ELSEVIER

Journal of Chromatography A, 715 (1995) 95–104

JOURNAL OF CHROMATOGRAPHY A

Chromatographic and chemometric investigation of the chemical defence mechanism of poplar tree genotypes against a bark fungine parasite

C. Baiocchi[a,*], E. Marengo[a], M.A. Roggero[a], D. Giacosa[a], A. Giorcelli[b], S. Toccori[b]

[a]*Dipartimento di Chimica Analitica, Università di Torino, 10125 Turin, Italy*
[b]*Istituto di Sperimentazione per la Pioppicoltura, SAF Str. Frassineto 35, Casale Monferrato (AL), Italy*

First received 14 October 1994; revised manuscript received 9 May 1995; accepted 16 May 1995

Abstract

Phenolic compounds in bark extracts of seven different clones of poplar trees were analysed by RP-HPLC with diode-array UV detection. The chromatographic results obtained were treated by means of chemometric methods to highlight any pattern in the data allowing one to discriminate poplars in terms of resistance to the fungus *Discosporium populeum*. The phenolic compounds turned out to be more related to the genetic origin of the clones than to their resistance.

1. Introduction

There is an increasing awareness that the phenolic constituents of plants are a key to understanding many plant–environment and, particularly, plant–parasite relationships. Poplar trees, like all plants, respond to various types of injuries with rather generalized metabolic responses. These frequently include the synthesis, oxidation and polymerization of phenolic compounds [1]. However, these substances probably represent only one of the many lines of defence that may be enforced by the plant. For this reason, it may be difficult to separate the defensive role of phenolics from that of alkaloids, terpenoids, coumarins or flavonoids. Anyway, considering phenolics as indicators of disease resistance may be justified on the grounds that the two principal routes for the formation of aromatic compounds from aliphatic precursors, the shikimate and the polyketide pathways, each contribute to the formation of phenolic compounds. The shikimate pathway gives the phenylpropanoid nucleus while the polyketide pathway yields the A-ring of flavonoids [2]. In practice, phenolics may be considered as a crossroads of different biosynthetic pathways, many of which are activated by a range of stress factors such as wounding, infection and low-temperature stress. The central role of phenolic compounds in the resistance of poplar trees to infection by the fungus *Discosporium populeum*, the main topic of this paper, has been suggested by many workers [3–5]. Even in the case where the

* Corresponding author.

0021-9673/95/$09.50
SSDI 0021-9673(95)00585-4

studies involved non-phenolic bark components such as the phenylalanine ammonia-lyase enzyme (PAL) [6], cytokinine [7] and β-glucosidase [8], these were related to phenolic substances and confirmed their importance.

In two previous papers [9,10], we tried to relate the qualitative and quantitative distributions of phenolics contained in bark extracts of poplar trees to their resistance to the fungus. Two clones of different genetic origin, one known, as a result of a biological test, to be resistant and the other susceptible to the fungal attack, were studied. The experimental data obtained were treated chemometrically so as to obtain the maximum information. As a final result it was possible to link the two clones to different classes of phenolics.

This paper constitutes an extension of the study to seven poplar clones belonging either to different species or to different genotypes of the same hybrid species. RP-HPLC with diode-array UV detection was used to analyse the bark extracts. Principal component analysis was applied to the chromatographic data in order to explore their structure. Subsequently KNN, a multivariate classification method, was used to investigate the kind and amount of information discriminating among the clones obtainable from the variation of the phenolic content of bark extracts.

2. Experimental

2.1. Instrumentation

The bark extracts were analysed using a Varian Star 9010 ternary gradient HPLC pump equipped with a Varian Star 9065 diode-array spectrophotometric detector. System control and data acquisition and processing were performed utilising the multi-task program Varian Star 9000 on an Epson 386 (33 MHz) PC.

The chromatographic column was Lichrosphere RP-18 (250 × 4.6 mm I.D.), 5 μm particle size (Merck, Darmstadt, Germany). The injection volume was 10 μl and the flow-rate was 1.0 ml/min.

2.2. Reagents

HPLC-grade methanol and acetonitrile and a 0.57% solution of acetic acid in water purified in a Milli-Q system (Millipore) were used as mobile phase constituents.

