Theor Appl Genet (2000) 101:1145–1154

K. Semagn · A. Bjornstad · B. Stedje · E. Bekele

Comparison of multivariate methods for the analysis of genetic resources and adaptation in *Phytolacca dodecandra* using RAPD

Received: 15 December 1999 / Accepted: 12 February 2000

Abstract The extent of genetic differentiation among 17 Ethiopian populations (249 individuals) of *Phytolacca dodecandra* (Endod) sampled along altitudinal gradients that varied from 1600 to 3000 m was investigated using random amplified polymorphic DNA (RAPD). The populations were classified into three altitude groups: lowland (1600–2100 m), central-highland (2101–2500 m) and highland (2500–3000 m). Seventy polymorphic loci scored from 12 RAPD primers, singly or in combination with ecogeographical variables (altitude, longitude, latitude, temperature and rainfall), were used for principal component, discriminant, correlation, and stepwise multiple regression analyses. Principal component analysis (PCA) clearly differentiated lowland and the central-highland populations from those of the highlands independent of their geographical regions. Canonical discriminant analysis separated the lowland plants from those of the highlands with the central-highland plants being intermediate. Classificatory discriminant analysis corrected classification of 92.8% of the 249 plants into their respective three altitude groups. Multiple regression analysis identified a strong association between some RAPDs and altitude, temperature and rainfall, while the variation in most RAPDs was explained by combinations of the different ecogeographical variables. It is hypothesised that the different altitude groups may be (1) chemical and/or physiological ecotypes produced as a result of complex interactions of altitude with climatic and/or edaphic factors, or (2) different in ploidy levels. The significant correlations obtained between population means from some RAPDs and altitude and temperature as well as the strong association of some RAPDs with the ecogeographical variables in the multiple regression analysis suggest that part of the RAPD polymorphism could be adaptive, and responsive to environmental selection.

Keywords Altitude differentiation · Environmental selection · *Phytolacca dodecandra* · RAPD

Communicated by P.M.A. Tigerstedt

K. Semagn (✉)
Addis Ababa University, Institute of Pathobiology,
P.O. Box 1176, Addis Ababa
Ethiopia; aau-ipb@telecom.net.et
Fax: +2511-755296

A. Bjornstad
Agricultural University of Norway,
Department of Horticulture and Crop Sciences, P.O. Box 5022,
N-1432, Ås, Norway

B. Stedje
University of Oslo, Botanical Garden and Museum,
Trondheimsveien 23B, N-0562,Oslo, Norway

E. Bekele
Addis Ababa University, Department of Biology, P.O. Box 1176,
Addis Ababa, Ethiopia

Introduction

The berries of *Phytolacca dodecandra* (locally called Endod in Amharic) have long been known for their traditional use for washing clothes in Ethiopia and elsewhere in Africa. Preliminary studies have also shown that the berries have spermicidal, larvicidal, hirudinicidal, fungicidal and microbial properties (Lemma *et al*. 1979). Extensive studies on the molluscicidal properties of Endod berries to control schistosomiasis and other snail-borne diseases are promising (Lemma et al. 1974; WHO 1983; Lambert et al. 1991). The local production of Endod berries as a means to control the snail intermediate hosts of schistosomiasis is considered to be cheap, readily available and environmentally acceptable (Pezzuto et al. 1984; Lambert et al. 1991). Endod is indigenous to many parts of Africa (Dalziel 1936). Its natural distribution in Ethiopia ranges between 1600 and 3000 m elevation (Wolde-Yohannes et al. 1988). Natural growth of local populations of Endod in Ethiopia, Zambia and Zimbabwe has been reported on acidic and humus-rich soils, high rainfall, low temperature and high altitude (Wolde-Yohannes 1983; Chimbelu and Shehata 1986; Ndamba and Chandiwana 1988).

Random amplified polymorphic DNA (RAPD) (Williams et al. 1990) has been frequently employed in assessing genetic diversity and relationships in various plant species (e.g. Baum et al. 1997; Fahima et al. 1999; Morden and Loeffler 1999). In a previous paper (Semagn et al. 1999a), we used RAPDs to investigate genetic diversity in 17 Ethiopian Endod populations. The results from that study revealed (1) a higher genetic differentiation among populations than would be expected for outcrossing plant species, (2) significantly higher genetic diversity in the lowland and the central-highland populations than those of the highlands and (3) distinct separation of the lowland and the central-highland populations from those of the *highland*s irrespective of their geographical regions and climatic zones.

The long debated controversy as to whether genetic variation in natural populations estimated from neutral markers is selectively meaningful or adaptively neutral (Lewontin 1974; Kimura 1983) is still an unresolved issue. According to the neutrality theory, most evolutionary change and intraspecific variability at the molecular level is caused not by natural selection but by random drift or mutant genes that are selectively equivalent (Kimura 1983). In this context, natural selection is regarded as a diversity-reducing force acting to sieve and save beneficial mutations by eliminating harmful ones. If natural selection is the main driving force, then different population-environmental combinations will tend to perpetuate individuals with complex assortments of genetic architecture (Dobzhansky 1970). According to Nevo (1978), the amounts of genetic polymorphism and heterozygosity vary non-randomly between loci, populations, species, habitats and life zones, and are correlated with ecological heterogeneity.

Endler (1986) has reviewed several methods which can be used to detect natural selection in natural populations. Correlation between geographic variation in the traits and environment is one of the methods used to suggest possible relationships between traits and environmental factors. If natural selection occurs, the geographic variation in the selective factor will give rise to parallel geographic variation in traits. The objectives of this paper were, therefore, to assess the extent of genetic divergence between altitude groups using different multivariate statistical methods and to test the hypothesis that ecogeographically varying selection results in a correlation between RAPDs and their selective ecogeographical factors in wild Ethiopian *Phytolacca dodecandra* populations.

