The origin of domesticated pejibaye (*Bactris gasipaes* var. *gasipaes*) – a research proposal

Charles R. Clement, INPA, Manaus, Amazonas, Brasil – cclement@vivax.com.br, crc@internext.com.br, cclement@inpa.gov.br

The origin of domesticated pejibaye (*Bactris gasipaes* var. *gasipaes*) from wild populations (var. *chichagui*) remains a matter of speculation, with three hypotheses currently under consideration: a single domestication event in southwestern Amazonia (Clement 1995) or northwestern Colombia (Morcote-Rios & Bernal 2001); or multiple domestication events in the distribution of the wild populations (Mora-Urpí 1993). Clement defended the parsimonious option, as most domesticates have been shown to have arisen from single events (Blumler 1992), and SW Amazonia because of Huber's (1904) proposal. Morcote-Rios & Bernal worked from archaeological information and the distribution of type 3 *chichagui*. Mora-Urpí defended multiple events given the wide distribution of wild pejibaye populations, and frequent phenotypic similarities between wild and adjacent domestic populations.

This essay examines the current state-of-the-art and outlines a proposal to attempt to identify the geographical origin(s) of the domestication event(s) using molecular genetic analyses. Criticism, comments and suggestions are requested. Note that archaeology is not included yet.

Bactris gasipaes

The recent revision of *Bactris* (Henderson 2000) gathered most wild populations into var. *chichagui*, and all domesticated populations and landraces into var. *gasipaes*. Within var. *chichagui*, Henderson identified three types, without describing their distribution or attributing synonyms. Clement & Evandro Ferreira propose (ms in prep.) the following approximate distributions (Figure 1) and attributions of synonymy (Table 1; full citations of synonyms in Henderson 2000).

Table 1. Probable synonymy of wild types of pejibaye (*Bactris gasipaes* var. *chichagui*) found in northern South America.

Type 1	Type 2	Type 3*
Guilielma mattogrossensis	*Guilielma macana*	*Bactris speciosa* v. *chichagui*
Martinezia ciliata ?	*Bactris caribaea*	*Martinezia ciliata* ?
Guilielma microcarpa ?		*Guilielma microcarpa* ?
		Guilielma insignis ?

* Numerous populations have been cited by Mora-Urpí (1999) and Couvreur et al. (2005).

The proposed transfer of *M. ciliata* from var. *gasipaes* to var. *chichagui* is worth comment. Andrew Henderson & Evandro Ferreira visited the type region in Peru, and found both type 1 and type 3 fruits (E. Ferreira, pers. com., 2005). These observations suggest that the original Ruiz & Pavon description may have included both types. Also, since Bernal's (1989) proposal was designed to conserve *B. gasipaes* before the wild populations were reorganized within var. *chichagui* by Henderson (2000), it seems reasonable to place *M. ciliata* within var. *chichagui* now.

Recent work in Brazilian Amazonia identified types 1 and 3 between the Purus and Madeira Rivers, allowing the suspicion that Huber (1904) may have considered these the same species when describing *G. microcarpa* along the Purus River. The material that Huber saw near Pucallpa, Peru, was almost certainly type 3 also, based on fruit data in Clement et al. (1989). Hence, both early botanists may have had wider species concepts than originally thought, without, however, realizing the synonymy with the original *B. gasipaes*, although Huber hypothesized that a cross between *G. microcarpa* and *G. insignis* may explain the origin of *B. gasipaes*.

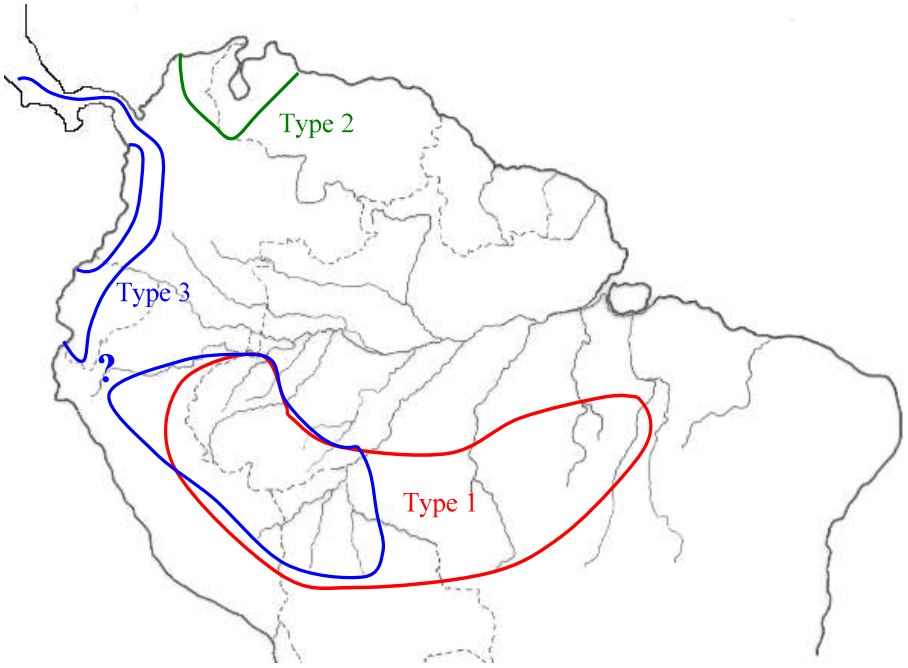

Figure 1. Approximate distributions of the three types of *Bactris gasipaes* Kunth var. *chichagui* (H. Karsten) Henderson. Note absence of *chichagui* in the Choco (Rodrigo Bernal, pers. com., 2006).

It is also worth mentioning the possible synonymy of *G. insignis* with type 3. Various authors have mentioned *G. insignis* as a possible error because it has never been recollected. Saldías-Paz' (1993) work near Santa Cruz, Bolivia, offers a possible explanation. Although working further south than the Beni type location, Saldías-Paz observed numerous inter-grading fruit types, from very small type 1, through type 3 to very small cultivated fruit (probably domesticated). Since type 3 is the most variable of Henderson's var. *chichagui* it is possible, and geographically likely, that *G. insignis* may be a large type 3 also. Mora-Urpí (1993, 1999) suspects that *G. insignis* may have given rise to the Tembe cultivated populations and contributed to the Pará landrace (see genetic evidence below).

