Classification of Fruit Trees — What Is the Problem ? What Is Important ? —

H. Iketani

National Institute of Fruit Tree Science, 2–1 Fujimoto, Tsukuba 305–8605, Japan

Summary

The classification of fruit trees and other cultivated plants has become an active field in agronomy, especially after the development of molecular markers, e.g., isozyme, RFLP, RAPD, AFLP, etc. However, some researchers do not understand the nature and limitation of these markers and they erroneously infer so-called “genetic distance” or phylogeny from them. In addition, some characters of cultivated plants like network phylogeny and lack of natural population confuse their classification with the methodology of wild interspecific or higher taxa. The research problems that we are confronted with are not necessarily resolved by genetic distance or inference of phylogeny. It is important to use the methods of orthodox genetics for pedigree levels and to use those of population genetics for higher but infraspecific levels.

Introduction

Classification, systematics, and phylogeny of cultivated plants including fruit trees are important fields of agronomy. Recent experimental and methodological progress in evolutionary biology instigates these studies for cultivated plants as well as for wild plants. However, cultivated species, especially fruit trees are different from wild species in many characteristics so that it is difficult to directly apply the methodology of wild plants. Unfortunately, many researchers ignore these problems and consequently they produce apparently correct, but in reality unreasonable inferences. What is worse, in many studies errors had been made in the data analysis before the results were interpreted. Here, I discuss the problems confronted with the classification of fruit trees and other cultivated plants. Then I review afresh on the necessity of classification and discuss what data or approaches are needed for agronomic purposes.

Methodological Problems

Inappropriateness of RAPD and AFLP techniques

The classification of fruit trees aims at infraspecific relationships, e.g., relatedness among cultivars, origin of cultivars and relationships to wild relatives, as well as interspecific relationships. In this lower level, molecules like chloroplast DNA do not necessarily show enough polymorphism for analysis. Thus more polymorphic molecules and/or techniques are desired and RAPD is most preferably adopted. However, RAPD has fatal

shortcomings for these analysis.

The first shortcoming is that apparently same-sized bands from different materials do not necessarily have sequence similarity as discussed by Hillis (1994). In RFLP analysis using southern hybridization, the similarity of compared bands are supported from molecular size and sequence similarity. However, the latter evidence is very weak in RAPD. Unlike the normal PCR using longer primers and with high annealing temperature, one intends to amplify the products with annealing by chance in RAPD. Thus the similarity of inner sequences is not strongly assured. According to Williams et al. (1993) the probability of annealing of primers by chance is very small and the expected number of bands is one or very few in higher plants. But in fact it is common to obtain several or more than 10 bands. This suggests that there may exist amplified products from partially matched annealing and that the probability by chance is perhaps larger. Thus it can be thought that the same primer sites may exist in nonhomologous but similar sized parts of the genome of different materials. These products seem to have same size under normal electrophoresis conditions by mini-gel. This possibility may increase when more remote materials are compared. If one hybridizes materials to be compared and performs genetic analysis, the homology of these bands is tested. If this is not done, when compared materials have close pedigree, the possibility of homology may be high. But when comparing very remote materials like different species, homology is not assured. Though most of the above objections are theoretical, there is some evidence (e.g., Smith et al., 1994).

The second shortcoming is non-Mendelian inheritance of RAPD bands in some cases. Riedy et al. (1992) found non-parental novel bands. Hunt and Page (1992) and Ayliffe et al. (1994) also found similar results and they revealed heteroduplex formation between different

This paper was presented at Session of Utilization of DNA-Marker for Breeding and Classification in Horticultural Plants, the Commemorative Symposium of the 75th Anniversary of Japanese Society for Horticultural Science held on 3rd and 4th of April, 1998.

bands. Halldén et al. (1996) inferred competition as a source of genotyping errors in RAPD analysis.

The third shortcoming is the weakness of reproducibility and comparability. Artificial variation due to minor changes of reaction conditions were reported by many authors (e.g., Ellsworth et al., 1993). Penner et al. (1993) showed irreproducibility among laboratories. Lowe et al. (1996) suggested the standardization of techniques with the recording technical conditions as much as possible, e.g., reaction conditions, hardware, thermocycler temperature profiles, analysis of fragments (staining, photographing, etc.). As insisted by many authors, RAPD bands can be reproduced by the same or different people in the same laboratory. However, it is practically impossible to reproduce all of experimental conditions in a different laboratory. In other methods like allozyme or RFLP of chloroplast DNA, even if one cannot completely reproduce the conditions, artifacts can be eliminated since the molecules to be compared are well defined. But in RAPD since what one compares is unknown, one cannot distinguish true bands from artifacts. Consequently, I can do only conclude that RAPD analysis using only band similarity lacks the possibility of reexamination which is absolutely required for empirical studies.