Materials and methods

Sampling and RAPD analysis

A total of 249 samples were collected from 17 localities that represent the major geographical sites and ecological characteristics of the natural distribution of Endod in Ethiopia (Table 1). The 5 years (1992–1996) of meteorological data (monthly temperature and rainfall) were obtained for 16 localities (except Bure) from the National Meteorological Service Agency, Addis Ababa. The procedures for leaf sample collection, DNA isolation, RAPD reactions, fragment separation and the RAPD primers used have been described in (Semagn et al. (1999a). Seventy polymorphic loci (RAPDs) that varied between 396 to 2000 bp each were scored from 12 random primers using a 1-kb ladder as a size marker. A locus was scored as polymorphic (P) when the frequency of the most common band (present or absent) was less than 0.98. The different polymorphic loci scored from each primer were designated by a name that consisted of the primer followed by a number starting from one for the heaviest locus (e.g. 29-1: UBC 29 locus 1; Table 2). Loci with the same mobility were considered to be identical fragments, receiving equal values, regardless of their staining intensity. The exact molecular size of each locus was not determined. All polymorphic loci were entered in a computer file as binary matrices where 0 coded for absence and 1 coded for presence.

Table 1 The geographical background of 17 Endod (*Phytolacca dodecandra*) populations sampled in Ethiopia

Population	Sample size	Locality	Region	Longitude	Latitude	Altitude (m)	Altitude group[a]
1	15	Wondo Genet	Sidamo	38.35	7.10	1900	LL
2	15	Kofele	Arsi	38.48	7.04	2700	HL
3	14	Zuquala	Shewa	38.52	8.32	2900	HL
4	15	Menagesha	Shewa	38.29	9.04	2400	CH
5	10	Debre Sina	Shewa	39.45	9.52	2800	HL
6	15	Dese	Welo	39.38	11.06	2500	CH
7	15	Korem	Welo	39.31	12.31	2450	CH
9	15	Woken	Gonder	37.53	13.10	2700	HL
10	15	Taragedam	Gonder	37.52	12.07	2300	CH
11	15	Bure	Gojam	35.05	10.43	2100	LL
12	15	Dima	Welega	36.56	9.52	2000	LL
13	15	Nugema	Welega	36.30	8.45	1800	LL
14	15	Gechi	Ilubabor	35.43	8.20	1600	LL
15	15	Gina	Ilubabor	35.30	7.36	2400	CH
19	15	Gerima	Kefa	36.50	7.40	2600	HL
21	15	Abakara	Bale	39.46	7.06	3000	HL
22	15	Entoto	Shewa	38.42	9.03	3000	HL
Total	249						

[a] LL, Lowland; CH, central-highland; HI, highland

The populations were classified into three altitude groups (lowland: 1600–2100 m; central-highland: 2101–2500 m; highland: 2501–3000 m) in the same way as reported in our previous paper (Semagn et al. 1999a). Two data matrices were used for the different multivariate analyses techniques: DATA SET-1 consisted of 249×70 (249 rows representing individuals and 70 columns representing RAPDs). DATA SET-2 included the Y-matrix of size 16×72 (16 rows representing populations and 72 columns corresponding to the total number of polymorphic loci, Shannon information measure and population mean for band presence in each of the 70 RAPDs) and the X-matrix of ecogeographical variables of size 16×27 (16 rows corresponding to the populations and 27 columns representing longitude, latitude, altitude, monthly average temperature and monthly total rainfall). Shannon's information measure (H) was calculated as: $H=-\sum pi \log_2 pi$, where p_i is the frequency of a given RAPD fragment at the ith locus (Lewontin 1972). Mean band presence (referred to here as population mean) per locus was preferred over allele frequency due to the uncertainty in the polyploidy level of Endod.

Statistical analyses

Four different multivariate analyses techniques were used. (1) Principal component analysis (PCA) can be used to investigate overall variation in data without the concern of specifically grouping observations into predefined classes (Rencher 1992). PCA was performed on all individuals (DATA SET-1) and on population means using THE UNSCRAMBLER software (Computer-Aided Modelling, CAMO, version7.5, Oslo, Norway, e-mail: camo@camo.no); the first two principal components were plotted for visual examination of the clustering pattern of individuals and populations (scores) and selected RAPDs (loadings). (2) Different discriminant analyses can be used to summarise variation between predefined classes or classification variables. Stepwise discriminant analyses try to find out a subset of linear combinations of variables that best reveals the difference among the classes; canonical discriminant analysis determines how best to separate known groups of individuals, and classificatory discriminant analysis predicts to which predefined class an observation belongs (SAS 1987). We performed stepwise discriminant analysis on DATA SET-1 using the STEPDISC procedure to select an optimal set of discriminating variables (RAPDs) that tended to separate the altitude groups to a maximum degree. Variables were chosen to enter or leave the discrimination model among groups based on the significance level of an F test (P=0.15) from an analysis of covariance, where the variables already chosen act as covariates and the variable under consideration is the dependent variable. PROC CANDISC was used to summarise variation between altitude groups, based on the RAPDs chosen by PROC STEPDISC. PRO DISCRIM with full cross-validation and the nearest-neighbour method, which is suitable for classification when the groups have a non-normal distribution, was used for correct classifying plants. (3) Spearman rank correlations (r_s) were calculated for DATA SET-2 to test for association between genetic indices (P, H and population means) and ecogeographical variables. (4) When data on ecogeographical variables are available, other statistical models, including stepwise multiple regression analysis (SMR) can be used to find out the most important ecogeographical variable(s) that influence the genetic variation (e.g. Nevo and Beiles 1989; Fahima et al. 1999). Stepwise multiple regression analysis, with ecogeographical variables as regressors, was then performed on DATA SET-2 using PROC REG to find out the best predictors of P, H and the population mean for each RAPD. The validity of the regression model was tested using the variance inflation factor (VIF) as a measure of multicollinearity (near-linear dependence) among the regressors. VIF for each term in the model measures the combined effect of dependencies among the regressors on the variance of that term. The VIF value that exceeds 10 is considered to be an indication of multicollinearity (Montgomery and Peck 1992). All discriminant, correlation and regression analyses were made using SAS, version 6.12.