Morphological analysis with respect to origins

Numerous studies by agronomists and geneticists analyzed the morphological variability and relationships among domesticated landraces of pejibaye during the 1980s, but only when the botanists decided to take on this spiny issue did related species get taken adequately into account. Sanders' (1991) cladistic analysis confirmed that *Bactris* is only monophyletic if *Guilielma* is included within it, resolving a long standing dispute about the validity of Martius' proposal. This analysis also showed the very close relationship between *B. gasipaes*, *B. dahlgreniana* (*G. microcarpa*) and *G. macana*.

Henderson (1995) and Henderson et al. (1995) were the first to join the wild species of *Guilielma* into a single entity. They recognized the synonymy of *B. caribaea* and *G. microcarpa* with *B. macana*, thus forming the basis for Henderson's later creation of var. *chichagui*.

Ferreira (1999) expanded the cladistic analysis of the species closest to *B. gasipaes*. He confirmed Sanders' (1991) conclusion about the relationship between *B. gasipaes*, *B. dahlgreniana* and *G. macana*, and observed that all cultivated pejibaye populations have seed shape and pore positioning similar to *B. dahlgreniana*. This is the first morphological evidence for a southern origin of the domesticated landraces. Remember, however, that *B. dahlgreniana* may represent a mixture of types 1 and 3 *chichagui*, in which case a western origin is also possible. During the preparation for this proposal, Evandro Ferreira (pers. com., 2006) suggested that this seed morphological evidence may

best be exemplified by type 3, rather than type 1, confirming the possibility of a southwestern-western Amazonia to Pacific NW South American to southern Mesoamerican origin (e.g., Morcote-Rios & Bernal 2001; Mora-Urpí 1993, 1999).

Couvreur et al. (2005) were the first to closely examine the morphological (and genetic, see below) relationships among wild and cultivated pejibaye after Henderson's revision, although on a limited geographic scale. They found type 3 *chichagui* along the Pacific coastal plain of Ecuador, including in southern, rather mesic environments (which is a range expansion), and in northern, more humid environments, but not in the super-humid Chocó of extreme northwestern Ecuador. They found good evidence for introgression among wild and cultivated types based on fruit size.

Molecular genetic analysis with respect to origins

Over the last decade, molecular markers have been brought to bear on the question of the origin of the domesticated landraces. The first study was by Rojas-Vargas et al. (1999), who found a dichotomy in their allozyme dendrogram (Figure 2). This dichotomy recurs in future studies.

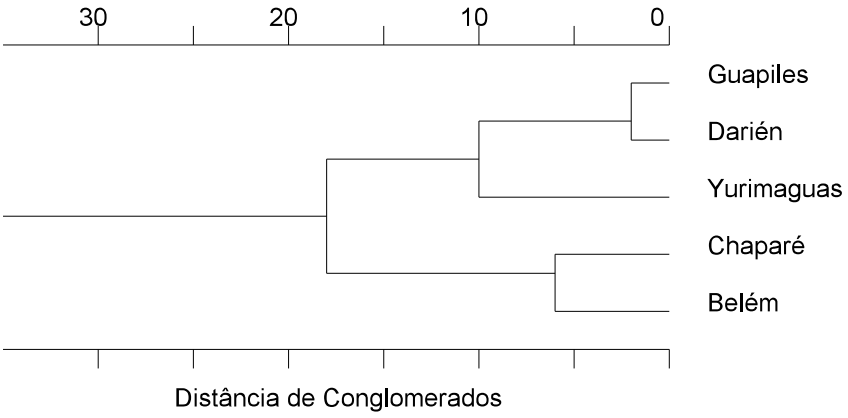

Figure 2. Phenetic distances among five populations (n = 5 plants each) analyzed with 10 enzyme systems by Rojas-Vargas et al. (1999). Some of the populations can be attributed to landraces: Belém to Pará; Yurimaguas to Pampa Hermosa; Guapiles to Utilis; Darién to Tuíra; Chaparé is close to Tembe in Bolivia.

Clement et al. (1997) observed that the spineless Central American Guatuso landrace was less variable than the Amazonian Pampa Hermosa and Putumayo landraces (Table 1), and contained a subset of Amazonian allozyme alleles with only two private alleles. They suggested that Guatuso was derived from Amazonia because of this, lower Ho, fewer polymorphic loci etc.

Table 1. Allozyme polymorphism at 16 putative loci in 9 enzyme systems in the Benjamin Constant (BC - Putumayo landrace), San Carlos (SC - Guatuso landrace) and Yurimaguas (Yu - Pampa Hermosa landrace) populations of var. *gasipaes* grown in Hawaii (Clement et al. 1997).

	Number of	Mean±SE Alleles /	% Loci	Mean±SE Heterozygosity	
Population	Alleles	Locus	Polymorphic[1]	Observed	H-W Exp.
BC	34	2.13±0.27	56.2	0.066±0.020	0.081±0.026
SC	25	1.56±0.18	43.7	0.051±0.023	0.149±0.056
Yu	33	2.06±0.23	68.7	0.141±0.035	0.191±0.047
over all	38	2.37±0.29	68.7	0.086	0.140

[1] A locus was considered polymorphic if the frequency of the most common allele did not exceed 0.99.

Rodrigues et al. (2004a) used RAPD markers to validate a wide set of landraces, and concluded that there is only one landrace in Central America and confirmed previous studies about the Solimões landrace (Figure 3, after fusions). They included samples of type 1 (Rio Branco, Acre) and type 3 (Benjamin Constant, Amazonas) in their study. Note first that the dendrogram has the same dichotomy as that of Rojas-Vargas et al. (1999). Note also how the *chichagui* samples cluster. Based on this dendrogram, they proposed a single domestication of pejibaye in southwestern Amazonia, with two expansions out of the region: one to the northeast down the Madeira River and throughout central and eastern Brazilian Amazonia; the other to the northwest down the Ucayali river, throughout western Amazonia, over the Andes into the Pacific coast of Ecuador and Colombia, and up into Mesoamerica.

