Theor Appl Genet (2002) 105:289–297
DOI 10.1007/s00122-002-0983-4

B. Van Droogenbroeck · P. Breyne · P. Goetghebeur
E. Romeijn-Peeters · T. Kyndt · G. Gheysen

AFLP analysis of genetic relationships among papaya and its wild relatives (Caricaceae) from Ecuador

Received: 30 October 2001 / Accepted: 4 March 2002 / Published online: 21 June 2002

Abstract The AFLP technique was used to assess the genetic relationships among the cultivated papaya (*Carica papaya* L.) and related species native to Ecuador. Genetic distances based on AFLP data were estimated for 95 accessions belonging to three genera including *C. papaya*, at least eight *Vasconcella* species and two *Jacaratia* species. Cluster analysis using different methods and principal co-ordinate analysis (PCO), based on the AFLP data from 496 polymorphic bands generated with five primer combinations, was performed. The resulted grouping of accessions of each species corresponds largely with their taxonomic classifications and were found to be consistent with other studies based on RAPD, isozyme and cpDNA data. The AFLP analysis supports the recent rehabilitation of the *Vasconcella* group as a genus; until recently *Vasconcella* was considered as a section within the genus *Carica*. Both cluster and PCO analysis clearly separated the species of the three genera and illustrated the large genetic distance between *C. papaya* accessions and the *Vasconcella* group. The specific clustering of the highly diverse group of *Vasconcella* × *heilbornii* accessions also suggests that these genotypes may be the result of bi-directional introgression events between *Vasconcella stipulata* and *Vasconcella cundinamarcensis*.

Keywords *Carica papaya* · *Vasconcella* · *Jacaratia* · AFLP · Genetic diversity · Phenetic analysis · Introgression · Interspecific hybrids

Communicated by J. Dvorak

B. Van Droogenbroeck · T. Kyndt · G. Gheysen (✉)
Department of Molecular Biotechnology,
Faculty of Agricultural and Applied Biological Sciences,
Ghent University, Coupure links 653, B-9000 Gent, Belgium
e-mail: Godelieve.Gheysen@rug.ac.be
Tel.: +32-9-264-5888, Fax: +32-9-264-6238

P. Breyne
Department of Genetics, Faculty of Sciences,
Flanders Interuniversity Institute for Biotechnology,
Ghent University, K.L. Ledeganckstraat 35,
B-9000 Gent, Belgium

P. Goetghebeur · E. Romeijn-Peeters
Department of Biology,
Faculty of Sciences Ghent University,
K.L. Ledeganckstraat 35, B-9000 Gent, Belgium

Present address:
P. Breyne, Institute for Forestry and Game Management,
Gaverstraat 4, P. Breyne, B-9500 Geraardsbergen, Belgium

Introduction

Papaya (*Carica papaya* L.) is a member of the Caricaceae, a small dicotyledonous family consisting of six genera of herbaceous, shrubby or arborescent plants (Badillo 1971, 1993, 2000). It is now the only species belonging to the genus *Carica*, since Badillo rehabilitated the *Vasconcella* group, until recently considered as a section within the genus *Carica* (Badillo 2000). *Vasconcella*, now comprising 21 species, is the largest genus of the family, followed by the genus *Jacaratia* with seven species. These two genera are predominantly South American in origin whereas the domesticated papayas appear to have originated from a small-fruited ancestor in Central America (Badillo 1993). The other genera include *Jarilla* with three species, from Mexico and Guatemala, *Horovitzia* with only one species indigenous to Mexico and *Cylicomorpha* with two species. *Cylicomorpha* is the only genus native to equatorial Africa (Badillo 1993).

Papaya is by far the best known and economically most important species of the family. Dioecious and hermaphrodite cultivars are grown in many tropical and subtropical countries for their edible, vitamin-rich fruits and to a lesser extent also for their milky latex (Drew et al. 1998). The different proteinases, present in the latex obtained from green unripe fruits, have a broad spectrum of activity and are therefore widely used in the food and pharmaceutical industries (Madrigal et al. 1980). Papaya production doubled over the last 5 years and was about 7.2 million metric tons in 2000, with almost half of this amount produced by Brazil (FAO 2001). Fifteen of the 21 species in the genus *Vasconella*, the so-called high-

land papayas, grow in the upland valleys of Ecuador at up to 3,000 m above sea level (Badillo 1971, 1983, 1997). Some of these highland papayas are regarded as unexploited species, while others possess interesting agronomic traits that may be useful in papaya breeding (National Research Council 1989). The Babaco (*Vasconcella* × *heilbornii* cv 'Babaco') is believed to be a natural sterile hybrid of *Vasconcella stipulata* and *Vasconcella cundinamarcensis* (formerly *Carica pubescens*) originating in the Ecuadorian highlands (Horovitz and Jimenez 1967). Thanks to its large seedless fruits and higher yields than papaya, Babaco cultivation is beginning in other countries (Kempler and Kabaluk 1996). So far, intergeneric hybridization of cultivated *C. papaya* with its wild *Vasconcella* relatives has been possible only with embryo or ovule rescue, since endosperm development does not occur (Manshardt and Wenslaff 1989; Manshardt and Drew 1998). However, new biotechnological approaches for gene transfer are increasing the potential for exploiting useful genes from these wild species (Drew et al. 1998).

In the course of previous ethnobotanical inventories of wild and semi-domesticated edible plants in southern Ecuador, an unrecognised variability among and within the species of the genus *Vasconcella* has been observed (Jiménez et al. 1999). Many of the wild relatives of papaya are intercompatible and can cross-pollinate to produce hybrids with varying degrees of fertility. These hybrids occur spontaneously in areas where species distributions overlap (Badillo 1971). A new *Vasconcella* species was discovered in 2000 (Badillo et al. 2000) with probably additional species yet to be discovered (National Research Council 1989). There also exists a need to monitor, collect and preserve the genetic resources in Ecuador, as certain habitats of these and other species are under great pressure for land clearance.