Table 2 The twelve 10-mer RAPD primers of University of British Columbia (UBC) used and the designation of the polymorphic loci scored in each primer

Primer	Locus	Prime	Locus
UBC 29	29–1 to 29–5	UBC 116	116–1 to 116–5
UBC 43	43–1 to 43–8	UBC 131	131–1 to 131–6
UBC 54	54–1 to 54–3	UBC 204	204–1 to 204–8
UBC 79	79–1 to 79–6	UBC 212	212–1 to 212–5
UBC 100	100–1 to 100–7	UBC 237	237–1 to 237–4
UBC 111	111–1 to 111–7	UBC 239	239–1 to 239–6

Results

Principal component analysis

The first five principal components (PCs) from principal component analysis of DATA SET-1 explained a total of

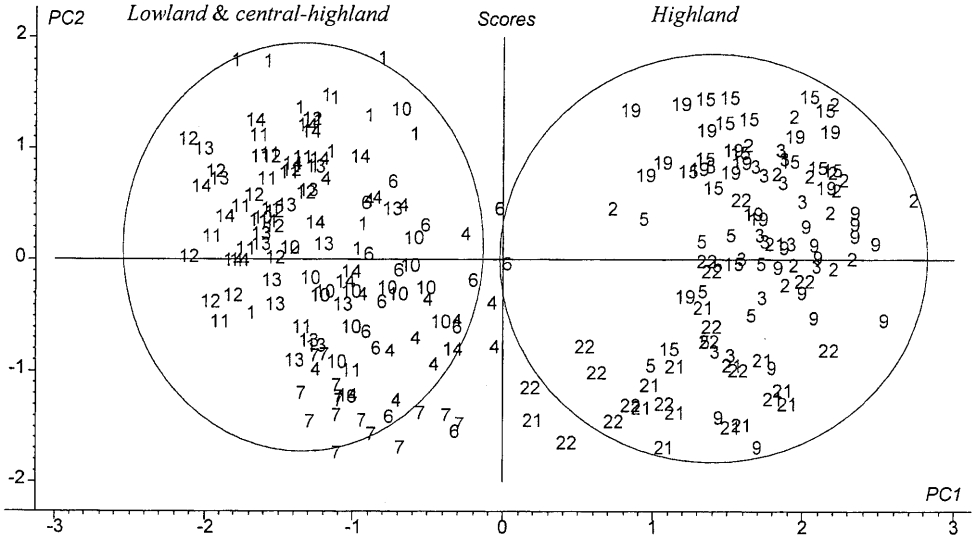

Fig. 1 Principal component analysis of 249 Endod plants from 17 populations using 70 RAPD markers. PC1 and PC2 explain 17% and 6% of the total variation, respectively. The *numbers* refer to the corresponding population numbers as indicated in Table 1

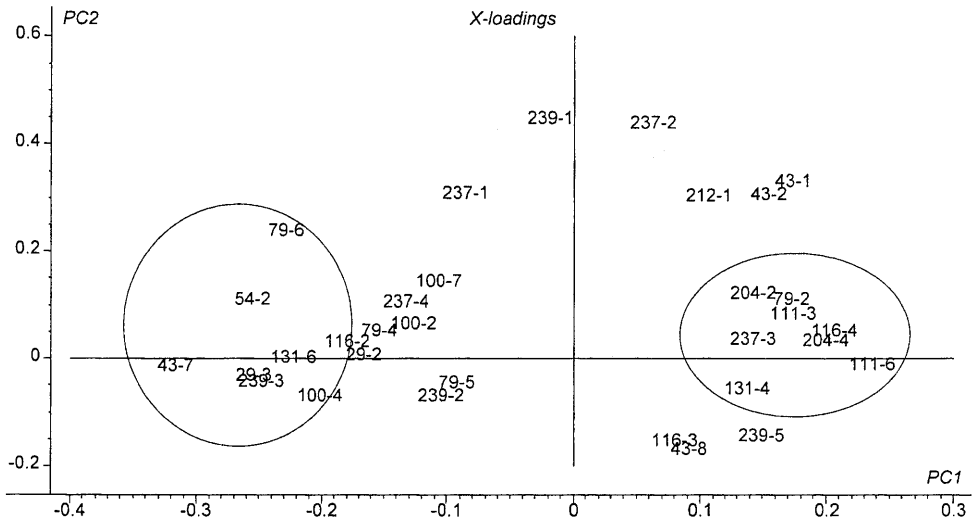

Fig. 2 The 32 selected RAPDs that contributed most to the PC1 and PC2 axes. The 16 RAPDs (indicated with *ellipse*) mainly contributed to separate the lowland and the central-highland populations from the highlands along PC1. The *numbers before* and *after* the *hyphen* refer to the primer and the particular locus for that primer, respectively (e.g. 239–1=UBC 239, locus 1)

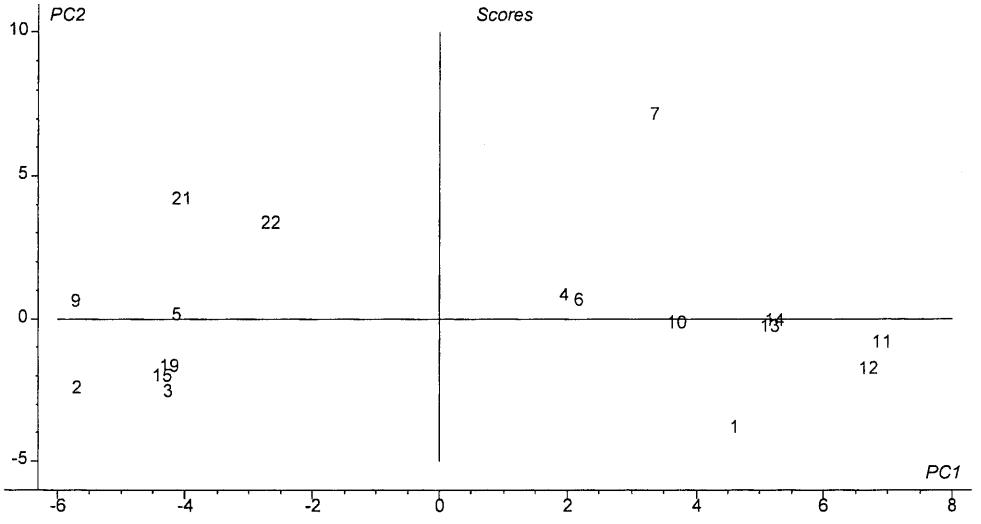

Fig. 3 Principal component analysis based on population means of 70 RAPDs. PC1 and PC2 explain 32% and 11% of the total variance, respectively. The *numbers* refer to the corresponding population numbers as indicated in Table 1

34% of the variations among samples. A plot of PC1 (17%) and PC2 (6%) revealed two major molecular groups (Fig. 1). The lowland and the central-highland populations were clearly separated from those of the highlands by PC1 with negative and positive co-ordinate values, respectively. The only exception was population 15 (altitude 2400 m) that was grouped with the highland populations. Figure 2 shows the 32 RAPDs that had relatively the highest loadings for PC1 and/or PC2. About 16 RAPDs scored from 11 primers were mainly responsible in separating the lowland and the central-highland populations from those of the highlands along PC1. When population means were used, instead of individuals, the first 5 PCs explained 66% of the total variance. A plot of PC1 and PC2 which explained 32% and 11% of the total variance, respectively, revealed two altitude groups as Fig. 1: the lowland and the central-highland populations (1, 4, 6, 7, 10–14) clustered to the right side of the quadrant while the highland plus population 15 clustered to the left side of the quadrant (Fig. 3). Populations 1 and 7 appeared more distant from the rest of the lowland and the central-highland populations.