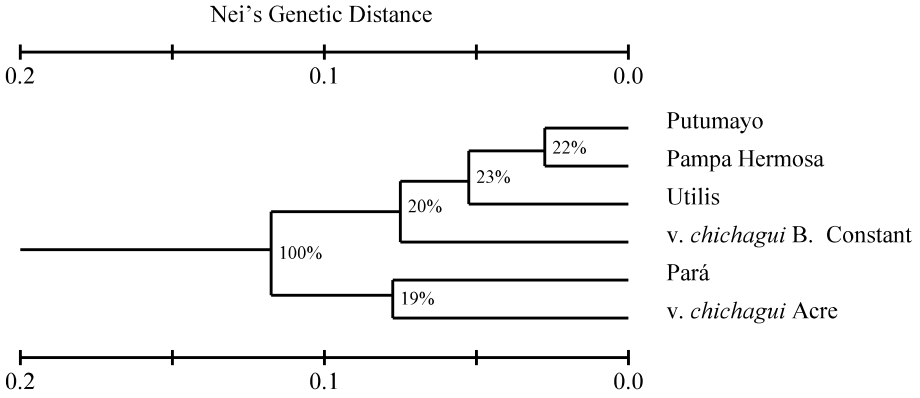

Figure 3. UPMGA dendrogram of Nei's genetic distances among four validated landraces and two populations of var. *chichagui* (Rodrigues et al. 2004a). B. Constant = type 3; Acre = type 1.

Rodrigues et al. (2004a) based their origin proposal on the distribution of heterozygosity and polymorphism between Amazonia and Central America (Table 2). Note that Central American pejibaye has 16% less heterozygosity than Amazonian landraces and 13% less polymorphism. Additionally, the less derived, i.e., more primitive, landraces are in the southwest (Tembe, Juruá) and east (Figure 4).

Table 2. Heterozygosity estimates and percentage of polymorphism of pejibaye (*Bactris gasipaes* var. *gasipaes*) landraces in Amazonia and Central America, and of populations of var. *chichagui* (Benjamin Constant and Acre) in Amazonia obtained from 113 RAPD markers (97 of which polymorphic) (Rodrigues et al. 2004a).

			% Polymorphism	
Variety, Region, Landrace	Sample (n)	Heterozygosity	95%	99%
var. *chichagui*	30	0.27	74.3	80.5
- type 1 – Acre[1]	15	0.22	60.2	67.2
- type 3 – Benjamin Constant[2]	15	0.22	59.3	68.1
var. *gasipaes* - Amazonia	133	0.30	83.0	86.0
- Pará	40	0.24	66.4	75.2
- Solimões	30	0.30	76.1	82.3
- Putumayo	33	0.27	73.4	77.0
- Pampa Hermosa	30	0.26	72.6	75.2
var. *gasipaes* - Central America	87	0.25	66.4	74.3
- Tuíra	30	0.22	62.0	64.6
- Utilis	30	0.24	62.8	64.6
- Guatuso	27	0.23	63.8	67.2
Overall	250	0.31	84.9	89.4

[1] type 1 - see Clement *et al*. (1989) for morphological details.
[2] type 3 - see Clement *et al*. (1999) for morphological details.

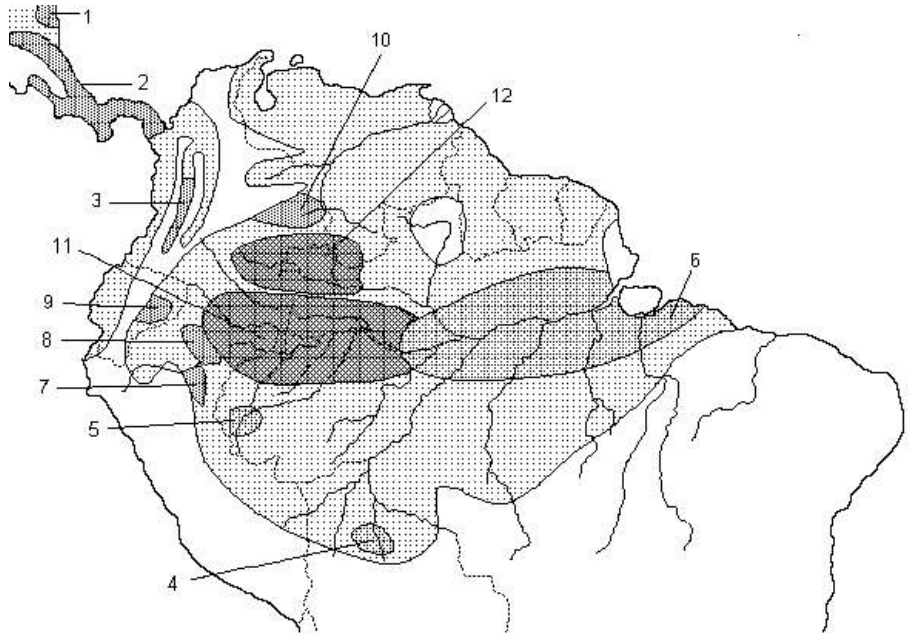

Figure 4. Approximate distribution of *B. gasipaes* var. *gasipaes* (light shading) in the lowland Neotropics, with the approximate distribution of valid (defined by molecular characterization and morphometric data) and still to be validated landraces (see Figure 3 above). Central America and northwestern South America landraces - 1. Rama (microcarpa), 2. Utilis (mesocarpa; now including Guatuso and Tuíra), 3. Cauca (mesocarpa). Amazonian landraces - microcarpa - 4. Tembé, 5. Juruá, 6. Pará; mesocarpa - 7. Pampa Hermosa, 8. Tigre, 9. Pastaza, 10. Inirida; macrocarpa - 11. Putumayo (now including Solimões), 12. Vaupés (Rodrigues et al. 2004a, modified from Mora-Urpí et al. 1997).

