Darwin, Mendel, and Genetics (Chapter 3)

The age old questions

- Who am I? In particular, what *traits* define me?
- How (and why) did I get to be who I am, that is, how were these traits passed on to me?

Pre-Science (and in fact, pre-history): Family trees

Science

1859: Charles Darwin publishes *On the Origin of Species*

1865: Gregor Mendel publishes *Experiments on Plant Hybridization*

Trees

Examples: Geneological trees, organizational charts, computer subdirectories, neighborhood utilities systems.

Comprised of

nodes: points of information in the tree (e.g., people, switches, intersections)

edges: link two nodes by some association (e.g., mother, boss, roadway)

root: node which is the source (or destination) of all associations in the tree

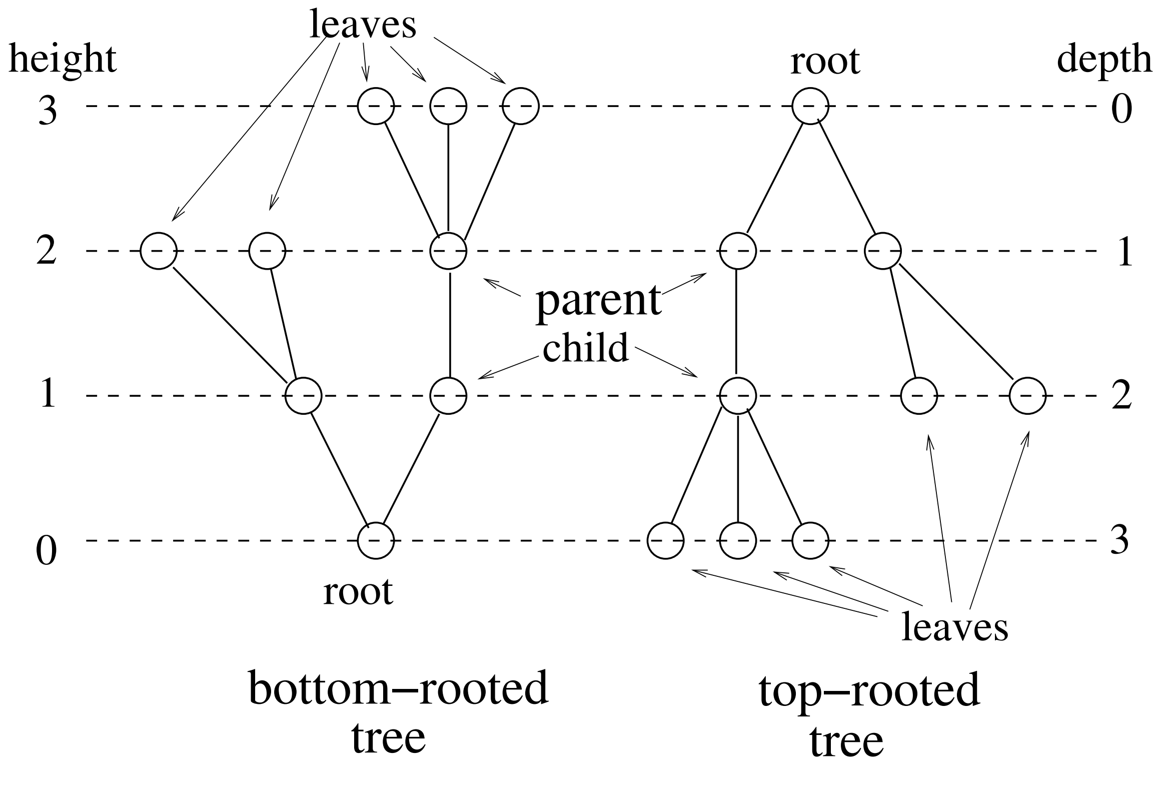
path: sequence of adjacent edges connecting one node to another.

Key property of trees: There is always a **unique** path between any two nodes in the tree.