Discriminant analyses

The discriminant analyses succeeded in differentiating, on the basis of band presence or absence, between the three altitude groups. The discrimination model obtained with the stepwise procedure of DATA SET-1 identified 37 RAPDs as the best differentiating factors (data not shown). A plot of CAN1 and CAN2 showed the distinct clustering of samples into their corresponding altitude groups as lowland, central-highland and highland (Fig. 4). CAN1 separated the highland plants (group mean =–3.67) from the lowland (group mean =4.38) with the central-highland plants positioned in between. CAN2 further separated mainly the central-highland plants (group mean =–1.77) from the others. *F*-statistics from pairwise comparisons of the three altitude groups

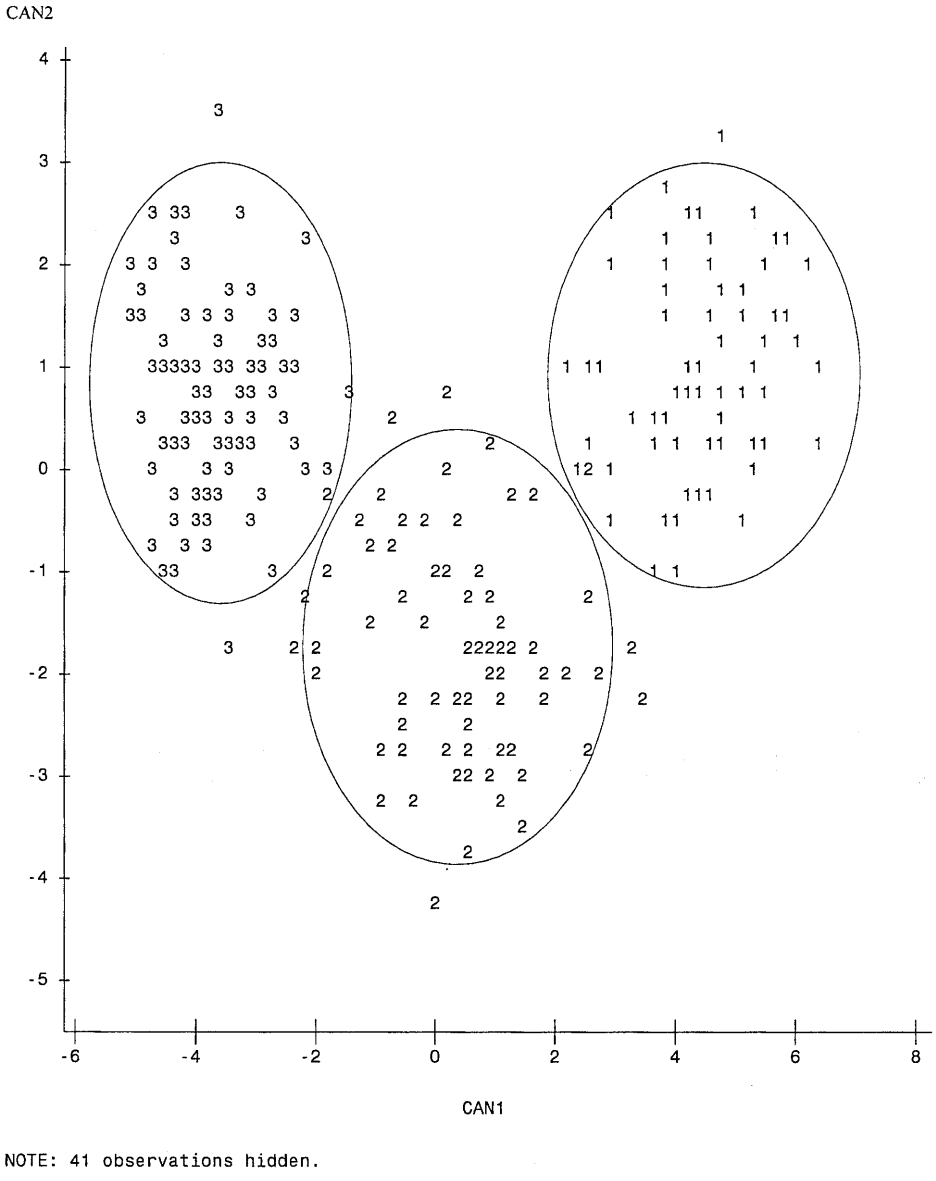

Fig. 4 Plot of canonical discriminant functions 1 and 2 based on 37 selected RAPDs. The numbers correspond to the altitude groups: *1* lowland, *2* central-highland, *3* highland

Table 3 Discriminant analyses, based on band presence or absence of 37 RAPDs, among three altitude groups of 17 populations of Endod

Pairwise comparisons of altitude groups; each *F*-statistic has 37 and 210 degrees of freedom

	Lowland	Central-highland	Highland
Lowland			
Central-highland	*F*=19.50, *P*=0.0001		
Highland	*F*=63.84; *P*=0.0001	*F*=22.63; *P*=0.0001	

Classification results for altitude groups

Actual group	*n*	Predicted group		
		Lowland	Central-highland	Highland
Lowland	75	70	5	0
Central-highland	75	6	64	5
Highland	99	0	2	97

231 plants (92.8%) were correctly classified

Table 4 Spearman rank correlation coefficients (r_s) of population means from all except population 11 with altitude, longitude, latitude, annual average temperature and annual rainfall. Only significant correlations were included in the table