Badillo has published extensively on the taxonomical classification of the Caricaceae using primarily morphological traits (Badillo 1971, 1993, 1997, 2000). Only a few molecular studies addressed inter- and intra-specific relationships among Caricaceae spp. Isozyme and random amplified polymorphic DNA analysis (RAPD) were used to study the genetic diversity among *C. papaya* varieties (Stiles et al. 1993; Morshidi 1998). Jobin-Décor et al. (1997) were the first to analyse the genetic relationships between six *Vasconcella* species and *C. papaya* using isozyme and RAPD marker data. They concluded that *C. papaya* is only distantly related to *Vasconcella* and revealed the close genetic relatedness among the South American relatives. These findings were supported by an analysis of chloroplast DNA diversity in *C. papaya* and 11 *Vasconcella* species (Aradhya et al. 1999). Based upon this work, combined with his own morphological observations, Badillo (2000) proposed the rehabilitation of the section *Vasconcella* as a genus.

Because the relationships within and between the genera *Carica* and *Vasconcella* are still incompletely understood, a more detailed analysis using other molecular techniques is necessary. The amplified fragment length polymorphism (AFLP) method (Vos et al. 1995) has been succesfully employed for fingerprinting varieties, cultivars and clones (e.g. Barrett and Kidwell 1998). Moreover, AFLP also proved effective to analyse inter- and intra-specific genetic diversity among highly diverse species, as illustrated for a wide range of crop species and their wild relatives (e.g. Aggarwal et al. 1999). In the present study, we used AFLP to determine the genetic diversity and the phenetic relationships among a group of 95 accessions of *Carica*, *Vasconcella* and *Jacaratia* species collected in Ecuador. This study is also the first to include *Vasconcella palandensis, Vasconcella weberbaueri, Vasconcella candicans* and *Jacaratia digitata* in a DNA-fingerprinting study.

Materials and methods

Plant material

All plant material was collected in Loja province, southern Ecuador. Fresh young leaves were cut off and dried with silica gel. A total of 95 accessions were included in the AFLP analysis, comprising six accessions of *C. papaya*, 83 accessions of at least eight different *Vasconcella* species, together with three accessions of both *Jacaratia spinosa* and *J. digitata* (Table 1).

DNA isolation and AFLP analysis

DNA extraction was performed on dried leaf tissue ground in liquid nitrogen. Total genomic DNA was extracted using the Qiagen Dneasy Plant Mini Kit (Qiagen GmbH, Hilden, Germany).

The AFLP analysis was carried out essentially as described by Vos et al. (1995) with minor modifications. All primers and adaptors were obtained from Genset (Paris, France). AFLP templates were prepared by simultaneous digestion of about 500 ng of DNA with *Eco*RI and *Mse*I. Ligation of the restriction fragments to the adapters was performed in the same step. A 1:5 dilution of the restricted and adapter-ligated DNA was used as a template in the pre-amplification reactions. Three different sets of pre-amplification products were generated using an *Eco*RI-primer carrying zero selective nucleotides (E-0) in combination with a *Mse*I-primer carrying two selective nucleotides (either M-AC, M-CT or M-GC). For the final selective amplification, the 1:10 diluted pre-amplified DNA was amplified using a γ-[^{33}P]-ATP labelled *Eco*RI-primer carrying two selective nucleotides (E-GA, E-CG or E-GT) in combination with a *Mse*I-primer containing four selective nucleotides (either M-ACAA, M-GCGT, M-CTGG or M-CTGT). Pre-selective and selective primer pairs were chosen based on the result of an initial screening for polymorphism among a limited number of samples. Following amplification, an equal volume of formamide loading dye was added to the PCR products. After denaturation, the products were separated electrophoretically on 5% denaturating polyacrylamide gels, and bands were visualized using autoradiography. The SequaMark DNA size marker (Research Genetics), ranging in size from 50 to 500 bp, was used to determine fragment sizes.

Data analysis

For each accession, the DNA fingerprints were scored by visual inspection for presence (1) or absence (0) of specific AFLP-bands. Only distinct, major bands were scored. The resulting data matrices were analysed using the Treecon (Version 1.3b; Van de Peer and De Wachter 1994) or NTSYS-pc program (Version 2.10L; Rohlf 2000). Genetic similarities based on the Dice coefficient

Table 1 List of accessions utilized in this study, their codes and origin. [1] This group represents 13 accessions that could not be assigned to any of the described varieties of *V.* × *heilbornii* Badillo. [2] These seven accessions include two different groups of unidentified *Vasconcella* sp

