COMMENT / COMMENTAIRE

Comment on "AFLP data and the origins of domesticated crops"[1]

F. Salamini, M. Heun, A. Brandolini, H. Özkan, and J. Wunder

Abstract: We review some concepts and methods of handling and using DNA fingerprinting in phylogenetic analyses related to crop domestication. Particular reference is made to AFLP markers and mode and place of einkorn, barley, and tetraploid wheat domestication in the Neolithic by human communities in the Fertile Crescent. The reconsideration of AFLP databases of domesticated and wild lines demonstrates that phylogenetic tree topologies, originally described for the three species, match closely the new results obtained by principle coordinate analyse.

Key words: AFLPs, discontinuous markers, crop domestication, einkorn wheat, barley, tetraploid wheat.

Résumé : Les auteurs passent en revue certains des concepts et des méthodes permettant de gérer et d'exploiter les empreintes génétiques pour des fins d'analyse phylogénétique dans le contexte de la domestication des espèces culti-vées. Une attention particulière est portée aux marqueurs AFLP ainsi qu'au mode et à l'endroit de la domestication de l'engrain, de l'orge et du blé tétraploïde lors du néolithique par des communautés humaines habitant le Croissant fer-tile. Un réexamen des bases de données sur les AFLP chez les lignées sauvages et cultivées montre que les topologies des arbres phylogénétiques décrites initialement pour ces trois espèces sont très proches des nouveaux résultats obtenus par analyse en coordonnées principales.

Mots clés : AFLP, marqueurs discontinus, domestication des cultures, engrain, orge, blé tétraploïde.

[Traduit par la Rédaction]

In a recent paper published in this journal, the point was taken that the use of a certain class of molecular markers (AFLPs; Vos et al. 1995) coupled to particular tree-building methods (like neighbor joining (NJ); Saitou and Nei 1987) may generate biased phylogenetic topologies (Allaby and Brown 2003). The issue in question concerned the possibil-ity of inferring the origin of crops from DNA fingerprinting. The paper made particular reference to the work done on three crops: einkorn wheat, barley, and tetraploid wheat (Heun et al. 1997; Özkan et al. 2002; Badr et al. 2000). The simulation studies by Allaby and Brown (2003) considered the NJ procedure and genetic distances calculated from AFLP fingerprinting. Basic assumptions underlying the sim-ulation study can be questioned. For instance, it is difficult to accept that the alleles of 200 molecular marker loci segre-gate independently, given that grass genome maps covering 10 000 cM are unrealistic. In addition, the NJ trees of Allaby and Brown (2003) are in part unclear because a small num-ber of nonetheless important bootstrap values concerning major tree branches are missing.

Here we argue that the intrinsic quality of the data used in our phylogenies on cereal domestication contribute more to understanding the scientific problem at hand than do simula-tions.

The paper by Heun et al. (1997) reported the site of einkorn wheat domestication. Site identification was possi-ble (*i*) because the domesticated forms represent a mono-phyletic group and (*ii*) because they are closely related to a group of wild lines still growing in the Karacadag range of southeastern Turkey. The genetic markers used were AFLP (288 loci) and the genetic distances among fingerprinted ge-notypes were calculated by the algorithms NEI 72, NEI-UB, Rogers-W, DICE and average, Euclidean, and Euclidean squared distances (citations are in the caption of Fig. 1). The phylogenetic analyses were performed by neighbor-joining, FITCH (from W.M. Fitch), and REML (unrestricted maxi-mum likelihood) procedures. Besides considering all single lines, phylogenies were derived from allelic frequencies at marker loci in geographical groups of wild and domesticated lines. All single line-based trees calculated as specified above revealed, with minor modifications, the same basic to-pology, thus showing a clear monophyletic origin of the do-mesticated genetic pool and its close relationship to the wild lines from the Karacadag. Minor topological modifications observed in trees constructed with the same or different

F. Salamini,[2] M. Heun, A. Brandolini, H. Özkan, and J. Wunder. Max-Planck-Institut für Züchtungsforschung, Carl-von-Linné-Weg 10, 50829 Köln, Germany.

