Mendel's Genetics

Floyd A. Reed

April 24, 2017

Classical genetics starts with the independent rediscovery of Gregor Mendel's work by three different people in 1900 (Carl Correns, Hugo de Vries, and Erich von Tschermak). In many ways genetics was an idea who's time had come; with enough active research by various labs in biology it became inevitable that genetics would be (re)discovered. Gregor Mendel was originally named Johann and attended the University of Olomouc where animal breeding—especially sheep and wool production—was studied. This was backed by research funding in the form of a tax that was collected from local wool merchants. However, his name was changed to Gregor when he joined an abbey in order to pay for his education. He went to the University of Vienna and studied physics from Christian Doppler, of the Doppler effect, then returned to the abbey to teach physics. Mendel had interests in beekeeping, astronomy, and meterology (most of his publications were in meterology). He conducted breeding experiments with mice, bees, and various species of plants. In 1854 Mendel began his famous pea (*Pisum sativum*) experiments in the abbey garden, which culminated in his publication on the laws of heredity in 1866.[1] In 1867 he became the abbot of the monastery, his experiments ended, and the rest of his life was consumed with administrative work.

It is often speculated what might have happened if his work was more appreciated and understood early on. He might have influenced Charles Darwin, who was developing his theories of evolution at the same time but without knowledge of the mechanism of inheritance, and others, and accelerated the fields of biology. However, his work was largely forgotten until 1900.

With hindsight some of the properties of genetics, such as the dominance of certain phenotypes, might seem obvious. However, it is worth keeping in mind that despite occurring in the world all around us, these properties

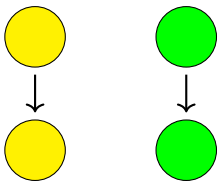
[1] `http://www.esp.org/foundations/genetics/classical/gm-65.pdf`

Figure 1: Mendel is apparently holding up part of a pea plant to explain genetics to his fellow monks. One monk is taking notes; however, many do not appear to be paying attention and will fail the exam. Author Hugo Iltis, published 1932, file link, license.

of genetics were not discovered until Mendel came along, and were subsequently almost forgotten. This is a good opportunity to talk about the scientific method. Science is a process in which we form the simplest possible ideas (hypotheses) about how nature works based on observations, test these hypothesis with new observations, and refine our ideas, often with added layers of complexity, motivated by experimental results. This is an iterative process that occurs round after round, of formulating hypotheses and testing them in new experiments, in order to refine our understanding. Finally—and this is very important—these results are then communicated to other scientists, and other scientist must also make time to seek out and listen, for independent testing and building additional hypotheses—sometimes integrating across disciplines.

Science is about keeping an open mind and to allow our thinking to be shaped by observations of the world around us with logical inference about the forces that drive the processes involved. If we are wrong we want to be able to discover that we are wrong. However, it is a mistake to think that human intuition and motivation do not influence science. If we were not creative in our thinking we couldn't build testable theories. Often the best scientists are the ones who use intuition and creativity to ask the *right kinds* of questions. It is important to keep in mind that there are an infinite number of possible hypotheses to explain a given data set. It is often most useful to focus on the simplest possible hypothesis to explain the data. Finally, no matter how simple or elegant a theory is, it is useless from a scientific perspective if it does not result in a testable hypothesis. It is most useful to focus on hypotheses that make different predictions for as-yet-unobserved but experimentally produceable data.

Mendel had different "true breeding" lines of plants. He observed that peas with green seed color had offspring with green seeds and that plants with yellow seeds had offspring with yellow seeds.



This was stable and continued generation after generation while these lines were maintained separately. Based on this we might hypothesize

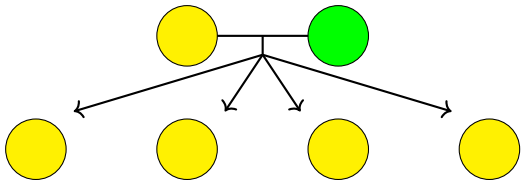
H_1: There is an "inheritance factor" inherited from the parents that determines seed color (versus a non-heritable effect such as differences in the soil they were grown in, *etc.*)

What would happen if we crossed the two lines together? We might suspect that the green and yellow colors were blended to an intermediate color or we might expect haphazard inheritance of the inheritance factor from one parent or the other—a mix of yellow and green seeded offspring. This divides H_1 into two sub-hypotheses.