Silva (2004) used the same RAPD markers to expand on Rodrigues et al.'s study, adding additional landraces and two new var. *chichagui* populations (Figure 5). Unfortunately, both *chichagui* populations had insufficient DNA for a full analysis, due to poor seed germination, but did allow for a preliminary assessment with fewer markers. Note that the dichotomy in the top of the dendrogram (above Pará) is similar to that of Rojas-Vargas et al. and Rodrigues et al., and that all western Amazonian landraces cluster on the same main branch. Perhaps due to small sample sizes, the Vaupés, Inirida and Cauca landrace samples do not cluster as expected given their geographic relationships (see also Hernández 2005 below), but the Juruá landrace clusters more or less where expected. Note also that a sample of type 3 *chichagui* (Magdalena) clusters at a great distance from the *B. gasipaes* cluster, and that the newly found (Silva & Clement 2005) Xingu population of type 1 *chichagui* clusters at an even greater distance, suggesting that they were not involved in the domestication of the crop.

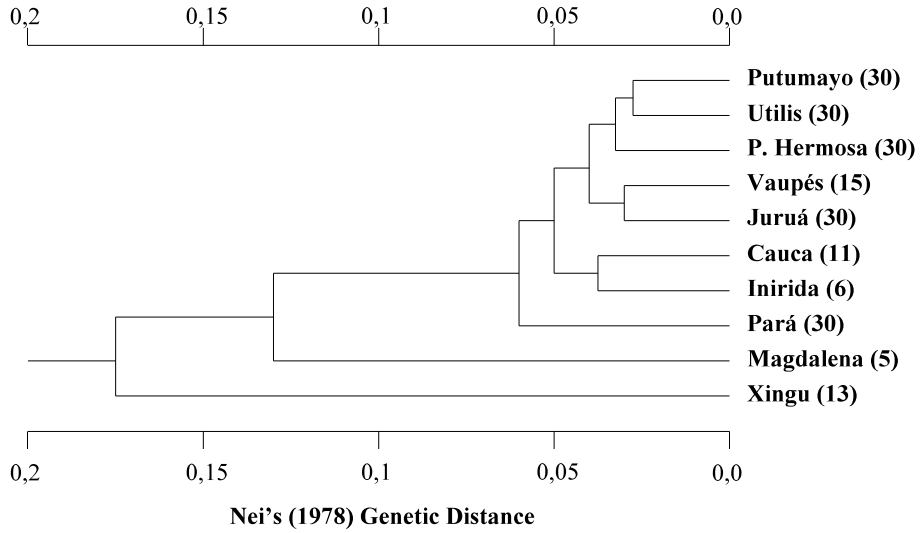

Figure 5. UPMGA dendrogram of Nei's genetic distances among five validated landraces, three non-validated landraces (due to small sample size) and two populations of var. *chichagui*, Magdalena (type 3) and Xingu (type 1) (Silva 2004).

Three sets of microsatellite markers have now been developed (Martinez et al. 2002; Rodrigues et al. 2004b; Billotte et al. 2004). Interestingly, several of Martinez et al.'s markers failed to transfer to type 2 var. *chichagui* (identification by Rodrigo Bernal, pers. com., 2005), although Hernández (2005) later obtained transferability with one of them. Billotte et al.'s set of markers was designed for transfer among *Bactris*, *Astrocaryum* and *Elaeis*, while Rodrigues et al.'s set was designed to work within the Pampa Hermosa landrace and has not yet been tested for transferability to var. *chichagui*.

Hernández (2005) was the first to use three of Martinez et al.'s (2002) microsatellite markers to study the relationships among numerous representatives of var. *gasipaes* and var. *chichagui* in Amazonia and north and west of the Andes. Although this is a very small number of loci, he identified 40 alleles in 18 populations, an overall genetic diversity of 0.67, and a significant difference between observed (0.52) and expected heterozygosity (0.88) (Table 3), suggestive of intensive selection and inbreeding. Twelve of the populations had sufficient individuals (n=20) for a reasonable genetic analysis of their relationships (Figure 6).

Table 3. Number of microsatellite alleles (3 loci), heterozygosity estimates, Wright's F statistics and gene flow estimates among pejibaye (*Bactris gasipaes* var. *gasipaes*) landraces in Amazonia and West-of-the-Andes, and various populations of var. *chichagui* (Hernández 2005). (*Chichagui* types renamed by Clement for this essay; see Hernández for precise geographic information.)

Region	Landrace or type *chichagui*	n	Alleles	Ho	He	Fis	Fit	Fst	Nm
Maracaibo			20	0.62	0.75	-	-	-	-
	type 2 var. veragua	7	-	-	-				
	type 2 var. arapuey	4	-	-	-				
	type 2	4	-	-	-				
West of Andes			34	0.51	0.81	0.28	0.37	0.12	1.86
	type 3 (Chontilla; W. Ecuador)	2	-	-	-				
	Valle de Cauca (W. Colombia)	20	24	0.55	0.76				
	Tuira (Panama)	20	21	0.53	0.78				
	type 3 (Azuero; NW Panama)	20	30	0.62	0.86				
	Tucurrique (Costa Rica)	20	12	0.35	0.60				
	Guatuso (N. Costa Rica)	20	14	0.48	0.62				
Western Amazonia			33	0.55	0.85	0.27	0.34	0.09	2.39
	Putumayo	20	19	0.42	0.75				
	Vaupés	20	21	0.60	0.82				
	Solimões	20	22	0.60	0.82				
	Pampa Hermosa	20	25	0.45	0.81				
	type 3? (Pucallpa, Peru?)	2	-	-	-				
	Ayacucho (Venezuela)	20	16	0.71	0.71				
Eastern Amazonia			31	0.44	0.79	0.38	0.43	0.09	2.52
	type 1 (Acre)	2	-	-	-				
	Tembe (*G. insignis*?; Bolivia)	20	22	0.45	0.72				
	Pará	20	18	0.43	0.73				
Overall		261	40	0.52	0.88	0.29	0.41	0.17	1.25

The levels of heterozygosity west-of-the-Andes are quite high, but certainly also influenced quite strongly by the Azuero type 3 *chichagui* population's very high heterozygosities. These very high levels of heterozygosity and allele numbers may be due to introgression between the Azuero population and surrounding populations of pejibaye (see Couvreur et al. 2005 for introgression in Ecuador), although the overall gene flow in the region is small. The levels of heterozygosity in eastern Amazonia are quite low; Rodrigues et al. (2004a) also observed that the Pará landrace had lower heterozygosities than the other Amazonian landraces.