[1]Appears in Genome, **46**: 448–453 (2003).
[2]Corresponding author (e-mail: salamini@mpiz-koeln.mpg.de).

Fig. 1. Unrooted trees based on different genetic distance algorithms and methods of tree building. 1, genetic distance NEI 72 is from Nei (1972) and the neighbor-joining method (NJ) from Saitu and Nei (1987); 2, NEI UN distance from Nei (1978) and NJ; 3, Rogers–Wright (RW) distance from Wright (1978) and NJ; 4, NEI 72 and the FITCH clustering program from Fitch and Margoliash (1967); 5, NEI UN and FITCH; 6, RW and FITCH; 7, average genetic distance (AD) from Rohlf (1982; the same source for Euclidean (ED) and Euclidean squared (ESD), distances) and FITCH; 8, ED and FITCH; 9, ESD and FITCH; 10, REML restricted maximum-likelihood estimation from Felsenstein (1981). Trees were generated by the PHYLIP (Felsenstein 1989) package and genetic distances by the computer program NTSYS (Rohlf 1993). The description of the AFLP procedure and the AFLP primer combinations used are in Heun et al. (1997). AFLP fragments scored were 288 in total. In the same paper, the precise locations of origin of the groups of lines indicated with the letters A (9 lines), B (29), C (7), D (19 from the Karacadag range), E (8), G (33), H (62), I (13), and L (14) are reported. Domesticated lines (Mono) and feral forms (Aegi) were, respectively, 68 and 9.

methods are inherent to the mechanics of tree building, and their relevance is taken care of quantitatively by bootstrapping procedures (Efron 1982). Moreover, it is commonly experienced that increasing the number of genotypes analyzed reduces bootstrap values at a tree's branches. Despite these shortcomings, all trees derived from the einkorn line collection fingerprinted by AFLP were coherent in indicating the monophyletic origin of the domesticated pool. The same data were used to derive allelic frequencies for geographic groups of lines, frequencies which have been entered in specific tree building programs. These einkorn results were available at the time of the original publication at a web site and are now reproduced in Fig. 1. The 10 trees pictured have been constructed with different distance algorithms and methods, all of which support the monophyletic derivation of the domesticated lines from the wild geographic group D collected in the Karacadag region (feral forms known as *T. aegilopoides* included). A similar logic of tree building and testing was followed in the papers describing the relations between wild and domesticated lines of barley and of tetraploid wheat. While for barley the monophyletic origin of domesticated lines was evident (Badr et al. 2000), tetraploid hulled emmer and hard wheat domesticated lines had two distinct topologies. However, in the NJ tree published, the two groups of lines had a common derivation from southeastern Turkey *T. dicoccoides* wild populations (Özkan et al. 2002).

To add further evidence in support of the mode of domestication of major crops, we have reconsidered our AFLP databases relative to wild and domesticated lines of einkorn wheat, barley, and tetraploid wheat. The data have been reanalyzed via principal coordinate analysis (PCA). Ordination methods such as PCA (Hotelling 1933), like tree-building procedures, simplify the presentation of multivariate data. The logic of this analysis is the existence of different degrees of correlation between allelic frequencies in pairs of populations. The extents of the correlation measure the history of the populations in terms of their common descent (Cavalli-Sforza et al. 1994). In an *n*-dimensional space formed by gene frequencies at *n* genes, a first principal axis is calculated and a straight line is obtained along which the populations are distributed. The second and further components can be extracted in the form of additional axes orthogonal to the previous ones. The information summarized by each component can be estimated as a proportion of the total variance. The connection between phylogenetic trees and PC analyses should be close: the first split in a tree would correspond to the separation of the populations operated by the first PC, and so on. Because genetic distances

among populations are correlated to geographical distances, the two principal components of a PC analysis align frequently with a geographic map. This is the reason why PC analysis has been extensively used to follow the geographical distribution of human genes (Cavalli-Sforza et al. 1994). The analysis is also adapted to assign a geographical origin to crops, their mode of domestication included. PCA maps supplement phylogenetic trees in presenting data in a different form: when PCA produces the same information as phylogenetic trees, the overall result is considered highly reliable (Cavalli-Sforza et al. 1994).

