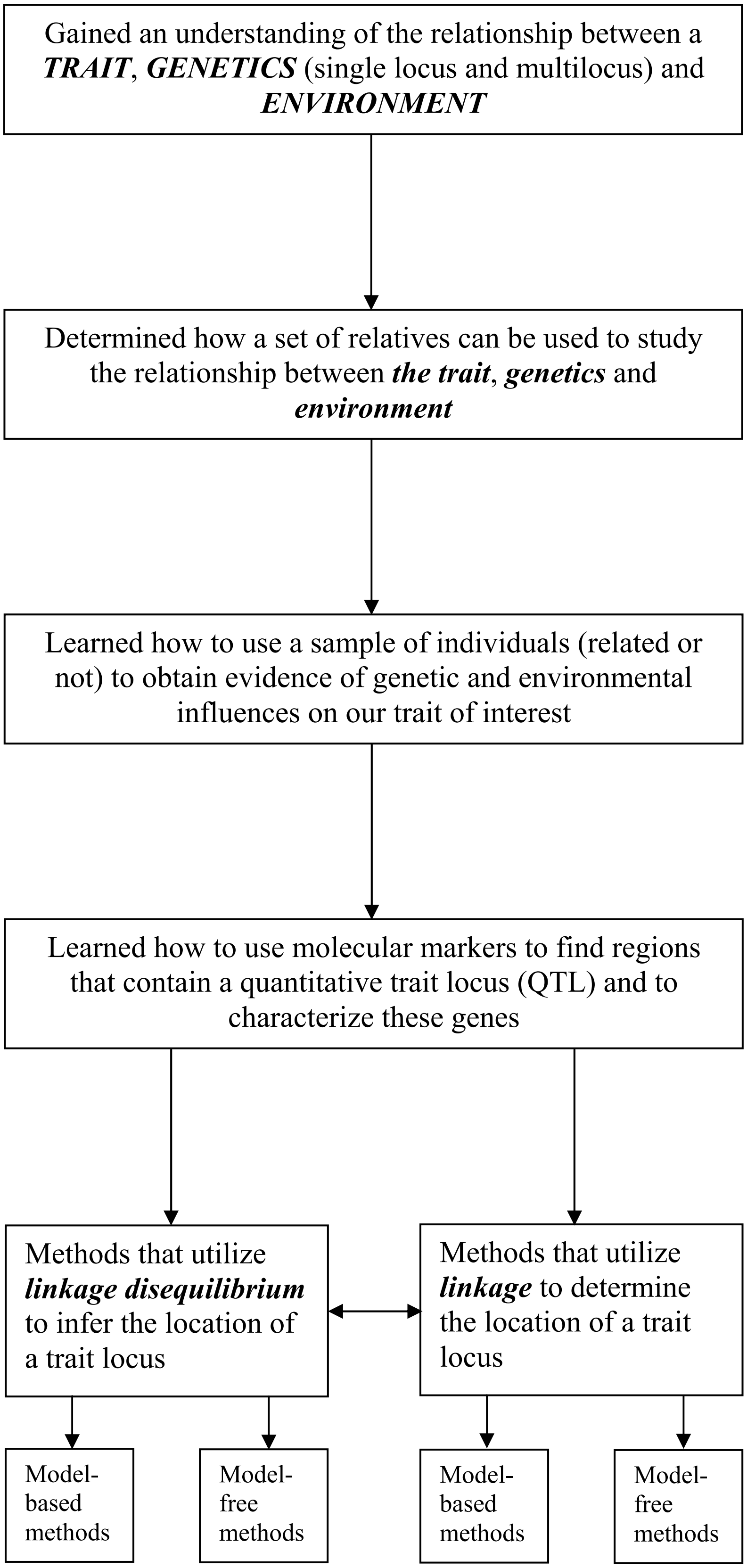
Summary for BIOSTAT/STAT551
Statistical Genetics II: Quantitative Traits



Gained an understanding of the relationship between a ***TRAIT***, ***GENETICS*** (single locus and multilocus) and ***ENVIRONMENT***

Theoretical Framework

Determined how a set of relatives can be used to study the relationship between ***the trait***, ***genetics*** and ***environment***

Additional Theoretical Framework

Learned how to use a sample of individuals (related or not) to obtain evidence of genetic and environmental influences on our trait of interest

What data do we have at this point???