Hernández (2005) observed that domesticated pejibaye generally had fewer alleles than var. *chichagui* (principally Azuero, because there were enough plants in the sample) and when a domesticate had a lot of alleles it was located adjacent to wild populations so that it could be experiencing introgression. Hernández suggested that Tuira, Cauca, Vaupés, Ayacucho, Yurimaguas and Solimões may be experiencing this introgression. However, no var. *chichagui* has yet been observed adjacent to or within the distributions of Vaupés, Ayacucho or Solimões (see Figure 1).

Hernández (2005) interpreted the genetic variability and these relationships in terms of Mora-Urpí's (1993, 1999) hypothesis of multiple domestication events from different populations of var. *chichagui*, but had problems with Pampa Hermosa, which clustered with the west-of-the-Andes populations (Figure 6). Also, the Ayacucho population clustered with the Solimões landrace, rather than the geographically closer Vaupés landrace, where gene flow would be expected. Hernández removed Pampa Hermosa from the analysis (because this may be a hybrid population [Mora-Urpí 1993]), and Azuero type 3 *chichagui* shifted to a basal position on that branch; the rest remained the

same. Nonetheless, the dendrogram suggests three origins: two east of the Andes and one west-of-the-Andes.

Hernández' difficulty suggests that we look at the dendrogram in a different way: western Amazonia and west-of-the-Andes cluster together, as in Rojas-Vargas et al. (1999), Rodrigues et al. (2004a) and Silva (2004); eastern Amazonia (Pará landrace) clusters with southwestern Amazonia (Tembe population), as in Rojas-Vargas et al. (1999) with the Chaparé population. This alternative interpretation highlights the original dichotomy, rather than a trichotomy.

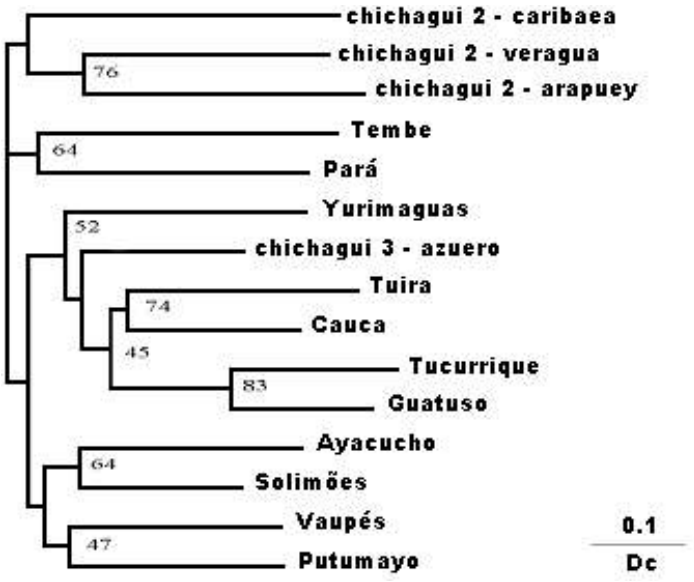

Figure 6. Neighbor-joining dendrogram of Cavalli-Sforza & Edwards' (1967) genetic distance chords (Dc) among 12 populations of *B. gasipaes* (Hernández 2005), with the *chichagui* type 2 set considered an out group. Var. *chichagui* renamed by Clement; Tembe is a pejibaye microcarpa population from Bolivia; Yurimaguas = Pampa Hermosa; Tucurrique is a mesocarpa population in central Costa Rica; Ayacucho is a mesocarpa population in southern Venezuela; others as in other figures and tables above.

Hernández (2005) also examined the relationships of various *chichagui* populations in this analysis, remembering that their sample sizes are to small for confidence. The clustering was not as expected, except for the Peruvian *chichagui* type 3, which clustered with the western Amazonian landraces. Chontilla type 3 *chichagui* (Ecuador) either acted as an additional out-group (when Yurimaguas = Pampa Hermosa was included) or clustered with the west-of-the-Andes populations (without Pampa Hermosa), but the removal of Pampa Hermosa caused the southeastern group (Pará, Tembe) to cluster with *chichagui* type 2. Clearly, better sampling is necessary for these *chichagui* populations so that their inclusion generates interpretable results.

Couvreur et al. (2005) used eight of Billotte et al.'s (2004) microsatellite loci to study the relationships among cultivated and wild pejibaye in Ecuador, looking especially at gene flow among these groups, given the potential for cultivated types to genetically swamp wild types if gene flow is high (Ellstrand 2003). They identified two groups of var. *chichagui* type 3, including a relatively homogenous small-fruited set of populations in the mesic environments of SW Ecuador, where pejibaye is not cultivated, and a much more heterogeneous set of populations adjacent to heart-of-palm plantations in north-central western Ecuador. They also sampled cultivated pejibaye in the palm heart plantations and determined that the seed had originated in Amazonia within the last 10-20 years, although some recent plantations used self-produced seed (with the potential for introgression with the local wild population and local pejibaye). They also used a sample of Amazonian and Central

American plants for contrast, and two *B. setulosa* individuals to act as an out-group in the dendrogram. [Ferreira 1999 had determined that *B. setulosa* is closely related to *B. gasipaes*.]

Couvreur et al. (2005) found higher genetic diversity in their materials (He = 0.77), certainly because of the greater number of loci sampled. The overall impression is of great variability with considerable introgression among populations (Figure 7). The southern *chichagui* type 3, *a priori* thought to be the most isolated from the rest, appears throughout the dendrogram, including a few plants immediately adjacent to the out-group (*B. setulosa*).