H_1a: There is an "inheritance factor" inherited from both parents that is blended together to determine seed color.

H_1b: There is an "inheritance factor" inherited randomly from either one parent or the other that determines seed color.

Importantly, we made these hypotheses before observing the new results. Now we cross the yellow and green lines together and look at the offspring ... which are all yellow. This does not fit into either hypotheses. So, we can reject both H_1a and H_1b as false.



We need to revise our hypothesis.

H_1c: There is an "inheritance factor" inherited from the parents, and one factor replaces the other, that determines seed color.

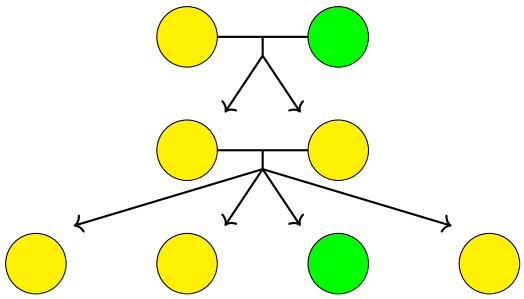
This goes along with another hypothesis to explain the data.

H_2: The yellow inheritance factor replaces the green inheritance factor when either can be inherited by offspring.

Okay, this new H_1c hypothesis can explain both the true breeding lines and the new data if we also have a rule that yellow replaces green.

However, constructing a hypothesis that can explain the data, after observing the data, is incomplete. The hypothesis has not yet been tested. The only way to test it is to set up a new experiment and produce new results that have not yet been observed.

So, we cross the yellow line to the green line and get yellow offspring. If the yellow inheritance factor has replaced the green inheritance factor among the offspring then their offspring must also be yellow. We try this and generate a second generation of crosses. ... and no; this is not what we see. There is a mix of yellow and green in the next generation. So, we can reject H_1c.



This may start to feel frustrating. We keep working to formulate new hypotheses based on new observations but they keep getting rejected by new data. However, this is how the scientific method works. Science and math work to determine the nature of reality from two different ends. Math uses logic to prove that ideas are true. Science uses data to prove that hypotheses are false. These two extremes meet each other in the field of probability and statistics.[2] Most hypotheses, that are not trivially simple, will probably ultimately be proven false in order to advance the field.[3]

Okay, so now we are at the most difficult point. We have a pattern that does not make sense in an easy way. All of our baseline hypotheses have been rejected. This is exciting because this is the point where the most important discoveries get made. They are not yet discovered precisely because they are not obvious and therefore frustrating and difficult. Somehow the green factor was present but hidden in the first generation cross and became uncovered in the next generation. How can this happen? Let's add an idea that there are actually two inheritance factors present in each organism—this starts to click; there are two parents to each organism so perhaps one inheritance factor is inherited from each parent.[4] So, in the true breeding parental

[2]The outcome of these crosses can be tested in order to quantify how far they deviate from expectations under various hypotheses and if this is consistent with random chance, but I do not want to use statistics here without going through how it is derived.

[3]Formal hypothesis testing is not the only form of science. There are more descriptive forms of science that record and communicate discoveries found in nature such as new species descriptions or new planet discoveries. This type of work is important, it is essentially data collection, but it is considered less powerful in terms of advancing the understanding of nature without hypothesis testing. (Albeit some data is very difficult and time consuming to collect and is a significant result in its own right.) However, unlike many other fields, all of science (should) shares a willingness to be tested and proven wrong in order to improve and the idea that just because something seems like it should be true or is appealing in some way (because of authority, intuition, or comforting ideas, for example), does not in any way mean that it is true.