Taxon[a]	Code	Origin[b]	Taxon[a]	Code	Origin[b]
C. papaya (6)	pap26	Vilcabamba	*J. spinosa* (3)	spin27	Zamora
	pap37	Zamora		spin28	Zamora
	pap38	Zamora		spin29	Zamora
	pap47	Catacocha	*J. digitata* (3)		
	pap48	Catacocha		dig30	Zamora
	pap57	Catacocha		dig35	Zamora
V. monoica (6)	mon31	Zamora		dig36	Zamora
	mon32	Zamora	*V. parviflora* (7)	parv40	Catacocha
	mon33	Zamora		parv41	Catacocha
	mon58	Valladolid		parv42	Catacocha
	mon60	Valladolid		parv43	Catacocha
	mon61	Valladolid		parv44	Catacocha
V. stipulata (15)	stip2	Loja		parv45	Catacocha
	stip7	Loja		parv46	Catacocha
	stip8	Loja	*V. weberbaueri* (6)	web3	Loja
	stip15	Chantaco		web4	Loja
	stip16	Chantaco		web5	Loja
	stip17	Chanaco		web6	Loja
	stip18	Chantaco		web9	Loja
	stip39	Loja		web10	Loja
	stip52	Catacocha	*V.* × *heilbornii* var.??[1] (13)	nm??11	Chantaco
	stip53	Catacocha		nm??12	Chantaco
	stip54	Catacocha		nm??13	Chantaco
	stip55	Catacocha		nm??14	Chantaco
	stip59	Catacocha		nm??71	Capur
	stip87	Loja BG		nm??74	Capur
	stip104	Loja BG		nm??78	Capur
V. × *heilbornii*. var. *chrysopetala* (6)	chrys19	Chuquiribamba		nm??79	Capur
	chrys75	Capur		nm??81	San Lucas
	chrys76	Capur		nm??88	Loja BG
	chrys83	San Lucas		nm??93	Loja BG
	chrys89	Loja BG		nm??95	Loja BG
	chrys102	Loja BG		nm??99	Loja BG
V. cundinamarcencis (8)	pub1	Loja	*V.* × *heilbornii* cv 'Babaco' (6)	bab77	Capur
	pub20	Chuquiribamba		bab73	Capur
	pub21	Chuquiribamba		bab84	San Lucas
	pub72	Capur		bab85	Loja
	pub80	Capur		bab98	Loja BG
	pub82	San Lucas		bab103	Loja BG
	pub90	Loja BG		carv22	Vilcabamba
	pub91	Loja BG	*Vasconcella* sp. [2] (7)	carv23	Vilcabamba
V. palandensis (7)	pal62	Palanda		carv24	Vilcabamba
	pal63	Palanda		carv25	Vilcabamba
	pal64	Palanda		carv34	Vilcabamba
	pal66	Palanda		carp65	Palanda
	pal68	Palanda		carp67	Palanda
	pal69	Palanda	*V. candicans* (2)	can50	Catacocha
	pal70	Palanda		can51	Catacocha

[a] For the correct and complete *Carica*, *Vasconcella* and *Jacaratia* nomenclature we refer to recent taxonomical revision published by Badillo (2000)

[b] All villages are located within Loja Province, southern Ecuador; BG = botanical garden

(Dice 1945), Jaccard's coefficient (Jaccard 1908) or the Simple Matching (SM) coefficient (Sokal and Michener 1958) were calculated using the SIMQUAL module of NTSYS-pc or the DISTANCE ESTIMATION option of Treecon. Correlations between distance matrices obtained from the different primer combinations were evaluated by a Mantel Test (Mantel 1967) using the MXCOMP module of NTSYS-pc to analyse the complementarity or redundancy of the information.

The generated similarity matrices were then analysed using the various clustering methods of NTSYS-pc (SAHN module) or Treecon (INFER TREE TOPOLOGY option): UPGMA (unweighted pair group method with arithmetic average; Sokal and Michener 1958), WPGMA (weighted pair group method; Sneath and Sokal 1973), complete linkage (Lance and Williams 1967) and single linkage (Lance and Williams 1967). In addition, the Neighbour-Joining (NJ) method (Saitou and Nei 1987), implemented in both Treecon and NTSYS-pc, was used to estimate dendrograms. All dendrograms were created with the TREE program of NTSYS-pc or the DRAW option of Treecon. The 'goodness of fit' of the clustering to the data matrix was determined by calculating the cophenetic correlation coefficient between the similarity matrix and the cophenetic matrix derived from the dendrogram, using the COPH and MXCOMP procedures of the NTSYS-pc (Rohlf 2000). Reliability of the dendrograms was tested by comparing dendrograms from different methods and by bootstrap analysis with 1,000 replications using Treecon. Some authors con-

sider that confidence limits obtained in bootstrap must be over 95% in order to consider the grouping of taxa at a branch to be statistically significant (Felsenstein 1985). Others use a lower limit (50% or higher) as indicating statistical support for the topology at a node (Highton 1993). In this study the grouping of taxa is considered as being statistically significant when both PCO and cluster analysis resulted in the same group, and bootstrap values reached the lower limit.

Additionally, a principal co-ordinate analysis (PCO; Gower 1966) based on the genetic similarity matrices was performed using the DCENTER and EIGEN algorithms of the NTSYS-pc software package (Rohlf 2000).

Results

AFLP analysis

A total of 65 *Eco*RI+2/*Mse*I+4 primer combinations were pre-screened for their ability to detect polymorphisms in five selected accessions representing one *C. papaya* genotype and four different *Vasconcella* species. Most selective primer combinations tested generated too many fragments for reliable scoring. Five primer pairs were selected (Table 2) based on the number of fragments amplified and the polymorphism rate observed. These primer pairs were applied to the complete set of accessions listed in Table 1.

When bands from all individuals were considered, the five primer combinations used revealed a total of 951 bands, ranging in length from 50 to 500 base pairs (Table 2). From these fragments, 512 unambiguous bands were scored. Only 16 bands were monomorphic across the complete germplasm set, resulting in 96.6% of the scored bands being polymorphic. Within the accessions of the genus *Vasconcella*, the number of scored bands generated by individual primer pairs ranged from 44 to 94. Out of the 330 markers scored, 35 were shared between all 83 *Vasconcella* accessions. As a result, the band polymorphism level within the genus *Vasconcella* was 89.1%.

Genetic similarity matrix and cluster analysis

The AFLP data were used to make pairwise comparisons of the genotypes based on both shared and unique amplification products to generate a similarity matrix using the Dice similarity coefficient. The correlation between the different genetic similarity matrices, obtained from the five separate primer pairs, was evaluated. The lowest correlation was observed between the similarity matrices generated with the primer pairs E-GA/M-GCGT and E-GT/M-ACAA (r = 0.60). This illustrates the fact that markers obtained from different sets of pre-amplification products provide complementary information. Although each primer combination individually could have given an approximation of the entire data set, sufficient differences existed between them. Therefore the data obtained from all five primer combinations was used in the analysis.

