Linkage relationships of loci controlling protein amounts in maritime pine (*Pinus pinaster* Ait.)

Nasser Bahrman and
Catherine Damerval

Laboratoire de Génétique des Systèmes Végétaux,
CNRS-INRA-UPS, La Ferme du Moulon, 91190
Gif-sur-Yvette, France.

The comparison of two-dimensional protein patterns of 56 haploid megagametophytes from a single maritime pine tree allowed the analysis of genetic determinism of qualitative and quantitative variation. One hundred and nineteen loci were localized in 12 linkage groups while two other loci were unlinked. Among the 81 qualitative variants, 58 could be interpreted as allelic products of structural genes. Their distribution in the linkage groups appeared to be non-random, since most of them clustered at a common point in two groups. The remaining 23 qualitative and 40 quantitative variants could best be explained by assuming monogenic determinism of protein amounts. Pleiotropic effects could be responsible for some of the variations in amounts of proteins which cosegregate.

INTRODUCTION

Abundant intra- and interspecific variability exists for enzyme quantity and activity (Laurie-Ahlberg *et al.*, 1982; Yamazaki and Matsuo, 1984). Numerous studies especially in *Drosophila*, mouse and also maize, have been undertaken to elucidate the genetic control of this variation. Numerous polymorphic regulatory elements have been found, which can be localized near or far from the structural gene. Such elements can operate at all stages of protein synthesis, from transcriptional (Sassone-Corsi and Borrelli, 1986) to post-translational levels (Beevers, 1982). The inheritance of enzyme activity (quantity) has been also studied in some cases (Gibson *et al.*, 1986; Hammer and Wilson, 1987).

Relatively abundant gene products are revealed by two-dimensional electrophoresis (2-D PAGE, O'Farrell, 1975) independently of possible enzymatic function. The genetic variability of protein sequences and also of protein amounts can be analyzed by this technique. The inheritance of protein amounts in hybrids has been studied in a few species (Anderson *et al.*, 1985a; Jungblut and Klose, 1985; Leonardi *et al.*, 1988; de Vienne *et al.*, 1988).

The haploid megagametophyte of conifer species provides a unique opportunity to analyze the genetic inheritance of enzyme activity or pro-

tein amounts. Ten to forty-three enzyme loci were studied in some species and linkage groups have been constructed. One of these linkage groups (associating leucine aminopeptidase, glutamate oxaloacetate transaminase and alcohol dehydrogenase loci) has been mapped in knobcone pine, lodgepole pine, loblolly pine, Jeffrey pine, sugar pine (Conkle, 1981; Strauss and Conkle, 1986), and Scots pine (Niebling *et al.*, 1987).

We have undertaken a molecular characterization of maritime pine (*Pinus pinaster* Ait., $2n = 2x = 24$), a mediterranean species for which, to our knowledge, no data on protein polymorphism are available. In this paper, the segregation of the protein characters as revealed by 2-D PAGE is studied by the analysis of haploid megagametophytes from a single tree. Such an approach to dissecting the genetic determinants of protein amounts and establishing linkage relationships has not been used before.

MATERIAL AND METHODS

Plant material

Seeds from a single tree of an Italian provenance of *Pinus pinaster* Ait., $2n = 2x = 24$, (Fontanin, Latitude North 43°52', Longitude East 4°45', Elevation 500 m) were provided by Dr. Ph. Baradat (INRA Cestas, Bordeaux, France). Fifty-six mega-

gametophytes were analyzed separately by 2-D PAGE.

Protein extraction, electrophoresis and staining

The megagametophytes were individually crushed in 6 µl/mg of 3M urea, 4 per cent FSN-100 (Fluorosurfactan), 2 per cent ampholytes (Pharmalyte pH 3-10) and 1 per cent dithiothreitol at room temperature (Anderson *et al.*, 1985*b*). Thirty-five microlitres of each sample were used. The isoelectrofocusing (IEF) and the sodium dodecyl sulfate (SDS) dimensions were performed as in Bahrman and Thiellement (1987). The gels bound to Gelbond supports (Granier and de Vienne, 1986) were silver stained according to Damerval *et al.* (1987).