2.3. Chromatographic conditions

Â ternary gradient programme was used involving (A) acetonitrile, (B) 0.57% acetic acid and (C) methanol. The starting conditions were 6% A–88% B–6%C and after 40 min the mobile phase composition was 6% A–48% B–46% C.

2.4. Sampling

The bark of seven clones known to differ in resistance to *D. populeum* was sampled. The sampling programme was scheduled in December, January, February and March. Several samples of each clone, for each period, were collected to give a total of 216 samples. The detailed structure of the data is given in Table 1.

Table 1
Description of data set

Total number of objects	216
Number of different clones	7 (A, B, C, D, E, F and G)
Number of samples collected for each clone	A = 29, B = 32, C = 30, D = 32, E = 31, F = 30, G = 32
Number of sample periods	4 (1 = December, 2 = January, 3 = February, 4 = March)
Number of samples collected for each period	1 = 55, 2 = 56, 3 = 53, 4 = 52
Number of resistance groups	4 (A, B, C, resistant, class 1; D, moderately resistant, class 2; E, F, moderately susceptible, class 3; G, very susceptible, class 4)

2.5. Sample preparation

About 30 mg of previously lyophilized bark were added to 2.0 ml of 1.0 *M* sodium hydroxide solution in a filter tube. The air was removed from the tube by flushing with nitrogen and the stopper secured. The suspension was shaken at 20°C for 20 h and subsequently filtered. The residue was washed with water (total volume of filtrate ca. 2.0 ml). The filtrate was acidified to pH 2.5 with 6.0 *M* hydrochloric acid and diluted with water to a final volume of 5.0 ml [9,10].

2.6. Standards

The standard substances 1,2,3-trihydroxybenzene, 3,4-dihydroxybenzoic acid, 1,2-dihydroxybenzene, 4-hydroxybenzoic acid, 4-hydroxy-3-methoxybenzoic acid, 4-hydroxybenzaldehyde, 3,5-dimethoxy-4-hydroxybenzoic acid, 4-hydroxy-3-methoxybenzaldehyde, 4-hydroxy-3,5-dimethoxybenzaldehyde, 4-hydroxycinnamic acid, 4-hydroxy-3-methoxycinnamic acid, 2-hydroxybenzoic acid, benzoic acid and 4-methoxybenzoic acid were purchased from Aldrich Chimica (Milan, Italy).

2.7. Chemometrics

Principal component analysis (PCA) is a well known technique which provides a significant insight into the structure of any numerical data matrix [11,12]. It generates a set of new orthogonal variables, the principal components (PCs), a linear combination of the original ones, so that the maximum possible amount of variance contained in the starting data matrix is concentrated in as few PCs as possible. These new variables can be used in place of the original ones for successive treatment.

The coefficients of the original variables defining each PC are called "loadings" and the projections of the experimental points on the new variables are called "scores". Since PCA effectively concentrates the variance of the data matrix in a smaller number of new variables, it is suitable to reduce the dimensionality of large data matrices by eliminating the non-significant PCs and facilitating successive treatments on the reduced data. PCA was computed through the diagonalization of the variance–covariance matrix by means of the Jacobi algorithm. The data were autoscaled before PC computation in order to assign the same numerical weight to each variable.

3. Results and discussion

Various bark samples of seven different poplar clones sampled in four different periods (December, January, February and March) were analysed. A detailed description of the samples (216 in total) is given in Table 1. The seven poplar clones were assigned to four different classes of resistance, determined by a biological test: LUX, DVINA (*P. deltoides*) and GHOY (*P.* × *euroamericana*) class 1, very resistant; B.L. COSTANZO (*P.* × *euroamericana*) class 2, resistant; BOCCALARI and FARSI (*P.* × *euroamericana*) class 3, moderately susceptible; and J. POURTET (*P. nigra*) class 4, very susceptible. In the following discussion these clones will be identified by the letters A, B, C, D, E, F and G, respectively.