The PCA results support the findings of Heun et al. (1997), Badr et al. (2000), and Özkan et al. (2002). For the three crops, the field defined by the first two or three principal axes is reported in Fig. 2. The following major observations emerge: as in phylogenetic trees, (*i*) the domesticated lines of einkorn and barley group together supporting a monophyletic origin; (*ii*) for einkorn, barley, and tetraploid wheat, the relative topologies of domesticated and related groups of wild lines match those reported in the papers mentioned; (*iii*) for tetraploid wheat, the domesticated groups consisting of hulled emmer and hard wheat lines occupy distinct positions in the principal coordinate space of Fig. 2C, supporting again the result of the NJ tree of Özkan et al. (2002). In this case, the wild lines of *T. dicoccoides* most related to both groups of domesticated tetraploid wheat were sampled in southeast Turkey. The AFLP database of tetraploid wheat, while supportive of a common geographical origin of hard and emmer domesticated forms, do not exclude the possibility that the domestication of this crop was biphyletic.

The major criticism of Allaby and Brown (2003) was that a sufficient linkage between the markers being used must be established before AFLPs should be considered in NJ analyses. However, since 1991, several barley maps have been established and the seven chromosomes of the species support total linkage values of 1096 cM (RFLP; Heun et al. 1991), 1873 cM (AFLP and RFLP; Becker at al. 1995), and 1597 cM (AFLPs and RFLPs; Castiglioni et al. 1998). For Einkorn, a first RFLP map of 714.2 cM was published by Dubcovsky et al. (1996) and a linkage map of 856 cM with AFLPs and RFLPs has been provided by Taenzler et al. (2002). Based on this, it seems that the demand by Allaby and Brown (2003) was abundantly fulfilled. For exemple, Heun et al. (1997) considered 288 AFLP polymorphic markers and, given a size of 856 cM for the A genome (Taenzler et al. 2002), linkage among those bands would be, on average, about 3.0 cM. At the time when the *Science* paper was published, already 75 out of 288 AFLP bands were mapped