Locus	Altitude	Longitude	Latitude	Annual average temperature	Annual rainfall
29–1	0.56*				
29–2	–0.68**			0.60*	
29–3	–0.75**			0.60*	
43–1			–0.60*		
43–3	0.66**				
43–5				0.54*	
43–7	–0.68**			0.56*	
43–8	0.50*				
54–2	–0.83**			0.66*	
79–2	0.51*				
79–4	–0.69**			0.63*	
79–6	–0.86*			0.65*	
100–1	–0.63*	–0.56*		0.64*	
100–2	–0.72**			0.75**	
100–3	–0.63**	–0.84**		0.71**	
100–4	–0.59*				
100–6	–0.50*				
111–1			–0.51*		
111–3	0.65*				
111–4	0.54*				
111–5					0.63*
111–6	0.53*			–0.50*	
116–2	–0.59*			0.61*	
116–4	0.51*				
131–1			0.64*		
131–3		0.58*			
131–6	–0.79**			0.78**	
204–1					
204–2			–0.55*		
204–4	0.61*			–0.65*	0.50*
212–3	–0.55*				
237–1	–0.71**	–0.56*	–0.53*	0.57*	
237–2		–0.67**			
237–3				–0.58*	
237–4	–0.70**	–0.51*		0.70**	
239–1	–0.61	–0.77**		0.72**	
239–2	–0.64**			0.51*	
239–3	–0.74**			0.62*	
239–4			0.57*	0.62*	

* $P<0.05$, ** $P<0.005$

[a] The number before and after the hyphen refer to the primer and the locus, respectively (e.g. 29–1: UBC 29, locus 1)

Table 5 Coefficient of multiple regressions (R^2) of genetic indices (P, H and population means) as dependent variables, and 17 ecogeographical variables (longitude, latitude, altitude, average temperature in July and October, and monthly rainfall) in 16 populations of *P. dodecandra* as independent variables. The following is a list of 7 RAPDs for which no variable met the 0.15 significance level for entry into the model: 29–4, 54–1, 79–3, 204–1, 204–6, 212–2 and 212–4. Each month is abbreviated by the first three letters (e.g. Jan for January)

Locus[a]	X_1	X_2	X_3	X_4	X_5	R^2_1	R^2_{12}	R^2_{123}	R^2_{1234}	R^2_{12345}
P	Jul T[b]					0.16				
H	Jul T					0.21				
Population means										
29–1	Altitude					0.49				
29–2	Altitude	Longitude				0.52	0.80			
29–3	Altitude	Longitude				0.38	0.70			
29–5	Jul RF	Aug RF	Sep RF[c]			0.21	0.44	0.57		
43–1	Jan RF	Altitude	Apr RF			0.44	0.58	0.69		
43–2	Aug RF	Longitude	May RF	Jan RF		0.22	0.34	0.50	0.67	
43–3	Altitude	Sep RF				0.32	0.43			
43–4	Altitude					0.16				
43–5	Oct RF	Altitude	Jul RF	Longitude		0.44	0.73	0.80	0.86	
43–6	Jul RF	May RF	Latitude			0.28	0.50	0.75		
43–7	Altitude	Longitude	Jan RF			0.62	0.80	0.88		
43–8	Altitude	Jul T				0.27	0.39			
54–2	Altitude	Longitude	Apr RF			0.78	0.88	0.90		
54–3	Sep RF	Feb RF				0.38	0.51			
79–1	Mar RF					0.46				

Table 5 Continued

Locus[a]	X_1	X_2	X_3	X_4	X_5	R^2_1	R^2_{12}	R^2_{123}	R^2_{1234}	R^2_{12345}
79–2	Jan RF	Altitude				0.31	0.43			
79–4	Altitude	Latitude				0.43	0.54			
79–5	Altitude	Oct T				0.46	0.63			
79–6	Altitude					0.86				
100–1	Altitude	Sep RF	Longitude	Jan RF		0.53	0.66	0.84	0.93	
100–2	Altitude	Longitude	Oct T	Jul T		0.61	0.79	0.89	0.92	
100–3	Oct T	Jul T	Oct RF			0.54	0.82	0.91		
100–4	Altitude	Aug RF	Longitude			0.34	0.65	0.72		
100–5	Apr RF					0.28				
100–6	Altitude	Jan RF				0.28	0.49			
100–7	Dec RF	Altitude				0.09	0.25			
111–1	Jan RF	Aug RF	Mar RF			0.54	0.68	0.76		
111–2	Aug RF	Altitude	Jan RF			0.31	0.48	0.58		
111–3	Altitude	Longitude	Jan RF	May RF		0.42	0.57	0.64	0.74	
111–4	Altitude	Mar RF	Oct T	Feb RF	Sep RF	0.38	0.58	0.75	0.81	0.87
111–5	Apr RF	Aug RF	Altitude	Dec RF	Feb RF	0.46	0.67	0.76	0.81	0.86[d]
111–6	Altitude	Longitude				0.33	0.72			
111–7	Nov RF	Altitude	Feb RF	Jan RF		0.45	0.62	0.71	0.77	
116–1	Apr RF	Oct RF				0.27	0.43			
116–2	Jul T	May RF	Altitude	Longitude	Jul RF	0.49	0.63	0.75	0.91	0.94
116–3	Oct T	Dec RF	Jan RF	Aug RF		0.19	0.35	0.67	0.77	
116–4	Altitude	Longitude	Jul RF			0.26	0.51	0.60		
116–5	Oct T	May RF	Mar RF	Dec RF	Aug RF	0.17	0.35	0.55	0.66	0.73
131–1	Latitude	Dec RF	Nov RF	Apr RF		0.26	0.60	0.71	0.83	
131–2	Jul RF	Altitude				0.45	0.54			
131–3	Longitude	Oct RF	Aug RF			0.10	0.36	0.40		
131–4	May RF					0.22				
131–5	Dec RF	Jul RF	Sep RF			0.17	0.36	0.58		
131–6	Jul T	Altitude	Jan RF	Feb RF	Latitude	0.52	0.63	0.69	0.81	0.87[e]
204–2	Latitude	Jul RF	Feb RF			0.33	0.44	0.54		
204–3	Mar RF					0.34				
204–4	Jul T	Nov RF				0.49	0.71			
204–5	Aug RF	Jan RF				0.22	0.38			
204–7	Sep RF					0.28				
204–8	Jul T	Altitude				0.27	0.45			
212–1	Jan RF	Oct T	Altitude			0.25	0.45	0.73		
212–3	Altitude	Aug RF				0.36	0.47			
212–5	Sep RF	Jul T				0.27	0.42			
237–1	Altitude	Latitude				0.59	0.75			
237–2	Longitude	Aug RF	Jul RF	Jun RF	Latitude	0.55	0.65	0.76	0.89	0.91[f]
237–3	Oct T					0.21				
237–4	Oct T					0.36				
239–1	Oct T	Jul RF	Mar RF	Jul T		0.57	0.72	0.83	0.89	
239–2	Altitude	Latitude	Longitude			0.45	0.63	0.74		
239–3	Altitude	Nov RF				0.49	0.62			
239–4	Latitude	Altitude				0.11	0.67			
239–5	Oct T	May RF				0.25	0.42			
239–6	Altitude					0.18				