Chromatographic results for the seven clones sampled in winter are reported in Fig. 1. The peaks were identified by comparing the retention times and UV spectra monitored between 220 and 330 nm with those for a standard mixture of sixteen phenolic substances (listed in Table 2). There are differences in the chromatographic profiles, but considering they account, in a complex way, for differences among clones, resistance and sample collection periods, it is evident that there is the need for a data treatment allowing one to separate the different sources of information and to organize them in a convincing pattern. For this purpose suitable statistical treatments were applied to a set of experimental data made up of 16 variables (peak areas) and 216 objects (all collected samples), giving a total number of 3456 data.

First a PCA was performed for a synthetic and informative visualization of the data. Table 3 shows the values of the variance explained by the

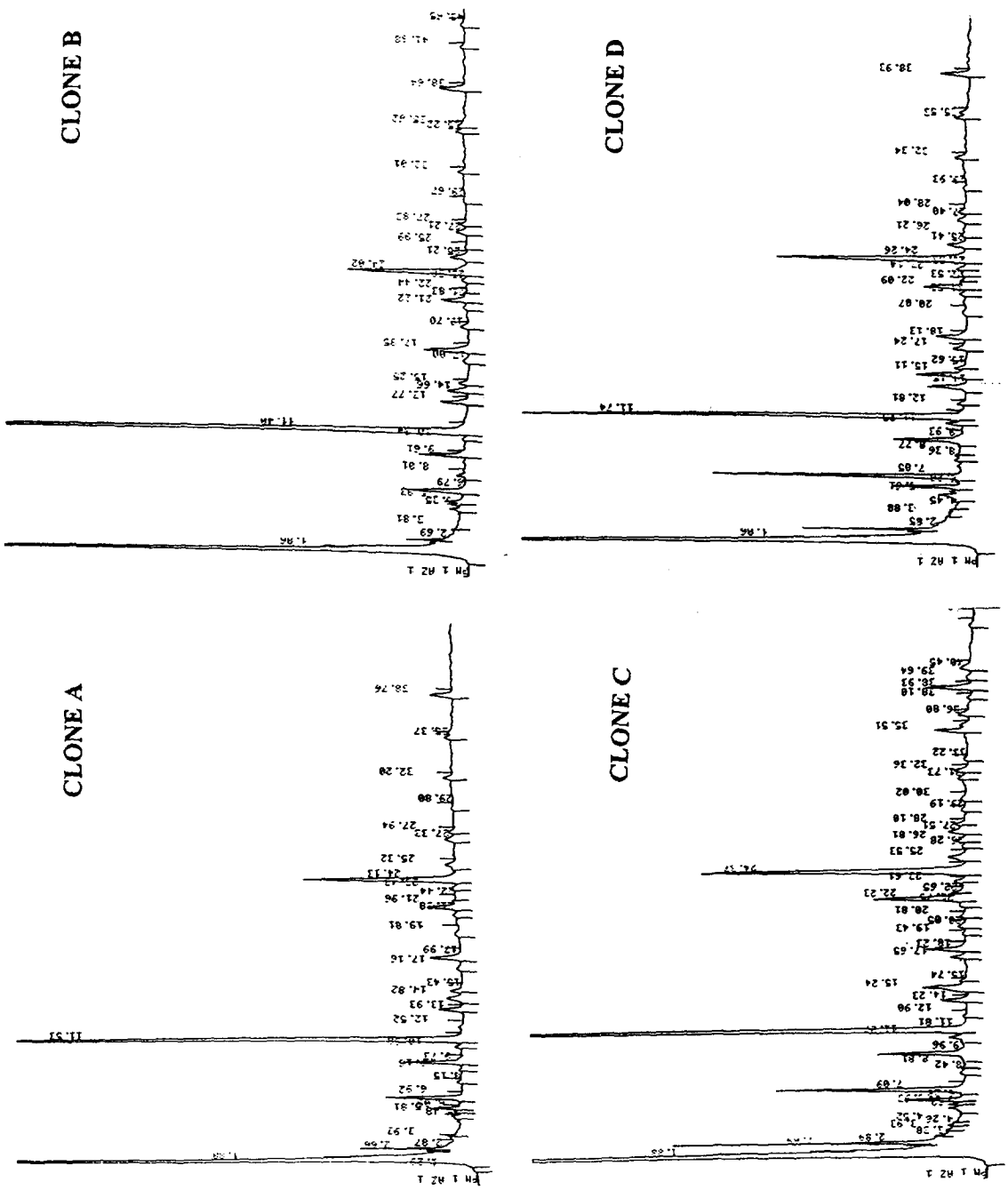

Fig. 1.

