Disertación del nuevo Académico Correspondiente, Dr. Michael T. Clegg.

“What can genetics tell us about plant domestication?”

INTRODUCTION

The invention of agriculture must surely be regarded as the most dramatic technological revolution in human history. This revolution did not happen suddenly, but played out over several millennia. It is likely that some external event(s) caused Neolithic peoples to abandon a nomadic hunter-gather economy and to gradually become tied to particular locations favorable for plant collection and animal domestication. There is considerable speculation about the external events that triggered this transition, but the climatic changes at the beginning of the Holocene seem the most likely trigger (Richerson et al., 2001). One reason for favoring a climatic cause is the more or less simultaneous invention of agriculture in the Near East, Mesoamerica and Asia. A global phenomenon must have driven these independent, but parallel developments. Presumably the initial process began with small human groups collecting nutritious plant materials in favorable locations, followed by the incidental reseeding of collected materials; this must have also have been associated with the unconscious selection of favorable traits (Zohary and Hopf, 1994) and ultimately with the fixation of traits adapted to human use.

Most of our knowledge of the initial steps in plant domestication rest on the analysis of archeological sites and these provide important information on the dates and localities where domestication began (Harris and Gosden, 1996; Wilcox, 1999). Archeology also provides important insights into the economy and social organization of early plant domesticators. Plant remains at early sites of cultivation reveal the signature of human selection through the presence of early domestication traits. A major domestication trait in cereal crops, such as wheat or barley, is the “non-shattering” trait that causes the grain to be retained on the stalk rather than abscising and being scattered on the ground or “shattering”. The non-shattering trait allows the human collector to gather the grain from the stalk easily instead of having to collect individual grains from the soil. Larger seed size provides further morphological evidence of early human selection.

Key Domestication Trait in Barley

Btr1 : brittle rachis
(Btr2 : brittle rachis)

btr1 : tough rachis
(btr2 : tough rachis)

What can the tools of molecular genetics add to this picture? There are two related categories of information from genetics that are useful in the study of plant and animal domestication. First, mutations in genes can discriminate among populations that appear identical at the morphological level, and second genetic data allows the reconstruction of the genealogical relationships that connect populations. This second point requires some further amplification. The coalescence theory of population genetics provides a framework for looking backwards in time and inferring genetic history (Hudson, 1990). Assuming that each mutation is unique (infinite sites model) we can trace the history of a gene backwards to its most recent common ancestor. Also if a mutation occurred just once, it must have arisen in a single individual at a single point in space, thereby revealing the spatial – temporal spread of the mutation.

The empirical task is to collect individuals from different points in space and to determine their genetic constitution. During recent years it has become possible to determine the precise DNA sequence of a gene in a single individual. As technology has advanced, the number of individuals and genes that can be sequenced has grown substantially. This provides an abundant source of data on mutational differences among individuals and has advanced the genetic study of plant domestication substantially.

My purpose in this essay is to describe the application of these ideas in two different regions of the world. I will first consider how genetics has provided new information on the domestication of barley in the Near East and then I will consider avocado domestication in Mesoamerica.

Barley a domesticate of the Near East

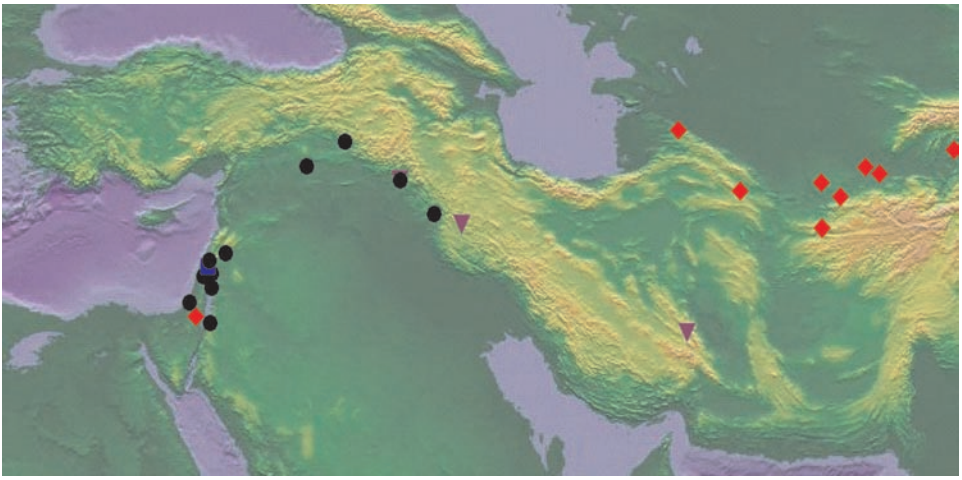
Barley (*Hordeum vulgare*), a cereal crop, is one of the first plants to have been domesticated by Neolithic humans. It has long been known that barley was domesticated in the Fertile Crescent region of the Middle East where sites at Abu Hureyra and Jericho date to approximately 10,500 years ago. But there is a long-standing question as to whether barley was domesticated a second time in southern Central Asia east of the Fertile Crescent (Harris and Gosden, 1996).