Scoring methods

The comparisons were made visually by superimposition of the dried gels upon a light box. The small amount of extract precludes repetitions of the electrophoresis for each of the 56 megagametophytes. Thus minor differences in spot intensity between individual patterns were not taken into account since they could be due to technical reasons. Continuous variation covering a great range of intensity, possibly revealing polygenic control of protein amounts, was never found for any spot. The spots were analysed as genetically variable when their intensity among the 56 megagametophytes was visibly distributed into few discrete classes. Actually only two classes were usually found. For two spots, three or four classes were suspected, but could not be unambiguously separated. They were not included in the analysis. Thus two kinds of variations were found: a given spot can be present or absent, and/or more or less intense according to the megagametophyte (fig. 1). Considering spots as protein characters, we partitioned them as qualitative (PC) and quantitative characters (QC), respectively. The loci involved in their determinism can be mapped. Following the notation used for enzymes, italics are for the locus names with lower case letters designating the alleles and capital letters are for the gene products.

Statistical analyses

The 1:1 Mendelian segregation was tested for all variable spots by chi-square analysis. The hypothesis of the independance of two loci was tested by the joint segregation chi-square analysis (Bailey, 1961; El-Kassaby *et al.*, 1982). Recombi-

nation frequency is:

$$R = r/N$$

where

r = Number of recombinant gametes

N = Number of total gametes.

Standard error is:

$$SE = \sqrt{R(1-R)/N}.$$

Since the parental genotype was unknown we assumed that the smallest class of progeny contained recombinants and the largest, parentals.

Finally using the Kosambi function (Kosambi, 1944) which takes into account the multiple crossing-over and the interferences, the recombination frequencies were converted into map distances in centimorgans:

$$D = 25 \ln [(1+2R)/(1-2R)].$$

The standard deviation of D (Owen, 1950) is:

$$SD = 0.25 SE / (0.25 - R^2).$$

RESULTS AND DISCUSSION

Genetics of protein characters

Among 637 spots reproducibly found in the 56 megagametophytes, 117 were variable (18.3 per cent). Eighty-one spots displayed presence/absence variation (12.7 per cent). Four of these spots showed also intensity variation. Thirty-six spots showed two intensity classes (5.6 per cent). For all the 121 characters (81 qualitative and 40 quantitative), the segregation did not significantly deviate from a 1:1 segregation.

The 81 qualitative characters can be divided into two categories, whose genetic bases may be different. First, 58 spots can be grouped in 29 pairs ($p1$ to $p29$, fig. 1 and fig. 2) according to the following criteria (i) the loci encoding each spot of the pair (e.g., $p1a$ and $p1b$) mapped at the same point (ii) the two spots concerned were not present or absent together (iii) they have similar characteristics of staining and molecular weight and differed mainly in their pI. The simplest hypothesis is that these pairs of polypeptides corresponded to allelic products of structural genes. Another possibility is that one polypeptide is a post-translational modification of the other, this phenomenon being controlled by two alleles at a regulatory locus. The action of such modifier loci had been put forward in the case of enzymes (Finnerty and

Figure 1 Two-dimensional electrophoregrams of two different megagametophytes. Examples of presence/absence and quantitative variations are indicated by P and Q, respectively. For the rectangles see figs 2, 3 and 5.

Johnson, 1979), but it cannot account for the extent of the polymorphism observed (Coyne *et al.*, 1979).