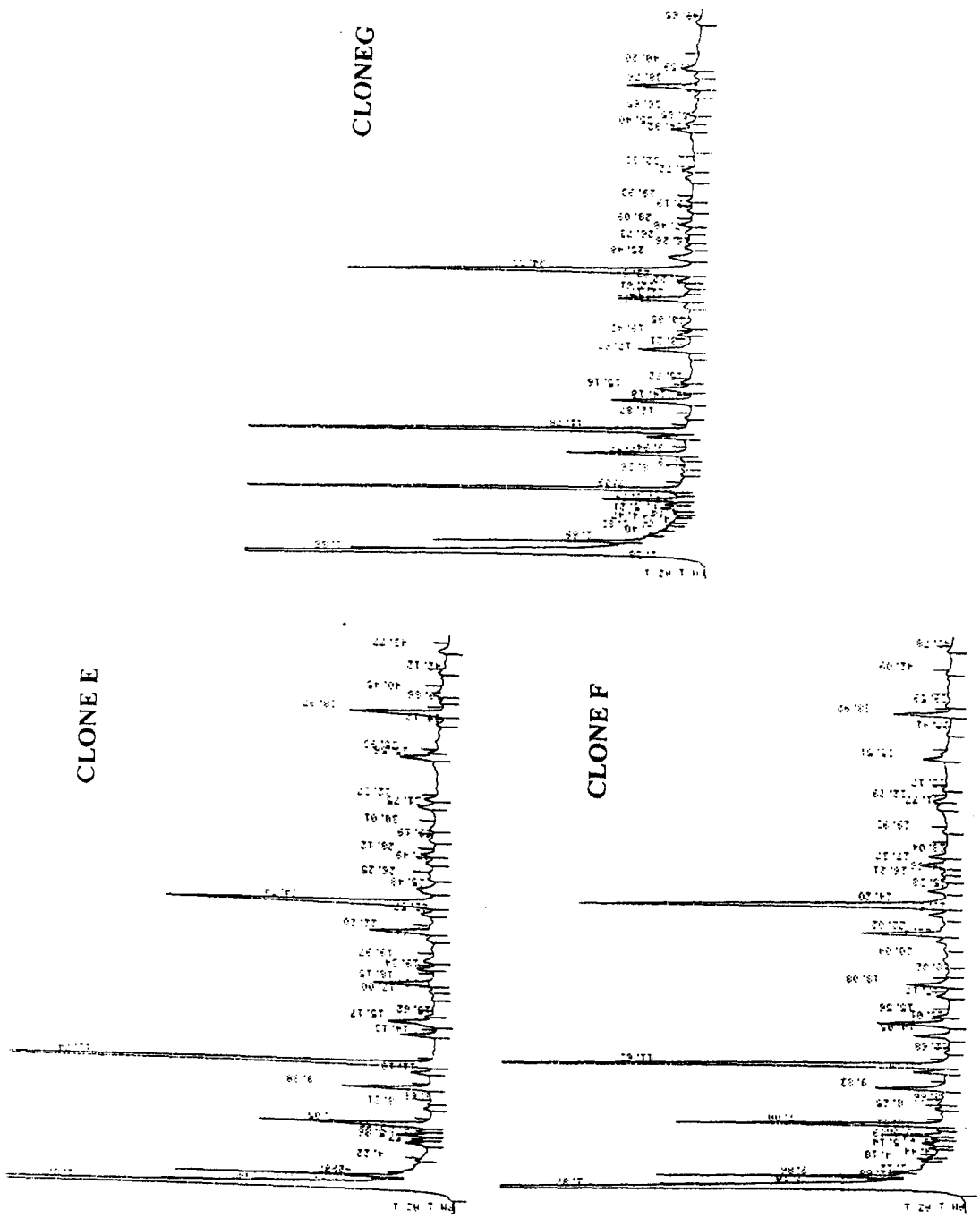

Fig. 1. Chromatographic results for the bark extracts of the seven different clones sampled in winter.