Two special types of trees

top-rooted tree: Trees with the root at the **top**, and all paths to the root going **upward**

bottom-rooted tree: Trees with the root at the **bottom**, and all paths to the root going **downward**



Other key concepts:

parents of a node: nodes *one step toward* from the root from that node.

children: nodes *one step away from* the root from that node.

Note: The root is a node with no parents.

leaves: nodes that have no children.

depth/height: of a node, is the number of steps to the root; the depth/height of the **tree** is the depth of the child farthest from the root

The Scientific Method: Search for the Replicable

Steps of the Scientific Method:

1. **Identify the phenomenon** for which you feel is important to explain in order to further the frontiers of science or humankind.

2. **Form a hypothesis** about what you think explains that phenomenon.

3. Decide on an **experiment** that will support or oppose your hypothesis.

4. **Perform this experiment** in as unbiased a manner as possible.

5. **Draw a conclusion** about whether your hypothesis is correct or not.

Remember: Hypotheses are never **proved**, only **supported**.

Darwin and the Concept of Natural Selection

Darwin's hypotheses:

- Organisms have a genetic inclination to change their inherent traits (mutation).
- The changes are *random*, in particular, they can be both beneficial and detrimental to the organism.
- Those organisms with a more beneficial set of traits are more likely to survive (selection).

His conclusion: Nature has a built-in experimental procedure for improving upon the species, *and in evolving new species where necessary*.

Mendel and Genetics

Still unanswered question for Darwinists: What is the mechanism by which these traits are passed on, and what is the mechanism by which mutation occurs?

Mendel's Experiments: Eight years of experiments on peas, involving seven varieties and over 30,000 plants.

The Goal: To determine exactly how the various traits (length of stems, types of flowers, shape of peas) were passed on from generation to generation. In particular, he compared a set of specific dissimilar traits of the parents with the corresponding trait for the offspring.

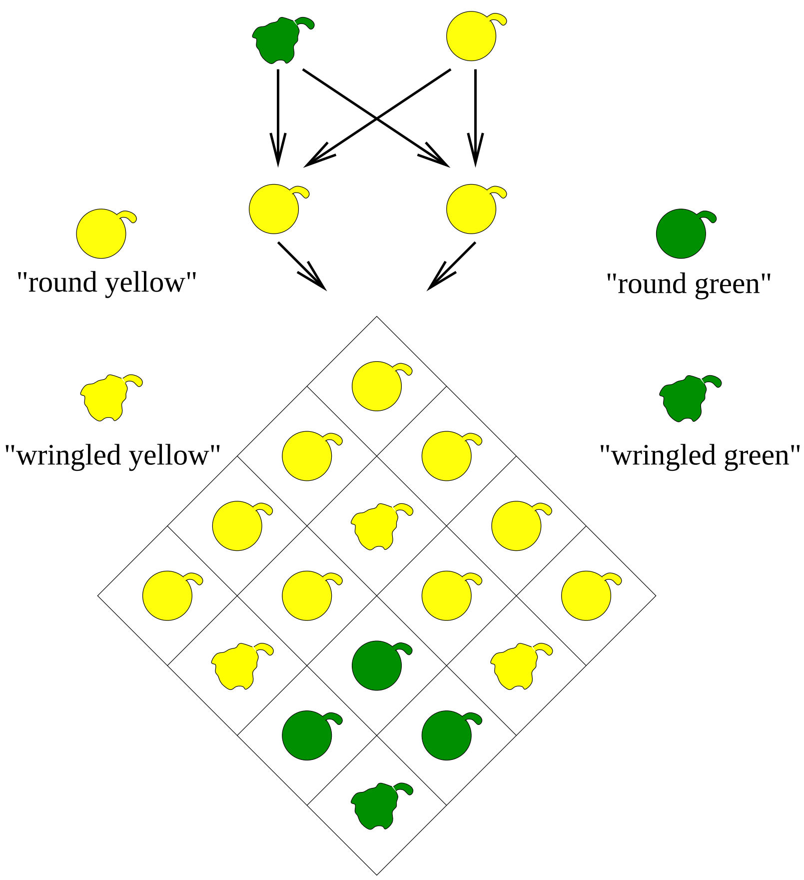
Observations:

- Pea plant reproduction involves pollen from one plant fertilizing an egg from another plant

- When two plants with different characteristics were bred, the offspring (generally) exhibited not *mixtures* of each trait, but rather *pure* traits from one or the other parent. The different traits, however, seemed to be mixed between parents.