1 NEI 72 – NJ   6 RW – FITCH
2 NEI UN – NJ   7 AD – FITCH
3 RW – NJ   8 ED – FITCH
4 NEI 72 – FITCH   9 ESD – FITCH
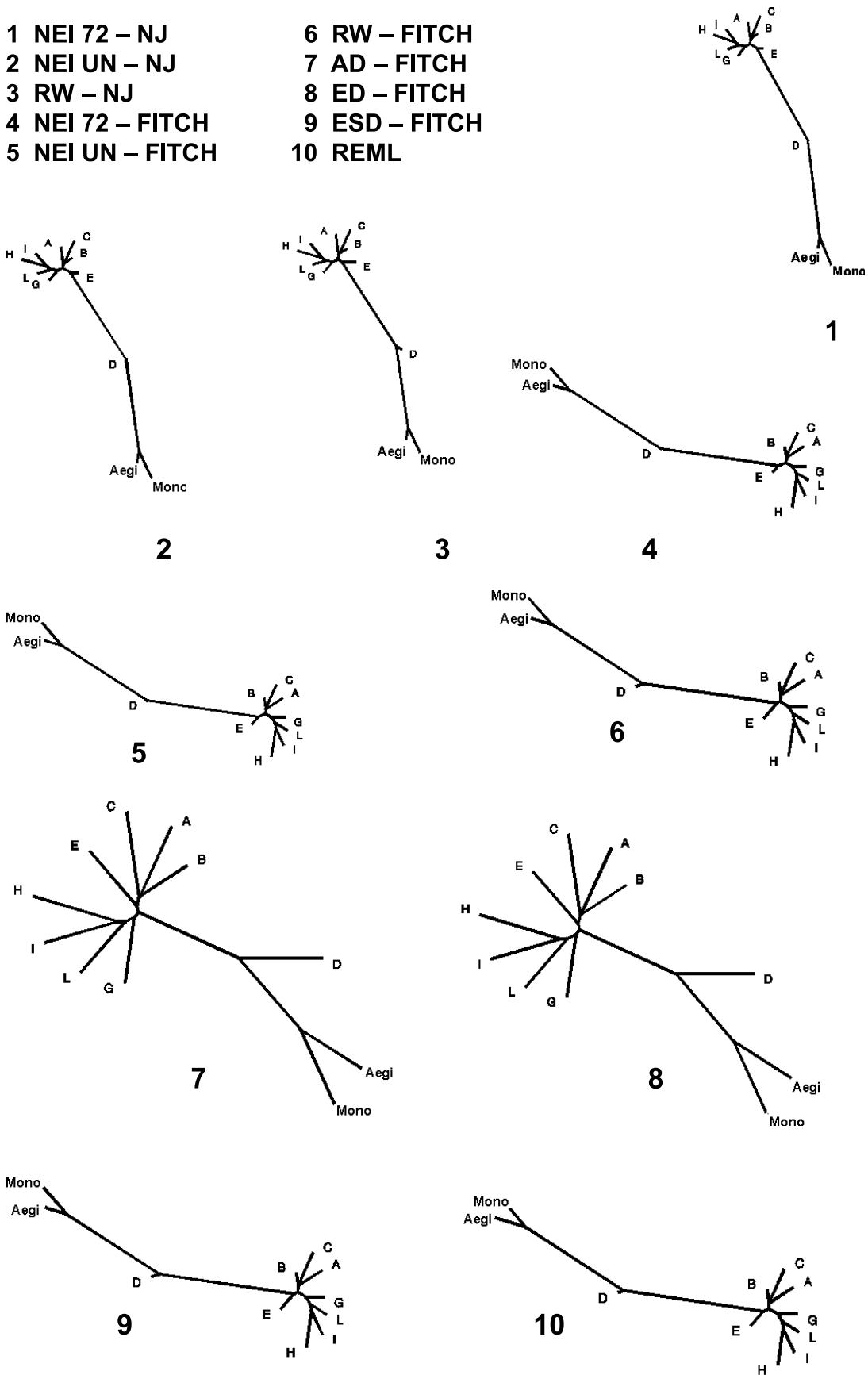5 NEI UN – FITCH   10 REML

