Implications for conserved genomic structure among plant species

Tim Helentjaris
Department of Plant Sciences, University of Arizona, Tucson, AZ 85721

While the theory of evolution was generally accepted early in this century, many implications of this theory, at least with respect to the issue of hereditary material, had yet to be considered in detail. In fact, the basic tenets of evolution were accepted well before the genetic repository was identified as DNA and many of the peculiarities of its chromosomal structure were elucidated. With the relatively recent advent of DNA sequencing on a practical scale, earlier observations from protein sequencing of similarity among homologous gene products from related (and not so related!) species have been considerably expanded. DNA sequencing is now preferred by many researchers as a measure of evolutionary relatedness over more traditional methods of morphological comparison. Although there is still much debate as to how the molecular clocks should be read and whether they keep the same time across genera and families, a molecular approach has significantly changed our understanding of how individual species relate to each other.

Why genes should share so much homology across species at the level of DNA sequence and not also in the arrangement of genes within related genomes is another issue that was simply awaiting a similar enhancement in technology. No technology was available that provided both the resolution and saturation to examine linkage relationships across species' boundaries until the development of DNA-based molecular markers (1). After the production of high-density linkage maps, it was only a matter of time before comparisons of maps between related species could reveal the extent of synteny of chromosomal segments. Nowhere has this observation been more utilized to date than in mammals, where the maps of many agriculturally useful species such as cows, sheep, and goats are now largely constructed as “derivatives” of the human genomic map (2). Given the conservation of DNA sequences across these species in both expressed as well as nongenic, repetitive sequences, it was not surprising that large areas of synteny were found that are proving invaluable in understanding the genomic structure of mammalian species. The less-studied species of this group benefit in that they gain access to the much more intensively characterized genetics of humans, whereas human genetics benefits in the development of model systems for disease in these related species.

As molecular markers are now being used to prepare detailed linkage maps for many plant species, syntenic relationships have also been described. Similarities between tomato, potato, and pepper (all members of the dicot family, Solanaceae) were first explored with tomato clones that were used to construct a high-resolution linkage map in this species (3). Subsequently using these same clones as molecular markers in both potato and pepper demonstrated that the former species retained a very similar genomic structure with relatively few rearrangements (only five inversions), whereas the latter species exhibited much conservation in gene repertoire but little conservation of gene order (3, 4). While tomato and potato are considered more related to each other than either is with pepper, with its almost 4-fold larger genome, the significant difference in conservation of genomic structure is inexcusable. These results have proven interesting to researchers working with solanaceous species; they have not evoked as much interest outside the tomato research community, perhaps due to the relatively small size of their agricultural markets (at least from the perspective of the economic value of their seed). It probably was also not clear how well the results from the solanaceous group could be extrapolated to other species. Similar studies comparing sorghum with its closely related but better-studied cousin, maize (both members of the monocot family, Graminaceae), showed that there was again considerable conservation of genomic structure between the two (5). While both sorghum and maize maintain identical haploid chromosome numbers of 10 and most randomly cloned genomic sequences of one also hybridize to the other species, sorghum does contain less than half as much DNA. Still the observation of significant synteny here was not too surprising and again did not elicit much interest in the agricultural research community because the decrease in genome size gained by studying sorghum instead of maize was only a factor of 2. Those crop plant researchers interested in positional cloning being perfected in humans and other model systems have been frustrated by the large genomes of so many of our agriculturally most important species. It appeared that even larger technical innovations were going to be required before these approaches could be applied on any significant scale to species such as wheat and corn.

The results of Ahn and Tanksley (6) may change this situation dramatically. In the former instances, whether in mammals or the dicot group of tomato-potato-pepper or the monocot groups of maize-sorghum or the Triticeae (7), members have usually been compared that were previously known to be closely related, both systematically as well as in molecular terms—i.e., where the majority of genomic sequences were known to cross-hybridize. In this study, however, two species, rice and maize, considered to differ significantly on both counts were compared, and a surprising amount of synteny was found. Although both are members of the Gramineae, they are thought to have diverged on either side of the grass family. It has also been shown that while expressed sequences of the two often cross-hybridize at moderate stringency, intergenic sequences rarely exhibit any cross-hybridization under low stringency, even to the intermediate Triticeae subfamily (personal observation). Because maize also possesses a different chromosome number from rice (10 vs. 12) and literally 25 times the DNA content, it would have been natural to assume that any remaining synteny would have diminished to only a microscopic level (less than a few centimorgans). By using random cDNA clones from rice that were then mapped in a pair of maize recombinant inbred populations previously characterized by over a thousand maize molecular markers, these investigators were able to show that extensive synteny still exists over very large regions. Some chromosomes appeared to have maintained most of their structure intact during the evolution of these two species over the last 50 million years. Others seem to possess the same gene complement but to have undergone extensive intrachromosomal rearrangement. Interestingly, those researchers also found that almost all of the rice clones present as single sequences in the rice genome
were found as two independent loci in the maize genome, in accord with earlier observations of extensive duplication within the maize genome and supporting its supposed amphidiploid origin (8). Even though the genome of maize is dramatically larger than the rice genome, the amount of recombination found within it is only greater by a factor of 2. As in other cases, recombination seems to be better correlated with the 2-fold increase in gene number and not the increase in total genomic size (4, 5), which is reminiscent of earlier suggestions that recombination is more closely associated with genic than with intergenic sequences (9).

