**Genetics in action.**

**Dr. Jessica Regan Lucas**

**Why study Arabidopsis?**

Arabidopsis thaliana is a flowering plant that many scientists across the globe use as a model system for studying plants. Model organisms are useful for scientists, because information gained from models can be extrapolated to other organisms. Model organisms are also useful because numerous scientists study the exact same organism, and they share their information and resources with one another which accelerates the rate of discovery.

Various model systems have been developed, and they all share a suite of characteristics that make them amenable for research. Of course, each model system has undesirable caveats that must be considered when designing an experiment. In general, all model organisms are easy and cheap to grow. Model organisms grow quickly and reproduce readily. Geneticists prefer model organisms that have a small genome and high reproductive output (meaning that they make lots of babies).

Arabidopsis is a good model system for biologists because you don’t need a green thumb to grow it. Arabidopsis is often considered a weed because it will grow with very little attention. Greenhouses are generally used to grow plants, and Arabidopsis grows easily in a small, indoor spaces without expensive climate controls. Depending upon the experiment, Arabidopsis does not require fancy soil, lighting, temperature or humidity controls. Arabidopsis researchers commonly convert closets to growth rooms by supplementing the closet with additional fluorescence lighting.

A mature Arabidopsis plant is about 16 inches tall, and nine plants can easily be grown side-by-side in a 3 cubic inch pot. The entire life cycle of an Arabidopsis plant is about 6 weeks. This means that a month and a half passes while a single Arabidopsis seed germinates, grows vegetatively, begins sexual reproduction, produces offspring, and disperses its progeny. For a plant, a 6-week life cycle is quite short as many trees live for 20 years before reproducing. Arabidopsis produces many offspring, and a single plant can produce up to 1000 seeds. This is important because genetic research often requires sorting through large numbers of offspring to determine segregation ratios.

Arabidopsis reproduces sexually with itself and this is a tremendous experimental advantage. Plant sexual reproduction requires pollination followed by the fusion of a sperm and egg cell. In Arabidopsis and other flowering plants, fertilization occurs in the flower. The eggs are inside the ovary / carpel / pistil, and the sperm are within the pollen grains. Fertilization requires pollination in which the pollen grains land on the stigma, germinate and grow down the style. Plants have evolved many ways to accomplish pollination. Some plants spread their pollen through the wind, while other plants entice insects to carry their pollen from flower to flower. Arabidopsis needs no outside help for pollination, as it pollinates and fertilizes itself. This is a boon for scientists because they do not need to spend time and effort ensuring that progeny will be produced.

Another reason that Arabidopsis was chosen as a model system is that it has a small genome. The small genome size enabled a consortium of scientists to fully sequence the Arabidopsis genome in 2000. The Arabidopsis genome is 135Mbp (megabase pairs, 1 Mbp = 1 million bases) separated into five chromosomes. Analysis of the genome revealed nearly 28,000 genes and currently only less than 6,000 of those genes has a predicted function. Many of the Arabidopsis genes are like genes in other organisms, however some genes are only found in plants. The Arabidopsis genome generally contains at least two copies of every gene.

Most crop plants of economic importance have significantly larger, more complicated genomes that confound scientists. Complete genomic sequences for many important food plants are not freely available to the public, however it is likely that agricultural companies have collected this data already.

The knowledge of complete genome sequences is useful because it enables people to study the effect of DNA mutations on the phenotype of the plant. Since humanity began to understand that DNA is the hereditary information in the early 1900’s, scientists have been striving to determine the function of every important gene (before this, there was significant debate on whether DNA or proteins was the heredity information).

**Why study knock-out mutants?**

Much information is gained by studying the anatomy, physiology, molecular biology and genetic basis of diseases. This information is important diagnosing, treating and developing cures for life-threatening conditions. In addition, we often learn new aspects about the human body by studying patients afflicted with diseases. For instance, much of what we know about neuroanatomy comes from people who have suffered tremendous brain damage and survived. In the case of depression, scientists have discovered that allelic differences in genes that produce proteins involved in the electron transport chain of mitochondria. This correlation between depression and less effective mitochondria suggests that energy homeostasis may be disrupted in depressed individuals. Whether depression can be alleviated by increasing mitochondrial activity is an open question at this point.