Fig. 2. Results of the principal coordinate analysis (PCA) of the AFLP data sets used for the phylogenetic reconstructions of relationships between domesticated and wild lines in einkorn, barley, and tetraploid wheats. The analyses were carried out using the computer package NTSYSpc (Rohlf 1998). The pairwise genetic distances were calculated according to the JACCARD algorithm (Jaccard 1908). (A) Einkorn. The AFLP database concerned 68 domesticated, 9 feral, and 248 wild lines out of those considered in the paper by Heun et al. (1997). Thirteen lines were not considered because their fingerprints were missing some AFLP bands. The first three axes explain 29% of the total variability. (B) Barley. Four hundred polymorphic AFLP fragments were scored in 374 lines (57 domesticated and 317 wild), as described in Badr et al. (2000). The first two axes explain 11% of the total variability. (C) Tetraploid wheats. Nineteen domesticated lines of hulled emmer and 23 of domesticated free-threshing hard wheat were studied based on 198 AFLPs from the database of Ozkan et al. (2002). Ninety-five wild emmer lines originating from several locations in the Fertile Crescent were included. The first two axes explain 20% of the total variability. Detailed methodological descriptions of the lines studied are reported in the three papers cited.
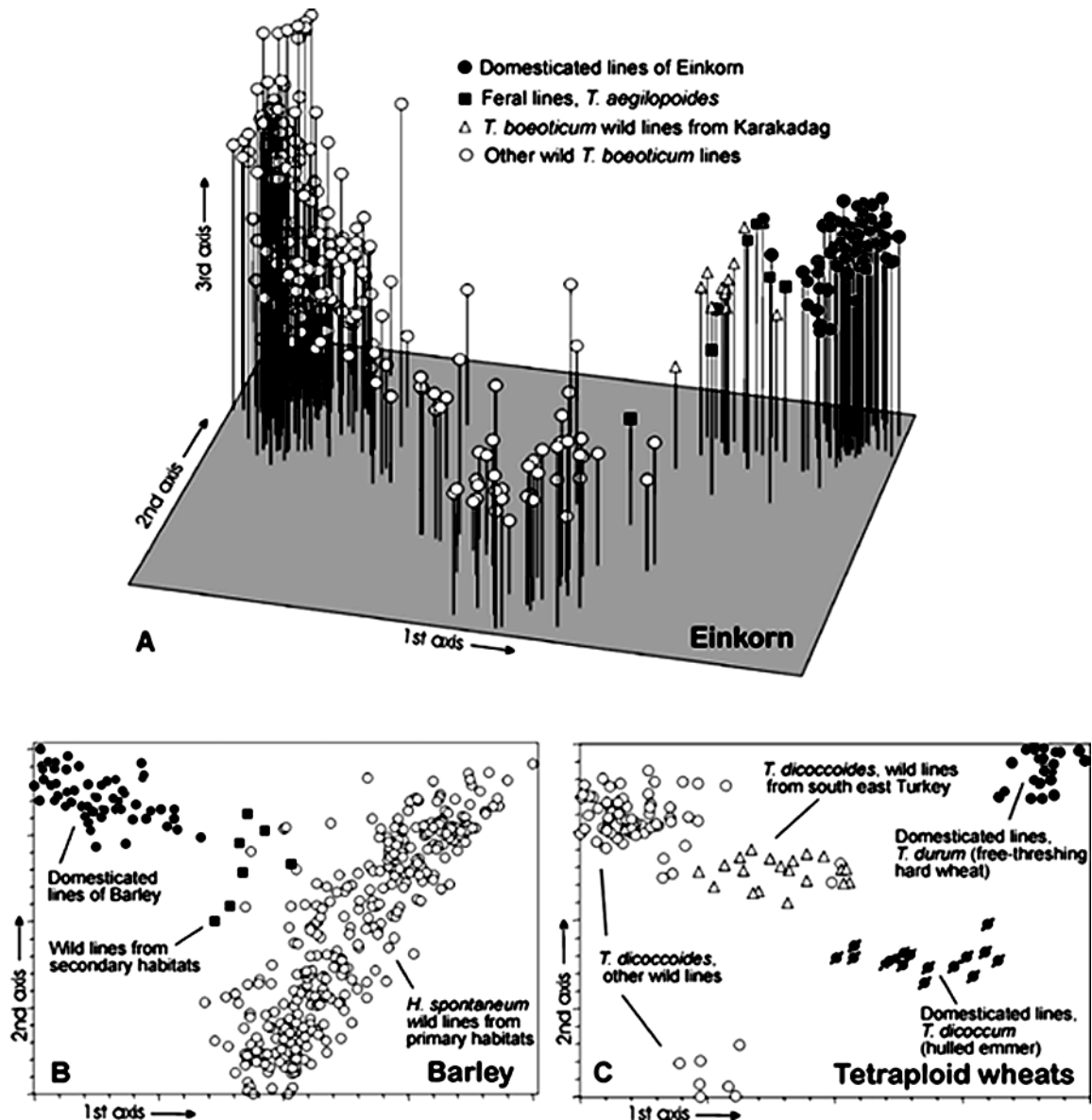
to loci of chromosomes 1 to 7 (Heun et al. 1997; see note 15). Thus, the basic assumption of Allaby and Brown (2003) of independent marker inheritance is not realistic.