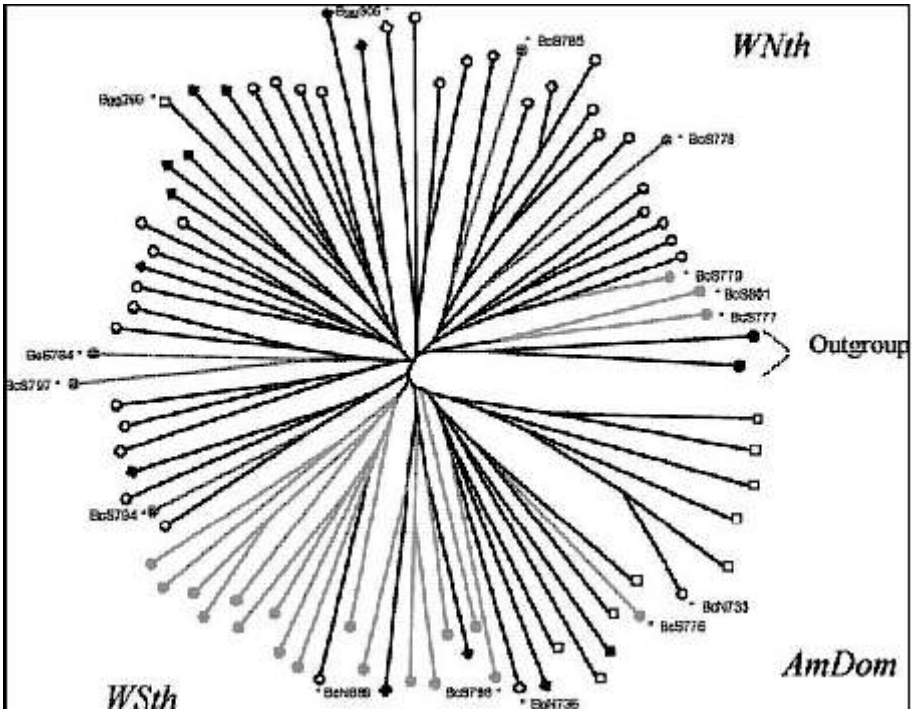

Figure 7. Unrooted neighbor-joining tree with microsatellite genotypes using Shared Alleles distance of the *Bactris gasipaes* complex in Western Ecuador (Couvreur et al. 2005). Each branch represents a single individual. Cultivated individuals in NW Ecuador are represented by solid diamonds, cultivated individuals from Amazonia and Central America (AmDom) by open squares, wild individuals from NW Ecuador (WNth) by open circles, and wild individuals from SW Ecuador (WSth) by gray circles. Outlier individuals are indicated by an asterisk and their names.

Couvreur et al. (2005) used various methods to attempt to discriminate between the two var. *chichagui* populations and the two pejibaye populations. Neither Bayesian method used for discrimination was efficient, although the first (using the Geneclass program and prior assignment) correctly assigned about 50% of the individuals to the wild vs cultivated groups, while the second methods (using the Structure program, with two populations expected) correctly assigned most cultivated individuals (76%) but only 47% of wild individuals to their respective groups. The inefficiency may be due to the small number of microsatellite loci (J.-C. Pintaud, IRD, pers. com., 2006).

Although Couvreur et al. (2005) mentioned the possibility that this data set may support a west-of-the-Andes domestication event, they preferred to concentrate their discussion on introgression. This is an important decision and one that has not been sufficiently considered in earlier work, except for that of Hernández (2005).

Limitations of these studies

Although a considerable number of studies have been produced in the last decade, the origin of cultivated pejibaye is still speculative. This is due principally to the fact that 'origin' has not been a primary objective and to deficiencies in the various studies.

Henderson's (2000) revision of *Bactris* needs more work in terms of var. *chichagui*, especially with a clearer definition and mapping of type 3, although types 1 and 2 also need more work (once there is agreement on synonyms, we can even give them names). Clement and E. Ferreira (ms in preparation) suggest that the types have very different sized fruits: **type 1** – 0.9 to 1.6 cm in length by 0.9 to 1.5 cm in diameter, weighing 0.5 to 1.5 g; **type 2** – 1.0 to 1.5 cm in length by 1.0 to 1.4 cm in diameter, weighing 0.5 to 1.5 g; **type 3** – 1.5 to 2.9 cm in length by 1.4 to 2.8 cm in diameter, weighing 3 to 10 g). Considering that type 3 is larger, more variable, more typical of cultivated fruit, this type needs more prospecting and analysis, which will be difficult given the current paranoia about biodiversity prospecting in Latin America (Gómez-Pompa 2004). Saldías-Paz' (1993) work suggests variation from type 3 into cultivated pejibaye, although Henderson's (2000) descriptions leave a gap between wild and cultivated fruit sizes: var. *gasipaes* – fruits broadly ovoid, 3.5-6.5 x 3-4.5(-6) cm, yellow, orange, red (p.72); var. *chichagui* – fruits subglobose to obovoid, rarely ovoid, 1.2-2.3 x 1.1 x 1.8 cm, orange (p.73). Couvreur et al.'s clear identification of introgression among wild and cultivated pejibaye suggests that morphological work must always take this into account, especially in areas where cultivated palms are numerous.

The molecular marker work has included dominant and co-dominant markers. The dominant RAPD markers are recognized as permitting uncertainty in heterozygosity estimates (Avise 2004). Combined with their difficulty of repeatability, they are gradually losing favor, although they are still useful for quick inexpensive surveys of genetic variability. AFLP markers have not yet been used for studies about origins in pejibaye, but have been widely used in the crop evolution literature in general (Ward 2006). Although generally dominant and more expensive, they are much more repeatable and a single pair of restriction enzymes can generate dozens of markers.

Co-dominant markers are the marker of choice for population genetics (Avise 2004). Recently, microsatellites have gained preference, although allozymes are also excellent, though less abundant, co-dominant markers. The two new studies mentioned here show the power of this technique and it will certainly be more widely used (a Colombian student is working on a project similar in design to the Couvreur et al. study; Yhon Jairo Acosta Perez, Univ. Cauca, Popayán, pers. com., 2005). To get good resolution, however, a reasonable number of loci must be sampled – e.g., 15 to 20 should give the resolution necessary to discriminate between wild and cultivated plants (J.-C. Pintaud, IRD, pers. com., 2006). With three sets of microsatellite markers published, the question of better resolution is clearly within our grasp.