[4]Often good scientific theories start to make sense in more than one way and start integrating with other theories and mechanistic hypotheses.

lines green/green individuals crossed together yielded green/green offspring because green was the only inheritance factor to choose from. The same goes for the yellow/yellow true breeding line. When these were crossed together green/green $\times$ yellow/yellow $\rightarrow$ green/yellow offspring because one inheritance factor was inherited from each parent. But the offspring were all yellow. Let's go back and modify H_2.

H_2b: The yellow inheritance factor masks the green inheritance factor in pea seeds when both are present in an organism.

If we crossed the offspring together we could get all combinations green/yellow $\times$ green/yellow $\rightarrow$ green/green, green/yellow, yellow/green, and yellow/yellow depending on which allele was inherited from which parent. Importantly, according to H_2b three of these types would appear yellow and only one would be green. If we look back at our data let's say that we observe an approximate three to one ratio of yellow to green.[5] We also realize that in order for everything to be consistent this implies that an individual cannot have more than two copies of an inheritance factor (i.e., they do not accumulate three, four, *etc.*) and an individual only passes on one copy to an offspring in order to maintain the total number of copies at two each generation (one from each parent).

And we write a new hypothesis to explain everything observed so far.

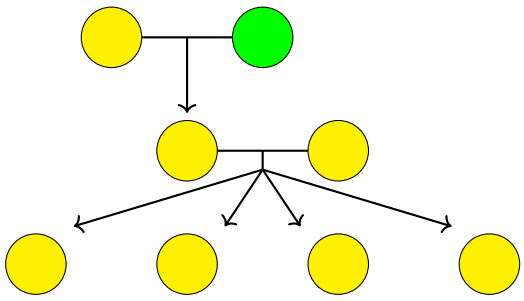
H_3: There is an "inheritance factor" inherited from each parent. These are maintained as distinct factors although the effects of one may be hidden by the other within an organism. Either one, but only one, can be passed on to the next generation.

Okay, we are feeling very good about ourselves. It may feel strange to think that we somehow contain two copies of an inheritance factor, as if we are simultaneously two individuals, and that one inheritance factor can mask another but science is about letting the data tell us what is going on under the surface of things. This new set of hypotheses (H_3 and H_2b) can explain a great deal of observed data without being overly complicated. However, this is dangerous. We formulated the hypothesis after observing the data. It has not yet been strongly tested. (It was weakly tested by comparing predictions to observed ratios of yellow and green but this was still after seeing the data and may have influenced our thinking in formulating the hypothesis.) In order to rigorously test this hypothesis it needs to hold up to unobserved data from a new type of experiment, not simply repeating

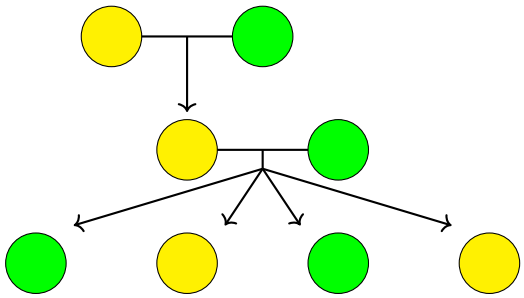
[5] This is one of Medel's greatest strengths. He used a quantitative approach to analyze data and paid attention to the predicted ratios of different crosses. No one had done this before.

the previous experiments.

There are two new types of crosses that we can make. We can cross the crossed offspring back to the true breeding lines. We are hypothesizing that the yellow offspring have green/yellow inheritance factors (remember that we cannot directly observe this) and that the two true breeding lines are composed of green/green and yellow/yellow individuals. If we cross green/yellow to yellow/yellow there are two types of offspring possible, yellow/yellow and yellow/green. In either case all of the offspring should appear yellow. We try this cross and it works.



