improving Water-use Efficiency and Drought Tolerance

Lai-Sheng Meng

Water is essential for plant growth and development by providing a medium for cellular functions. Globally, water deficiency is increasingly contributing to a significant decline in plant production. Water deficiency causes reduced leaf cell turgor, restricts cell elongation, and postpones development. Therefore, improving water-use efficiency (WUE) and drought tolerance are effective strategies for addressing this problem. In many cases, a key strategy is to genetically engineer genes involved in regulating the stomata number or density, or the root systems. Regulatory genes that activate or deactivate suites of drought-responsive genes or related gene cascades are of particular interest to biotechnologists.

Discovery of YODA regulator

Plants respond to global climate factors such as water, CO_2, and light availability by sensing environmental conditions via mature leaves and adjusting plant stomatal density, stomatal index, and stomatal distribution. Altering the stomata number or density is a longer-term process than regulating stomata movement (opening and closing).

YODA (*YDA*), a MAPKK kinase gene, regulates stomata number (**Fig. 1**) by modulating a fundamental cell fate decision in the Arabidopsis epidermis. YDA acts as a central molecular switch controlling the pathway that determines the number of cells destined to become stomata. BR-INSENSITIVE 2 (bin2) phosphorylates and inhibits YDA, suggesting that YDA is regulated at the post-translation level. However, the molecular basis of *YDA* expression is unclear. Identification of the factors that drive *YDA* transcriptional expression will elucidate the underlying mechanism that restricts *YDA* expression to meristemoid cells that initiate division. Some factors regulating the integration of environmental signaling and stomatal development are still unidentified.

ANGUSTIFOLIA3 (AN3), encoding a homologue of the human transcription co-activator, is reported to be a positive regulator of stomatal development. *YDA* is a target gene of AN3, and AN3 suppresses *YDA* expression at the transcriptional level. AN3-YDA forms a gene cascade. Moreover, AN3 is an integrative point of environmental signaling. Thus, environmental signaling can be integrated into regulation of stomatal development via AN3-YDA gene cascade (**Fig. 1**).

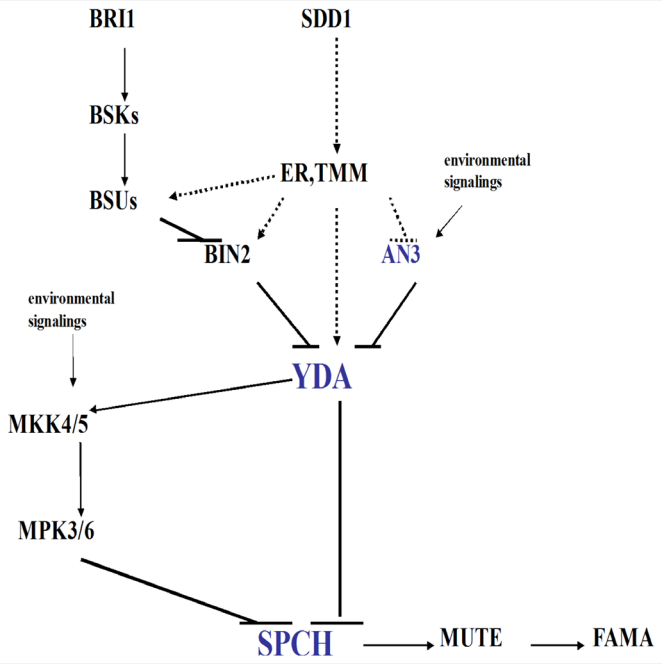

Figure 1. Model illlustrating a proposed AN3-YDA signalling cascade. Brassinosteroid regulates glycogen synthase kinase 3 (GSK3)-mediated inhibition of MAPKK kinase YDA, and then MAPKK kinase YDA regulates inhibition of bHLH SPCH; which forms a BIN2-YDA-MPK3/6-SPCH signaling cascade for regulation of stomatal initiation.

Improving Water-Use Efficiency and drought tolerance

Gas exchange occurs primarily via stomata surrounded by two guard cells. Plants modulate transpiration and water use by adjusting stomatal conductance, which is determined by stomatal movement and density. While the regulation of stomata density is well understood, it is not completely known how water-use efficiency is increased by the modification of stomata density.