There are two pieces of evidence that hint at a second domestication of barley, one genetic and the other anthropological. The anthropological evidence is that domesticated barely remains are found in Central Asian sites of human habitation dating back more than 8,000 years BP. Such early evidence of barley cultivation may suggest an independent origin in Central Asia. The genetic evidence comes from two key domestication traits. Both the tough rachis trait that determines non-shattering and the six-rowed versus two-rowed phenotype are each controlled by two different loci (Komatsuda et al., 2004; Tanno et al., 2002; Azhaguvel and Komatsuda, 2007). How can we explain this redundancy? The most plausible scenario is that primitive human domesticators selected the first mutation encountered that produced these useful phenotypes and then built on this innovation. According to this scenario the genes for these two traits should trace back to just two fundamental mutations in all domesticated barley, but this is not the case. One way out of this dilemma is to assume a second domestication where entirely different foundational mutations were seized upon by human plant domesticators (Zohary, 1999). How can spatial temporal data on many barley genes help test the hypothesis of two domestications? To answer this question we need to consider spatial-genetic patterns in the wild progenitor of cultivated barley.

Patterns of genetic differentiation in wild barley

Wild barely (*Hordeum vulgare* spp *spontaneum),* an annual grass of semi arid savannah regions, is the progenitor of cultivated barley. The native range of wild barley extends from the eastern Mediterranean across the Zagros Mountains, and into adjacent southwest Asia, a distance west to east of 3,500 km. Lin et al (2001) found that one of three triplicate alcohol dehydrogenase loci (*Adh* 3) is characterized by strong genetic differentiation between populations east and west of the Zagros mountains (Figure 1). Subsequently, Morrell et al. (2003) determined the DNA sequence of nine different genetic loci from a sample of 25 accessions drawn from across the 3500 km range of wild barley and found that approximately 50% of the loci in the sample showed strong geographic patterns of genetic differentiation all defined by the Zagros range.

Figure 1
Adh 3 haplotype distribution from the Mediterranean across the Zagros mountains and east across the Central Asian Plateau. The symbols show the geographic location of different common haplotypes.

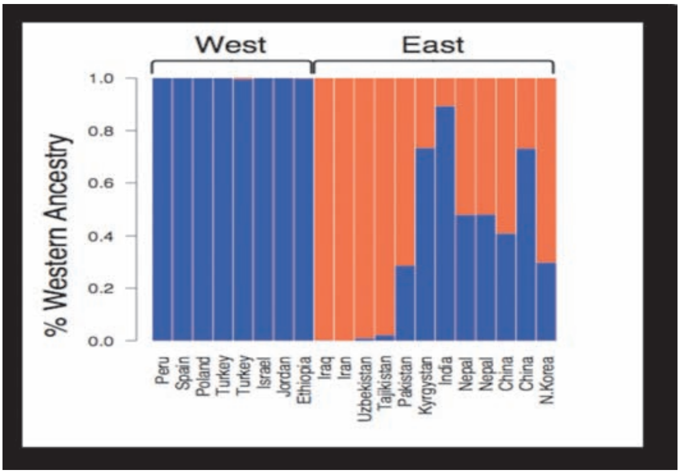


The fact that some wild barley loci show strong genetic differences east and west of the Zagros Mountains provides a means to ask whether barley was domesticated separately east of the Fertile Crescent. If barley was domesticated only once in the Fertile Crescent region, we would expect domesticated barley to have mutations that derive from this region and not from the quite different wild barley populations east of the Zagros. If, on the other hand, barley was domesticated separately east and west of the Zagros Mountains we would expect primitive domesticates (known as land races) to reflect the genetic characteristics of their region of origin. To investigate these possibilities (Morrell et al., 2007) sequenced seven of the genes that showed the most dramatic geographic patterns from 20 land races and 12 modern cultivars.

