An Introduction to Quantitative Genetics I
Heather A. Lawson
Advanced Genetics
Spring 2017

Outline
• What is Quantitative Genetics?
• Genotypic Values and Genetic Effects
• Heritability
• Linkage Disequilibrium and Genome-Wide Association

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 measurable 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

Different genetic architectures
Different effects on the phenotype
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

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

Phenotypic Distribution of a Trait
Consider a Specific Locus Influencing the 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
• GWA – genome-wide association
• Deduce molecular basis for genetic trait variation

Basic Model of Quantitative Genetics

\[ P = G + E \]

- Phenotypic Value = the value observed when a trait is measured on an individual
- 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 (A1A1, A1A2, A2A2) you can estimate two genetic parameters (additive and dominance).
- To analyze parent-of-origin dependent effects, you need ordered genotypes (A1A1, A1A2, A2A1, A2A2).
  - Adds another parameter.
Arbitrarily assigned genotypic values

<table>
<thead>
<tr>
<th>Genotype</th>
<th>-a</th>
<th>d</th>
<th>+a</th>
</tr>
</thead>
<tbody>
<tr>
<td>A_1A_1</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A_1A_2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A_2A_2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Genotypic values from Example 7.1, Falconer and Mackay

<table>
<thead>
<tr>
<th>Genotypic Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>weight(g)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>A_1</td>
</tr>
<tr>
<td>A_2</td>
</tr>
<tr>
<td>A_3</td>
</tr>
</tbody>
</table>

Arbitrarily assigned genotypic values

- Genotype
- A_1A_1
- A_2A_2
- Genotypic value
- 0
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 \( (d) \)

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

\( d \) 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 \((A_1 A_2 \neq A_2 A_1)\).

\[
a = \frac{g(LL) - g(SS)}{2} = -1.63
\]

\[
b = \frac{g(LL) - g(SS)}{2} = -0.49
\]

\[
c = g(SS) - \left(\frac{g(LL) + g(SS)}{2}\right) = 0.75
\]

Genotypic Values

<table>
<thead>
<tr>
<th>LLL</th>
<th>LLS</th>
<th>LSL</th>
<th>SSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\text{weight (g)})</td>
<td>6</td>
<td>8</td>
<td>12</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.

$pg$ pg (6g) pg+ (8g) +pg (12g) ++ (14g)

$r=10g$

$\pm i = 2g$

$T_{a} = \frac{g(LL)-g(SS)}{2} = 7.5$

$T_{i} = \frac{g(LS)-g(SL)}{2} = 7.5$

$T_{a} = \frac{g(LL)-g(SS)}{2} = -7.5$

$T_{i} = \frac{g(LS)-g(SL)}{2} = -7.5$
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

\[ \text{Heritability} = \frac{\text{Genetic}}{\text{Phenotype}} = \frac{\text{Additive} + \text{Dominance} + \text{Epistasis}}{\text{Phenotype}} \]

\[ \text{Var}(G) = \text{Var}(A) + \text{Var}(D) + \text{Var}(E) + \text{Cov}(A,D) + \text{Cov}(D,E) + \text{Cov}(E,D) \]

\[ \text{H}^2 = \frac{\text{Var}(G)}{\text{Var}(P)} \]

- Broad-sense heritability includes additive, dominance, epistatic variance, and parent-of-origin effects
- Narrow-sense heritability includes only additive genetic variance

\[ \text{Narrow-sense heritability} = \frac{\text{Var}(A)}{\text{Var}(P)} \]

\[ \text{Narrow-sense heritability} = \frac{\text{Additive}}{\text{Total phenotypic variance}} \]
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
- 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

Heritability From Twins

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

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

Where the components of variance are $A$ for additive, $C$ for common environment, and $E$ for unique environment.
Heritability of Human Traits

Missing Heritability

<table>
<thead>
<tr>
<th>Disease</th>
<th>Number of loci</th>
<th>Proportion of heritability explained</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age-related macular degeneration</td>
<td>5</td>
<td>50%</td>
</tr>
<tr>
<td>Crohn’s disease</td>
<td>32</td>
<td>20%</td>
</tr>
<tr>
<td>Systemic lupus erythematosus</td>
<td>6</td>
<td>15%</td>
</tr>
<tr>
<td>Type 2 diabetes</td>
<td>38</td>
<td>0%</td>
</tr>
<tr>
<td>HDL cholesterol</td>
<td>7</td>
<td>5.2%</td>
</tr>
<tr>
<td>Height</td>
<td>60</td>
<td>3%</td>
</tr>
<tr>
<td>Early onset myocardial infection</td>
<td>5</td>
<td>2.8%</td>
</tr>
<tr>
<td>Fasting glucose</td>
<td>4</td>
<td>1.3%</td>
</tr>
</tbody>
</table>

GWAS have been successful in identifying common variants involved in complex trait aetiology. 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
  - Repeat 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

Non-random association of alleles at different loci.

Decay of Linkage Disequilibrium

Human Chromosome 22

Linkage Disequilibrium

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>Frequency</th>
<th>Allele</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>A/B'A/A</td>
<td>X11</td>
<td>A1A1</td>
<td>p1 = X11 + X12</td>
</tr>
<tr>
<td>A/B'B/b</td>
<td>X12</td>
<td>A1B1</td>
<td>p2 = X11 + X12</td>
</tr>
<tr>
<td>A/B'B/b</td>
<td>X22</td>
<td>A1B2</td>
<td>q1 = X11 + X12</td>
</tr>
<tr>
<td>B/b'B/b</td>
<td>X21</td>
<td>B2B2</td>
<td>q2 = X11 + X12</td>
</tr>
</tbody>
</table>

Decay of Linkage Disequilibrium

Human Chromosome 22

Correlation coefficient \( r = \frac{D}{\sqrt{p_1 q_1 p_2 q_2}} \)

\( D = X_{11} - p_1 q_1 \)
Linkage and Linkage Disequilibrium

Linkage and Association

Whole Genome Association (Linkage Disequilibrium Mapping)

Indirect Association
GWAS Methodology

GWAS Workflow

Case-Control Association Plot and LD Map
Human Chromosome 22