The second issue arising from Allaby and Brown (2003) concerns nature and genetic behavior of AFLP markers. In AFLP analysis (Vos et al. 1995), two restriction enzymes are used to digest DNA and adapters are linked to the resulting DNA cutting sites. A PCR amplification follows based on oligonucleotide primers complementary to adapters. A further PCR is carried out using primers with 2–3 nucleotide extensions from both annealing sites into the pre-amplified DNA fragments. This reduces the number of amplified fragments, thus leading to AFLP fingerprints where single bands are clearly scorable. In a literature review carried out for the period from 1996 to August 2003, 1790 papers have been found that make use of AFLPs. They deal with genetic di-

versity in crop and wild plants, crop and plant genome evolution, intra- and interspecific comparisons, genetic–geographical structure of populations, phylogenetic analyses, molecular map construction, gene cloning, QTL mapping, and crop domestication. The last group of papers reports, for the period 2000–2003, on the domestication of eggplant, chicory, radish, common bean, wheat, coffee, *Vigna unguiculata*, cotton, rice, cassava, *Allium*, and *Setaria italica*. In DNA fingerprinting, AFLP markers are attractive because, depending on the species considered, in an autoradiography up to 50 bands appear to be polymorphic in intra- or interpopulation comparisons. Moreover, it is known that the very large majority of segregating AFLP fragments map at biallelic loci. Our laboratory has located 438 such loci on an einkorn linkage map (Taenzler et al. 2002) and 819 on an integrated map of barley (Pozzi et al. 2003). In barley, it has been clarified that comigrating AFLP bands produced by amplifying DNA of different genotypes derive from identical DNA sequences. It can be concluded that an AFLP polymorphism serves as an unequivocal specific marker in different genotypes or populations of the same species. Even more, a recent study based on AFLP band sequencing indicates that, in interspecific comparisons within the genus *Hordeum*, AFLP fragment size, together with band intensity, is a reliable indicator of band DNA homology (El Rabey et al. 2002).

In the last paragraph of their paper, Allaby and Brown (2003) state that, like for AFLPs, "anonymous markers such as RFLPs and RAPDs would behave in an identical way in these simulations. Therefore, our conclusions regarding the possible invalidity of the results of phylogenetic analysis apply to any dataset obtained by anonymous band scoring." The statement introduces a misunderstanding of the nature of "anonymous" markers: AFLPs result from nucleotide polymorphisms within a restriction enzyme recognition site, or within the annealing site of the selective nucleotides, as well as from insertions–deletions, all possibilities that make them quite related to RFLPs. Microsatellites derive from simple sequence repeat variations, and SNPs reveal nucleotide polymorphisms precisely located in genomes. Therefore, it is evident that these markers as AFLPs describe variations of a DNA sequence scattered along the genomic DNA, whereas when gene sequences are compared, the nucleotides are contiguous. This is why in a short EMBO report we have indicated the former as "discontinuous" DNA markers (Martin and Salamini 2000). In support of their capacity to describe nucleotide variations, it is known that discontinuous markers can provide a measure of π, the genomic nucleotide diversity index of Nei and Li (1979) (for AFLP data see Mougel et al. 2002; Innan et al. 1999). Discontinuous markers, traditionally the tools of geneticists and plant breeders, are diffusing into studies of natural diversity and evolution, of human populations included (Nei and Takezaki 1996). This because gene sequence comparisons of one or few loci from nuclear or organelle genomes may not provide suitable measures of overall genome similarity, particularly at the transitional territory between populations and species, where nuclear alleles tend to be older than speciation processes. This point has been recently discussed by Rokas et al. (2003). Because each gene may give a different tree of alleles, only the sum of all trees comes close to explaining the

separation of emerging biological entities. This is also the significance of using, in such studies, molecular markers scattered along a genome providing, however, that their number is as high as possible(Cavalli-Sforza et al. 1994).

