the adipocyte[13], whereas the inhibition of a serine proteinase (plasminogen) inhibits adipocyte differentiation[14]. The cleavage of different components of the ECM by these two classes of enzymes may explain their different effects on the cell. It is possible that a balance between the local activity of serine proteinases and MMPs regulates the differentiation of osteoblasts, osteoclasts and fibroblasts.

In contrast with another class of ECM enzymes, the cysteine proteinases (which are necessary for the removal of cartilage and bone ECM), the MMPs seem to be required for the recruitment of cells to their sites of action. For example, MMP-9 recruits osteoclasts to the hypertrophic zone of the cartilage anlage to initiate the process of endochondral bone formation[15]. MMP-2 may also help to recruit osteoblasts. If so, its deficiency could result in decreased bone formation and increased bone resorption through the unopposed action of the cysteine proteinases.

It is apparent that the organism uses many and diverse mechanisms to ensure a precise control of the ECM microenvironment, and that a delicate balance among these mechanisms is essential to normal physiology. Too much or too little of any one component may lead to disease. Whereas it is tempting to consider interventionist strategies to treat diseases caused by imbalance of ECM components, such strategies could have unpredictable consequences. It is imperative that we fully understand the biology of the proteins that modify the ECM before considering them as therapeutic targets. □

1. Martignetti, J. *et al. Nature Genet.* **26**, 261–265 (2001).
2. Werb, Z. *Cell* **91**, 439–442 (1997).
3. Sternlicht, M.D. *et al. Annu. Rev. Cell. Dev. Biol.* (in press).
4. Itoh, T. *et al. Cancer Res.* **58**, 1048–1051 (1998).
5. Holmbeck, K. *et al. Cell* **99**, 81–92 (1999).
6. Zhou, Z. *et al. Proc. Natl. Acad. Sci. USA* **97**, 4052–4057 (2000).
7. Erlebacher, A. *et al. Mol. Biol. Cell* **9**, 1903–1918 (1998).
8. Erlebacher, A. & Derynck, R. *J. Cell Biol.* **132**, 195–210 (1996).
9. Filvaroff, E. *et al. Development* **126**, 4267–4279 (1999).
10. Serra, R. *et al. J. Cell Biol.* **139**, 541–552 (1997)
11. Yang, X. *et al. J. Cell Biol.* **153**, 35–46 (2001).
12. Yu, Q. & Stamenkovic, I. *Genes Dev.* **14**,163–176 (2000).
13. Alexander, C.M. *et al. J. Cell Biol.* **152**, 693–703 (2001).
14. Selvarajan, S. *et al. Nature Cell Biol.* **3**, 267–275 (2001).
15. Engsig, M.T. *et al. J. Cell Biol.* **151**, 879–890 (2000).

Deconstructing maize population structure

Jonathan K. Pritchard

Department of Statistics, University of Oxford, Oxford OX1 3TG, UK. e-mail: pritch@stats.ox.ac.uk

Association studies have rarely been used in plant genetics, in part because of the risk of false positives caused by population structure. A study of flowering time in maize makes the first use of recent 'structured association' methods—statistical approaches that use independent loci to control for the effects of structure and admixture.

Association mapping has the potential to play a central role in the task of identifying genes for complex traits, both in humans and in other organisms[1]. Unfortunately, the possibility of spurious associations due to population structure has restricted the utility of these methods. The last two years have seen the development of statistical techniques that can overcome these problems; on page 286 of this issue, Jeffry Thornsberry and colleagues[2] present the first empirical demonstration of these new approaches, identifying a gene involved in the determination of flowering time in maize.

The difficulty with association mapping is that population structure can lead to highly significant associations between a marker and a phenotype, even when the marker is not physically linked to any causative loci. (Here, population structure refers to the presence of subgroups within the population.

One of the clearest examples of this comes from a study of type 2 diabetes in Native Americans of the Pima and Papago tribes from southern Arizona[3], groups that have an extremely high rate of diabetes. It was found that diabetic individuals in the study population were less likely than controls to have a particular haplotype at the immunoglobulin G locus. It would be tempting to conclude from this that the haplotype is protective against diabetes. But there is a snag: many of the individuals in the study population had some fraction of recent European ancestry. The authors showed that the controls had a higher proportion of European ancestry than the cases, and that the frequency of the haplotype was higher in Europeans than in Native Americans. When comparisons were made between cases and controls with similar amounts of European ancestry, the association disappeared. The original association was due to the confounding effect of population admixture.