The remaining 23 PCs (P30 to P52) could have various genetic interpretations. They could correspond to quantitative variation where the non-visible polypeptide was actually absent or was below the level of detection by silver staining. The analysis of other developmental stages or organs might help to determine this (Leonardi *et al.*, 1988; de Vienne *et al.*, 1988). Another possibility is that just one allelic product was detected because the other one was concealed by another polypeptide, or because the shift in pI was such that it fell outside the pI gradient used (five to seven). A further possibility is that one of the alleles is indeed "silent", *i.e.*, never encoding any product. The proportion of loci with silent alleles, though difficult to determine, is however, not likely to be very high due to genetic load. Eventually it is probable that many of the 23 qualitative characters will actually be shown to be due to quantitative variation of gene products. In this case, the simplest and most generally applicable hypothesis is that the presence/absence is determined by two alleles of a regulatory locus.

For all the quantitative characters observed, the intensity fell into two discrete classes (fig. 3, which could be explained by a major gene responsible for the determinism of polypeptide amount. This result was rather unexpected, given the numerous elements controlling enzyme activity or

Figure 2 Example of presence/absence variation of spots (P5A to P10A and P5B to P10B) interpreted as allelic variation of structural genes (*p5-p10*). The proportion of individuals in each class of pattern is indicated on the right.

quantity found in natural populations (Kozak, 1985; Elsholtz *et al.*, 1986).

Linkage relationships

One hundred and nineteen loci were arranged in 12 linkage groups and two loci (*p46* and *q27*) remained unlinked (table 1 and fig. 4). The structural gene loci (*p1* to *p29*) were distributed in seven groups and many of them mapped at common points (fig. 4). In such cases, the pairs of polypeptides could be produced by different genes whose genetic distance was such that no recombination event was detected in the sample (≤ 5 cm), $P = 0.05$). This could also be interpreted as translational or post-translational modifications affecting both allelic products of the same structural gene. Such mechanisms were probably operating in the case of polypeptides of similar aspects and close

Figure 3 Example of quantitative variation of spots Q1-Q16. a and c are two individuals displaying the two different patterns observed; b is a coelectrophoresis (1:1) of these two individuals. The proportion of individuals in each class of pattern is indicated on the right.

Table 1 Recombination frequencies (R) and genetic distances (D) between loci. The standard deviations and linkage groups (LG) are also indicated. The loci for which no recombinants were found are not included. Only one member of such clusters of loci was used to calculate recombination frequencies with other loci