An alternative to studying diseases in affected people is to determine the phenotypic outcome of mutations in genes of interest. In other words, scientists purposely create mutant organisms to study the outcome of the mutation on the organism. This genetic strategy is akin to reverse engineering, meaning that if you want to understand a machine you carefully remove single pieces of the machine one at a time, and then you determine the impact on the machine (does it still work?). Much information has learned through this process of “reverse genetics”. Scientists are only able to do reverse genetics, mutating specific genes, because of recent major technical advances in genome sequencing that has revealed the full genetic code of many organisms. Before full genome sequences were available, scientists would perform forward genetics by randomly inducing gene mutations by exposing organisms to DNA damaging agents like UV light, gamma radiation, or chemicals. There are advantages and disadvantages to reverse and forward genetics.

Reverse genetics is powerful because we scientists know that there are thousands of genes that we are clueless about their function. This information has compelled many scientists to use reverse genetics to uncover the role of these genes. Genome sequencing across the tree of life has also shown that most organisms have many of the same genes. This fact has allowed scientists to use model organisms like yeast, rats and plants to make inferences about human biology. While it may seem absurd to think that studying brain development in a zebrafish would produce any useful information about humans, it is true. Many discoveries about human biology have been first elucidated in Arabidopsis, such as the importance of epigenetics in development and reproduction. Keep mind that humans are not comfortable doing genetic experiments on other humans.

In reality, every scientist works in a group to understand a specific biological question. Some scientists prefer to study particular genes such as RNA-binding proteins, while others choose to study certain process like asexual reproduction. Each lab group makes discoveries and then communicates their research by delivering seminars and writing papers describing their work that are published in scientific journals. A huge amount of information is discovered every year, and it is very difficult to keep up with all the news! Bioinformatics is a field of science that uses computers and software to organize and catalog huge amounts of experimental data to help scientists sort through relevant information to understand biological systems.

**Arabidopsis Knock-out Library**

A group of California scientists were funded by the federal government to create a community resource: a set of mutant Arabidopsis seeds in which every seed line contains a genetic mutation that knocks out a single gene. These mutant plants were created by randomly inserting a large piece of DNA into the Arabidopsis genome using the bacterium Agrobacterium tumefaciens. The inserted DNA does not code for a protein, and instead this DNA interferes with transcription and translation of the genes in which it is inserted within. The scientists catalogued the location of each piece of inserted DNA and made this information freely available on the T-DNA Express Website (signal.salk.edu/cgi-bin/tdnaexpress?gene). Anyone can order this mutant seeds lines for reverse genetics experiments. These seed lines are insertion mutants, and many of them are knock-outs mutants. Knock-out means that there is no protein produced in the mutant. However, it is possible that the insertion alleles are actually knock-downs (a lesser amount of protein is made or a truncated version is produced).

**Additional Information Reading:**

**Arabidopsis as a model organism.**

From The National Science Foundation, a federal organization that supports science.

https://nsf.gov/bio/pubs/reports/arabid/chap1.htm

Plants are vital to our existence. They provide the oxygen we breathe, the food we eat, the fibers for our clothes, the materials to build our homes, and the raw goods for our industries. A quarter of our medicinal drugs comes from plant species. The paper on which this report is printed is a plant product.

Yet, despite the important contributions of plants to our standard of living, far less is known about them than about mice, flies, or the bacteria that inhabit our intestines. We need to learn more about how plants grow and develop; how they produce useful chemicals; how they protect themselves from pests; and how they sense, respond to, and even alter our environments. One way to learn these things is through study of a plant's genes. The information that plants use to grow and develop, and to interact with their environment, is coded in their genomes. To fully understand plants, we need to read and interpret their genomic information. In the 1980s, there was a growing awareness that significant investments in studies of many different plants, such as corn, oilseed rape, and soybean, were diluting efforts to fully understand the basic properties of all plants. Scientists began to realize that the goal of completely understanding plant physiology and development is so ambitious that it can best be accomplished by turning to a model plant species that many scientists then study. Fortunately, because all flowering plants are closely related, the complete sequencing of all the genes of a single, representative, plant species will yield much knowledge about all higher plants. Similarly, discovery of the functions of the proteins produced by a model species will offer much information about the roles of proteins in all higher plants.