Stomata density has a sophisticated influence on plant fitness as well as water use efficiency. The utility of the YDA system (a MAPK signaling cascade) is in identifying new stomatal cell fate regulators. YDA regulates the SPEECHLESS transcription factor at the post-translation level to modulate stomatal development. In addition, utilizing the AN3-YDA gene cascade to adjust stomatal density in a model plant species help tos elucidate the mechanisms by which stomatal density is controlled and reveals the physiological relevance of the large variation in density found in nature.

Arabidopsis thaliana AN3 loss-of-function mutations exhibit enhanced water deficit tolerance and higher integrated WUE through declining daytime transpiration, without a demonstrable reduction in biomass accumulation. *an3* plants have higher instantaneous WUE that is attributable to ~44% lower transpiration and stomatal conductance but equivalent CO2 assimilation. Lower transpiration is closely related to higher YDA expression and an ~48% reduction in abaxial stomatal density. *AN3* expression occurs in abaxial stomata but not in epidermal cells. In epidermis, the protein is localized to the nucleus of subepidermal cells, and its expression is downregulated via water stress.

Altering the *an3* stomatal density directly produces a significant decline in stomatal conductance, which is positively associated with transpiration. Transpiration in the *an3* plants decreases as stomatal conductance falls. Another transpiration factor, Vapor Pressure Deficit (VPD), was similar for all lines under all conditions, which indicates that decreased transpiration in the *an3* plants is not because of differences in VPD. Thus, the decrease in transpiration is attributed to the decline in stomatal density.

The decrease in stomatal density of *an3* plants is proportion to the WUE in plants of the same biomass due to decreased water consumption. Epidermal properties, stomatal conductance, and total leaf area determine total water loss. On the other hand, the carbon gain in photosynthetic tissues in *an3* plants may determine biomass, and the carbon gain is modulated via the net assimilation rate and functional leaf area.

In general, decreasing stomatal conductance in *an3* plants is a water conservation strategy; however, it may result in reduced cumulative photosynthetic activity and limited biomass production. However, in normally unstressed conditions, *an3* plants can grow and develop normally, though their leaf blades display a narrow shape. This characteristic could be very intriguing for crop improvement. Also, *an3* plants do not exhibit growth retardation. We thus suggest that *AN3-YDA* is a good candidate gene cascade to improve drought tolerance in transgenic poplar through the modulation of stomatal development.

Improving root system for drought tolerance

In addition to stomata, plant drought tolerance is affected

by root systems, including traits such as root depth, primary root length, lateral root number, root length distribution, root volume, and root biomass. These traits can be targeted in molecular breeding. The mechanisms whereby root systems affect drought tolerance in plants are complex. For example, in common beans, improving drought adaptation involves a deep and balanced rooting system that increases extraction of soil moisture from greater depths. Efficient water uptake by increasing root growth and development is crucial to plants adapting to drought environments. This water uptake has been attributed to root size (either length or mass), activity, and spatial distribution. A gain-of-function mutant, enhanced drought tolerance1 (edt1), has altered root architecture with deeper roots and more lateral roots. These large root systems enhance the accessibility of water, which might positively contribute to drought tolerance.

The an3 plants display altered root architecture comprised of long primary roots and more abundant lateral roots. The root characteristics of *an3* plants at the heading stage of reproduction show an increase in root diameter compared to wild-type (Col-0). This improvement is based on elongation of root cells, which may be due to an enlarged xylem, and larger cortical cells and epidermis. The elongated root cells in *an3* plants are mainly caused by light signaling. Arabidopsis expresses 14 photoreceptors, most of which are also present in roots. The vessel diameter is closely and positively correlated with better water flux, and larger xylem poses a lower risk of cavitation. Also, improved root cortical aerenchyma enhances drought resistance by decreasing root metabolic costs, permitting greater root growth and water acquisition from drying soil. The relative water content of mid-day leaf in the high root cortical aerenchyma lines is 10% greater than in the low root cortical aerenchyma lines under water stress.

In summary, these findings demonstrate that *an3* plants boost root diameter due to a large root system, which may be due to an enlarged xylem and augmented cortical cell size relative to that of the wile-type (Col-0). These changes in the root architecture systems are also coupled with the down-regulation of the *YDA* genes involved in signal transduction (for example, light signaling), barrier formation (lignin and suberin biosynthesis), and cell development and morphogenesis, which aid the plants in surviving drought stress. These results reveal that *an3* plants enhance drought resistance at the reproduction stage, leading to greater grain yield under drought conditions.