Combination	n	r	χ^2	R	SE	D	SD	LG
$p1:p49$	27-28	1-0	52.1***	0.02	0.02	1.8	0.02	I
$p1:p45$	22-25	4-5	26.0***	0.16	0.05	16.7	0.05	I
$p1:p37$	27-26	0-3	44.9***	0.05	0.03	5.4	0.03	I
$p1:q28$	22-23	6-5	20.6***	0.20	0.05	20.8	0.06	I
$p37:p45$	22-22	8-4	18.6***	0.21	0.05	22.9	0.06	I
$p37:p49$	28-26	0-2	48.3***	0.04	0.02	3.6	0.02	I
$p37:q28$	23-21	5-7	18.3***	0.21	0.05	22.9	0.06	I
$p49:q28$	22-22	6-6	18.3***	0.21	0.05	22.9	0.06	I
$p11:q22$	19-18	12-7	6.1*	0.34	0.06	41.3	0.11	II
$q37:q18$	19-18	9-10	5.8*	0.34	0.06	41.3	0.11	III
$p23:p34$	29-25	0-2	47.9***	0.04	0.02	3.6	0.02	IV
$p23:p53$	25-17	6-8	12.9***	0.25	0.06	27.5	0.08	IV
$p20:p34$	20-25	7-4	20.3***	0.20	0.05	20.8	0.06	IV
$p34:p52$	23-17	6-10	10.0**	0.29	0.06	32.5	0.09	IV
$p23:p40$	29-25	0-2	47.9***	0.04	0.02	3.6	0.02	IV
$p23:p20$	26-19	5-6	18.6***	0.20	0.05	20.8	0.06	IV
$p20:p40$	25-20	4-7	20.3***	0.20	0.05	20.8	0.06	IV
$p14:q19$	28-25	1-2	44.4***	0.05	0.03	5.4	0.03	V
$p14:q24$	19-17	7-13	4.9*	0.36	0.06	44.8	0.12	V
$q29:q24$	21-15	9-11	4.2*	0.36	0.06	44.8	0.12	V
$q29:p42$	20-16	10-10	4.4*	0.36	0.06	44.8	0.12	V
$p31:p38$	21-15	10-10	4.2*	0.36	0.06	44.8	0.12	VI
$q20:q21$	29-22	0-5	37.1***	0.09	0.04	9.0	0.04	VII
$q23:q20$	22-27	7-0	32.1***	0.12	0.04	12.8	0.04	VII
$q21:q25$	27-23	0-6	35.1***	0.11	0.04	10.9	0.04	VII
$q20:q25$	22-23	11-0	23.4***	0.20	0.05	20.8	0.06	VII
$q23:q25$	19-25	8-4	17.9***	0.21	0.05	22.9	0.06	VII
$q21:q23$	25-25	4-2	34.7***	0.11	0.04	10.9	0.04	VII
$p4:p22$	27-26	1-2	44.6***	0.05	0.03	5.4	0.03	VIII
$p35:p36$	19-20	10-7	8.7**	0.30	0.06	35.2	0.09	VIII
$p22:p35$	20-20	9-7	10.3**	0.29	0.06	32.5	0.09	VIII
$p22:p36$	29-26	0-1	51.8***	0.02	0.02	1.8	0.02	VIII
$p4:p48$	27-26	1-2	44.6***	0.05	0.03	5.4	0.03	VIII
$p4:p36$	27-25	1-3	41.1***	0.07	0.03	7.2	0.03	VIII
$p39:q26$	28-27	0-1	52.1***	0.02	0.02	1.8	0.02	IX
$p30:q26$	27-27	1-1	48.3***	0.04	0.02	3.6	0.02	IX
$p39:p30$	27-26	1-2	44.6***	0.05	0.03	5.4	0.03	IX
$p44:q30$	26-25	2-3	37.8***	0.10	0.04	9.0	0.04	X

χ^2 = Joint segregation chi-square analysis.

* $P > 0.05$.

** $P > 0.01$.

*** $P > 0.001$.

n = Number of parental gametes.

r = Number of recombinant gametes.

SE = Standard deviation of R .

SD = Standard deviation of D .

position on the gels, as is the case for P1A-P2A-P3A, P1B-PSB-P3B (allelic products of one structural gene and 2 post-translational modifications for each); P5A to P10A, P5B to P10B (fig. 2); P12A-P13A, P12B-P13B; P15A-P16A, P15B-P16B (fig. 5), in group I; P23A-P24A, P23B-P24B; P26A to P29A, P26B to p29B, in group IV and P17A-P18A, P17B-P18B, in group XII. Even if this hypothesis allowed the number of different

cosegregating loci to be reduced, the distribution of the varying structural genes appeared to be not random, most of them being found in group I and IV at a common point.

It is worth noting that if this post-translational modification hypothesis is correct, both allelic products were identically modified. Moreover both allelic products of 27 out of 29 structural genes appeared in the same amount. The only exception

Figure 4 Linkage map of the eighty qualitative character loci (*p*) and 39 quantitative character loci (*q*).

was the P15A, P15B, P16A and P16B products whose amount was controlled by distant genetic element(s) localized in group III (*q*37 to *q*40, fig. 4 and 5). Thus allelic and quantitative variation generally do not affect the same gene products.

Several linkage groups (I, III, IV, VI, VIII, XI and XII) show quantitative and/or qualitative character loci mapping at the same point (fig. 3). There are two possible explanations for this observation (i) these could correspond to different linked regulatory loci (ii) this could result from a pleiotropic effect of one regulatory locus. It has been shown in pea that a single mutation can affect more than 10 per cent of the gene products in some organs (Gottlieb and de Vienne, 1988). Thus, in group I and IV, one qualitative and five quantitative character loci, and one qualitative and 16 quantitative character loci respectively mapped at

the same point as the groups of structural genes mentioned above. It is then possible that the genetic element responsible for one or more of these variations is one of these structural genes.

