An Introduction to Quantitative Genetics I

Heather A. Lawson
Advanced Genetics
Spring 2021

Outline

• What is Quantitative Genetics?
• Genotypic Values and Genetic Effects
• Heritability
• Linkage and Linkage Disequilibrium

Quantitative Genetics

• The theory of the statistical relationship between genotypic variation and phenotypic variation.

1. What is the cause of phenotypic variation in natural populations?

2. What is the genetic architecture and molecular basis of phenotypic variation in natural populations?
• **Genotype**
  • The genetic constitution of an organism or cell; also refers to the specific set of alleles inherited at a locus

• **Phenotype**
  • Any measureable characteristic of an individual, such as height, arm length, test score, hair color, disease status, migration of proteins or DNA in a gel, etc.

**Nature Versus Nurture**

• Is a phenotype the result of genes or the environment?
  
  • False dichotomy
  
  • If NATURE: my genes made me do it!
  
  • If NURTURE: my mother made me do it!

• The features of an organisms are due to an interaction of the individual's **genotype and environment**

**Genetic Architecture**: "sum" of the genetic effects upon a phenotype, including additive, dominance and parent-of-origin effects of several genes, pleiotropy and epistasis
Types of Traits

• Monogenic traits (rare)
  • Discrete binary characters
  • Modified by genetic and environmental background

• Polygenic traits (common)
  • Discrete (e.g. bristle number on flies) or continuous (human height) or binary (disease status)
  • Modified by genetic and environmental background

Frequency Distribution of a Phenotype

As the number of loci controlling a trait increases, the phenotype frequency becomes increasingly continuous

The Normal Distribution Can Be Completely Described by Just 2 Numbers: The Mean (μ) and Variance (σ²)
Consider a Specific Locus Influencing a Normally Distributed Polygenic Trait

For this locus, mean phenotype = 0.15, while overall mean phenotype = 0

Goals of Quantitative Genetics

• Partition total trait variation into genetic (nature) and environmental (nurture) components
• Predict resemblance between relatives
  • If a sib has a disease/trait, what are your odds?
• Find the underlying loci (genes, nucleotides) contributing to this variation
  • QTL – quantitative trait loci
  • GWAS – genome-wide association
• Deduce molecular basis for genetic trait variation

Basic Model of Quantitative Genetics

\[ P_{ij} = \mu + G_i + E_j \]

Phenotypic Value = the value observed when a trait is measured on an individual
Mean = the mean phenotype for the entire population
Genotypic Value = the average phenotype of those carrying the specified genotype
Environmental Deviation = the deviation of the observed phenotype in an individual from the genotypic value
If we measure the genotype and the phenotype in the same individuals, we can estimate genotypic values for various phenotypes.

**Genotypic Values**
- At a single locus, it is the average phenotype of those carrying the specified genotype.
  - Can be decomposed into:
    - Additive effects
    - Dominance effects
    - Parent of Origin effects

**Parent-of-Origin Genetic Effects**
- With 3 genotypes ($A_1A_1$, $A_1A_2$, $A_2A_2$) you can estimate two genetic parameters (additive and dominance).
  - To analyze parent-of-origin dependent effects, you need ordered genotypes ($A_1A_1$, $A_1A_2$, $A_2A_1$, $A_2A_2$).
    - Adds another parameter.
Arbitrarily assigned genotypic values

Genotype

\[
\begin{array}{c}
\text{Genotypic value} \\
-\alpha & 0 & +\alpha
\end{array}
\]

Genotypic Values

<table>
<thead>
<tr>
<th>weight(g)</th>
<th>6</th>
<th>12</th>
<th>14</th>
</tr>
</thead>
</table>

from Example 7.1
Falconer and Mackay
Reference point \((r)\)

\[
r = \frac{g(++) + g(pg \; pg)}{2}
\]

\(r\) is the mid-point between the two homozygotes

---

Additive genotypic value \((a)\)

\[
a = \frac{g(++) - g(pg \; pg)}{2}
\]

\(a\) is the half-the difference between the two homozygotes
Dominance genotypic value \((d)\)

\[
d = g(+pg) - \left(\frac{g(++) + g(pg \, pg)}{2}\right)
\]

\(d\) is the deviation from the midpoint between the two homozygotes
**Genomic Imprinting:** Different expression of alleles inherited from the mother and father results in phenotypic differences between reciprocal heterozygotes ($A_1 A_2 \neq A_2 A_1$)

![Diagram showing genomic imprinting]

**Genotypic Values**

<table>
<thead>
<tr>
<th>Genotype</th>
<th>$A_1 A_1$</th>
<th>$A_1 A_2$</th>
<th>$A_2 A_2$</th>
<th>$A_2 A_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>weight (g)</td>
<td>6</td>
<td>8</td>
<td>12</td>
<td>14</td>
</tr>
</tbody>
</table>

adapted from Example 7.1
Falconer and Mackay
Parent-of-origin imprinting genotypic value ($i$)

$$i = \frac{g(pg^+)-g(+pg)}{2}$$

$i$ is half the difference between the two heterozygotes.

\[ r = 10g \]

\[ \pm i = 2g \]

