Chapter 13
Applications of Linkage Disequilibrium and Association Mapping in Maize

Elhan S. Ersoz, Jianming Yu, and Edward S. Buckler

13.1 Introduction

Association mapping, also known as linkage disequilibrium mapping, is a relatively new and promising genetic method for complex trait dissection. Association mapping has the promise of higher mapping resolution through exploitation of historical recombination events at the population level, that may enable gene level mapping on non-model organisms where linkage-based approaches would not be feasible (Risch and Merikangas 1996; Nordborg and Tavare 2002).

Association mapping utilizes ancestral recombinations and natural genetic diversity within a population to dissect quantitative traits and is built on the basis of the linkage disequilibrium concept (Geiringer 1944; Lewontin and Kojima 1960). One of the working definitions of linkage disequilibrium (which here on will be referred to as LD) is the non-random co-segregation of alleles at two loci.

In contrast to linkage-based studies, LD-based genetic association studies offer a potentially powerful approach for mapping causal genes with modest effects (Hirschhorn and Daly 2005). While linkage analysis is based upon detection of non-random association between a genotype and a phenotype in well-characterized pedigrees, association mapping focuses on associations within populations of *unrelated* individuals. In general, chromosomes sampled from *unrelated* individuals in a population will be much more distantly related than those sampled from members of traditional pedigrees. In other words, the time to most recent common ancestor

Elhan S. Ersoz
Institute for Genomic Diversity, Cornell University, Ithaca, NY 14853, USA

Jianming Yu
Department of Agronomy, Kansas State University, Manhatten, KS 66506, USA

Edward S. Buckler
USDA-ARS, US Plant, Soil and Nutrition Laboratory, Ithaca, NY 14853, USA
Institute for Genomic Diversity, 159 Biotechnology, Cornell University, Ithaca, NY 14853, USA
e-mail: esb33@cornell.edu

A.L. Kriz, B.A. Larkins (eds.), *Molecular Genetic Approaches to Maize Improvement*
Biotechnology in Agriculture and Forestry, Vol. 63

(MRCA) of any given two individuals from a population of unrelated individuals would be greater than that of a pedigree population. This is what makes LD mapping suitable for fine-scale mapping: there will have been more opportunities for recombination to take place over several generations, between many alleles, in a species, while there can be only a few generations of recombination present in pedigree populations. Increase in the rate of recombination will lead to reshuffling of the chromosomal segments into smaller pieces. This will lead to reduction of the LD in short distances around loci, and lead to significant co-occurrence (i.e. LD) between only loci physically close, allowing high resolution. Whereas pedigree studies work with recombination events in few generations that enable exchange between chromosomes at the order of megabases, association studies deal with segmental exchanges measured in kilobases (Paterson et al. 1990; Stuber et al. 1992; Thornsberry et al. 2001).

13.2 What is Linkage Disequilibrium and How is it Related to Association Mapping Studies

The term *linkage disequilibrium* was first introduced back in the late 1940s to describe the degree of non-random association between pairs of loci. In the absence of demographic effects that might confound the LD patterns, LD summary statistics such as r^2 can be used to define the level of co-occurrence of alleles at two loci (Hill and Robertson 1968). When r^2 is zero, alleles at two loci do not co-occur more frequently than would be expected under random sampling. r^2 approaches its maximum of 1 as alleles at two loci show more frequent co-occurrence within the population sample examined. There are various other LD statistics that can be used for this purpose (Hedrick 1987) all of which aim to estimate the predictive value of a marker locus on another locus that is displaying non-zero LD with it (if LD statistic is zero, two loci examined have zero predictive value for each other).

Association mapping uses these properties of the measures of pairwise LD statistics to infer the predictive value of a marker locus for the association of the chromosomal region where it resides with the phenotype. The high-LD chromosomal region around a marker locus defines the predictive range of a certain genetic marker. If LD within this genomic range is complete, any polymorphism within this range will have the same predictive value for the association with the phenotype. Hence, as a result of a significant marker–phenotype association, it can be concluded that the causative polymorphism resides within this high LD region around the marker locus.

With respect to association mapping, the most significant aspect of LD is its predictive properties over the haplotype it resides in. However, the extent of LD (in base pairs) within species and even within individual genomes is highly variable, and therefore most reliably estimated empirically (Long and Langley 1999).