Gametic selection seemed to occur in group VII, since four of the ten gametic associations involving *q*20 and/or *q*25 and another of the two remaining loci mapped in this group were not found (table 1).

Thus, under the hypotheses of pleiotropy and post-translational modifications, the 121 protein characters could correspond to about 50-60 genes heterozygous in the parental tree, grouped in ten linkage blocks and four unlinked loci. Our work thus clearly reveals that 2D-PAGE technique can be successfully applied to analyse linkage relationships of protein characters, both quantitative as well as qualitative. The results concerning genome

Figure 5 Presence/absence and quantitative variations of spots P15A, P15B, P16A and P16B. P15A and B (P16A and B) were interpreted as allelic products of structural gene locus *p15* (*p16*) localized in group I. P15A and P16A (P15B and P16B) were further interpreted as post-translationally modified products of the same gene locus with two alleles (see text). The loci controlling intensity variation of these products (*q37-q40*) were mapped in linkage group III. The proportion of individuals in each class of pattern is indicated on the right.

markers (restriction fragment length polymorphisms, isozymes) or genes controlling phenotypic variation.

Acknowledgements We thank our colleagues and especially D. de Vienne for critical reading and discussion of the manuscript. This research was supported in part by Grant AIP 4570 from the I.N.R.A.

REFERENCES

- ANDERSON, N. L., NANCE, S. L., TOLLAKSEN, S. L., GIERE, F. A. AND ANDERSON, N. G. 1985a. Quantitative reproducibility of measurements from Coomassie Blue-stained two-dimensional gels: analysis of mouse liver protein patterns and a comparison of BALB/c and C57 strains. *Electrophoresis*, **6**, 592-599.
- ANDERSON, N. G., TOLLAKSEN, S. L., PASCOE, F. H. AND ANDERSON, N. L. 1985b. Two-dimensional electrophoretic analysis of wheat seed proteins. *Crop Sci.*, **25**, 667-674.
- BAHRMAN, N. AND THIELLEMENT, H. 1987. Parental genome expression in synthetic wheats (*Triticum turgidum* sp. $\times$ *T. tauschii* sp.) revealed by 2D electrophoresis of seedling proteins. *Theor. Appl. Genet.*, **74**, 218-223.
- BAILEY, N. T. J. 1961. *Introduction to the Mathematical Theory of Genetic Linkage*. Oxford University Press, London.
- BEEVERS, L. 1982. Post-translational modifications. In Boulter, C. and Parthier, B. (eds.) *Encyclopedia of Plant Physiology*, New series, vol. 14A. Springer-Verlag, Berlin, pp 136-168.
- CONKLE, M. T. 1981. Isozyme variation and linkage in six conifer species. In: *Proc. Symp. Isozymes of North American Forest Trees and Forest Insects*. USDA For. Serv. Gen. Tech. Rep., PSW-48, pp 11-17.
- COYNE, J. A., EANES, W. F. AND LEWONTIN, R. C. 1979. The genetics of electrophoretic variation. *Genetics*, **92**, 353-356.
- DAMERVAL, C., LE GUILLOUX, M., BLAISONNEAU, J. AND DE VIENNE, D. 1987. A simplification of Heukeshoven and Dernick's silver staining of proteins. *Electrophoresis*, **8**, 158-159.
- EL-KASSABY, Y. A., SZIKLAI, O. AND YEH, F. C. 1982. Linkage relationships among 19 polymorphic allozyme loci in coastal Douglas-fir (*Pseudotsuga menziesii* var. *menziesii*). *Can. J. Genet. Cytol.*, **24**, 101-108.
- ELSOLTZ, H. P., MANGALAM, H. J., POTTER, E., ALBERT, V. R., SUPOWIT, S., EVANS, R. M. AND ROSENFELD, M. G. 1986. Two different *cis*-active elements transfer the transcriptional effects of both EGF and phorbol esters. *Science*, **234**, 1552-1557.
- FINNERTY, V. AND JOHNSON, G. 1979. Post-translational modification as a potential explanation of high levels of enzyme polymorphism: xanthine dehydrogenase and aldehyde oxidase in *Drosophila melanogaster*. *Genetics*, **91**, 695-722.
- GIBSON, J. B., WILKS, A. V., CAO, A. AND FREETH, A. L. 1986. Dominance for sn-glycerol-3-phosphate dehydrogenase activity in *Drosophila melanogaster*: evidence for differential allelic expression mediated via a trans-acting effect. *Heredity*, **56**, 227-235.
- GOTTLIEB, L. D. AND DE VIENNE, D. 1988. Assessment of pleiotropic effects of a gene substitution in pea by two-dimensional polyacrylamide gel electrophoresis. *Genetics*, **119**, 705-710.