Table 2
Variables taken into consideration in the chromatographic and chemometric analysis

Variable	Compound
1	1,2,3-Trihydroxybenzene
2	3,4-Dihydroxybenzoic acid
3	1,2-Dihydroxybenzene
4	4-Hydroxybenzoic acid
5	4-Hydroxy-3-methoxybenzoic acid
6	4-Hydroxybenzaldehyde
7	3,5-Dimethoxy-4-hydroxybenzoic acid
8	4-Hydroxy-3-methoxybenzaldehyde
9	4-Hydroxy-3,5-dimethoxybenzaldehyde
10	4-Hydroxycinnamic acid
11	Degradation product of 10
12	4-Hydroxy-3-methoxycinnamic acid
13	Degradation product of 12
14	2-Hydroxybenzoic acid
15	Benzoic acid
16	4-Methoxybenzoic acid

more significant principal components and Table 4 lists the loadings of the sixteen variables on the first three PCs. Fig. 2 shows a plot of the scores of PC1 vs. PC2 for all the objects. In spite of the complexity due to the large number of classes present, it is evident that the clusters of clones A, B and G are separated, whereas all the others are more or less in a cluster lying between the previous ones. In fact the distinction of the clones belonging to pure species. (*P. deltoides*, clones A and B, *P. Nigra*, clone G) is very sharp. The discrimination among the hybrids (*P.* × *euroamaricana*, clones C, D, E and F) which are sensitive to a more or less genetic proximity to their parents is instead more difficult. PC1 mainly discriminates clones A and B with respect to the others. As the list of loadings reported in Table 4 shows, the variables contributing substantially to the PC1 definition are variables 2 (3,4-dihydroxybenzoic acid) 4 (4-hydroxybenzoic acid) and 12 (4-hydroxy-3-methoxycinnamic acid). Variables 5 (4-hydroxy-3-methoxybenzoic acid) and 8 (4-hydroxy-3-methoxybenzaldehyde), the main constituents of PC2, seem instead effective in the separation of clones G and D from the group of clones B, E and F. Given that, from the resistance point of view, the expected classification ought to be A, B, C–D–E, F–G, there is evidence of a clear decrease in the contribution of this kind of information to the discrimination of the objects in favour of one related to the genetic diversity of the clones. PC3 does not provide any further contribution in this direction, but seems rather to contain some information about the possible influence of seasonal factors.

Table 3
Variance explained by the PCs with eigenvalue greater than 1.0

PC	Variance explained	Cumulative variance
1	28.7	28.7
2	15.2	43.9
3	13.9	57.8
4	8.6	66.4
5	6.7	73.1

Fig. 3 shows the plot of the scores of PC2 and PC3, defining a trend to discriminate winter samples (1 and 2) from early spring samples (3 and 4). The variables which seem to exert the most discriminating effect are 1 (1,2,3-trihydroxybenzene), 3 (1,2-dihydroxybenzene), 6 (4-hydroxybenzaldehyde), 8 (4-hydroxy-3-methoxybenzaldehyde) and 11 (degradation product of 4-hydroxycinnamic acid). These variables are different from those discriminating the clones, so there is no apparent mixing between the information related to clone classification and to the sample collection period classification.

From the variance–covariance matrix (not reported) used to perform PCA calculations, it is also possible to visualize the complete pattern of correlations among the variables important for this study. A positive correlation among the variables may be interpreted in terms of a strict link characterizing the presence of the substances that they represent in the poplar bark. For instance, they may be final products of correlated biosynthetic pathways, or even of an individual pathway, and may contribute to define the same sample features. This is the case with the variables 2 (3,4-dihydroxybenzoic acid) and 12 (4-hydroxy-3-methoxycinnamic acid), important constituents of PC1. A negative correlation