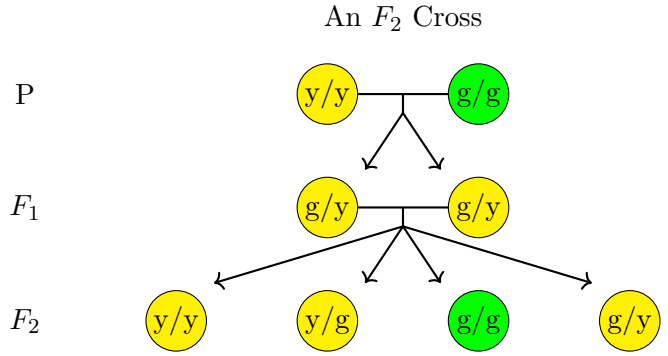
All of the offspring are yellow in color. The other cross is green/yellow to green/green. The two types of offspring from this cross are green/yellow and green/green. In this case there are two different seed colors predicted among the next generation, yellow (green/yellow) individuals and green (green/green) individuals. We also predict them at an equal 1 to 1 frequency (rather than the 1:3 frequency from crossing green/yellow to green/yellow before). We do this cross and it works, there are two types of offspring, and they are approximately 50/50.



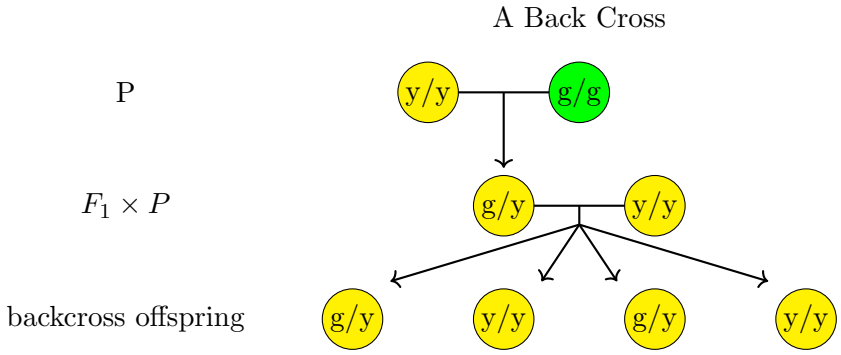
Now we have a very strong hypothesis on its way to becoming an established theory (if it can be verified and built upon by independent labs). It explains previous results and, because we designed a *testable* hypothesis,

we were able to see that it is consistent with previously unobserved data resulting from new experiments designed to test it.[6]

Here I will give the crosses again but will label the type of cross, the generations, and individuals with the inheritance factor (g for green and y for yellow).

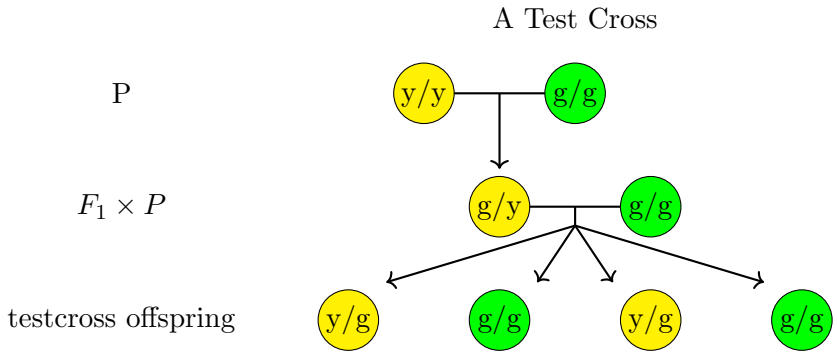


P indicates the parental generation; these are individuals from the true breeding lines. F_1 indicates the first filial generation and F_2 the second filial generation. The entire cross and in particular the last generation are referred to as an F_2 cross in short.



A back cross is where the F_1 generation is crossed back to a P or parental line.

[6] One big component of the scientific method that is missing in this example is the use of controls. Controls are very important when testing the effects of presence or absence of a factor in an experiment, but they are not always necessary, as in this case, depending on the nature of the hypotheses being tested.



A type of back cross is known as a test cross, when the F_1 is crossed back to the parental type that is masked in the F_1 generation. It can be thought of as testing the F_1 individual. The color of the offspring only depend on which inheritance factor was inherited from the F_1 parent.