To conclude, the interpretation of the mode of origin of the majority of founder crops of the Neolithic agriculture in the Near East is not the outcome only of AFLP studies. Based on (*i*) presence or absence of patterns indicative of founder effects in the domesticated pool, (*ii*) uniformity of alleles governing principal domestication traits, and (*iii*) wild species diversity in the area under consideration, Zohary and Hopf (2000) support a monophyletic origin of founder crops. Even authors frequently citing the diphyletic domestication of barley (based on the finding of varieties mutated either at one of two genetic loci supporting the non-brittle rachis character; Abbo et al. 2001) forget to mention that recombination among these two tightly linked loci has never been reported. This would, at least, admit an alternative hypothesis: the existence of two complementing alleles of one locus, with one of them derived from the other by a mutation occurring in the cultivated genetic pool. In the review by Salamini et al. (2002), a point is made concerning the evidence that, at least at the excavated archaeological site of Abu Hureyra, cultivation of rye preceded its domestication, which followed as a gradual process. This kind of step-by-step man-guided evolution would be in contrast with the existence of the founder effect induced by a mono- or biphyletic domestication; a conclusion, on the other hand, strongly supported by genetic data. A multiregionality of cultivation of wild species can, however, be reconciled with a restricted regionality of domestication: superior landraces of founder crops may have emerged in a core area from which they moved, thus displacing local races.

References

Abbo, S., Lev-Yadun, S., and Ladizinsky, G. 2001. Tracing the wild genetics stocks of crop plants. Genome, **44**: 309–310.

Allaby, R.G., and Brown, T.A. 2003. AFLP data and the origins of domesticated crops. Genome, **46**: 448–453.

Badr, A., Müller, K., Schäfer-Pregl, R., El Rabey, H., Effgen, S., Ibrahim, H.H., Pozzi, C., Rohde, W., and Salamini, F. 2000. On the origin and domestication history of barley (*Hordeum vulgare*). Mol. Biol. Evol. **17**: 499–510.

Becker, J., Kuiper, M., Vos, P., Salamini, F., and Heun, M. 1995. Combined mapping of AFLP and RFLP markers in barley. Mol. Gen. Genet. **249**: 65–73.

Castiglioni, P., Pozzi, C., Heun, M., Terzi, V., Müller, K.J., Rohde, W., and Salamini, F. 1998. An AFLP-based procedure for the efficient mapping of mutations and DNA probes in barley. Genetics, **149**: 2039–2056.

Cavalli-Sforza, L.L., Menozzi, P., and Piazza, A. 1994. The history and geography of human genes. Princeton University Press, Princeton, N.J.

Dubcovsky, J., Luo, M.-C., Zhong, G.-Y., Bransteitter, R., Desai, A., Kilian, A., Kleinhofs, A., and Dvorak, J. 1996. Genetic map of dipoid wheat, *Triticum monococcum* L., and its comparison with maps of *Hordeum vulgare* L. Genome, **143**: 983–999.

Efron, B. 1982. The jackknife, bootstrap, and other resampling plans. Society for Industrial and Applied Mathematics, Philadelphia, Penn.

El Rabey, H.A., Badr, A., Schäfer-Pregl, R., Martin, W., and Salamini, F. 2002. Speciation and species separation in *Hordeum* L. (Poaceae) resolved by discontinuous molecular markers. Plant Biol. **4**: 567–575.

Felsenstein, J. 1981. Evolutionary trees from gene frequencies and quantitative characters: finding maximum likelihood estimates. Evolution, **35**: 1229–1242.

Felsenstein, J. 1989. PHYLIP – phylogeny interenc package (version 3.2). Cladistics, **5**: 164–166.

Fitch, W.M., and Margoliash, E. 1967. Construction of phylogenetic trees. Science (Washington, D.C.), **155**: 279–284.