The Couvreur et al. and Hernández studies highlight the importance of introgression between wild and cultivated pejibaye, and the difficulty of clearly discriminating segregants from legitimate wild or cultivated individuals. The importance of introgression has not been sufficiently incorporated into origin studies to date, although many of the studies cited above may contain cases of introgression that have not been adequately examined (e.g., Rodrigues et al. 2004a; Hernández 2005). For example, the type 1 *chichagui* used by Rodrigues et al. occurs in the range of the Pará landrace (based on a recent range expansion of this landrace done by Santos & Farias Neto 2005), so introgression can not be ruled out as an explanation for this wild type clustering with the Pará landrace (Figure 3). Similarly, the type 3 *chichagui* used by Rodrigues et al. occurred close (5 km) to type 1 *chichagui* and immediately adjacent (20 m) to pejibaye plants in Benjamin Constant, providing the opportunity for multiple introgression.

Given that most extant var. *chichagui* populations are within the range of pejibaye populations and often adjacent to human populations that cultivate significant numbers of pejibayes, the question of genetic purity of remnant populations of var. *chichagui* is an important consideration (Couvreur et al.

2005). Especially in regions thought to be involved in the initial domestication of pejibaye, for example in Bolivian Amazonia, the probability of purity is reduced by gene flow from cultivated populations (cf. Saldías-Paz 1993). When heart-of-palm plantations, with seed production plantations, are in the immediate vicinity of wild populations, such as in Ecuador, purity can not be expected, as shown by Couvreur et al. This raises another question: is better resolution enough to identify origins?

Molecular techniques and the origins of other Neotropical crops

Motley et al. (2006) organized a collection of studies using molecular techniques to identify crop origins, study crop-to-wild and modern crop-to-traditional crop gene flow, and related topics. Some of these studies suggest alternative analyses and new directions for our research on *B. gasipaes*.

Reiseberg & Harter (2006) reviewed work on sunflower (*Helianthus annuus*), an allogamous annual, whose wild populations are spread from northern Mexico to north central continental USA and occur as weeds in disturbed areas. Early allozyme work could not identify an origin, but did find that all domesticates tend towards monomorphism of a subset of wild alleles, nor did sequencing of chloroplast DNA resolve the issue, although all domesticates had the same cpDNA haplotype and narrowed the number of geographic areas for further study. A recent study has now resolved previous doubts and located the center of domestication in the east-central USA. Harter et al. (2004) carefully sampled 21 wild populations throughout the species' wild range, taking care to avoid populations with nearby cultivated sunflower, with the expectation of reducing or eliminating the possibility of gene flow and introgression. They compared these with 8 Native American landraces, 10 domesticated lineages and 2 commercial lines, using 18 microsatellite loci. They analyzed this data set in three ways: 1) pairwise genetic distances (D_A, Nei et al. 1983) were calculated and used to construct an unrooted majority rule consensus neighbor-joining tree, with no out group; 2) model-based clustering was implemented with the Structure program (Pritchard et al. 2000; Falush et al. 2003) to infer population structure in wild sunflower and then to assign the domesticates to inferred populations; 3) the Structure program was used to make inferences about ancestral allele frequencies in the common ancestor of wild and domesticated sunflower and the degree of drift away from the ancestral genomic composition of each population. The interaction among these analytical methods was synergistic and provided good confidence in the inferences about origin.

Buckler & Stevens (2006) reviewed work on maize (*Zea mays*), an allogamous annual, whose wild populations are spread from central Mexico to Honduras. As the most important grain crop, maize has long been the subject of studies attempting to define its origin. The earliest studies used citogenetic analysis, specifically the number and location of chromosomal knobs (highly repetitive sections of DNA), to identify three subspecies of maize that are most closely related. RFLP analysis of chloroplasts was used to further refine the phylogeny of *Zea* and add another subspecies within maize. Thirty allozymes were used to determine which subspecies was most likely to have given rise to domesticated maise and even to propose the exact geographic origin. Sequencing of nuclear ribosomal internal transcribed spacers permitted a further refinement of the phylogeny of the genus and, specifically, of maize. Finally, 99 microsatellite loci were used to clinch the issue, confirming all previous work and defining *Z. mays* ssp. *parviglumis* as the wild progenitor and the Balsas River basin as the theater of domestication.

Schaal et al. (2006) reviewed work on manioc (*Manihot esculenta*), an allogamous annual, whose wild populations are spread across southern Amazonia, with a distribution similar to *B. gasipaes* var. *chichagui* type 1 and a little further south. Although generally propagated vegetatively, seedlings are important for generating new diversity and maintaining disease resistence in manioc plantations (Martins 2001). Allem (1999) identified manioc's closest relatives based on morphological systematics, after finding apparently wild manioc in southern Amazonia. Sequencing of the G3pdh enzyme was used to create a haplotype network, which found that domesticated manioc contains only six of the 23 haplotypes found in wild manioc, strongly suggesting that Allem's hypothesis is correct. The haplotype

network also identified a set of wild populations in Mato Grosso, Rondonia and Acre (Brazil) as the origin of the domesticated haplotypes. Five microsatellite loci were used for a population genetic analysis, which found that domesticated manioc contains only 15 of the 73 alleles found in wild manioc, again confirming Allem's hypothesis. Additionally the 15 alleles in the domesticate are from the same geographic region as the six G3pdh haplotypes. Olsen (2004) also used single nucleotide polymorphisms (SNPs) in two low copy nuclear genes (*BglA, Hnl*) to confirm the other molecular studies.

Recently, Miller & Schaal (2005) examined the origin of jocote (*Spondias purpurea*), an allogamous perennial of Mesoamerica, and determined that it was brought into cultivation at least twice. The authors collected 96 samples of wild and cultivated jocote from 11 areas in Mesoamerica, as well as numerous out groups (other sympatric and allopatric *Spondias* spp.). They sequenced the *trnG-trnS* chloroplast spacer and created haplotype networks, which contained two groups, only one of which contained jocote. Jocote had 17 distinct haplotypes, 12 in wild populations and 9 in cultivated populations, concordant with wild vs domesticated expectations, although several haplotypes were not found in the wild, perhaps because of local extinctions or insufficient sampling. They used coalescent theory (Posada & Crandall 2001) to identify two groups of ancestral jocote haplotypes in the network, one in western central Mexico and one in Guatemala and El Salvador.