Sources

Meng, L.S and Yao, S.Q. (2015). Transcription co-activator Arabidopsis ANGUSTIFOLIA3 (AN3) regulates water-use efficiency and drought tolerance by modulating stomatal density and improving root architecture by the transrepression of YODA (YDA). ***Plant Biotechnol. J.***.13:893-902.

Meng, L.S. (2015) Transcription Coactivator Arabidopsis ANGUSTIFOLIA3 Modulates Anthocyanin Accumulation and Light-Induced Root Elongation through Transrepression of Constitutive Photomorphogenic1.***Plant, Cell & Envir.*** 38: 838-851.

Meng LS and Liu A. (2015). Light signaling induces anthocyanin biosynthesis via AN3 mediated COP1 expression. ***Plant signaling & behav.*** 10:9, e1001223,DOI: 10.1080/15592324.2014.1001223.

Lai-Sheng Meng
Key Laboratory of Economic Plants and Biotechnology
Kunming Institute of Botany, Chinese Academy of Sciences, 132 Lanhei Road Kunming 650201, PR China.
menglaisheng@mail.kib.ac.cn.

Towards the Pangenome

Agnieszka Golicz and David Edwards

DNA sequencing technology has revolutionised biology, and there are an increasing number of reference genome assemblies being produced for an array of animals, plants, and microbes. The availability of a reference genome for your favourite species is, however, only the first step towards understanding how the hereditary material is related to the observed phenotype of the organism. The field of genomics is busy gaining new insights into the role of genes as well as their regulatory elements. Much has been learnt from the study of genome variation, where differences between individuals can be attributed to particular variations in phenotype.

In this way we are gaining an understanding of the roles and functions of diverse genes in human disease as well as the agronomic performance of crop plants[1]. This research has focused on allelic variation, often in the form of single nucleotide polymorphisms which make relatively minor changes to the amino acid sequence of the resulting protein or the expression pattern of the gene. With the increase in genome sequencing projects, it is becoming clear that these minor variations are only one of many forms of inherited variation resulting in changes in phenotype. One surprising finding is that plants within a species often demonstrate significant differences in gene content and that many of the genes within the genome are dispensable[2]. The concept that individuals within a species all have the same genome is now clearly incorrect, and studies are currently underway in a range of species to understand the complete gene content for the species, known as the species pangenome. The evolution of dispensable genes and their role in crop performance is currently under scrutiny, as this knowledge may greatly assist in improving crop varieties to face the challenge of feeding an increasing global population under a more unpredictable and changing climate.

The pangenome concept was established by Tettelin et al.[3] who produced a pioneering pangenome analysis for a prokaryote *Streptococcus agalactiae*. Since then, pangenome studies have become increasingly popular, with many examples from bacteria[4-7], other microorganisms[8], and humans[9]. Recently pangenomes of rice[10], soybean[11], maize[12], and Brassica[13] were investigated. However, the establishment of plant pangenomes comes not only from dedicated pangenomic studies but also from examination of structural variants, especially copy number variants (CNVs) and presence/absence variants (PAVs).

In an early comparative genomic investigation in *A. thaliana*, analysis of two ecotypes identified 3.4 Mb of sequence which was extremely diverged, missing, or had undergone duplication relative to the reference genome[14]. Subsequent sequencing of eighteen *A. thaliana* genomes identified between 2.1 and 3.7 Mb of sequence that was present in the reference but missing in these new accessions, with on average ~300 novel genes or gene fragments per accession[15]. A broader survey of 80 accessions suggested that a tenth of the genes was absent in one or more accessions compared to the reference sequence[16]. In general, the genes that were subject to presence/absence variation were shorter, younger, and had fewer paralogues than non-variable genes[17]. A comparison of assembled cultivated and wild soybean genomes revealed several thousand PAVs that were absent in both the reference and the wild soybean[18]. A second study demonstrated that presence of several hundred CNVs and over a hundred PAVs in four diverse soybean genotypes[19]. Subsequent *de novo* assembly and comparison of seven diverse accessions of wild soybean relatives found that 80% of the pangenome was present in all accessions, while the remaining 20% was variable and displayed greater sequence variation[11].

Comparison of 2.3 Mb region of two rice accessions *Oryza sativa* ssp *japonica* and ssp *indica* revealed that they differed by 27 genes[20]. A whole genome comparison between *O. sativa* ssp *japonica* and *indica* identified CNVs which ranged in size from one to over 180 Kb, and amounted to 7.6 Mb of sequence[21]. A pangenome study of three divergent rice varieties found that 92% of the genes were core and the remaining 8% were variable[10].