- Traits that did not appear in one generation often showed up in later generations.

Mendel's Experiments: Organized Summary



(Boxes represent approximately equal size groups)

Mendel's hypotheses:

1. **Pairing of Traits:** Each organism has a *pair* of **factors** for each trait.

2. **Independent Segregation:** Each offspring obtains its trait pair as a result of inheriting exactly one factor from each parent. Each factor of a parent's pair is equally likely to be passed to the offspring.

3. **Independent Assortment:** The pair of factors obtained for each trait is not affected by what happens to the other traits.

4. **Dominance and Recession:** Each pair of factors is either **dominant** or **recessive**. An organism exhibits a recessive trait *only if both factors* passed on by the parents are recessive, otherwise it exhibits the dominant trait.

The Terminology of Genetics

gene: genetic representation of a trait

phenotype: outward manifestation of these traits

allele: a particular form of the gene, that could contribute to different manifestations of that trait

genotype: genetic representation of complete organism — written as a list of **paired alleles**, one for each gene, chosen from one of two allele for that trait, a **dominant allele** and a **recessive allele**.

gamete: genetic representation of the genetic blueprint contributed by one parent of a mate. Represented by a list of **single alleles**, one for each trait, chosen from that parent's genotype.

genotype from gametes: The genotype of the new organism is obtained by matching the gametes from each parent gene for gene to obtain the new allele pairs.

phenotype from genotype: Each trait of the new organism is the determined by the **dominant allele** if **either** of the alleles in the pair is dominant, and by the **recessive allele** if **both** alleles in the pair are recessive.

Example

Consider two traits, with two alleles each:

Trait 1 (shape):
$$\begin{array}{lcll} R & = & \text{“round”} & \text{(dominant)} \\ r & = & \text{“wrinkled”} & \text{(recessive)} \end{array}$$

Trait 1 (color):
$$\begin{array}{lcll} Y & = & \text{“yellow”} & \text{(dominant)} \\ y & = & \text{“green”} & \text{(recessive)} \end{array}$$

There are nine possible genotypes and four phenotypes:

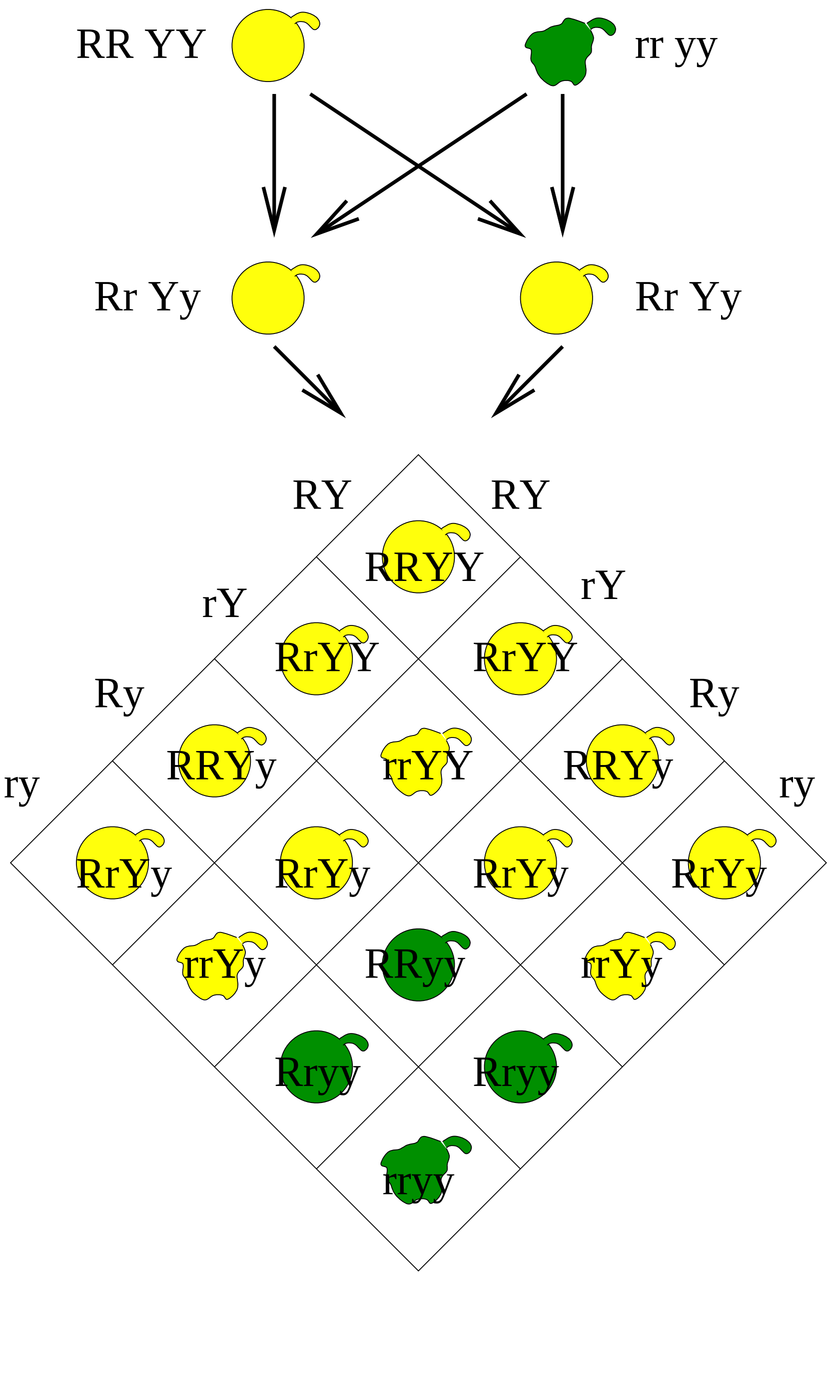
texture-color phenotype	**associated genotypes**
”round-yellow”	$RR\ YY,\ RR\ Yy,\ Rr\ YY,\ Rr\ Yy$
”round-green”	$RR\ yy,\ Rr\ yy$
”wrinkled-yellow”	$rr\ YY,\ rr\ Yy$
”wrinkled-green”	$rr\ yy$

A parent with genotype, say, $RR\ Yy$ could contribute two possible gametes, RY and Ry.

If this parent contributed Ry and another parent contributed gamete RY, then the offspring would have genotype $RR\ Yy$.

The phenotype of this offspring would then be "round-green".

Mendel's Experiments with Allele Labels



The Algebra of Cross-Breeding

Example: Suppose two $RR\ Yy$ parents mate. We know the possible gametes for each parent are RY and Ry

algebraic listing of gametes: write the gamete as a **polynomial** $(R+R)(Y+y)$, which expands as $RY+Ry+RY+Ry = 2RY+2Ry$.

possible genotypes from these gametes: Any pair of the above terms could combine for three possible combinations: $RR\ YY$, $RR\ Yy$, and $RR\ yy$.

algebraic listing of offspring genotypes: Simply multiply the two polynomials for the parents together. In the above example:

$$
\begin{aligned}
&(R+R)(Y+y)(R+R)(Y+y) \\
&\quad = (2RY+2Ry)(2RY+2Ry) \\
&\quad = 4RRYY + 8RRYy + 4RRyy
\end{aligned}
$$

Technique for Multiplying Polynomials

1. Place the polynomials along the top and right-hand-sides of a square grid.

2. At each square of the grid, place the grouped set of letters from the top and r-h-s in that row or column, rearranged in order of trait, dominant allele first.