For dendrograms resulting from any of the different similarity matrices (Dice, Jacard's or Simple Matching) in combination with the various clustering methods (UPGMA, WPGMA, complete linkage or single linkage), the 'goodness of fit' of the clustering to the data matrix was evaluated. Therefore the co-phenetic correlation coefficient for each of the generated dendrograms was determined and compared (Table 3). Application of the UPGMA clustering technique gave the highest co-phenetic correlation scores, where $r > 0.9$ indicates a very good fit; $0.8 < r < 0.9$ indicates a good fit; and $r < 0.8$ indicates a poor fit. The use of Jaccard's as well as the Dice similarity coefficient resulted in very high and comparable correlation values, and these were consistently higher than the SM coefficient. Furthermore, the Dice similarity matrix derived from the entire data set (512 scored markers) was highly correlated (r = 0.98)

Table 3 Comparison of co-phenetic correlation values obtained from three similarity coefficients and four clustering methods employed for analysing the generated AFLP data

Method	Similarity coefficients		
	Dice	Jaccard's	SM
UPGMA	0.953	0.958	0.926
WPGMA	0.951	0.956	0.914
Complete linkage	0.924	0.932	0.901
Single linkage	0.915	0.913	0.879

Table 2 Selected primer combinations and polymorphism rates for AFLP analysis of the 95 accessions belonging to the genera *Carica* (one species), *Vasconcella* (eight species, one variety and one cultivar) and *Jacaratia* (two species)

Primer pairs	Total number of bands	Complete dataset (95 accessions)			Genus *Vasconcella* (83 accessions)		
		Scored bands	Polymorphic bands	Polymorphism (%)	Scored bands	Polymorphic bands	Polymorphism (%)
E-GA/M-ACAA	195	101	98	97.0	63	58	92.1
E-GT/M-ACAA	127	72	69	95.8	44	38	86.4
E-GA/M-GCGT	208	95	92	96.8	65	59	90.8
E-GA/M-CTGT	241	155	152	98.1	94	85	90.4
E-CG/M-CTGG	180	89	85	95.5	64	55	85.9
Total	951	512	496		330	295	
Mean	190	102	99	96.6	66	59	89.1

Table 4 Genetic similarity values between (upper triangle) and within (diagonal) the different groups of accessions of the genera *Jacaratia*, *Carica* and *Vasconcella*, based on all pairwise similarities between individuals

Group	*spin* (*n*=3)	*dig* (*n*=3)	*pap* (*n*=6)	*pub* (*n*=8)	*chrys* (*n*=6)	*bab* (*n*=6)	*nm*? (*n*=13)	*stip* (*n*=15)	*can* (*n*=2)	*mon* (*n*=6)	*pal* (*n*=7)	*parv* (*n*=7)	*web* (*n*=6)	'carv' (*n*=5)	'carp' (*n*=2)
J. spinosa (*spin*)	0.99	0.33	0.18	0.23	0.25	0.25	0.25	0.32	0.30	0.23	0.24	0.23	0.24	0.22	0.22
J. digitata (*dig*)		0.99	0.21	0.22	0.21	0.22	0.21	0.25	0.27	0.20	0.21	0.24	0.21	0.21	0.20
C. papaya (*pap*)			0.99	0.21	0.20	0.21	0.21	0.23	0.25	0.25	0.25	0.24	0.24	0.25	0.23
V. cundinamarcensis (*pub*)				0.90	0.47	0.54	0.69	0.49	0.44	0.55	0.61	0.42	0.40	0.59	0.59
V. × *heilb.* var. *chrys.* (*chrys*)					0.94	0.78	0.74	0.81	0.51	0.41	0.44	0.59	0.64	0.43	0.45
V. × *heilb.* cv 'Babaco' (*bab*)						0.78	0.70	0.69	0.49	0.45	0.48	0.57	0.59	0.48	0.48
V. × *heilb.* var.?? (*nm*?)							0.78	0.68	0.48	0.48	0.51	0.53	0.54	0.50	0.51
V. stipulata (*stip*)								0.94	0.58	0.46	0.47	0.63	0.60	0.47	0.51
V. candicans (*can*)									0.94	0.51	0.44	0.53	0.46	0.45	0.47
V. monoica (*mon*)										0.99	0.56	0.47	0.40	0.65	0.71
V. palandensis (*pal*)											0.95	0.47	0.42	0.60	0.65
V. parviflora (*parv*)												0.94	0.65	0.44	0.51
V. weberbaueri (*web*)													0.97	0.39	0.42
'carv'														0.97	0.70
'carp'															0.95