Studies in maize suggest that as much as 20% of genomic regions are not shared between the inbred lines B73 and Mo17[22]. Another comparison identified ~3,800 CNV or PAV sequences between these two maize lines, with several hundred genes present in the B73 genome but absent from Mo17[23]. Comparison of six inbred lines with

the B73 reference identified 296 genes that are present in B73 but absent from at least one of the six other lines, while 570 new genes were found to be absent from B73[24]. A more detailed survey of 27 lines suggested that the reference B73 genome may only capture around 70% of the maize pangenome[25], while a further study of over five hundred diverse maize inbred lines identified 8681 representative transcript assemblies which were not present in B73[12].

The production of a pangenome for a species is not straight forward. One obvious approach would be to assemble the genomes of multiple individuals and compare their predicted gene content. The major challenges to this approach are the difficulties in producing high quality assemblies and the variation in gene prediction results. Draft genome assemblies can be produced, even for complex species, relatively quickly and at low cost; however, these assemblies rarely reflect the total gene content as repetitive regions frequently collapse during the assembly process, leaving only a subset of the genes. This is particularly challenging for polyploid genomes and for gene families that share a high degree of sequence identity. The observed difference in gene content between individuals may therefore be due to differences in assembly rather than true biological variation. The failure of consistent genome annotation also introduces error in pangenome analysis, as assemblies of very similar genomes often result in differences in gene prediction. This is observed when a gene appears to be missing from the annotated gene content of one individual, but comparison at the genome sequence level identifies the missing gene as present but unannotated.

Given the challenges of whole genome assembly comparison, there are several alternative strategies for eukaryote pangenome assembly, including iterative assembly and k-mer based methods. For k-mer based methods, the pangenome sequence is stored as a coloured *de Bruijn* graph[26,27]. The genomes are decomposed into segments and the relationships between segments can be traced by following edges of the graph. For the iterative mapping and assembly approach, a single genome is used as a starting reference. Then the reads from other genomes are sequentially mapped, assembled, and added to the reference, producing a non-redundant pangenome[2].

The selection of appropriate individuals that capture the majority of diversity within a species is crucial to successful pangenome study design. The selection of a small number of closely related individuals will result in significant underestimation of the pangenome size. On the other hand, the selection of diverse individuals gives more realistic estimates[2]. While most studies examine the pangenome across a defined species, it is not inconceivable for future projects to aim to identify the presence and distribution of all genes and allelic variants across all plant species. While this may seem ambitious given the current technology, the rapid and continued advances in genome sequencing will make achieving this goal inevitable.

Dispensable genes are increasingly associated with important heritable traits. In humans, several associations between CNVs, PAVs, and health and disease have been described[28-30]. Less is known about how CNVs and PAVs influence phenotype in plants, though some examples include metabolite production, flowering time, submergence tolerance, phosphorus uptake, and biotic stress response[10,19,31-37].

The opium poppy genome hosts a 10 gene cluster that is only present in plants producing noscapine, an antitumor alkaloid[31]. Wheat genes that regulate flowering by changing photoperiod response (Ppd-B1 alleles) or vernalization requirement (Vrn-A1 alleles) demonstrate copy number variation[32]. In rice, the Sub1A gene is absent in varieties that are not submergence tolerant[10], and the selection for the presence of this gene has led to the production of diverse varieties that tolerate extended periods of submergence, supporting food security in developed and developing countries. Another gene in rice encoding a protein kinase Pstol1 is associated with phosphorus uptake efficiency and demonstrates presence/absence variation; the gene is absent in cultivars intolerant to phosphorus starvation[10]. In addition to PAVs associated with abiotic stress, a significant number of biotic stress response genes demonstrate presence/absence variation in a range of species[19,33-37].

It has been suggested that many genes that demonstrate CNV/PAV in plants may not have a large phenotypic effect, because they are members of large multigene families, where at least partial redundancy of function is expected between paralogues[38]. However a

different study demonstrated lack of enrichment of large gene family members within CNVs[19].

The presence of CNVs/PAVs within gene families may have implications for heterosis[38]. Each CNV/PAV gene in the family may contribute to the functionality, so the loss of one family member may result in a small effect on phenotype as other gene copies provide compensatory function. However, the additive effect of many absent genes could result in decreased vigour, which would be alleviated in hybrids, resulting in hybrid vigour[38]. In addition, unique genes displaying PAV may also contribute to heterosis on a single gene rather than gene family level. Considering this, an understanding of CNVs/PAVs in the entire population available for the breeding of hybrid crops may accelerate the production of improved higher yielding varieties[38].