So far I purposely avoided some of the terms used in modern genetics in order to approach it in a manner that someone discovering this for the first time might have. Genetics (and evolution) can be difficult to teach to biologists because so many feel like they already understand it. Sometimes it is useful to shake up that understanding a bit. When I talked about yellow versus green "inheritance factors" I was really talking about different alleles of the same gene. When I talked about green versus yellow seed colors I was talking about the *phenotypes*—what is observed by a human conducting the experiments. The combinations of alleles in each individual are referred to as *genotypes*. For simplicity I used y and g for the two different allele designations. If an individual has two copies of the same allele they are referred to as a *homozygote*. If an individual has two different copies of an allele they are a *heterozygote*. A y/y homozygote has a yellow phenotype. A g/g homozygote has a green phenotype. A g/y heterozygote has a yellow phenotype.

You may have seen capital versus lower case allele designation nomenclature. In this example yellow is the normal phenotype encountered in the wild and green is a mutant phenotype. Yellow is dominant to green (heterozygotes are yellow). We can use a "g" (standing for the green mutant) to represent the allele. g/g homozygote genotypes have a green phenotype. G/G and G/g genotypes have yellow phenotypes. We can reasonably guess that the normal function of the *green* gene is to produce yellow seeds and that this phenotype is disrupted in mutant homozygotes.

... and it turns out that this is the case. Today we call this gene *Stay Green 1* with an abbreviation *Sgr1*. $Sgr1/Sgr1$ homozygotes are the commonly encountered wildtype form in a wide range of plant species. Mutant $sgr1/sgr1$ homozygotes have seeds that remain green. The normal function of $Sgr1$ is to break down chlorophyll, which gives plant cells a green color. $Sgr1/sgr1$ heterozygotes are also yellow because only one working copy of the gene is required to break down the chlorophyll. It turns out that the continued presence of chlorophyll reduces the viability of seeds. So, a plant is more likely to reproduce, in the long run, if it can remove the chlorophyll from its seeds.[7]

Why are a lot of ripe fruit and seeds not green? A range of non-exclusive hypotheses can be written down for this:

1. to promote preservation of the cells by removing chlorophyll,
2. to attract frugivores which spread the seeds,
3. for camouflage against the soil to help seeds avoid being eaten,
4. or by selective breeding by humans. In fact as a counter example, many fruit and seeds that are green when ripe are due to new mutations, such as $sgr1$, that were selected by humans.

This might have implications for the preservation of fruit and vegetables, which leads to more testable hypotheses. Vegetables that are normally green, such as celery stalks or asparagus, might store for longer without spoiling if sand is piled around the growing stalks to prevent sunlight from stimulating chlorophyll production (etiolation). Because the conditions of storage and mechanisms of spoilage are so complex this is best done with a control, uncovered asparagus, grown and stored in parallel. This might explain why this is done in some parts of Europe, apart from the change in appearance and flavor. The storage of fruit and vegetables were very important in the past before powered refrigeration and rapid shipping.

[7] See `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4443279/`

Mendel's results have stood the test of time and have been shown in a remarkably wide range of organisms—human inheritance follows the same rules as Mendel's peas.

Mendel studied plants and found a rule
that works for many animals in a zoo.
    The children of F_1s
    are three to one,
and are known as the F_2.

Today we know much more, such as the molecular structure of a gene and how it interacts with other molecules. To be complete, Mendel studied many traits in addition to the color of pea seeds (flower color, plant height, *etc.*) and also discovered that they were inherited independently of each other—I haven't gone into this here in order to focus one just one aspect of Mendelian genetics. The passing on of either one allele or the other from heterozygotes, with equal probability, and the independence of inheritance of traits controlled by alleles at different genes (purple flower color does not affect the probability of green seeds if a plant is heterozygous at both genes ...) are so well established they became known as Mendel's laws. Fortunately, there are exceptions where Mendelian genetics does not work as expected, the "laws" are broken, and these cases have led to refinements of genetic theory with additional discoveries into how genetics works.