Table 4
Loadings of the original variables on the first three PCs

Variable	PC1	PC2	PC3
1. 1,2,3-Trihydroxybenzene	−0.18	−0.11	−0.11
2. 3,4-Dihydroxybenzoic acid	**−0.38**	**0.34**	−0.21
3. 1,2-Dihydroxybenzene	−0.15	−0.31	**−0.38**
4. 4-Hydroxybenzoic acid	**0.50**	−0.07	0.06
5. 4-Hydroxy-3-methoxybenzoic acid	−0.24	**0.40**	**−0.32**
6. 4-Hydroxybenzaldehyde	**−0.31**	−0.22	−0.10
7. 3,5-Dimethoxy-4-hydroxybenzoic acid	0.08	0.22	0.11
8. 4-Hydroxy-3-methoxybenzaldehyde	0.12	**−0.44**	**−0.31**
9. 4-Hydroxy-3,5-dimethoxybenzaldehyde	0.05	**−0.34**	−0.17
10. 4-Hydroxycinnamic acid	**−0.32**	−0.03	**0.32**
11. Degradation product of 10	−0.03	0.05	0.29
12. 4-Hydroxy-3-methoxycinnamic acid	−0.43	−0.12	0.05
13. Degradation product of 12	−0.05	−0.09	−0.11
14. 2-Hydroxybenzoic acid	−0.19	−0.13	**0.34**
15. Benzoic acid	−0.13	−0.26	**0.48**
16. 4-Methoxybenzoic acid	−0.15	−0.30	0.03

may instead be interpreted in terms of a reciprocal transformation of the substances involved, as if one or more of them were key compounds in the biosynthesis of the others (this may be the case with variables 5 and 8, relevant in the PC2 definition). A variable negatively correlated with the majority of the others is variable 4 (4-hydroxybenzoic acid). In previous papers [9,10], the role of this substance in defining the data structure was found to be extremely important.

At this stage, a classification study was performed in order to check thoroughly the possibility of discriminating the clones as species or even as genotypes.

A *K*th nearest neighbours (KNN) classification study was performed. This method determines for each object the *k*th nearest ones, and assigns it to the most represented class among the *K* nearest samples. In the case of an equal population of two or more classes, the sample is assigned to the nearest among them. It would be

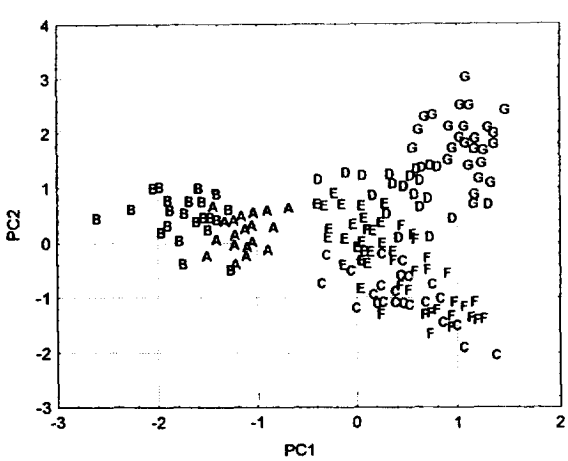

Fig. 2. Scatter plot of samples vs. PC1 and PC2.

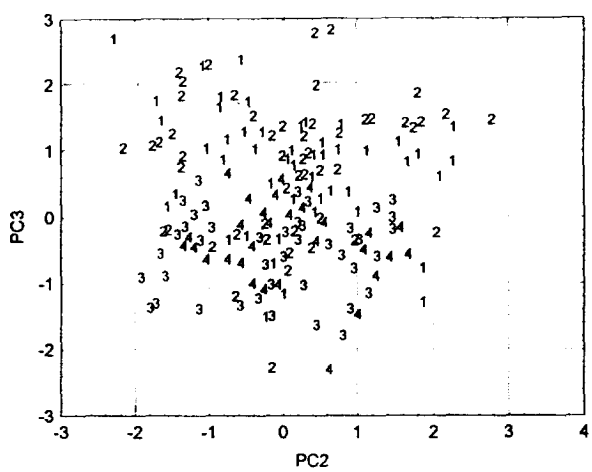

Fig. 3. Scatter plot of samples vs. PC2 and PC3.