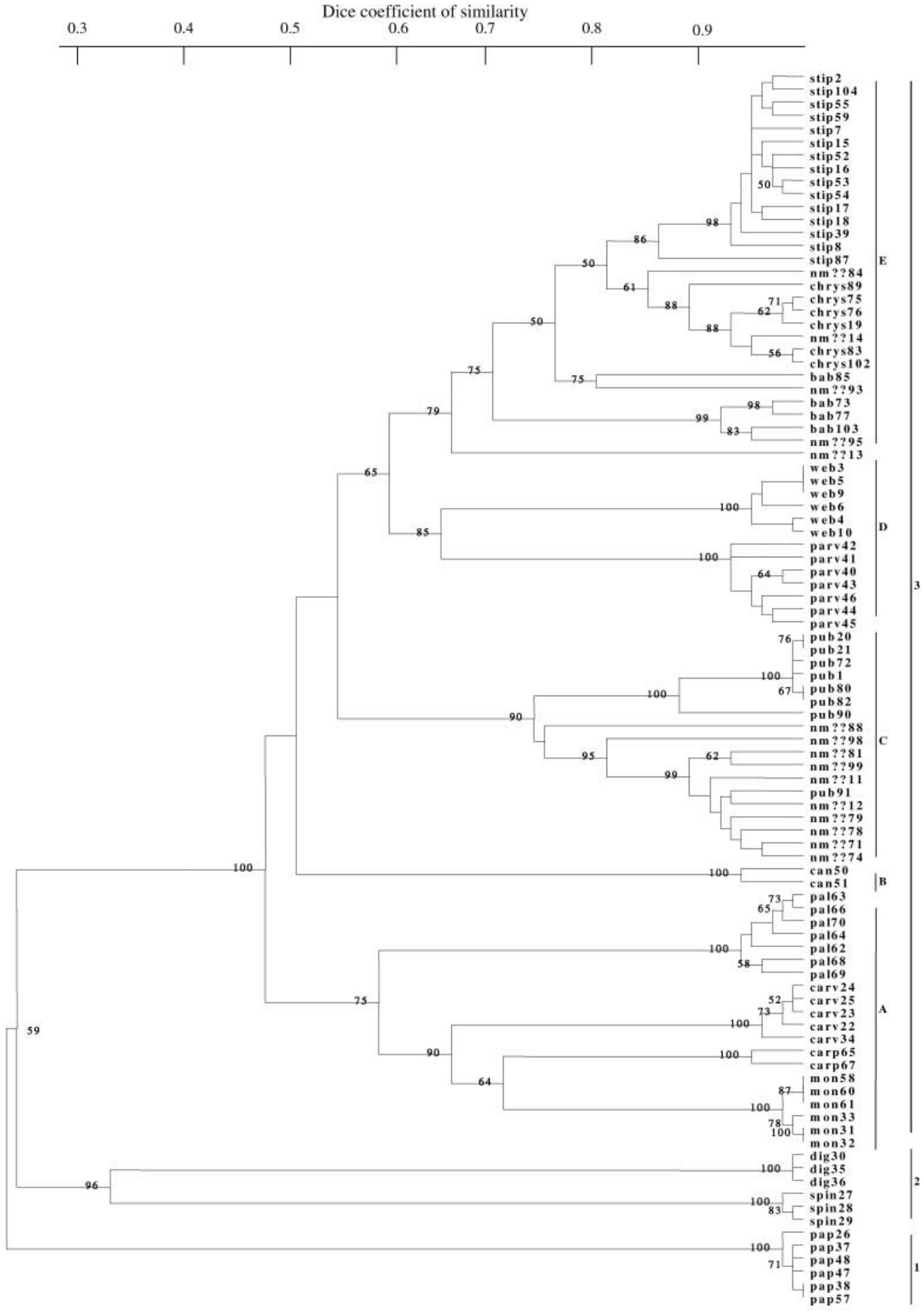
with the one calculated using Jaccard's coefficient. Comparing all the dendrograms produced, and selecting the cut off points depending on the similarity coefficient and the clustering method employed, the three genera clearly clustered separately and the same five main clusters within the genus *Vasconcella* could be identified. Also, the species relationships obtained from the NJ-method were similar to those obtained using UPGMA (Fig. 1), except for minor differences in branch lengths and a topological rearrangement where cluster 3A clustered within cluster 3C (data not shown).

Figure 1 shows the dendrogram generated using the Dice coefficient and the UPGMA clustering method. Cluster 1 contains only *C. papaya* genotypes, clearly separated from cluster 2 containing both *Jacaratia* sp. and cluster 3, containing all species of *Vasconcella*. Within the latter well-supported clade (bootstrap value 100%), there are five subclusters that can be identified at the 60% similarity level. Cluster 3A contains two species (*Vasconcella monoica* and *V. palandensis*) and two groups of unidentified genotypes ('carv' and 'carp' group of accessions). *V. candicans* is separated from all other *Vasconcella* spp. in the monophyletic cluster 3B. Within cluster 3C, a discrete group was formed with one of the putative parent species of *V.* × *heilbornii*, *V. cundinamarcensis* (formerly *C. pubescens*). The rest of this cluster consists of unidentified specimens of *V.* × *heilbornii*, exhibiting more variation. Cluster 3D consists of *V. weberbaueri* and *Vasconcella parviflora*, supported by a bootstrap value of 85%. Cluster 3E contains the other species believed to be parent of the *V.* × *heilbornii* hybrid: *V. stipulata*. All *V. stipulata* specimens are grouped in a distinct subcluster, supported by a bootstrap value of 86%. Another subgroup, consisting of predominantly *V.* × *heilbornii* var. *chrysopetala* accessions is linked to the *V. stipulata* subgroup at about 80% similarity. A final subgroup, enclosing more unidentified *V.* × *heilbornii* genotypes, intermingled with 'Babaco'-accessions, makes this cluster complete at a similarity level of 75% or lower. Within cluster 3E, the accessions representing *V. stipulata* also appear as a compact group, with greater similarity among its members than within the other subclusters of cluster 3E.

Genetic similarities among all *Vasconcella* taxa ranged from 0.39 (between the unidentified 'carv' genotypes and *V. weberbaueri)* to 0.81 (between *V. stipulata* and *V.* × *heilbornii* var. *chrysopetala*), with an average similarity of 0.54 (Table 4). When all genera are considered, the mean genetic similarity value, based on the pairwise comparison between all species from *Jacaratia* and *Vasconcella* (0.24), was slightly higher than the average similarity value derived from the comparison of all accessions of *Carica* and *Vasconcella* (0.23). Within species, all groups displayed very high genetic similarity values, except for the group of unidentified varieties of *V.* × *heilbornii* (0.78) and the 'Babaco' group (0.78).