To ask whether the land races show region specific genetic patterns, assignment analyses were employed (Prichard et al., 2000) to assign land races to either the western or eastern regions based on their genetic composition using a probabilistic model. It is important to note that the assignment analyses were carried out without regard to the geographic origin of the land races. If the genetic analyses assign land races to separate populations and these correspond to the regions of origin of the land races, then the hypothesis of two domestications is supported. The resulting analyses are shown in Figure 2.

The results in Figure 2 reveal a sharp divide in land race genetic composition at about the boundaries of the Zagros Mountains, suggesting that land races from east of the Zagros are derived from wild barley east of the Zagros. This would appear to support the two-domestication hypothesis. However, the patterns are more complex, because as we consider eastern land races from outside the natural range of wild barley (east of western Pakistan), genes from western sources become common. What might explain this pattern?

Figure 2
Blue represents western ancestry and red eastern ancestry. The geographic origins of land races are given on the X-axis



Ancient traders, beginning thousands of years ago, moved domesticated barley along the Silk Road between the Near East and Far East. Barley reached China at least three thousand years ago and Japan soon after. This represents an expansion of barley cultivation, because the range of wild barley ends near western Pakistan, so domestication could not have occurred *in situ* east of this region. It appears that both the transport of western barley and the introgression of western genes into eastern barley accounts for the patterns in Kyrgyzstan, India and eastward depicted in Figure 2.

The genetic data also show that virtually all European and US cultivars of barley derive from the Fertile Crescent domestication (Morrell and Clegg, 2007). (The term cultivar refers to products of modern plant breeding where the plant exhibits distinct traits that are stable upon propagation.) Remarkably the genetic traces of the historical spread of barley agriculture are still retained in both the west (US and Europe) and in Asia. These populations are still genetically distinct, despite the widespread use of modern cultivars around the world. Clearly, the two semi distinct barley gene pools have much to offer to modern plant breeders in the future.

Avocado a New World subtropical tree crop

Avocado (*Persea ameriana*) is a major fruit crop with a world production value of more than 600 million US dollars. The largest producer is Mexico followed by Indonesia and the United States. The fruit is prized for its high nutritive value and for its rich flavour (Chanderbali *et al.*, 2008). Humans have utilized the avocado for at least 9,000 years and archaeological investigations suggest human selection and domestication dating back at least 4,000 years ago in the Oaxaca Valley of Mexico (Smith, 1966, 1969).

There are three botanical races of cultivated avocado known as the Guatemalan (*P. americana* var. *guatemalensis* Williams), Mexican (*P. americana* var. *drymifolia* Blake) and West Indian (*P. americana* var. *americana* Mill.) races. Each race appears to have been domesticated independently by Neolithic peoples in Mesoamerica (summarized in Davis et al, 1999; Ashworth and Clegg, 2003). The appellation West Indian is a misnomer, as avocado is not native to the West Indies; instead the West Indian race is believed to be a domesticate of the lowlands of Central America. There are substantial morphological differences among the races of avocado that include fruit shape and size, tree architecture and frost tolerance. Figure 3 illustrates morphological variation among fruit of wild Criollo avocados and contrasts these with a modern cultivar.

Figure 3. Wild Criollo Mexican avocado with green skinned cultivar for scale.

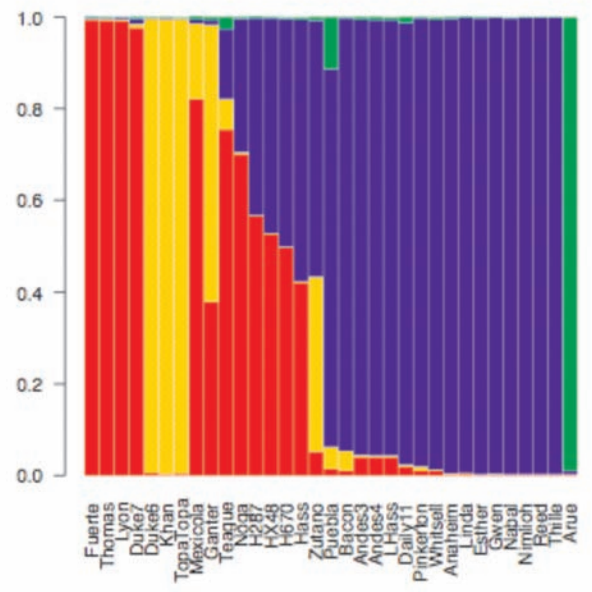
Applying genetic markers to unraveling of avocado genealogical relationships