Heun, M., Kennedy, A.E., Anderson, J.A., Lapitan, N.L.V., Sorrells, M.E., and Tanksley, S.D. 1991. Construction of a restriction fragment length polymorphism map for barley (*Hordeum vulgare*). Genome, **34**: 437–447.

Heun, M., Schäfer-Pregl, R., Klawan, D., Castagna, R., Accerbi, M., Borghi, B., and Salamini, F. 1997. Site of Einkorn wheat domestication identified by DNA fingerprinting. Science (Washington, D.C.), **278**: 1312–1314.

Hotelling, H. 1933. Analysis of a complex of statistical variables into principal components. J. Educ. Psychol. **24**: 417–441; 498–520.

Innan, H., Terauchi, R., Kahl, G., and Tajima, F. 1999. A method for estimating nucleotide diversity from AFLP data. Genetics, **151**: 1157–1164.

Jaccard, P. 1908. Nouvelle recherches sur la distribution florale. Bull. Soc. Vaud. Sci. Nat. **44**: 223–270.

Martin, W., and Salamini, F. 2000. A meeting at the gene. Biodiversity and natural history. EMBO Rep. **1**: 208–210.

Mougel, C., Thioulouse, J., Perrière, G., and Nesme, X. 2002. A mathematical method for determining genome divergence and species delineation using AFLP. Int. J. Systematic Evol. Microbiol. **52**: 573–586.

Nei, M. 1972. Genetic distance between populations. Am. Nat. **106**: 283–292.

Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics, **89**: 583–590.

Nei, M., and Li, W. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. USA, **76**: 5269–5273.

Nei, M., and Takezaki, N. 1996. The root of the phylogenetic tree of human populations. Mol. Biol. Evol. **13**: 170–177.

Özkan, H., Brandolini, A., Schäfer-Pregl, R., and Salamini, F. 2002. AFLP analysis of a collection of tetraploid wheats indicates the origin of emmer and hard wheat domestication in southeast Turkey. Mol. Biol. Evol. **19**: 1797–1801.

Pozzi, C., di Pietro, D., Halas, G., Roig, C., and Salamini, F. 2003. Integration of a barley (*Hordeum vulgare*) molecular linkage map with the position of genetic loci hosting 29 developmental mutants. Heredity, **90**: 390–396.

Rohlf, F.J. 1982. Consensus indices for coparing classifications. Math. Biosci. **59**: 131–144.

Rohlf, F.J. 1993. NTSYS-pc. Numerical taxonomy and multivariate analysis system. Applied Biostatistics Inc., New York, N.Y.

Rohlf, F.J. 1998. NTSYSpc. Numerical taxonomy and multivariate analyis, version 2.02i. Exeter Software, New York, N.Y.

Rokas, A., Williams, B.L., King, N., and Carroll, S.B. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature (London), **425**: 798–804.

Saitou, N., and Nei, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**: 406–425.

Salamini, F., Özkan, H., Brandolini, A., Schäfer-Pregl, R., and Martin, W. 2002. Genetics and geography of wild cereal domestication in the Near East. Nat. Rev. Genet. **3**: 429–441.

Taenzler, B., Esposti, R.F., Vaccino, P., Brandolini, A., Effgen, S., Heun, M., Schäfer-Pregl, R., Borghi, B., and Salamini, F. 2002. Molecular linkage map of Einkorn wheat: mapping of storage – protein and soft-glume genes and bread-making quality QTLs. Genet. Res. Camb. **80**: 131–143.

Vos, P., Hogers, R., Bleeker, M., Rijans, M., Van der Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M., and Zabeau, M. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. **23**: 4407–4414.

Wright, S. 1978. Evolution of the genetics of population. Vol. 4. Variability within and among natural populations. University of Chicago Press, Chicago, Ill.

Zohary, D., and Hopf, M. 2000. Domestication of plants in the old world: the origin and spread of cultivated plants in West Asia, Europe, and the Nile Valley. Oxford University Press, New York, N.Y.