While the sequencing of the first plant genomes have moved plant biology into the genomic era, we still have much to learn about the natural variation found in the genomes of diverse plant species, knowledge that can greatly assist in the production of new crops to support food security for future generations.

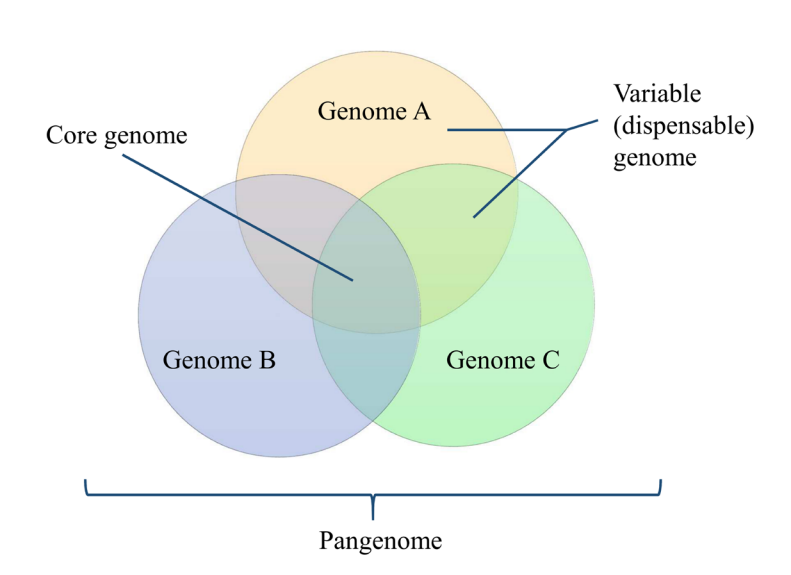

Figure 1. Schematic depiction of pangenome. Core genome is present in all samples, while variable (dispensable) genome is present in some samples only.

References

1. Huang, X. and B. Han, Natural Variations and Genome-Wide Association Studies in Crop Plants. *Annual Review of Plant Biology,* 2014. **65**(1): p. 531-551
2. Golicz, A.A., J. Batley, and D. Edwards, Towards plant pangenomics. *Plant Biotechnology Journal*, 2015: p. n/a-n/a.
3. Tettelin, H., et al., Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome". *Proc Natl Acad Sci USA*, 2005. 102(39): p. 13950-13955.
4. Baddam, R., et al., Genome dynamics and evolution of Salmonella Typhi strains from the typhoid-endemic zones. *Sci. Rep.*, 2014. 4.
5. Liu, F., et al., Comparative genomic analysis of Acinetobacter baumannii clinical isolates reveals extensive genomic variation and diverse antibiotic resistance determinants. *BMC Genomics*, 2014. 15(1): p. 1163.
6. Zhou, Y., et al., Phenotypic and Genotypic Analysis of Clostridium difficile Isolates: a Single-Center Study. *Journal of Clinical Microbiology*, 2014. 52(12): p. 4260-4266.
7. Donati, C., et al., Structure and dynamics of the pan-genome of Streptococcus pneumoniae and closely related species. *Genome Biology*, 2010. 11(10): p. R107.
8. Read, B.A., et al., Pan genome of the phytoplankton Emiliania underpins its global distribution. *Nature*, 2013. 499(7457): p. 209-213.
9. Li, R., et al., Building the sequence map of the human pan-genome. *Nat Biotech*, 2010. 28(1): p. 57-63.
10. Schatz, M., et al., Whole genome de novo assemblies of three divergent strains of rice, Oryza sativa, document novel gene space of aus and indica. *Genome Biology*, 2014. 15(11): p. 506.