The races of avocado evidently came into contact following the European invasion of Mesoamerica. Beginning in the late 19th century, hybridization among the races created the basis for modern avocado cultivars. Today's cultivars are the result of more than a century of interracial hybridization, but owing to a complex system of mating the precise genealogical relationships among cultivars are based on speculation and anecdotal information. Up until the advent of molecular markers it was impossible to determine the pollen donor (male parent) that gave rise to particular cultivars. The problem is that a mature avocado tree produces more than 10^6 flowers while only around 400 will lead to mature fruit. This makes controlled pollination impractical and other techniques such a caging trees have proved unreliable. So the only way to determine the pollen parent is retrospectively, using genetic markers specific to each parent. But this still does not resolve more complex pedigrees. For example, the cultivar Hass (the most widely grown cultivar) is speculated to be a Guatemalan X (Mexican X Guatemalan) backcross. How can we unravel these more complex relationships? The answer lies in the use of molecular markers and statistical assignment algorithms. But to proceed we once again need spatial-genetic information from the wild progenitor of cultivated avocado.

Chen et al, (2008) sequenced four loci from a geographic sample of wild avocado that spanned the three racial groupings. The sequence data included information from 5960 base pairs of DNA revealing a total of 176 SNPs. Assignment analyses using these data verified that wild avocados clustered into three genetically distinct populations that corresponded to the three races of avocado (Chen et al., 2008). Each race was found to be genetically distinct, with the Mexican race showing the greatest differentiation from the Guatemalan and West Indian races. These analyses are consistent with prior morphological results and provide direct genetic confirmation of the existence of three distinct populations of avocado. In addition, Chen et al (2008) discovered that the Mexican race could be further subdivided into two genetic populations that are separated into populations above and below 2000 m in elevation, a distinction that had not been previously recognized.

Following this a set of 33 cultivars were resequenced for the same loci and these data were subjected to assignment analyses to investigate the geographic origins of modern cultivars of avocado. Figure 4 illustrates the results of the assignment analyses.

Figure 4: The geographic origins of the genomes of various avocado cultivars are depicted by colored bars. Red represents high elevation and yellow low elevation Mexican, blue depicts Guatemalan and green represents West Indian. The height of the colored bar represents the percentage of the genome contributed by a particular source. The X-axis gives cultivar names.



The results confirm that ‘Hass’ is a hybrid between Guatemalan (G) and Mexican (M) races, with inferred probability of assignment split as 60% Guatemalan and 40% Mexican to the two ancestries. Cultivars ‘H287’, ‘H670’ and ‘HX48’ are genetically very close to ‘Hass’. Cultivar ‘Gwen’, considered to be a hybrid of Mexican and Guatemalan origins, has a very high assignment to Guatemalan ancestry (99.2%), perhaps consistent with the suggestion based on RFLP data that its paternal parent is ‘Thille’ (Davis et al, 1998). In a similar case, cultivar ‘Whitsell’ also shows a high probability of Guatemalan origin. In contrast, cultivar

‘Fuerte’ appears to be a pure Mexican type rather than a Mexican X Guatemalan hybrid as had been thought previously. We can now assign Guatemalan origins for cultivars ‘Andes 3’, ‘Andes 4’ and ‘Leavins’ Hass’; while cultivar ‘Puebla’ appears to derive from all three races (Mexican, Guatemalan and West Indian). The popular rootstock ‘Duke7’ which was considered to be of Mexican origin is a hybrid between Mexican and Guatemalan ancestors. While these results greatly increase our understanding of cultivar ancestry, they are based on a small number of loci and future studies with more loci and wider geographic sampling will be needed to refine the picture of genetic ancestry.

CONCLUSIONS

Modern genetic and statistical techniques reveal much about the past that cannot be inferred using any other approach. These methods are powerful because they exploit the pattern of mutational differences between copies of a gene (alleles) to reveal the temporal order of change and they provide a natural measure of genetic distance. Genetic spatial information can be combined to provide a trace of the spatial/temporal dynamics of change. We use these ideas to investigate the question of multiple domestications of barley and we show that the evidence strongly favors a minimum of two independent domestications east and west of the Zagros Mountains. We also show how a complex history of hybridization can be dissected in the case of modern avocado. Approaches like the ones described here are likely to be widely employed in the future to enrich our understanding of historical change.