desirable to eliminate from the data the information related to the sample collection period and to non-discriminant variables, but any procedure we can use for this purpose may have some side-effect on the KNN method, which is based on the calculation of distances. However, since the presence of non-discriminant information may introduce confusion in the results, we introduced two correction mechanisms. A stepwise algorithm of selection of the discriminant variables, after autoscaling, searching iteratively for the best set of variables by adding or deleting one variable at a time, repeated until no further improvement of the percentage non-error rate (%NER), defined as the percentage of the correctly classified objects, was achieved. A further variable scaling operation, centred on sample collection time, was performed in order to "clean up" the data from this kind of information. Hence the following four different data sets were defined: set 1 (all original variables, autoscaled), set 2 (all original variables, centred on sample collection period), set 3 (selected variables 2, 4, 5, 7, 8, 10, 12 and 15, autoscaled) and set 4 (selected variables 2, 5, 6, 10, 12, 13 and 15, autoscaled and centred on sample collection period). Calculations were then performed and the results were compared.

The results obtained using $K = 5$, summarized in Table 5, show that the best set of variables corresponds to set 4. In particular, the results obtained using all the variables (untreated or class centred) are markedly worse than those obtained with the stepwise refinement algorithm. This confirms that the variables which do not contain information on the clone discrimination introduce a lot of confusion (noise) in the pattern of distances, which makes the KNN classification much less effective.

Table 5
KNN results for: set 1, all autoscaled variables; set 2, all variables centred on the sample collection time after autoscaling; set 3, variables 2, 4, 5, 7, 8, 10, 12 and 15, autoscaled; set 4, variables 2, 5, 6, 10, 12, 13 and 15, centred on the sample collection time after autoscaling

Clone	%NER			
	Set 1	Set 2	Set 3	Set 4
A	82.7	86.2	96.5	89.7
B	84.4	87.5	84.4	93.7
C	73.3	73.3	90.0	93.3
D	65.6	68.7	93.7	90.6
E	80.6	90.3	87.1	96.7
F	86.7	76.7	93.3	93.3
G	100.0	100.0	100.0	100.0
Overall	81.9	83.3	92.1	94.0

The differences between the class-centred and the untreated variables are smaller. It is interesting that the variables which emerged from the PCA treatment are present among the selected variables of sets 3 and 4, with the noteworthy exception of variable 4 in the case of the data set centred on the sample collection period. This variable was eliminated during the last step of the stepwise procedure because its elimination does not change the %NER. This can be easily justified if we remember the mentioned high and negative correlation of this variable with all other variables and particularly with variables 2 (a possible hydroxylation product) and 10 (a possible precursor). Variable 4 may be considered no longer necessary since its information is already present.

The remaining variables, i.e., 6 and 13, or 7 and 8, increase the %NER in both cases by about 2% and are added only during the last refinement steps. Their statistical significance is dubious.

The %NER values for the clones show that the best results are achieved by performing the calculation with the reduced dataset, after eliminating the information about the sample collection time. The consideration of the confusion matrix for this calculation (Table 6) better describes the pattern of the classification results. This is a matrix containing categorical classification results. The rows of the matrix correspond to true classes and the columns to predicted classes. The number of correctly classified objects in each class appears on the main diagonal, while the number of misclassified objects are the off-diagonal elements. From Table 6, it can be concluded that the classification can be effectively performed and that the highest number of errors is due to exchanges between clones 1 and

Table 6
Confusion matrix for the KNN classification of clones, using set 3

Class	A	B	C	D	E	F	G
A	**26**	3	0	0	0	0	0
B	2	**30**	0	0	0	0	0
C	0	0	**28**	0	0	2	0
D	0	0	0	**29**	1	0	2
E	0	0	1	0	**30**	0	0
F	1	0	1	0	0	**28**	0
G	0	0	0	0	0	0	**32**

2 (belonging to the same species and not mixed with the others). This means that the selected variables are effectively responsible for the discrimination between the clones.

An analogous study by means of the KNN method was performed on the classification of the samples on the basis of their different collection times. Only the best results are reported (Table 7) obtained with autoscaled variables, selected using the stepwise procedure.