3. Count all groups of letters that look alike, and put the number next to the group.

4. Sum these up and you have your expanded product.

Example: $(XY+Xy+xY+xy)(XY+xY+xy)$

	XY	Xy	xY	xy
XY	$XXYY$	$XXYy$	$XxYY$	$XxYy$
xY	$XxYY$	$XxYy$	$xxYY$	$xxYy$
xy	$XxYy$	$Xxyy$	$xxYy$	$xxyy$

The polynomial: 1XXYY + 1XXYy + 2XxYY + 3XxYy + 1xxYY + 2xxYy + 2Xxyy + 1xxyy

Probability

We obviously cannot know exactly how many offspring there will be of each genotype. What we can do is determine the **average fraction** of each genotype or phenotype that would appear if the population were big enough. Another way to say this is that we are computing the **probability** that a particular genotype or phenotype will occur in an arbitrarily chosen offspring.

Assumptions of a Mating:

1. Each allele of a gene pair in a parent is **equally likely** to appear in the new genotype, and
2. The particular allele chosen for any one trait will not affect the choice of alleles for any other trait.

Conclusion:

- Each genotype in the algebraic lists or grids made above is **equally likely** to appear in a gamete or offspring genotype. In particular,
- If we divide the polynomial by the **sum of the coefficients**, then the probability of any particular genotype is equal to the coefficient next to that genotype.
- The probability of any phenotype is found by adding up the probabilities of all genotypes associated with that phenotype.

Example

If two $RR\ Yy$'s mate, the genotype polynomial for the offspring will be distribution of the offspring will be $4RRYY + 8RRYy + 4RRyy$, and so the distribution polynomial is gotten by dividing the coefficients by 4+8+4=16, that is, $RRYY$ and $RRyy$ will appear 25% of the time and $RRYy$ will appear 50% of the time.

The resulting phenotypes will be 75% "round-yellow" offspring and 25% "round-green" offspring.

The Probability of Population Cross-Breeding

Suppose that each genotype appears with a certain percentage in a population. If we assume that each male-female pair is equally likely to mate, then the fraction of male-female pairs with given genotypes mating is the **product** of the fraction of the population each mate comprises.

Example: If $RR\ Yy$'s comprise one-third of the population, then the fraction of each offspring from all $RR\ Yy$ pairings ($\frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}^{th}$ of the population) will be

$$\frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{16}(R+R)(Y+y)(R+R)(Y+y)$$
$$= \frac{1}{36}RRYY + \frac{1}{18}RRYy + \frac{1}{36}RRyy$$

That is, among the offspring from this pairing

$RRYY$'s will comprise $\frac{1}{36}$ of the population, $RRYy$'s will comprise $\frac{1}{18}$ of the population, and $RRyy$'s will comprise $\frac{1}{36}$ of the population.

(See *The Algebra of Mating* pages)

Adding It All Up: Use the Computer!

If we were given a starting list of all of the population genotype distributions, then we could conceivably produce the list of each succeeding generation by putting together all pairs of genotypes and computing the distributions of all of the offspring. Even for two traits this is an enormous task. Thus we will use the computer to do this.

IDEAS

IDEAS is a software database that solves a large number of operations research models, and which is available in all of the labs and dorms on campus. Simply go to *Start, Programs, Math Applications, MATLAB* and when **MATLAB** starts, type "ideas". Click on *STOR072* and then *Genetics*. Then simply put in the decimal (or fractional) amounts representing the portion of each genotype in the current population. (*Make sure that they sum to* 1!) The computer will compute the succeeding population distributions from generation to generation, so you can see how your population is progressing. You can also get the distributions the corresponding actual phenotypes. Click on "Help" if you need instructions.

Enter Darwin

Darwin claimed that organisms with better genes — or more precisely, better phenotypes — are more likely to survive. Let's test this out.

