

Module 4: Measuring and Interpreting Marker Variation

Introduction

This module moves from the technical side of marker detection (Module 3) to the analytical side of marker data. Now that we have seen how marker data can be generated, what do the data tell us? This module includes a collection of widely used diversity statistics as well as divergence and distance metrics to assess how allelic variation is distributed within and among populations. Theoretical expectations based on the absence of evolutionary forces (i.e. neutrality theory) form baseline comparators against which observations are tested. In this way, historical population behavior (such as admixture or selection) can be inferred from contemporary observations. We also shift gears to include analyses of data from multiple loci simultaneously, which brings into play the role of genetic recombination and linkage disequilibrium (LD). LD plays a crucial role in the ability of markers to predict behavior of linked functional genes which might otherwise remain undetected. LD is the basis for mapping QTL using either genetic linkage maps (Module 5) or association mapping (Module 6).

Key Messages

- Measures of genetic variation within populations include counts of polymorphic alleles, heterozygosity, and proportions of polymorphic loci. These measures (or their analogs) can be estimated for a variety of marker types, including DNA sequence polymorphisms
- Variation among populations can be measured in terms of variance (F_{st} , G_{st}) or distance metrics.
- Neutral theory provides a means of formulating “null hypotheses” against which observed data can be compared
- Linkage disequilibrium measures the correlation among alleles at different loci (linked or unlinked) and can be influenced by population history
- Haplotypes represent clusters of linked genes that tend to be inherited together, and their size is highly variable among organisms

Outcomes

Course attendees will:

- learn how genetic variation within and among populations is measured in a variety of ways, including using DNA polymorphisms
- learn how to estimate genetic variation using several types of marker data
- understand how population history affects measures of genetic variation
- realize that linkage disequilibrium can affect both linked and unlinked loci, and how LD behaves over generations
- gain hands-on familiarity with several types of diversity estimators and data types

Module 4: Measuring and Interpreting Marker Variation

Outline

- I) Historical Perspective and Background
 - A) Early observations from allozymes
 - B) Background on neutrality

- II) “Classical” metrics (timeframe 1:30-1:50)
 - A) Levels of polymorphism: polymorphic alleles (A), proportion of polymorphic alleles (P), heterozygosity (H)
 - B) Organization of variation: hierarchical populations
 - C) Measures of Genetic Divergence

- III) What happens with multiple loci?
 - A) Linkage disequilibrium (LD, or D)
 - B) Alternative metrics
 - C) Other background

- IV) DNA Sequence Polymorphisms
 - A) New dimensions and challenges
 - 1) Evolutionary and functional considerations: homologs, orthologs, paralogs
 - 2) SNPs vs. “indels” (defined in Mod 3)
 - 3) Synonymous vs. non-synonymous substitutions
 - B) Nucleotide diversity and metrics
 - C) Neutrality Tests: Tajima’s D, others
 - D) HKA Test
 - E) MK Test
 - F) Haplotypes, haploblocks, and LD decay
 - 1) Visualizing LD
 - 2) Human HapMap Project

- V) Examples and case studies
 - A) Human
 - B) Maize
 - C) Conifers
 - D) Others??

- VI) Lab (several concurrent)
 - A) FASTA files and DNAsp
 - B) Hands-on calculations (custom spreadsheet examples) of
 - 1) Estimates of LD, given
 - (a) recombination
 - (b) admixture
 - 2) Nucleotide diversity estimators