Multifarious maize. Assorted Guatemalan maize landraces. Image kindly provided by Hugh Iltis (University of Wisconsin, Madison)

In general, one might not know about the presence of population structure or admixture when carrying out an association study, and so results from such studies are often viewed with some suspicion. In response to this, the 1990s saw the development of family-based tests of

association such as the transmission disequilibrium test[4], which avoid the problem of population structure. Often, however, there are major practical advantages to standard association studies that sample unrelated individuals. More recently, statistical methods have been developed for standard association studies that use independent marker loci to detect problems with population structure[5] and to correct for them[6–9]. The key idea behind the recent methods is that population structure has a similar effect on all loci. Thus, when population structure is a problem, we can expect to see associations at random marker loci all across the genome.

Flowering time in maize

Thornsberry *et al.*[2] set out to test whether polymorphism in the maize gene *Dwarf8* is associated with quantitative variation in two correlated phenotypes: flowering time and plant height. *Dwarf8* was a candidate gene for these traits based on previous mutagenesis and quantitative trait loci studies. The authors collected phenotype data, and sequenced *Dwarf8* in 92 inbred lines; they also genotyped an additional 141 simple sequence repeat (SSR) loci from across the maize genome. From the outset, it was clear that population structure might be a severe problem, as the 92 lines included members of three major groups, representing both tropical and North American lines. Moreover, the trait of interest, flowering time, is quite likely to be under divergent selection in different parts of the geographic range. Indeed, an excessive proportion of SSR alleles were found to be significantly associated with the two phenotypes, indicating severe confounding with population structure (for example, 15% of SSR alleles were associated with flowering time at the 5% level).

They then applied one of the new methods for testing for association in the presence of population structure[7], modified to allow for quantitative phenotypes. This method uses multilocus genotype data (in this case, from the SSR loci) to infer the population structure and to estimate the ancestry of individuals in the sample. The goal is reminiscent of the diabetes study[3] described above: when we compare allele frequencies just among individuals with similar ancestry, only true associations should remain. In this case, the rejection rate at the SSR loci was reduced to close to the nominal levels. Final *P* values at the *Dwarf8* locus were estimated from the empirical distribution of SSR *P* values[8]. The results indicate that, after accounting for the effects of population structure, a block of polymorphic sites near the 5′ end of *Dwarf8* is strongly associated with flowering time. Of these sites, a 6bp deletion flanking a conserved SH2 domain may be the functional polymorphism. The associations were replicated across five field studies.

The study by Thornsberry *et al.* makes several contributions. One is the demonstration of the utility of association methods in plant genetics, where they have so far received little attention. Association methods (including the transmission disequilibrium test, as well as the one used here) would seem to be immediately useful for testing candidate genes and for fine mapping of quantitative trait loci in plants. Remington *et al.* (manuscript submitted) show that linkage disequilibrium in maize seems to decay rather rapidly, which is favorable for fine mapping. Finally, the authors have provided the first empirical example of structured association methods in action. Their positive results should encourage the further testing of these methods in diverse genetic systems. □

1. Risch N. *Nature* **405**, 847–856 (2000).
2. Thornsberry J.M. *et al. Nature Genet.* **28**, 286–289 (2001).
3. Knowler W.C., Williams R.C., Pettitt D.J. & Steinberg A.G. *Am. J. Hum. Genet.* **43**, 520–526 (1988).
4. Spielman R.S., McGinnis R.E. & Ewens W.J. *Am. J. Hum. Genet.* **52**, 506–513 (1993).
5. Pritchard J.K. & Rosenberg N.A. *Am. J. Hum. Genet.* **65**, 220–228 (1999).
6. Devlin B. & Roeder K. *Biometrics* **55**, 997–1004 (1999).
7. Pritchard J.K., Stephens M., Rosenberg N.A. & Donnelly P. *Am. J. Hum. Genet.* **67**, 170–181 (2000).
8. Reich D.E. & Goldstein D.B. *Genet. Epid.* **20**, 4–16 (2001).
9. Satten G.A., Flanders W.D. & Yang Q. *Am. J. Hum. Genet.* **68**, 466–477 (2001).