Suppose that certain traits were more conducive to producing offspring. In particular, suppose that for each trait we know what is the **average number of offspring** having that trait that will reach reproductive age. We will assume (not very realistically) that these numbers are **multiplicative** across traits, that is, if "round" peas tend to produce offspring in a 2-to-1 ratio, and "green" peas tend to produce offspring in a $\frac{1}{3}$-to-1 ratio, then "round-green" peas will produce offspring in a $2 \cdot \frac{1}{3} = \frac{2}{3}$-to-1 ratio.

We therefore adjust the resulting fractions by multiplying them by the factor associated with the phenotype of that term. For example, consider the mating of two $RR\ Yy$'s as we did above, only now suppose that "round" reproduces at a 2-to-1 ratio, "green" produces at a $\frac{1}{3}$-to-1 ratio, and "wrinkled" and "yellow" each produce at a 1-to-1 ratio. Then the resulting fractions $\frac{1}{36}RRYY + \frac{1}{18}RRYy + \frac{1}{36}RRyy$ would be adjusted by multiplying each term by the appropriate factor

$$\begin{aligned} &2 \cdot 1 \cdot \tfrac{1}{36}RRYY + 2 \cdot 1 \cdot \tfrac{1}{18}RRYy + 2 \cdot \tfrac{1}{3} \cdot \tfrac{1}{36}RRyy \\ &\quad = \tfrac{1}{18}RRYY + \tfrac{1}{9}RRYy + \tfrac{1}{54}RRyy \end{aligned}$$

Normalizing to Represent True Proportions

Although these now represent the *relative* sizes of the populations, they do not necessarily add to 1. We must now **normalize** these numbers by dividing each one by the sum of all the population coefficients to get the actual *fraction* of the offspring population of each type.

In our example, we add

$$\frac{1}{18}+\frac{1}{9}+\frac{1}{54}=\frac{3}{54}+\frac{6}{54}+\frac{1}{54}=\frac{10}{54}$$

So we multiply each fraction by $\frac{54}{10}$ — or simply replace 54 by 10 in the denominators — to get the normalized genotypes population of 30% $RRYY$, 60% $RRYy$ and 10% $RRyy$, or 90% "round-yellow" and 10% "round-green".

A Simple Example

We will use a one-trait example to illustrate how the whole thing works. Suppose the population is made up of 40% "pure round" (RR) peas and 60% "pure wrinkled" (rr) peas. Further, suppose that "round" peas (RR or Rr) will produce double the number of offspring in a generation, while "wrinkled" peas (rr) only produce an equal number of offspring. Then the fraction of each type of offspring available for breeding in the next generation is computed by the following table:

Mating	**Polynomial**	**Darwin Factor**	**Normalized Fraction**
$RR\&RR$	$\frac{2}{5}\cdot\frac{2}{5}\cdot\frac{1}{4}(R+R)(R+R)=\frac{4}{25}RR$	$\frac{8}{25}RR$	$\frac{8}{41}$
$RR\&rr$	$2\cdot\frac{2}{5}\cdot\frac{3}{5}\cdot\frac{1}{4}(R+R)(r+r)=\frac{12}{25}Rr$	$\frac{24}{25}Rr$	$\frac{24}{41}$
$rr\&rr$	$\frac{3}{5}\cdot\frac{3}{5}\cdot\frac{1}{4}(r+r)(r+r)=\frac{9}{25}rr$	$\frac{9}{25}rr$	$\frac{9}{41}$
	Sum	$\frac{41}{25}$	

That is, $\frac{32}{41}$th of the population will be "round" and $\frac{9}{41}$th of the population will be "wrinkled".

(Our complete example with Darwin factors is also on *The Algebra of Mating* pages)

IDEAS and Darwin Factors

IDEAS can also take into account Darwinian factors as well. In the boxes marked "Reproductive Rates", you put in the multiple of number of offspring having a certain genotype that will reach reproductive age in the next generation. (This number can be greater than 1 or fractional.) **IDEAS** will then take these factors into account when computing the fractions of the genotypes and phenotypes in each succeeding generation.

The interesting thing about this is that you can perform "natural selection" over many generations to see how a particular trait comes to either dominate a population or die out, and exactly how fast this occurs.