The overall and partial %NER are not as large as in the previous case. Anyway, it should be noted that most of the wrong assignments are due to exchanges between sampling periods 1 and 2 on one side and 3 and 4 on the other. This means that the four sampling periods can be grouped into two main behaviours: winter and early spring, as already suggested by the PCA results. Also, the fact that the selected variables are partially different from those found by PCA is due to the fact that KNN tries a priori to discriminate period 1 from period 2 samples and period 3 from period 4 samples, while during the principal component analysis the similarity between the two couples has been accepted a posteriori after the analysis of the results. If the four sample collection periods are grouped together into two couples, winter (W) and spring (S) samples, the best results obtained by the stepwise KNN will consist of an overall %NER of 93.5. The corresponding confusion matrix is reported in Table 8. The classification is highly satisfactory. Several variables among those selected contribute very poorly to the improvement of the classification; their statistical significance is small. The most important variables seem to be 3 (1,2-dihydroxybenzene), 6 (4-hydroxybenzaldehyde) and 9 (4-hydroxy-3,5-dimethoxybenzaldehyde), which alone leads to an overall %NER of about 90.

Reporting all the information obtained to a phytochemical basis, we can draw important and maybe definitive considerations about the possibility of using phenolics as indicators of poplar tree resistance towards *D. populeum* infection. The statistical methods adopted constituted different approaches to the analysis of data, so all possible, if any, underlying structures were highlighted. It is then clear that the analysis of phenolics provides information not about the plant resistance, but about the genetic origin of clones. In our previous work [9,10] we attributed resistance discrimination ability to phenolic compounds only because there occurred a fortuitous coincidence of the two features. In fact, these substances seem to account also for seasonal effects and the variables responsible are different from those discriminating among the clones. This confirms that phenolics are involved in many ways in the biochemical production of poplars,

Table 7
Confusion matrix for the KNN classification on the basis of the sample collection periods, variables selected 2, 3, 6, 9 and 12

Class	1	2	3	4	%NER[a]
1	**49**	9	3	0	78.2
2	9	**37**	4	6	66.1
3	2	1	**35**	15	66.0
4	1	0	10	**41**	78.8

[a] Overall %NER: 72.2.

Table 8
Confusion matrix for the KNN classification on the basis of the sample collection periods, variables selected 2, 3, 4, 6, 9, 12, 13 and 15

Class	W	S	%NER[a]
1	99	12	89.2
2	2	103	98.1

[a] Overall %NER: 93.5.

maybe also as pre-infection factors, but that they can be better defined as multi-purpose compounds and are not easily related to the defence mechanism against *D. populeum*.

Further developments in this direction must take into account new data and new substances in the attempt to characterize the mechanism of defence. Work of this kind is in progress.

References

[1] M.J.C. Rhodes and L.S.C. Wooltorton, in G. Kahl (Editor), The Biochemistry of Wounded Plant Tissues, Walter de Gruyter, Berlin, 1978, pp. 243–246.
[2] K. Hahlbrock and H. Grisebach, Annu. Rev. Plant. Physiol., 30 (1979) 105.
[3] S. Pucacka, Arbor. Kornickie, 20 (1975) 227.
[4] S. Pucacka, Arbor. Kornickie, 25 (1980) 257.
[5] Y. Chuanhe, Y. Wang and Z. Zhongming, presented at International Poplar Commission, XVIIIth Session, Beijing, September 5–8, 1988.
[6] S. Pucacka, Arbor. Kornickie, 24 (1979) 209.
[7] M. Rudawska, Annual Reports, Institute of Dendrology, Kornik, Poland, 1980, p. 195.
[8] B. Kieliszewska-Rokicka, Annual Reports, Institute of Dendrology, Kornik, Poland, 1980, p. 203.
[9] C. Baiocchi, E. Marengo, G. Saini, M.A. Roggero and D. Giacosa, J. Chromatogr., 644 (1993) 259.
[10] C. Baiocchi, E. Marengo, M.A. Roggero, D. Giacosa, L. Vietto and S. Toccori, Chromatographia, 39 (1994) 7.
[11] E.R. Malinowski and D.G. Howery, Factor Analysis in Chemistry, Wiley, New York, 1978.
[12] S. Wold, Pattern Recognition, 8 (1976) 127.