During the last 8 to 10 years, *Arabidopsis thaliana* has become universally recognized as a model plant for such studies. Although it is a non-commercial member of the mustard family, it is favored among basic scientists because it develops, reproduces, and responds to stress and disease in much the same way as many crop plants. What's more, *Arabidopsis* is easy and inexpensive to grow, and produces many seeds; this allows extensive genetic experiments, often involving tens of thousands of plants. Also, *Arabidopsis* has a comparatively small genome, thereby simplifying and facilitating genetic analysis. Compared to other plants, it lacks the repeated, less-informative DNA sequences that complicate genome analysis.

Initially, there was much debate about whether an improved understanding of *Arabidopsis* would help in the breeding of commercial crops, and much controversy over decisions to devote limited resources to this non-commercial species. However, the many advances reported over the past few years offer clear evidence that this plant is not only a very important model species for basic research, but also extremely valuable for applied plant scientists and plant breeders. Publications on *Arabidopsis* in top-quality journals are increasing exponentially, following substantial increases in investment by many governments. In the United States, for example, the U.S. Department of Agriculture, the Department of Energy, the National Institutes of Health, and the National Science Foundation collectively supplied US$7.5 million in 1990 for *Arabidopsis* research and US$22 million in 1993. And the European Community has invested a significant portion of its biotechnology research resources to *Arabidopsis* genome research over the last 5 years. In fact, many of the world's leading laboratories in plant science have initiated programs using *Arabidopsis*, and many young plant scientists have chosen to start their careers using this species.

But how can discoveries with *Arabidopsis* contribute to the development of improved crops? Simply put, once a gene has been discovered in *Arabidopsis*, the equivalent gene may be found more easily in other plants. Thus, the function of many genes isolated from crop plants can be better understood via study of their *Arabidopsis* homologues. So knowledge gained from *Arabidopsis* on the defense mechanisms against pathogens, for example, can be used directly to develop disease-resistant plants in other species.

Genetic comparisons between *Arabidopsis* and crop species are increasing, as shown by the large number of *Arabidopsis* publications cited for 1993 that also involved studies of crop plants such as soybean, rice, maize, wheat, barley, rye, pepper, tomato, potato, cotton, or sorghum. There is ample reason to believe that, in the coming years, *Arabidopsis* will serve more and more as a resource base for breeders of crop plants and as a model plant species that furthers the knowledge of plant scientists worldwide.

**Utility of knock out organisms and reverse genetics.**

From http://www.nature.com/scitable/topicpage/scientists-can-analyze-gene-function-by-deleting-6526138

Scientists Can Analyze Gene Function by Deleting Gene Sequences

One way to understand the function of a gene is to observe a biological system that lacks that gene. But what is the best system to use? When studying human genes, researchers typically employ biological systems that approximate these genes and their functions as closely as possible. In particular, researchers often turn to mice because of all the various model organisms most commonly used in the lab (e.g. fruit flies and yeast), mice have the genome that most closely resembles that of humans. Consequently, manipulation of genes within the mouse genome has proven an effective method for learning about human gene functioning. Indeed, experimentally removing or altering certain genes within a mouse allows for the examination of a biological system with specific gene alterations.

What can mice reveal about human gene function?

On the outside, humans and mice look nothing alike. However, human and mouse chromosomes share many of the same genes. In fact, 99% of the 20,000 to 25,000 genes in humans have a similar mouse counterpart. This high degree of genetic similarity between humans and mice offers researchers a unique approach for understanding human gene functioning.