Learned how to use molecular markers to find regions that contain a quantitative trait locus (QTL) and to characterize these genes

What data do we have at this point???

Methods that utilize ***linkage disequilibrium*** to infer the location of a trait locus

- Model-based methods
- Model-free methods

Methods that utilize ***linkage*** to determine the location of a trait locus

- Model-based methods
- Model-free methods

Theoretical Framework

An understanding of the genetic basis of traits is important not only

- in model-based methods (methods that require the specification of a particular mode of inheritance for the trait under study)

- but also in model-free methods (methods that do **NOT** require the specification of a particular mode of inheritance for the trait under study)

Major components of genetic models

- Allele frequencies
  - genotype frequencies as they relate to allele frequencies (Hardy-Weinberg disequilibrium/equilibrium)
  - multilocus frequencies as they relate to allele frequencies (linkage disequilibrium/equilibrium)

- Trait distribution

  Expressing our trait or phenotype as a function of genotype, environment and their interaction

  $$z_{ijk} = G_i + I_{ij} + E_j + e_{ijk} \text{ with } \sigma_z^2 = \sigma_G^2 + \sigma_I^2 + 2\sigma_{G,E} + \sigma_E^2 + \sigma_e^2$$

  - can characterize the different kinds of genetic effects
    - additive (effects of individual alleles)
    - dominance (effects of genotypes)
    - both single and multilocus genotypes

- Resemblance between relatives

  It is important to be able to describe the resemblance between relatives both in terms of allele sharing and trait covariance.

  - (condensed) **coefficients of indentity** Δ_1 to Δ_9 were used to describe the possible identity by descent configurations
  - these were then used to derive the **coefficient of kinship,** Θ_{xy}, the probability that a random gene from individual *x* and a random gene from individual *y* are identical by descent
  - also derived the **coefficient of fraternity**, Δ_{xy}, the probability that individuals *x* and *y* share both of their alleles identical by descent
  - and the **genetic covariance** for a pair of relatives was then expressed as a function of our genetic variances, Θ_{xy} and Δ_{xy}.

Determining whether there is a genetic basis to a trait

Commingling analysis

Established a likelihood framework for testing whether the distribution of a trait is a mixture of more than one distribution. Data consisted of a sample of *n* independent individuals with trait measurements.

- testing different modes of inheritance
- allowing for nonnormality

(Complex) Segregation Analysis

Expanded the likelihood framework to data that consisted of related individuals. That is, we now have a collection of *n* pedigrees. Tests were obtained for

- a major gene component and its mode of inheritance
- a polygenic component
- environmental effects
- non-mendelian transmission
- for a quantitative as well as a dichotomous trait

The following paper provides a nice review of complex segregation analysis and good resource for additional references:

> Jarvik GP (1998) Complex segregation analyses: uses and limitations. Am J Hum Genet 63:942-946

Using markers to map QTLs

Using markers to find a region of linkage

Measures of informativeness for a marker: heterozygosity, polymorphism information content (PIC) and proportion of fully informative matings (PFIM). For a parent to have the possibility of being informative for linkage, they must be heterozygous at the marker AND the QTL. We discussed

- how to pick a marker so that a specific (expected) proportion of families are informative for at least one parent
- how to determine the expected proportion of families that will be informative for linkage

Model-based linkage analysis

"Hybrid" Model-based/Model-free linkage analysis

Model-free linkage analysis

Using association/linkage disequilibrium/gametic phase disequilibrium to fine-map

- Rational for using association to fine-map

- Model-based association analysis (in reading)

- Model-free association analysis:

- "Hybrid" Model-based/Model-free association analysis

Some useful reviews

1. Elston RC (1998) Linkage and association. Genet Epidemiol 15:565-576

2. Olson JM, Witte JS, Elston RC (1999) Tutorial in biostatistics: genetic mapping of complex traits. Stat in Med 18:2961-2981

3. Terwilliger JD, Goring HHH (2000) Gene mapping in the 20th and 21st centuries: Statistical methods, data analysis and experimental design. Human Biology 72(1): 63-132

4. Feingold E (2002) Regression-based quantitative-trait-locus mapping in the 21st century. Am J Hum Genet 71:217-222

5. Doerge RW (2002) Mapping and analysis of quantitative trait loci in experimental populations. Nature Reviews Genetics 3:43-52

6. Kraft P, Horvath S (2003) The genetics of gene expression and gene mapping. Trends Biotechnol. 21(9):377-8