Parent of Origin Patterns: Parental

Paternal Expression

\[ a = \frac{g(LL) - g(SS)}{2} = 7.5 \]

\[ i = \frac{g(LS) - g(SL)}{2} = -7.5 \]
Parent-of-Origin Patterns: Dominance

Bipolar Dominance

Chances of T2D
Bipolar Dominance Associated with Type 2 Diabetes

Opposite effect PO-GWAS result for age of menarche
Heritability

• The proportion of phenotypic variation that is attributable to genotypic variation

• Penetrance: the percentage of individuals with a specific genotype that possess an associated disease
  • Categorical, e.g. Huntington's disease

• Expressivity: the degree to which individuals with a specific genotype display the associated phenotype
  • Quantitative, e.g. psoriasis

• Phenotype (P) = Population Mean (µ) + Genotype (G) + Environment (E)

• Partition the variance of these components
  • Phenotypic variance = genetic variance + unexplained variance
    • This partitioning can only be performed at the level of a population.

• Broad-sense heritability
  • $H^2 = \frac{\text{Var}(G)}{\text{Var}(P)}$

• Broad-sense heritability includes additive, dominance, epistatic variance, and parent-of-origin effects

Heritability

• Genetic (G) = Additive (A) + Dominance (D) + Imprinted (I) + Epistasis (E)

• $\text{Var}(G) = \text{Var}(A) + \text{Var}(D) + \text{Var}(I) + \text{Var}(E)$

• Narrow-sense heritability
  • $h^2 = \frac{\text{Var}(A)}{\text{Var}(P)}$

• Narrow-sense heritability is the ratio of the additive genetic variance to the total phenotypic variance
Heritability

- Ranges from 0 (all environmental) to 1 (all genetic)
- Specific to a population in specific environmental circumstances
- Highly inbred population with no genetic variation
  - Heritability of 0
- No environmental variance in an outbred population if all genetic variance is additive
  - Heritability of 1
- Most traits have a heritability between 0.2 and 0.8
- Can decrease by either a decrease in additive variance ($V_A$) or by an increase in environmental variance ($V_E$)

Estimating Heritability

Analysis of Variance or Regression

(using phenotypic covariance)

Heritability From Twins

For comparison of monozygotic (MZ) and dizygotic (DZ) twins:

- $r_{\text{MZ}} = A + C$
- $r_{\text{DZ}} = 0.5 \times A + C$
- $A = 2 \times (r_{\text{MZ}} - r_{\text{DZ}})$
- $C = r_{\text{MZ}} - A$
- $E = 1 - r_{\text{MZ}}$

Where the components of variance are A for additive, C for common environment, and E for unique environment.
GWAS have been successful in identifying common variants involved in complex traits. However, for the majority of complex traits, <10% of genetic variance is explained by common variants. Thus only a small part of heritable variation in a trait can be explained – most is missing!

What Explains Missing Heritability?

- Genomic regions or classes of variation are not well covered by existing markers
- LD between markers and causal variants
- Repeats or heterochromatic regions
- Rare variants
- Local heritability around associated SNPs
- Heritability estimates (narrow-sense, i.e. additive) are overestimated
- Power
- De novo mutations
- Explains a subset of cases of autism and other psychiatric disorders
  - May contribute to disease incidence but not to missing heritability
Linkage Disequilibrium (LD)

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>AaBi</td>
<td>X_{11}</td>
</tr>
<tr>
<td>AaBb</td>
<td>X_{12}</td>
</tr>
<tr>
<td>AaBu</td>
<td>X_{21}</td>
</tr>
<tr>
<td>AaBu</td>
<td>X_{22}</td>
</tr>
</tbody>
</table>

- $p_1 = X_{11} + X_{12}$
- $q_1 = X_{11} + X_{21}$
- $p_2 = X_{12} + X_{22}$
- $q_2 = X_{12} + X_{22}$

- $D = X_{11} - p_1q_1$
- Correlation coefficient $r = \frac{D}{\sqrt{p_1q_1q_2q_3}}$

Decay of Linkage Disequilibrium

Human Chromosome 22
Linkage and Linkage Disequilibrium

Linkage and Association

Theoretical Decay of LD in a Random-Mating Population

In a genomic region with no recombination, the LD created by mutation never dissipates.
Genomic Linkage Mapping

Need Pedigree Data (humans) or crosses between strains. Obtain sufficient polymorphic markers scattered throughout the genome such that any locus is ideally no more than 10 cM from a marker locus. Genetic resolution is limited by the number of recombination events in a single generation in your sample.

E.g., A Genome Linkage Scan for Persistence of Fetal Hemoglobin in 3,765 Sardinians (Uda et al. 2008) Scanned for Nearly 10,000 SNP’s (Single Nucleotide Polymorphisms), With Interval Mapping at Every 0.5 CM Position

Whole Genome Association
(Linkage Disequilibrium Mapping)

Indirect Association

![Diagram showing linkage disequilibrium mapping with chromosome, region of high linkage disequilibrium, disease risk SNP, and genotyped SNP.]
Linkage Disequilibrium Mapping
Obtain sufficient polymorphic markers scattered throughout the genome such that adjacent markers display significant linkage disequilibrium in the population of interest. The more markers, the finer the genetic resolution.

E.g., A Genome Linkage Scan for Persistence of Fetal Hemoglobin in 3,765 Sardinians vs. a GWAS on 4,305 (overlapping) Sardinians (Ulla et al. 2008)