Finally, the fourth shortcoming is that the inheritance of RAPD bands is inappropriate for phylogenetic analysis based on the principle of parsimony as noted by Backeljau et al. (1995). Because of the dominant nature, a heterozygous individual can reproduce recessive homozygous descendants and this result deranges phylogenetic inferences. Thus, application of cladistic methods is completely misused by many researchers (e.g., Hoey et al., 1996).

AFLP (Vos et al., 1995), which has very rapidly spread in the last two or three years because it is much more polymorphic than RAPD, is thought to be more reproducible than RAPD. Surely, since this method combines the RFLP technique and the normal PCR technique with high annealing temperature, the possibility of partial matched annealing is perhaps rarer. It is also very much easier to distinguish apparently very similar-sized different bands because AFLP is usually performed with a sequence gel. But as with RAPD, sequence similarity except for primers is not assured a priori. In addition, AFLP bands also usually inherits dominantly. From the above reasons, both AFLP and RAPD are completely inappropriate for phylogenetic analysis.

Misused "genetic distance"

In evolutionary biology the term "genetic distance" originally meant the difference of genetic structure among populations or species (Nei, 1987). Later, it was also adopted to mean the difference of the sequences of amino acids or nucleotides. In this case it can be estimated by comparing individuals. It is calculated

either directly from sequence data or indirectly from site mutations inferred from RFLP. More simple methods involving the direct comparison of fragment polymorphism is also proposed (e.g., Nei and Li, 1979) and several authors adopted this method with RFLP, RAPD or isoenzyme data (e.g., Wilde et al., 1992). However, those applications are misused in most cases. This fragment polymorphism method can be applied only for RFLP. In addition, it is valid only when almost all of the restriction fragment can be inferred to be site mutations. When applying it to anonymous polymorphism, the cause of which is not clearly inferred, some values can be calculated, but they have no genetic meaning. It is only "phenetic distance". Phenetic distance has a serious shortcoming compared with genetic distance. The value (distance) calculated from phenetic methods is specific to the particular experiment. Because it is not theoretically standardized to an absolute index like substitution rate of nucleotides or evolutionary time, it cannot be compared with other experiments performed with other organisms or other molecular markers.

Problems Due to the Nature of Cultivated Plants

Network phylogeny among cultivars

Unlike the evolution of species or higher rank taxa, the phylogeny of infraspecific populations become more or less like a network because they are cross-capable, much more so among individuals. The true phylogeny must become a network, especially in cultivars of artificial cross hybridization. Unfortunately since phylogenetic studies have been developed only with species or higher rank taxa, the methodology for inferring network phylogeny has been scarcely developed. In infraspecific groups characters including molecular ones are distributed dispersively and continuously. Prentice (1986) pointed out these problems and he strongly denied the usefulness of infraspecific classification as it is utopian.

Lack of natural populations

To infer the relationship among infraspecific groups or closely related species, population genetics approaches should be undertaken. However, in cultivated plants it is difficult to recognize natural populations. Some cultivated plants, especially fruit trees, consist of only cultivars and strains that have been already artificially selected. In addition, fruit trees have been often propagated by grafts and they have been sometimes transported over a long distance. In this case the natural geographical genetic variation might have been distorted. This becomes a serious problem in population genetics. It may distort the gene frequency and other derivative estimates.

Can anything be estimated from genetic distance and phylogeny ?

The reason why researchers in agronomy estimate genetic distances or variations is usually not explicitly stated in their work. Perhaps they think it can be used as indices for the estimation of differences and variations of agronomic characters. It is certainly important for collecting and conserving genetic resources (Dullo et al., 1997). However, its predicting power is doubtful in other cases. For example, taxonomic and genetic relationships have been thought to be useful as index of cross-ability (Hawkes 1981). Certainly, cross-ability and genetic kinship may have positive correlation in quality, but in quantity it is impossible to predict the former from the latter. The prediction of other characters may also be the same. In order to research specific useful characters, it is necessary to investigate themselves.