[a] The number before and after the hyphen refer to the primer and the locus respectively (e.g. 29–1: UBC 29, locus1)
[b] T: Average temperature
[c] RF: Rainfall
[d] Jul RF and May RF increased the R^2 to 0.94
[e] Apr RF and May RF increased the R^2 to 0.94
[f] Altitude, May RF and Nov RF increased the R^2 to 0.98

have also indicated significant differences, with the difference between the lowland and the highland groups being very large (Table 3). The correct classification of plants into their respective altitude groups, based on the 37 selected RAPDs, was 92.8% (Table 3); the addition of the 33 RAPDs which were removed by STEPDISC increased the percentage of correct classification only by 0.4%.

Spearman rank correlation (r_s)

Correlations between genetic indices (P, H and population means) and ecogeographical variables were computed with the aim of determining whether the clinal variation in PCA and discriminant analyses is associated with any of the ecogeographical variables. The results indicated the lack of correlation of P and H with any of the ecogeographical variables. Table 4 summarises the relationship between population means of selected RAPDs and ecogeographical variables. Out of the 70 RAPDs tested,

a total of 28 RAPDs showed significant positive (9 RAPDs) and negative (19 RAPDs) correlations with altitude. Population means from 21 RAPDs were also correlated with annual average temperature. All RAPDs, except 3 (43–5, 237–3 and 239–4), that were correlated with annual average temperature were also correlated with altitude; the converse was not true. Six and 7 RAPDs were also correlated with latitude and longitude, respectively, against only 2 RAPDs (111–5 and 204–4) with annual rainfall.

Stepwise multiple regression (SMR) analyses

SMR was first performed on DATA SET-2 using genetic indices (P, H, and population means of the 70 RAPDs) as dependent variables and the 27 ecogeographical variables (longitude, latitude, altitude, monthly average temperature and monthly total rainfall) as independent variables. Each of the genetic indices identified 1 or more of the ecogeographical variables. There was, however, severe multicollinearity among the selected ecogeographical variables as indicated by the high variance inflation factors (VIFs) that ranged between 1 to 457 (data not shown). Multicollinearity among the regressors was observed when two or more monthly temperatures were selected by the model. Regression analyses were repeated several times by removing temperature until VIF <10 for all selected regressors. This criterion reduced the number of regressors into 17 by removing all monthly temperatures except July and October.

Temperature in July was the only explanatory variable selected for P (R^2=0.16) and H (R^2=0.21). None of the 17 ecogeographical variables were selected in 7 RAPDs. Population means from the remaining 63 RAPDs identified 1 or more explanatory variable(s) (Table 5) and can broadly be classified into seven categories based on their primer association with the ecogeographical variables: (1) geographical factors (altitude, latitude and longitude) (11 RAPDs), (2) geographical and rainfall factors (25 RAPDs), (3) geographical and temperature factors (3 RAPDs), (4) rainfall only (11 RAPDs), (5) rainfall and temperature factors (7 RAPDs), (6) temperature only (2 RAPDs) and (7) geographical, rainfall and temperature factors (4 RAPDs). Altitude was the most important of the geographical variables. It singly explained 40–57% of the variance in 9 RAPDs (29–1, 29–2, 79–4, 79–5, 100–1, 111–3, 239–2, 239–3 and 239–4) and 58–86% of the variance in 5 other RAPDs (43–7, 54–2, 79–6, 100–2, 237–1). Rainfall in particular months singly explained about 40–57% of the variance in 7 RAPDs (43–1, 43–5, 79–1,111–1, 111–5, 111–7 and 131–2). Temperature in July and/or October singly explained 40–57% of the variance in 5 RAPDs (100–3, 116–2, 131–6, 204–4 and 239–1), against only 1 RAPD (237–2) for longitude and none for latitude. Most of the variation in the other RAPDs was explained by 2 or more combinations of the different ecogeographical variables. The best two-variable combinations was altitude and longitude in 29–2, 29–3 and 111–6; altitude and latitude in 237–1 and 239–4; July temperature and November rainfall in 204–4 (Table 5).

It is also evident from Table 5 that most of the variance in 13 of the 21 RAPDs (29–2, 29–3, 43–7, 54–2, 79–4, 79–6, 100–1, 100–2, 111–6, 237–1, 239–2, 239–3 and 239–4) which were significantly correlated with annual average temperature was primarily explained by altitude. In fact in all except 100–2 of the latter, temperature was not selected as an explanatory variable, clearly indicating an indirect association of temperature with altitude. Temperature explained most of the variance only in the other 6 RAPDs (100–3, 116–2, 131–6, 204–4, 237–4 and 239–1).

Discussion

Our results support the earlier conclusion drawn by Semagn (1999) concerning the existence of strong genetic differentiation among Ethiopian *Phytolacca dodecandra* populations along altitudinal gradients. Genetic differentiation between populations along altitudinal gradients has been demonstrated in several plant species because important environmental conditions change rapidly with elevation (e.g. Clausen et al. 1940; Nevo et al. 1983; Fahima *et al*. 1999). As pointed out by Bradshaw (1965), genetic differentiation is predicted for traits that are adaptations to environmental conditions likely to remain fairly constant during the lifespan of an organism. In Endod, evidence supporting genetic differentiation, either into two- or three- altitude groups, comes from principal component and discriminant analyses and from environmental correlates of RAPD polymorphisms. Principal component analyses performed on all individuals and population means (Figs. 1, 3) revealed distinct differentiation of the lowland and the central-highland populations from the highlands. When altitude was used as a classification variable, discriminant analyses resulted in a distinct separation between the lowland and highland populations, with the central-highland plants being intermediate between the former (Table 3; Fig. 4). No single plant from the lowland populations was classified into the highland and vice versa (Table 3).