Principal co-ordinate analysis

Principal co-ordinate analysis (PCO) based on the Dice genetic similarity matrix was used to visualize the genetic relationships among investigated taxa. The first three eigenvectors accounted for 45% of the variation observed and were able to separate the different groups. PCO separations support the results obtained with the cluster analysis. The PCO-plot generated with the first two eigenvectors (Fig. 2) clearly distinguishes the genera *Carica*, *Jacaratia* and *Vasconcella* from each other. Remarkably, the *C. papaya* genotypes lie proximal to *Jacaratia* taxa but significantly distanced from the *Vasconcella* taxa. The five principal subclusters of the genus *Vasconcella* identified with the UPGMA-analysis are recognised in the PCO-plot as well. Furthermore, the PCO-plot evidently illustrates the genetic relationships between *V. stipulata*, *V. cundinamarcensis* and their putative hybrid progeny. The suggested parent species are placed at the outskirts of a group (Fig. 2, cluster I and


Dice coefficient of similarity
0.3
0.4
0.5
0.6
0.7
0.8
0.9
stip2
stip104
stip55
stip59
stip7
stip15
stip52
stip16
stip53
stip54
stip17
stip18
stip39
stip8
stip87
nm??84
chrys89
chrys75
chrys76
chrys19
nm??14
chrys83
chrys102
bab85
nm??93
bab73
bab77
bab103
nm??95
nm??13
web3
web5
web9
web6
web4
web10
parv42
parv41
parv40
parv43
parv46
parv44
parv45
pub20
pub21
pub72
pub1
pub80
pub82
pub90
nm??88
nm??98
nm??81
nm??99
nm??11
pub91
nm??12
nm??79
nm??78
nm??71
nm??74
can50
can51
pal63
pal66
pal70
pal64
pal62
pal68
pal69
carv24
carv25
carv23
carv22
carv34
carp65
carp67
mon58
mon60
mon61
mon33
mon31
mon32
dig30
dig35
dig36
spin27
spin28
spin29
pap26
pap37
pap48
pap47
pap38
pap57
E
D
C
B
A
3
2
1

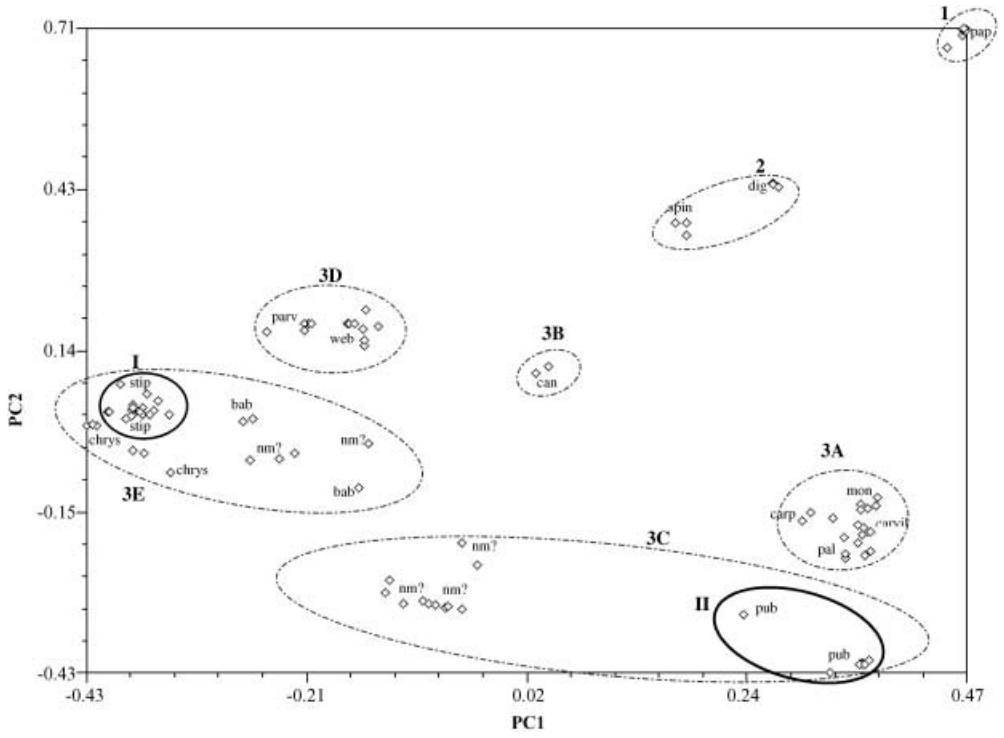


Fig. 2 Principal co-ordinate plot of Caricaceae genotypes for the first and second principal co-ordinates estimated with 512 AFLP markers, using the Dice similarity matrix. Not all accessions plotted are labelled, instead the common codename is placed next to the accessions to indicate their identity. The *numbers* on the dotted circles refer to the corresponding cluster in the UPGMA-dendrogram (Fig. 1), while *full circles* I and II enclose *V. stipulata* and *V. cundinamarcensis* respectively, the putative parents of *V.* × *heilbornii*

II), while all *V.* × *heilbornii* accessions are to be found between the two parent species. In agreement with the cluster analysis, the *V.* × *heilbornii* var. *chrysopetala* and the 'Babaco' accessions are positioned closer to *V. stipulata*. The loose grouping of the *V.* × *heilbornii* accessions, illustrated in the PCO-plot, reflects the diversity present in this set of accessions which was also revealed with the cluster analysis.

Discussion

Phenetic relationships

The *C. papaya* accessions were shown to be very distinct from *Vasconcella*. This is illustrated in the cluster analysis and supported by the bootstrap analysis (59%; Fig. 1, cluster 1), as well as in the PCO analysis (Fig. 2). In a recent study using RAPD and isozyme markers, Jobin-Decor et al. (1997) found about the same level of similarity (30%) between *Vasconcella* and *C. papaya*. Moreover, both *J. digitata* and *J. spinosa* displayed a higher level of similarity to *Vasconcella* than to *C. papaya*, indicating that these species may even be closer related to this group of highland-papayas. This has also been noticed for *Jacaratia mexicana* by Aradhya et al. (1999) using RFLP to study the variation in an cpDNA intergenic spacer region. Previous research on ovary morphology (Badillo 1993) and interspecific hybridisation barriers (e.g. Manshardt and Wenslaff 1989) also indicate that papaya is only distantly related to *Vasconcella*. Results obtained in this AFLP analysis also support the suggestion that *C. papaya* must have separated relatively early from the South American *Vasconcella* species and evolved in isolation more northwards on the American continent (Aradhya et al. 1999). Our data further confirms the rehabilitation of *Vasconcella* as a genus by Badillo (2000). Only limited variation was detected within the *C. papaya* group (Fig. 1, cluster 1). The small number of *C. papaya* accessions in this study prevent any meaningful discussion about the intraspecific variation, but the high level of similarity among these genotypes is in accord with other studies (Stiles et al. 1993; Morshidi 1998).