The predictive power of phylogeny is clearer than genetic distance as follows. Many authors have argued that the best classification system should be general purpose, containing maximum information, and with a high predictive power (e.g., Gilmour, 1940; Sneath and Sokal, 1973). Farris (1979) proved that from these viewpoints phylogenetic classification is the best one compared with phenetics and evolutionary systematics. For example, if a character which seems to be the same is inferred to be not homologous from phylogeny, we can suppose that there may exist differences in that character if we investigate in more detail, e.g., anatomy, physiology, molecular biology, etc. However, as discussed above, we cannot infer network phylogeny of infraspecific groups. Even if forcibly inferring tree-like phylogeny, it may be completely different from the truth and the predictive power is, if any, much less.

What Can We Do ? What Should We Do ?

In the classification of cultivated plants, the analyzed level expands from interspecific level to pedigree level. In the latter both phylogenetics and population genetics are powerless. This level is the field of genetics. However, research in this area has progressed much less in fruit trees than in other crops, especially in Japan. The subjects in this level, e.g., inference of pedigrees, correlation between useful characters and molecular markers, should be resolved with methodology of genetics. Genome mapping will provide much useful and important information in this field. RAPD, AFLP and other molecular polymorphism approaches are very powerful tools in this level if the data are analyzed genetically. But anonymous polymorphism has scarcely any value except for the identification of individuals. SSR, or microsatellites may also become a powerful tool. Because of the laborious nature of finding species-specific primers, this method has spread more slowly than RAPD and AFLP. However, since it has many

advantages, e.g., very polymorphic, a large number of loci, codominance, etc., in evolutionary biology it has enabled inferences to be performed that were otherwise impossible. In fruit trees there have been very few applications of this method (Guilford et al., 1997; Thomas and Scott, 1993; Thomas et al., 1994). But now, almost all research groups who engage in the genome mapping of fruit trees have begun to use it. It will also become a powerful tool for inferring pedigree, the degree of relatedness, and possibly network phylogeny.

In population level, I think the appropriate method is population genetics. Though artificial selection pressure and other natural or artificial phenomena must have disturbed the structure of populations, I think that those disturbances can be estimated from the information provided by genome analysis. Unfortunately, population genetics of fruit trees is also far behind. It is perhaps impossible to analyze those plants that completely lack natural populations, still we have many subjects and materials to do.

The inference of network phylogeny of infraspecific level will also become important if the methodology develops. Felsenstein (1985) asserted that the characters of organisms are not independent from phylogeny and that phylogeny must be considered when comparing organisms. This opinion has been very rapidly accepted in ecology (e.g., Harvey and Pagel, 1991), and the methodology of comparison among species or higher rank taxa has completely changed in this decade.

In comparing among infraspecific organisms, these new methods are not applicable at present because the inference of network phylogeny is impossible. But in the future, comparative methods in agronomy where the comparison of infraspecific cultivars is common will drastically change.

Acknowledgments

I wish to thank Dr. Nobuhiro Minaka of the National Institute of Agro-Environmental Sciences and the members of the Mailing List "EVOLVE" for their helpful advises and discussions. I am also grateful to Dr. Tadashi Kajita of the University of Tokyo for critical reading of the manuscript.

Literature Cited

- Ayliffe, M. A., G. J. Lawrence, J. G. Ellis and A. J. Pryor. 1994. Heteroduplex molecules formed between allelic sequences cause nonparent RAPD bands. *Nuc. Acids Res.* 22 : 1632–1636.
- Backeljau, T., L. De Bruyn, H. De Wolf, K. Joprdans, S. Van Dongen, R. Verhagen and B. Winnepeninkx. 1995. Random amplified polymorphic DNA (RAPD) and parsimony methods. *Cladistics* 11 : 119–130.
- Dullo, M. E., L. Guarino and B. V. Ford-Lloyd. 1997. A bibliography and review of genetic diversity studies of African germplasm using protein and DNA markers. *Genet. Res. Crop Evol.* 44 : 447–470.