Correlation of traits with environmental factors is one of the commonest and oldest method for detecting selection. Although a correlation does not necessarily imply a cause-and-effect relationship, it is useful in determining just what traits and what environmental variables are suspected of being engaged in a selective process (Endler 1986). In the present work, population means from 28 RAPDs were significantly correlated with altitude, against 2 RAPDs with annual rainfall, 6 RAPDs with latitude, 7 RAPDs with longitude and 21 RAPDs with annual average temperature. Stepwise multiple regression analysis indicates altitude, singly or in combination with the other ecogeographical variables, as an explanatory variable(s) in 36 RAPDs against only 16 for longitude, 16 for temperature and 10 for latitude (Table 5). Hence, it is highly likely that altitude (and/or other

correlated environmental variables) exerts a differential selective pressure on closely linked genes or co-adapted gene blocks. We hypothesise that the genetic differentiation at or linked to RAPD-DNA loci between altitude groups is partly adaptive and responsive to natural ecological selection, as has previously been reported in Drosophila (e.g. Brussard 1984), mole-rats (e.g. Nevo 1990), Trimerotropis pallidipennis (Orthoptera) (Confalonieri and Colombo 1989) and wild barley (e.g. Nevo et al. 1983; Baum et al. 1997; Fahima et al. 1999).

Alternative hypotheses to explain RAPD differentiation

Data on morphological traits in Endod did not support the differentiation of populations into morphological ecotypes (Semagn et al. 1999b), but the possibility that they may be chemical and/or physiological ecotypes (Stace 1989) remains to be fully determined. There are two indications for the presence of chemical ecotypes. First, the Ethiopia Endod populations are recognised by two names, depending on the colour of the berries: arabe with pinkish berries and ahiyo with greyish berries (Lemma 1983). The former is the case for population 15 and most of the *highland* plants. Preliminary data from high-performance liquid chromatography (HPLC) indicated that the pinkish colour in the arabe-type berries is due to phenolic compounds that belong to the betacyanin family (F. Vinicieri et al., unpublished). Secondly, Parkhurst et al. (1990), based on their studies on variation in triterpene saponins from 15 populations collected over a wide geographical range in Africa, have also reported the presence of low (<66%) and high (>80%) levels of oleanolic acid groups. An indication of physiological or phenological ecotypes may be the observed distinct differences in growth habit, morphology and flowering time among 11 clones cultivated in a common garden experiment in Addis Ababa (K. Semagn, unpublished).

Habitat differentiation between other polyploid species and their diploid progenitors has been demonstrated in a number of field and laboratory studies. For example, allozyme and cytological analyses suggested that natural populations of *Heuchera grossulariifolia* (Wolf et al. 1990), *Heuchera micrantha* (Ness *et al*. 1989) and *Tolmiea menziesii* (Soltis 1984; Soltis and Rieseberg 1986) consist of either diploid or tetraploid plants; no mixed populations were detected. Similarly, four chromosomal species (2n=52, 54, 58, 60) were reported in mole-rats, *Spalax ehrenbergi* superspecies complex, each of which were associated to a different temperature, precipitation and humidity (Nevo 1990). The ploidy level of *P. dodecandra* is unclear. Chromosome counts made from Rwandan plant material indicated that the species is tetraploid with 2n=4x=36 (Auquier and Renard 1975), while preliminary chromosome counts from Ethiopian material indicated a much higher ploidy level (B. Stedje; K. Dagne, unpublished) although detailed investigation was found difficult due to the small size of chromosomes (<1 µm). Stomatal size was measured from 48 herbarium specimens of Endod to see if the data could be used to predict differences in ploidy level between altitude groups. These preliminary data, however, did not indicate any significant difference between altitude groups (B. Stedje, unpublished). With the same analogy as that of diploid/tetraploid relationships mentioned above or that of mole-rats, the existence of habitat differentiation between altitude groups due to different ploidy levels in Endod is one possible hypothesis, which can be tested by crossing the lowland and the highland genotypes and assesssing the viability and seed set of the F_1 plants. The latter, however, has not yet been tested due to the long time required obtain F_1 plants (about 2 years) in order to assess their fertility.

Acknowledgements We gratefully acknowledge the financial support of the Norwegian Universities Committee for Development Research and Education (NUFU), project number 39B/96. The experiments described herein comply with the current laws in Norway.

References

Auquier P, Renard R (1975) Nombres chromosomiques de quelques Angiospermes du Rwanda, Burundi et Kivu (Zaire). Bull Jard Bot Nat Belg 45:421–445

Baum BR, Nevo E, Johnson DA, Beiles A (1997) Genetic diversity in wild barley (*Hordeum spontaneum* C. Koch) in the Near East: a molecular analysis using random amplified polymorphic DNA (RAPD) markers. Genet Res Crop Evol 44:147–157

Bradshaw AD (1965) Evolutionary significance of phenotypic plasticity in plants. Adv Genet 13:115–155

Brussard PE (1984) Geographical patterns and environmental gradients: the central-marginal model revisited. Annu Rev Ecol Syst 15:25–64

Chimbelu EG, Shehata MA (1986) A review of research in Zambia on *Phytolacca dodecandra* (Endod). In: Makhubu L, Lemma A, Heyneman D (eds) Endod II (*Phytolacca dodecandra*). Report of the second international workshop on Endod. Mbabane, Swaziland, Council on International and Public Affairs' Press, New York, pp 109–126

Clausen JD, Keck D, Hiesey WM (1940) Experimental studies on the nature of species. I. The effect of varied environments on western North American plants. Carnegie Inst Washington Publ 520

Confalonieri VA, Colombo PC (1989) Inversion polymorphisms in *Trimerotropis pallidipennis* (Orthoptera): clinal variation along an altitudinal gradient. Heredity 62:107–112

Dalziel JH (1936) The useful plants of West Africa. Crown Agents, London

Dobzhansky T (1970) Genetics of the evolutionary process. Columbia University Press, New York

Endler JA (1986) Natural selection in the wild. Princeton University Press, Princeton, N.J.