Within the species complex represented by the compact *Vasconcella* cluster 3A, *V. monoica*, *V. palandensis* and the two groups of unidentified species, 'carv'- and 'carp'-accessions, all were clearly delimited and supported by maximum bootstrap values (Fig. 1). The general morphology of the 'carv' accessions, particularly of the fruit, is very similar to that of the recently discovered *V. palandensis*. However, significant differences in leaf morphology were recorded during collection. AFLP analysis confirmed that, while morphologically similar, the collected 'carv' genotypes were genetically distinct from *V. palandensis*. More accessions of these unknown genotypes need to be located and investigated to deter-

◀ **Fig. 1** Dendrogram showing the genetic relationships among 95 Caricaceae accessions based on AFLP data, using the Dice coefficient of similarity and UPGMA clustering. The seven clusters are discussed in the text. Accessions are labelled according to the codes listed in Table 1. Numbers shown at the different nodes indicate bootstrap confidence values (1,000 bootstrap replicates). Nodes without numbers had bootstrap values of less than 50

mine if these groups of plants represent a new species. Our results support this since each of these groups has distinct morphological characteristics and intra-/interspecific genetic distances that are comparable with those seen among related species.

Although only two *V. candicans* accessions were included in this study, it was clear from the dendrogram in Fig. 1 and the PCO in Fig. 2 that these genotypes did not associate with any other species complex within the genus *Vasconcella*. These results corroborate the morphological observations: *V. candicans* is the only true arborescent species in our sample set, bearing heart-shaped leaves and having multiple lateral branches.

Within subcluster 3C (Fig. 1) all accessions except one (pub91), determined as *V. cundinamarcensis*, grouped together. The collection of unknown *V.* × *heilbornii* accessions placed in this group displayed almost twice as much variation as among the *V. cundinamarcensis* genotypes. A similar organisation of the genetic diversity was noticed within cluster 3E. *V. stipulata* and *V.* × *heilbornii* var. *chrysopetala* are the only species that clearly could be distinguished in discrete subclusters. The tight clustering of accessions nm??14 and nm??84 to the moderately supported subcluster of *V.* × *heilbornii* var. *chrysopetala* accessions (bootstrap value of 61%) suggests that these genotypes probably can be identified as *V.* × *heilbornii* var. *chrysopetala*. The other accessions in this group, determined as Babaco's or unknown varieties of *V.* × *heilbornii* clustered intermingled, and at a lower level of similarity to both *V. stipulata* and *V.* × *heilbornii* var. *chrysopetala*. In contrast to our results, Jobin-Décor et al. (1997) with both RAPD and isozymes found a similarity level of 85% between *V. stipulata* and *V. cundinamarcensis*. Considering that the species-characteristic stipules were not found on their "*V. stipulata*", our results suggest that their accession probably represents a *V.* × *heilbornii* genotype.

Finally, cluster 3D included the only two species of the genus *Vasconcella* producing small fruits: *V. parviflora* and *V. weberbaueri*. The accessions of both species clustered into a maximum-supported species-specific subcluster (Fig. 1, bootstrap 85%; Fig. 2).

V. stipulata, *V. cundinamarcensis* and their putative hybrid, *V.* × *heilbornii*

Horovitz and Jimenez (1967) concluded that *V.* × *heilbornii* is a natural hybrid between *V. stipulata* and *V. cundinamarcensis*. In the AFLP-analysis presented here, all *V.* × *heilbornii* accessions clustered together with either one of the putative parent species (Fig. 1, cluster 3C and 3E, and Fig. 2). It appears that introgression in both directions divided the hybrids into two groups, each displaying a higher level of similarity to one of the putative parent species. The described varieties *V.* × *heilbornii* var. *chrysopetala* and *V.* × *heilbornii* cv 'Babaco' cluster together with *V. stipulata*, suggesting *V. stipulata* as the backcross parent. Results are in agreement with the observations of Horovitz and Jimenez (1967) who suggested introgression of *V. cundinamarcensis* (formerly *C. pubescens*) into *V. stipulata*. They also reported that the experimental pollination of *V. stipulata* with *V. cundinamarcensis* pollen sometimes resulted in viable seeds. Considering the allogamous nature of *V. stipulata* and *V. cundinamarcensis*, together with their sympatric distribution in some regions of Ecuador (Badillo 1983), introgression of one species into another is plausible. Further introgression may have blurred the distinction among the described *V.* × *heilbornii* varieties. This is illustrated in our analysis by the anomalous clustering of pub91. While this accession exhibited a *V. cundinamarcensis* phenotype, it clustered among the well-supported group *V.* × *heilbornii* (bootstrap 95%).

The high level of intraspecific diversity (Table 4) found among the unidentified *V.* × *heilbornii* accessions, reflected in both the dendrogram (Fig. 1) and the PCO (Fig. 2), was expected since they represented a morphologically diverse group. For 'Babaco' accessions the high degree of intra-varietal diversity is remarkable. Since these plants produce parthenocarpic fruit and therefore are propagated in a vegetative way (Badillo 1993; Kempler and Kabaluk 1996) only limited genetic variation within this group was expected. However, this unexpected diversity has also been observed by Aradhya et al. (1999) even though they included only two Babaco genotypes, obtained from the same source. They suggested that hybridisation may have occurred more than once and involved reciprocal crosses between the two parent species (Aradhya et al. 1999). Other explanations include the incorporation of volunteer plants originating from seeds produced by unrecognized sexual recombination or the accumulation of somatic mutations through vegetative propagation. This intra-varietal variation does not support the proposed status of the Babaco, *V.* × *heilbornii* cv 'Babaco', as a cultivar (Badillo 2000).