Rice and maize present a situation of two species that differ 25 times in DNA content. Intergenic sequences have diverged so much that homology is no longer evident by nucleic acid hybridization, and yet the genomic structure itself seems to have been conserved at a level not anticipated. This observation has also been extended to the Triticeae, which also possess a genomic structure similar to rice (10). This general observation will spark many questions among plant researchers, such as how does substantial genomic synteny impact on systematic classification or what exactly does this tell us about the evolution of different species from ancient ancestors. However, I would like to consider a different topic. Most plant biologists tend to classify themselves in terms of the primary species they study—i.e., as maize geneticists, barley cytogeneticists, tomato physiologists, Arabidopsis molecular geneticists, etc. Perhaps this distinction is an artificial one. Clearly the vast repertoire of genes are going to be common between related species, and in the future, if not now, researchers need to be prepared to take advantage of the substantial commonality among species. Molecular-marker detection of synthetic relationships provides researchers with the potential for collating the genetics of many different plant species, such that studies of any one member can be considerably enhanced by examination of the results of the group as a whole.

Nowhere is this more obvious than in the Gramineae, which includes all of the world’s major grain crops—such as rice, corn, wheat, barley, and sorghum. Each of these crops deserves study from an economic standpoint, but all possess both significant advantages and disadvantages as model genetic systems. Corn has excellent classical genetics and cytogenetics, a high degree of sequence polymorphism that facilitates molecular-marker analysis, and an excellent system for cloning genes through transposon tagging, but it also possesses too large a genome for many positional-cloning approaches and is not yet transformable on a desirable scale. Rice possesses a very small genome and is much more easily transformed but does not possess the history of both genetic and physiological study of some of the other species. Even wheat, which has a very large genome and very little sequence polymorphism, still possesses an excellent system of cytogenetics and the ability to add and subtract both whole chromosomes and segments from the genome almost at will. As Bennetzen and Freeling (11) have proposed, perhaps the time has come to take combined advantage of the attributes of the individual members of the Gramineae and to consider them as a single genetic system for study based upon the synteny of their genomes as determined through molecular-marker analysis. Genes defined as important in one species could be both studied and isolated in related Gramineae species, and the results can be extrapolated across many other species based upon both genomic and genic homology.

An obvious beneficiary of this strategy is the Plant Genome Initiative, which has lacked an overriding philosophical rationale, in contrast to the Human Genome Project. To date, projects have been funded based upon scientific significance but also, in part, because of the economic importance of individual species (with corn and soybeans garnering a large part of the funding). A better tack would be to treat the Gramineae as one major system and to examine individual proposals on the basis of how they contribute to the larger understanding of this group. Fine-scale maps should be completed in both rice and corn to provide the foundation map. Lower-resolution maps could then be produced in other species as derivatives of this “generic” Gramineae map. The development of techniques to target specific regions of the genome, such as “bulk segregant analysis” (12), could supplant much of the map development in other species. Individual researchers would saturate the region surrounding a gene of interest and then relate their derived markers to the generic map. Researchers could then take advantage of species that offer the best advantages both for cloning and for the study of a target gene, eventually moving back to their preferred species when appropriate.

Obviously this strategy is limited by the amount of available synteny, eventually breaking down as one moves further from the monocots. Nevertheless, even considering just the Gramineae as a single group offers tremendous benefit to a large body of plant researchers. It remains to be seen just how far one might expand this approach within even larger groups of monocots or dicots. On a microscopic level one might see linked groups of genes across even greater evolutionary distances, perhaps not by hybridization techniques, which fail at lower levels of homology, but with DNA sequencing, which is even now being significantly improved in its throughput and beginning to be used for large-scale projects. By combining both sequencing and mapping, researchers might be able to collate more and more genetic data from larger families of plant species and provide a stronger basis for interaction of plant researchers.