11. Li Y.-h et al., De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nat Biotech,* 2014. **32**(10): 1045-1052.
12. Hirsch, C.N., et al., Insights into the Maize Pan-Genome and Pan-Transcriptome. *The Plant Cell,* 2014. **26**: p. 121-135.
13. Lin, K., et al., Beyond genomic variation - comparison and functional annotation sof three Brassica rapa genomes: a turnip, a rapid cycling and a Chinese cabbage. *BMC Genomics*, 2014. **15**(1): p. 250.
14. Ossowski, S., et al., Sequencing of natural strains of Arabidopsis thaliana with short reads. *Genome Research*, 2008. **18** p. 2024-2033.
15. Gan, X., et al., Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*. *Nature*, 2011. **477**(7365): p. 419-423.
16. Tan, S., et al., Variation of presence/absence genes among Arabidopsis populations. *BMC evolutionary biology,* 2012. **12**: p. 86.
17. Bush, S.J., et al., Presence/absence variation in A. thaliana is primarily associated with genomic signatures consistent with relaxed selective constraints. *Molecular Biology and Evolution*, 2014. **31**: p. 59-69.
18. Lam, H.M., et al., Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nature Genetic*s, 2010. **42**: p. 1053-U41.
19. McHale, L.K., et al., Structural Variants in the Soybean Genome Localize to Clusters of Biotic Stress-Response Genes. *Plant Physiology,* 2012. **159**(4): p. 1295-1308.
20. Han, B. and Y. Xue, Genome-wide intraspecific DNA-sequence variations in rice. *Current Opinion in Plant Biology*, 2003. **6**(2): p. 134-138.
21. Yu, P., et al., Detection of copy number variations in rice using array-based comparative genomic hybridization. *BMC Genomics*, 2011. **12**(1): p. 372.
22. Morgante, M., et al., Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. *Nature Genetic*s, 2005. **37**: p. 997-1002.
23. Springer, N.M., et al., Maize Inbreds Exhibit High Levels of Copy Number Variation (CNV) and Presence/Absence Variation (PAV) in Genome Content. *PLoS Gene*t, 2009. **5**(11): p. e1000734.
24. Lai, J., et al., Genome-wide patterns of genetic variation among elite maize inbred lines. *Nat Gene*t, 2010. **42**(11): p. 1027-1030.
25. Gore, M.A., et al., A first-generation haplotype map of maize. *Scienc*e, 2009. **326**(5956): p. 1115-1117.
26. Marcus, S., H. Lee, and M.C. Schatz, SplitMEM: a graphical algorithm for pan-genome analysis with suffix skips. *Bioinformatics*, 2014. **30**(24): p. 3476-3483.
27. Iqbal, Z., et al., De novo assembly and genotyping of variants using colored de Bruijn graphs. *Nat Genet,* 2012. **44**(2): p. 226-232.
28. Henrichsen, C.N., E. Chaignat, and A. Reymond, Copy number variants, diseases and gene expression. *Human Molecular Genetics*, 2009. **18**(R1): p. R1-R8.
29. Zhang, F., et al., Copy Number Variation in Human Health, Disease, and Evolution. *Annual Review of Genomics and Human Genetics*, 2009. **10**(1): p. 451-481.
30. Conrad, D.F., et al., Origins and functional impact of copy number variation in the human genome. *Nature*, 2010. **464**(7289): p. 704-712.
31. Winzer, T., et al., A Papaver somniferum 10-Gene Cluster for Synthesis of the Anticancer Alkaloid Noscapine. *Scienc*e, 2012. **336**(6089): p. 1704-1708.
32. Díaz, A., et al., Copy Number Variation Affecting the Photoperiod-B1 and Vernalization-A1 Genes Is Associated with Altered Flowering Time in Wheat (Triticum aestivum). *PLoS ONE*, 2012. **7**(3): p. e33234.
33. Bakker, E.G., et al., A Genome-Wide Survey of R Gene Polymorphisms in Arabidopsis. *The Plant Cell*, 2006. **18**(8): p. 1803-1818.
34. Cook, D.E., et al., Copy Number Variation of Multiple Genes at Rhg1 Mediates Nematode Resistance in Soybean. *Science*, 2012. **338**(6111): p. 1206-1209.
35. Gonzalez, V., et al., High presence/absence gene variability in defense-related gene clusters of Cucumis melo. *BMC Genomic*s, 2013. **14**(1): p. 782.
36. Shen, J., et al., Unique evolutionary mechanism in R-genes under the presence/absence polymorphism in Arabidopsis thaliana. *Genetics*, 2006. **172**: p. 1243-50.
37. Xu, X., et al., Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. *Nature Biotechnology*, 2012. **30**: p. 105-111.
38. Swanson-Wagner, R.A., et al., Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. Genome Res, 2010. **20**(12): p. 1689-1699.

Professor Dave Edwards
School of Plant Biology
University of Western Australia
Perth, 6009 Australia
Dave.Edwards@uwa.edu.au