Acknowledgements The authors thank Ir. Xavier Scheldeman for his continued support during our stay in Ecuador, for valuable advice and useful discussions. This research project was funded by the Flemish Fund for Scientic Research (FWO-Vlaanderen Project no. 3G005100). Appreciation for assistance during sample collection in Ecuador is extended to Ing. José Paramon Romero Motochi, presently working at the Fundación San Francisco, Ecuador. The whole staff of the AFLP-group of the Laboratory of Genetics, Faculty of Sciences, Ghent University, is very much appreciated for their skilful assistance. The experiments comply with current Belgian laws and regulations.

References

Aggarwal RK, Brar DS, Nandi S, Huang N, Khush GS (1999) Phylogenetic relationships among *Oryza* species revealed by AFLP markers. Theor Appl Genet 98:1320–1328

Aradhya MK, Manshardt RM, Zee F, Morden CW (1999) A phylogenetic analysis of the genus *Carica* L. (Caricaceae) based on restriction fragment length variation in a cpDNA intergenic spacer region. Genet Res Crop Evol 46:579–586

Badillo VM (1971) Monografia de la familia Caricaceae. Publicada por la Asociacion de Profesores, Univ Centr Venezuela, Maracay, Venezuela

Badillo VM (1983) Caricaceae. In: Harling G, Sparre B (eds) Flora of Ecuador. Balogh Scientific Books, Illinois, pp 27–47
Badillo VM (1993) Caricaceae. Segundo esquema. Rev Fac Agron Univ Centr Venezuela 43:1–111
Badillo VM (1997) Neotipificación de *Carica pubescens* Lennén et Koch y de *Carica quercifolia* (St. Hil.) Hieron y nuevos registros de la familia para Ecuador. Ernstia 6:201–205
Badillo VM (2000) *Carica* L. vs *Vasconcella* St. Hil. (Caricaceae): con la rehabilitación de este último. Ernstia 10:74–79
Badillo VM, Van den Eynden V, Van Damme P (2000) *Carica palandensis* (Caricaceae), a new species from Ecuador. Novon 10:4–6
Barrett BA, Kidwell KK (1998) AFLP-based genetic diversity assessment among wheat cultivars. Crop Sci 38:1261–1271
Dice LR (1945) Measures of the amount of ecologic association between species. Ecology 26:297–302
Drew RA, O'Brien CM, Magdalita PM (1998) Development of interspecific *Carica* hybrids. Acta Hort 461:285–292
FAO (2001) Statistical databases of the Food and Agriculture Organization of the United Nations. http://apps.fao.org
Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791
Gower JC (1966) Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53:325–338
Highton R (1993) The relationship between the number of loci and the statistical support for the topology of UPGMA trees obtained from genetic distance data. Mol Phylog Evol 2:337–343
Horovitz S, Jimenez H (1967) Cruziamentos interespecificos e intergenericos en Caricaceas y sus implicaciones fitotecnicas. Agron Trop 17:323–344
Jaccard P (1908) Nouvelles recherches sur la distribution florale. Bull Soc Vaud Sci Nat 44:223–270
Jiménez Y, Romero J, Scheldeman X (1999) Colección, caracterización y descripción de *Carica* ×*heilbornii* nm. *pentagona* B.; *Carica pubescens* (A.DC.) Solms-Laub y *Carica stipulata* B. en la provincia de Loja. Revista de Difusión Técnica y Cientifica de la facultad de Ciencias Agricolas de la Universidad Nacional de Loja 29:43–54
Jobin-Decor MP, Graham GC, Henry RJ, Drew RA (1997) RAPD and isozyme analysis of genetic relationships between *Carica papaya* and wild relatives. Genet Res Crop Evol 44:471–477
Kempler C, Kabaluk T (1996) Babaco (*Carica pentagona* Heilb.): a possible crop for the greenhouse. Hortscience 31:785–788
Lance GN, Williams WT (1967) A general theory of classificatory sorting strategies. 1. Hierarchical systems. Computer J 9:373–380
Madrigal L, Ortiz AN, Cooke RD, Fernandez RH (1980) The dependence of crude papain yields on different collection ('Tapping') procedures for papaya latex. J Sci Food Agric 31:279–285
Manshardt RM, Drew RA (1998) Biotechnology of papaya. Acta Hort 461:65–73
Manshardt RM, Wenslaff TF (1989) Inter-specific hybridization of papaya with other species. J Am Soc Hort Sci 114:689–694
Mantel NA (1967) The detection of disease clustering and a generalized regression approach. Cancer Res 27:209–220
Morshidi M (1998) Genetic control of isozymes in *Carica papaya* L. Theor Appl Genet 103:89–94
National Research Council (1989) Highland papayas. In: Ruskin FR (ed) Lost crops of the Incas: little-known plants of the Andes with promise for worldwide cultivation. National Academy Press, Washington D.C. pp 252–261
Rohlf FJ (2000) NTSYS-pc. Numerical taxonomy and multivariate analysis system. Version 2.1. Exeter Software, Setauket, New York, USA
Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425
Sneath PH, Sokal RR (1973) Numerical taxonomy. WH Freeman and Co, San Francisco
Sokal RR, Michener CD (1958) A statistical method for evaluating systematic relationships. Univ Kansas Sci Bull 38:1409–1438
Stiles JI, Lemme C, Sondur S, Morshidi MB, Manshardt RM (1993) Using randomly amplified polymorphic DNA for evaluating genetic relationships among papaya cultivars. Theor Appl Genet 85:697–701
Van de Peer Y, De Wachter R (1994) TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. Comput Appl Biosci 10:569–570
Vos P, Hogers R, Bleeker M, Reijans M, Van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP a new technique for DNA fingerprinting. Nucleic Acids Res 23:319–32

Note added in proof: During the revision of the manuskript Badillo published a corrective note for the generic name of the genus *Vasconcella*. He sugggests it should be *Vasconcellea* instead. The reference in which he declares the change is the following: Badillo VM (2001) Nota correctiva *Vasconcellea* St. Hill. y no *Vasconcella* (Caricaceae). Ernstia 11:75–76