organisation and regulation obtained here from a single progeny population need to be confirmed by analysis of other progenies. The linkage map for *Pinus pinaster* Ait. can be further extended by the localization of other loci such as molecular

- GRANIER, F. AND DE VIENNE, D. 1986. Silver staining of proteins: standardized procedure for two-dimensional gels bound to polyester sheets. *Anal. Biochem.*, 155, 45-50.
- HAMMER, M. F. AND WILSON, A. C. 1987. Regulatory and structural genes for lysozymes of mice. *Genetics*, 115, 521-533.
- JUNGBLUT, P. AND KLOSE, J. 1985. Genetic variability of proteins from mitochondria and mitochondrial fractions of mouse organs. *Biochem. Genet.*, 23, 227-245.
- KOSAMBI, D. D. 1944. The estimation of map distances from recombination values. *Ann. Eugen.*, 12, 172-175.
- KOZAK, L. P. 1985. Interacting genes control glycerol-3-phosphate dehydrogenase expression in developing cerebellum of the mouse. *Genetics*, 110, 123-143.
- LAURIE-AHLBERG, C. C., WILTON, A. N., CURTSINGER, J. W. AND EMIGH, T. H. 1982. Naturally occurring enzyme activity variation in *Drosophila melanogaster*. I. Sources of variation for 23 enzymes. *Genetics*, 102, 191-206.
- LEONARDI, A., DAMERVAL, C. AND DE VIENNE, D. 1988. Organ-specific variability and inheritance of maize proteins revealed by two-dimensional electrophoresis. *Genet. Res. Camb.*, 52, 97-103.
- NIEBLING, C. R., JOHNSON, K. AND GERHOLD, H. D. 1987. Electrophoretic analysis of genetic linkage in Scots pine (*Pinus sylvestris* L.). *Biochem. genet.*, 25, 803-814.
- O'FARRELL, P. H. 1975. High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.*, 250, 4007-4021.
- OWEN, A. R. G. 1950. The theory of genetical recombination. *Adv. Genet.*, 3, 117-157.
- SASSONE-CORSI, P. AND BORRELLI, E. 1986. Transcriptional regulation by *trans*-acting factors. *Trends Genet.*, 2, 215-219.
- STRAUSS, S. H. AND CONKLE, M. T. 1986. Segregation, linkage, and diversity of allozymes in knobcone pine. *Theor. Appl. Genet.*, 72, 483-493.
- DE VIENNE, D., LEONARDI, A. AND DAMERVAL, C. 1988. Genetic aspects of variation of protein amounts in maize and pea. *Electrophoresis*, 9, 742-750.
- YAMAZAKI, T. AND MATSUO, Y. 1984. Genetic analysis of natural populations of *Drosophila melanogaster* in Japan. III. Genetic variability of inducing factors of amylase and fitness. *Genetics*, 108, 223-235.