Fahima T, Sun GL, Beharav A, Krugman T, Beiles A, Nevo E (1999) RAPD polymorphism of wild emmer wheat populations, *Triticum dicoccoides*, in Israel. Theor Appl Genet 98:434–447

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge

Lambert JDH, Temmink JHM, Marquis J, Parkhurst RM, Lugt CB, Schoonen AJM, Holtze K, Warner JE, Dixon G, Wolde-Yohannes L, Desavigny D (1991) Endod: safety evaluation of a plant molluscicide. Reg Pharm Toxicol 14:189–201

Lemma A (1983) Background and historical review. In: Lemma A, Heyneman D, Silangwa SM (eds) *Phytolacca dodecandra* (Endod). Final report of the international workshop. Lusaka,

Zambia, TYCOOIY International publishing Ltd., Dublin, Ireland, pp 12–44
Lemma A, Parkhurst RM, Skinner WA (1974) Method of producing molluscicide from Endod. United States Patent Office, Patented Procedures, Patent No. 3,813,3
Lemma A, Heyneman D, Kloos H (1979) Studies on the molluscicidal and other properties of the Endod plant *Phytolacca dodecandra*. Special Publication, Institute of Pathobiology, Addis Ababa University, and Department of Epidemiology and International Health, University of California, San Francisco, Calif
Lewontin RC (1972) The apportionment of human diversity. Evol Biol 6:381–398
Lewontin RC (1974) The genetic basis of evolutionary change. Columbia University Press, New York
Montgomery DC, Peck EA (1992) Introduction to linear regression analysis. 2nd edn. John Wiley and Sons, New York
Morden CW, Loeffler W (1999) Fragmentation and genetic differentiation among subpopulations of endangered Hawaiian mint *Haplostachys haplostachya* (Lamiaceae). Mol Ecol 8:617–625
Ndamba J, Chandiwana SK (1988) The geographical variation in the molluscicidal potency of *Phytolacca dodecandra* in Zimbabwe. Trop Geogr Med 40:34–38
Ness BD, Soltis DE, Soltis PS (1989) Autopolyploidy in *Heuchera micrantha* (Saxifragaceae). Am J Bot 76:614–626
Nevo E (1978) Genetic variation in natural populations: patterns and theory. Theor Pop Biol 13:121–177
Nevo E (1990) Evolutionary theory and processes of active speciation and adaptive radiation in subterranean mole rats, *Spalax ehrenbergi* superspecies in Israel. Evol Biol 25:1–125
Nevo E, Beiles A (1989) Genetic diversity of wild emmer wheat in Israel and Turkey: structure, evolution, and application in breeding. Theor Appl Genet 77:421–455
Nevo E, Beiles A, Storch N (1983) Microgeographic edaphic differentiation in hordein polymorphisms of wild barley. Theor Appl Genet 64:123–132
Parkhurst RM, Thomas DW, Adams RP, Makhubu LP, Mthupha BM, Wolde-Yohannes L, Mamo E, Heath GE, Stobaeus JK, Jones WO (1990) Triterpene aglycones from various *Phytolacca dodecandra* populations. Phytochemistry 29:1171–1174
Pezzuto JM, Swanson SM, Farnsworth NR (1984) Evaluation of the mutagenic potential of Endod (*Phytolacca dodecandra*), a molluscicide of potential value for the control of schistosomiasis. Toxicol Lett 22:15–20
Rencher AC (1992) Interpretation of canonical discriminant functions, canonical variates, and principal components. Am Stat 46:217–225
SAS (1987) SAS/STAT guide for personal computers, version 6 edition. SAS Institute, Cary, N.C.
Semagn K, Stedje B, Bjornstad A, Bekele E (1999a) Analysis of genetic diversity in wild populations of *Phytolacca dodecandra* in Ethiopia using RAPDs: an aid in conservation and breeding programs. In: Semagn K, Morphological and molecular diversity in Ethiopian populations of *Phytolacca dodecandra* (Endod): Contributions to the understanding of Endod genetics for Future Endod improvement. Ph. D. thesis, Agricultural University of Norway in collaboration with Addis Ababa University, Agricultural University of Norway Press, ÅS, Norway.
Semagn K, Stedje B, Bjornstad A, Bekele E (1999b) Patterns of morphological and molecular variation in *Phytolacca dodecandra* in Ethiopia. In: Semagn K, Morphological and molecular diversity in Ethiopian populations of *Phytolacca dodecandra* (Endod): Contributions to the understanding of Endod genetics for Future Endod improvement. Ph. D. thesis, Agricultural University of Norway in collaboration with Addis Ababa University, Agricultural University of Norway Press, ÅS, Norway.
Soltis DE (1984) Autopolyploidy in *Tolmiea menziesii* (Saxifragaceae). Am J Bot 71:1171–1174
Soltis DE, Rieseberg LH (1986) Autopolyploidy in *Tolmiea menziesii* (Saxifragaceae): genetic insights from enzyme electrophoresis. Am J Bot 73:310–318
Stace CA (1989) Plant taxonomy and biosystematics. 2nd edn. Chapman and Hall, New York
WHO (1983) The control of schistosomiasis. Report of a WHO Expert Committee. WHO Tech Rep Ser 728, WHO, Geneva
Williams JGK, Kublelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphism's amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 18:6531–6535
Wolde-Yohannes L (1983) Past and ongoing agrobotanical studies of *Phytolacca dodecandra* (Endod) in Ethiopia. In: Lemma A, Heyneman D, Silangwa SM (eds) Phytolacca dodecandra (Endod). Final report of the international workshop. Lusaka, Zambia, TYCOOIY International publishing Ltd., Dublin, Ireland, pp125–129
Wolde-Yohannes L, Demeke T, Mamo E, Retta N, Bulbula A (1988) The natural distribution and traditional uses of *Phytolacca dodecandra* in Ethiopia. IPB Res Report 4:20–22
Wolf PG, Soltis DE, Soltis PS (1990) Chloroplast-DNA and allozymic variation in diploid and autotetraploid *Heuchera grossularifolia* (Saxifragaceae). Am J Bot 77:232–244