- Ellsworth, D. L., K. D. Rittenhouse and R. L. Honeycutt. 1993. Artificial variation in random amplified polymorphic DNA banding patterns. *Biotechniques* 14 : 214–217.
- Farris, J. S. 1979. The information content of the phylogenetic system. *Syst. Zool.* 28 : 483–519.
- Felsenstein, J. 1985. Phylogenies and the comparative method. *Amer. Nat.* 125 : 1–15.
- Gilmour, S. L. 1940. Taxonomy and philosophy. p.404–424. In : J. Huxley (ed.). *The new systematics*. Oxford Univ. Press, Oxford.
- Guilford, P., S. Prakash, J. M. Zhu, E. Rikkerink, S. Gardiner, H. Bassett and R. Forster. 1997. Microsatellites in *Malus x domestica* (apple) : abundance, polymorphism and cultivar identifications. *Theor. Appl. Genet.* 94 : 249–254.
- Harvey, P. H. and M. D. Pagel. 1991. *The comparative method in evolutionary biology*. viii+239 pp. Oxford Univ. Press, Oxford.
- Halldén, C., M. Hansen, N.-O. Nilsson, A. Hjerdin and T. Säll. 1996. Competition as a source of errors in RAPD analysis. *Theor. Appl. Genet.* 93 : 1185–1192.
- Hawkes, J. G. 1981. Biosystematic studies as an aid to breeding research and plant breeding. *Kulturpflanze* 29 : 327–336.
- Hillis, D. M. 1994. Homology in molecular biology. p. 339–368. In : B. K. Hall (ed.). *Homology, the hierarchical basis of comparative biology*. Academic Press, San Diego.
- Hoey, B. K., K. R. Crowe, V. M. Jones and N. O. Polans. 1996. A phylogenetic analysis of *Pisum* based on morphological characters, and allozyme and RAPD markers. *Theor. Appl. Genet.* 92 : 92–100.
- Hunt, G. J. and R. E. Page, Jr. 1992. Patterns of inheritance with RAPD molecular markers reveal novel types of polymorphism in the honey bee. *Theor. Appl. Genet.* 85 : 12–20.
- Lowe, A. J., O. Hanotte and L. Guarino. 1996. Standardization of molecular genetic techniques for the characterization of germplasm collections : the case of random amplified polymorphic DNA (RAPD). *Plant Genet. Res. Newsl.* No. 107 : 50–54.
- Nei, M. 1987. *Molecular evolutionary genetics*. x+512pp. Columbia Univ. Press, New York.
- Nei, M. and W.-H. Li. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. USA* 76 : 5269–5273.
- Penner, G. A., A. Bush, R. Wise, W. Kim, L. Domier, K. Kasha, A. Laroche, G. Scoles, S. J. Molnar and G. Fedek. 1993. Reproducibility of random amplified polymorphic DNA (RAPD) analysis among laboratories. *PCR Method Appl.* 2 : 341–345.
- Prentice, H. C. 1986. Continuous variation and classification. p.21–32. In : B. T. Styles (ed.). *Intraspecific classification of wild and cultivated plants*. Clarendon Press, Oxford.
- Riedy, M. F., W. J. Hamilton III and C. F. Aquadro. 1992. Excess of non-parental bands in offspring from known primate pedigrees assayed using RAPD PCR. *Nuc. Acids Res.* 20 : 918.
- Smith, J. J., J. S. Scott-Craig, J. R. Leadbetter, G. Y. Bush, D. L. Roberts and D. W. Fulbright. 1994. Characterization of random amplified polymorphic DNA (RAPD) products from *Xanthomonas campestris* and some comments on the use of RAPD products in phylogenetic analysis. *Mol. Phylog. Evol.* 3 : 135–145.
- Sneath, P. H. and R. R. Sokal. 1973. *Numerical taxonomy : the principles and practice of numerical classification*. xvi + 573 pp. W. H. Freeman, San Francisco.
- Thomas, M. R. and N. S. Scott. 1993. Microsatellite repeats in grapevine reveal DNA polymorphisms when analyzed as sequence-tagged sites (STS). *Theor. Appl. Genet.* 86 : 985–990.
- Thomas, M. R., P. Cain and N. S. Scott. 1994. DNA typing of grapevines : A universal methodology and database for describing cultivars and evaluating genetic relatedness. *Plant Mol. Biol.* 25 : 939–949.
- Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper and M. Zabeau. 1995. AFLP : a new technique for DNA fingerprinting. *Nuc. Acids Res.* 23 : 4407–4414.
- Wilde, J., R. Waugh and W. Powell. 1992. Genetic fingerprinting of *Theobroma* clones using randomly amplified polymorphic DNA markers. *Theor. Appl. Genet.* 83 : 871–877.
- Williams, J. G. K., M. K. Hanafey, J. A. Rafalski and S. V. Tingey. 1993. Genetic analysis using random amplified polymorphic DNA markers. *Methods Enzymol.* 218 : 704–740.