Ethical considerations prevent researchers from genetically engineering humans that lack a given gene for the sole purpose of learning how that gene functions. However, because so many mouse genes are similar to human genes, geneticists can generate **knockout mice** in which the mouse counterpart of a human gene of interest is deleted or disrupted. The term "knockout mouse" may at first conjure images of a mouse boxing champion or beauty queen. However, geneticists use the term to refer to a mouse that has been genetically engineered such that at least one of its genes is functionally inactivated (i.e. the inactivated gene is "knocked out"). If a mouse gene has a high degree of similarity to a human gene, researchers can predict that the two genes carry out related functions. Therefore, by generating a knockout mouse without a gene of interest, these scientists may be able to determine the functions carried out by the related human gene.

In some cases, knockout mice exhibit phenotypes that mimic symptoms associated with human disease, including cancer, diabetes, obesity, cardiovascular disease, and neurodegenerative disease. In these cases, a knockout mouse is referred to as a **mouse model**of that form of human disease.

**Plant research is important for human health.**

Excerpted from http://www.sciencedirect.com/science/article/pii/S0092867408007022

*Arabidopsis thaliana*, the reference species for plant biology and a key model system for all of biology, has had a greater impact on human health research than may seem evident at first glance. In this Essay, we highlight examples where research using *Arabidopsis* has informed the identification of a prototype protein or domain involved in a human disease, human development, or other important aspects of human biology. These examples do not include the enormous impact that *Arabidopsis* research has had on plant biology, improving food security, and alleviating malnutrition, which is a major threat to human health and is the basis of 50% of human disease worldwide. We make the case that well-chosen, deep investment into research using plant genetic models will help to elucidate basic life processes and to illuminate the evolutionary plasticity of cellular pathways and networks.

The *Arabidopsis thaliana* genome sequence was completed in 2000. Three years later, with the reporting of the annotated human genome sequence, it became evident that a majority of human genes that were suspected or known to play a role in disease had orthologs in *Arabidopsis* (<http://mips.gsf.de/proj/thal/db/tables/disease.html>). This marked degree of similarity in “disease genes” is comparable to that observed in other model organisms. For example, among cancer genes, 70% of genes implicated in cancer have *Arabidopsis* orthologs with E-value cutoffs of less than Eˆ10, whereas the percentage of orthologs in the fruit fly *Drosophila melanogaster* is 67%, in the worm *Caenorhabditis elegans* is 72%, and in the budding yeast *Saccharomyces cerevisiae* is 41%. In one sense, this high percentage of shared genes is not surprising given that development and disease follow the normal and abnormal activities, respectively, of proteins that serve basic cellular functions. But with 1.6 billion years for *Arabidopsis* and humans to have diverged, is it too much to expect that the original functions of these gene products survived intact? Are the distinct body plans and life strategies of humans and plants so different that the respective human and plant orthologs were not similarly constrained in their evolution and could not conceivably function in a similar way? We answer “no” to both questions by highlighting several examples where *Arabidopsis* research led the way in the discovery or analysis of genes and processes of importance in human health. Many model systems greatly impacted human health, and we argue that *Arabidopsis* is a part of this diversified portfolio of tools needed to understand basic cellular processes.

Arabidopsis research helped to understand:

Innate Immunity and Intracellular Receptors

Light Signaling, Protein Degradation, and Cancer

Cryptochromes and the Circadian Clock

The Ubiquitin Cycle, and Alzheimer's Disease

Argonautes and RNA Silencing

DNA Methylation

Leucine-Rich Receptor Kinases, Ion Transport, and G Protein Signaling

Natural Genetic Variation

**Links**

List of all model organisms:

<https://en.wikipedia.org/wiki/List_of_model_organisms>

Comparison of land plant and algal genome sizes: https://en.wikipedia.org/wiki/List\_of\_sequenced\_plant\_genomes

Arabidopsis genomes annotation data: <https://www.arabidopsis.org/portals/genAnnotation/gene_structural_annotation/annotation_data.jsp>