LITERATURE CITED

Ashworth, V. E. T. M. and M. T. Clegg. (2003). Genealogical relationships among cultivated avocado (*Persea americana* Mill) genotypes inferred from microsatellite markers. J. Heredity 94: 407-415.

Azhaguvel P. and T. Komatsuda. (2007) A phylogenetic analysis based on nucleotide sequence of a marker linked to the brittle rachis locus indicates a diphyletic origin of barley. *Annals of Botany* 1–7.

Chanderbali, A. S., V. A. Albert, V. E. T. M. Ashworth, M. T. Clegg, R. E. Litz, D. E. Soltis and P. S. Soltis. (2008) *Persea americana* (avocado): bringing ancient flowers to fruit in the genomics era. Bioessays 30:386-396.

Chen, H., Morrell, P. L., de la Cruz, M. and M. T. Clegg. (2008). Nucleotide Diversity and Linkage Disequilibrium in Wild Avocado (*Persea americana* Mill.). (J. Heredity, in press)

Chen, H., P. L. Morrell, V. E. T. M. Ashworth, M. de la Cruz, and M. T. Clegg (2008). Tracing the geographic origins of major avocado cultivars. (submitted).

Davis, J., D. Henderson, K. Kobayashi, M. T. Clegg and M. T. Clegg. (1998). Genealogical relationships among cultivated avocoado as revealed through RFLP analyses. J. Heredity 89: 319-323.

Gepts P. (2004) Crop domestication as a long-term selection experiment. Plant Breed Rev 24 (Part 2): 1-44.

Harris D. R. and C. Gosden, (1996) "The Beginnings of Agriculture in Western Central Asia," in *The Origins and Spread of Agriculture and Pastoralism in Eurasia: An Overview*, ed. David R. Harris, London, 1996, pp. 370-89.

Hudson, R. R. 1990. Gene genealogies and the coalescent. Oxford Surveys in Evolutionary Biology. 7: 1-44.

Komatsuda T, Maxim P, Senthil N, Mano Y. (2004) High-density AFLP map of nonbrittle rachis 1 (*btr1*) and 2 (*btr2*) genes in barley (*Hordeum vulgare* L.). Theoretical and Applied Genetics 109:986–995

Lin, J.-Z., A.H. D. Brown and M. T. Clegg. (2001). Heterogeneous Geographic Patterns of Nucleotide Sequence Diversity between Two Alcohol Dehydrogenase Genes in Wild Barley (*Hordeum vulgare* ssp. *spontaneum*). Proc. Natl. Acad. Sci. USA. 98:531-536.

Morrell, P. L., K. E. Lundy and M. T. Clegg. (2003). Distinct geographic patterns of genetic diversity are maintained in wild barley (*Hordeum vulgare ssp. spontaneum)* despite migration. Proc Natl Acad Sci USA100: 10812-10817.

Morrell, P. L. and M. T. Clegg. (2007). Genetic evidence for a second domestication of barley (*Hordeum vulgare*) east of the Fertile Crescent. Proc Natl Acad Sci 104: 3289-3294.

Prichard J. K., M. Stephens and P. Donnelly. (2000). Inference of Population Structure Using Multilocus Genotype Data. Genetics 155: 945–959.

Richerson P, Boyd R, Bettinger R (2001) Was agriculture impossible during the Pleistocene but mandatory during the Holocene? Amer Antiq 66:387-411

Smith C. E. (1966) Archaeological evidence for selection in avocado. *Economic Botany*, **20**, 169-175.

Smith C. E. (1969) Additional notes on pre-conquest avocados in Mexico. *Economic Botany*, **23**, 135-140.

Willcox G. (1999). Agrarian change and the beginnings of cultivation in the Near East: Evidence from wild progenitors, experimental cultivation and archaeobotanical data. In: Gosden C, Hather J, editors. Prehistory of Food. London: Routledge, p. 479-500.

Tanno K, Taketa S, Takeda K, Komatsuda T. (2002) A DNA marker closely linked to the *vrs1* locus (row-type gene) indicates multiple origins of six-rowed cultivated barley (*Hordeum vulgare* L.). Theoretical and Applied Genetics 104:54–60

Zohary, D. and Hopf, M.(1994) *Domestication of Plants in the Old World.* Second edition. Oxford: Oxford Science Publications, Clarendon Press.