Several things are clear from these four studies. First, it is important to clearly identify the phylogenetic relationships closest to the target domesticate. In pejibaye we have a phylogenetic hypothesis (Henderson 2000) that has not yet been clearly articulated within the species, i.e., what exactly are the three types, and what are their relations to each other and other closely related species (e.g., *B. setulosa*, *B. riparia*, and the Caribbean *Guilielma* group as defined by Ferreira 1999 and Henderson 2000), remembering also that among species gene flow is common in palms? Ferreira's cladistic analysis offers a starting point for the *Guilielma* group, but the within pejibaye analysis needs work. This might best be approached by sequencing one or several chloroplast and/or nuclear gene segments, as we have numerous accessions of these wild types in the various national germplasm collections, and these genes tend to have slower mutation rates than microsatellites, hence are likely to be conserved over Holocene time frames.

Second, it is important to determine which var. *chichagui* type is the progenitor of pejibaye. This can also be approached with the sequencing mentioned above, as pejibaye is likely to contain a subset of the haplotypes in one of the var. *chichagui* types. Here it will be important to define a representative set of pejibaye accessions for the comparison. Our project at INPA is currently working to define a nuclear collection (Johnson & Hodgkin 1999) within our germplasm bank (CNPq Universal project no. 476189/2003-9), but we do not have good samples of several landraces (e.g., Vaupés, Inirida, Pastaza), nor even accessions of some critical populations (e.g., Tembe). Hence, it would be best to create a nuclear collection in the much more representative Costa Rican germplasm bank for this study.

Third, it is important to increase precision for analytical purposes. This can probably best be done by selecting the 20 best microsatellite loci published (in terms of number of alleles and transferability among var. *chichagui* types and pejibaye) and using these for future studies.

Fourth, it is important to agree on analytical methodology, so that all studies can easily be compared with each other. The number of methods and computer programs available is growing constantly, and their quality is improving constantly also. In this respect, the proposed sequences of chloroplast and nuclear genes should be analyzed as haplotype networks, with full geographical referencing, and then analyzed with phylogenetic software (e.g., PAUP?) at the type-level and above. The within pejibaye analysis must decide on the best distance measure (e.g., Nei's D_A of 1983?), the most appropriate cluster algorithm (N-J vs UPMGA vs ?), the use of several closely and more distantly related species to root the trees (e.g., *B. setulosa* or *B. riparia* for within pejibaye analyses, and an *Astrocaryum* or *Acrocomia* for within group-*Guilielma* analyses?), the use of modeling to best describe

wild and cultivated populations and their most likely original allele frequencies (e.g., Structure?), the use of modeling of genetic drift to get an idea of degree of change after domestication (e.g., Structure?).

The proposal

General objective: Determine the origin of domesticated pejibaye and the relationships among its various landraces.

Specific objectives:
1) Determine the phylogenetic relationships among the three types of wild pejibaye and closely related species;
2) Determine the wild type(s) that was(were) domesticated to create cultivated pejibaye;
3) Identify the approximate region(s) of origin of the domestication event(s);
4) Determine the phylogenetic relationships among the various landraces.

Material (assuming that only INPA is involved, since I only have full data on the INPA germplasm bank) [most accessions have 9 plants] (ideally this will be a multinational project):

var. *chichagui*
- Type 1: 6 accessions Rio Branco; 1 accession Rio Xingu; 1 accession Benjamin Constant; others can be obtained relatively easily in Brazil, less easily in Bolivia and Peru
- Type 2: 0 accessions
- Type 3: 3 accessions Pucallpa; 1 accession Contamana; 1 accession Benjamin Constant; others can be obtained in Brazil, less easily elsewhere

var. *gasipaes* – microcarpa landraces
- Pará: 53 accessions + new Madeira River material currently at Embrapa Amazônia Oriental
- Juruá: 4 accessions; others can be obtained relatively easily
- Other SW Amazonia: Puerto Maldonaldo – 3 accessions

var. *gasipaes* – mesocarpa landraces
- Pampa Hermosa: 70 accessions
- Pastaza: 1 accession
- Inirida: 3 accessions
- Cauca: 3 accessions
- Utilis: 18 accessions + 10 Guatuso + 6 Tuira

var. *gasipaes* – macrocarpa landraces
- Putumayo: 113 accessions + 72 Solimões
- Vaupés: 4 accessions; others can be obtained relatively easily

var. *gasipaes* – **nuclear collection** = logarithmic sampling in over-represented landraces and proportional in under-represented landraces

out groups: *B. riparia* – 1 accession; *B. ferruginea* – 1 accession; *Astrocaryum aculeatum* – 1 accession

Methods – objectives 1 + 2 + 3

Sequence one chloroplast spacer (as in *Spondias*?) and one nuclear gene (G3pdh?) from one plant of each var. *chichagui* accession, 5 geographically diverse plants of each var. *gasipaes* nuclear collection, one *B. riparia* and one *B. ferruginea* plant. Create geographically referenced haplotype network and identify probable domestication event(s).

Analyze all available var. *chichagui* plants, 30 geographically diverse plants of each var. *gasipaes* landrace in nuclear collection, and all available plants of *B. riparia, B. ferruginea* and 20 *A. aculeatum* with 20 microsatellite loci. Estimate Nei's D_A, create a rooted UPMGA cluster diagram, use Structure to infer wild populations, relate each to domestic populations and estimate genetic drift among appropriate wild and domesticated pejibayes.

Methods – objective 4

Analyze all available plants of the most likely var. *chichagui* population(s), 30 plants of each var. *gasipaes* landrace in nuclear collection, all available plants of non-representative var. *gasipaes* collection, and all available plants of *B. riparia* with 20 microsatellite loci. Estimate Nei's D_A, create a rooted UPMGA cluster diagram, use Structure to estimate genetic drift among them. Relate this dendrogram to Morcote-Rios & Bernal's (2001) archaeological data set and propose new archaeological sites for analysis.

This proposal is designed to be multinational in scope, as the INPA collection is insufficient for an ideal analysis. I believe that we have enough var. *chichagui* samples in the collections to make a good start. If we can get national financing for studies in Brazil, Colombia, Costa Rica and Peru, agree upon methodologies, and pool analytical results, we should be able to identify the origin of var. *gasipaes* within a two-year project. Criticism, comments and suggestions are requested.

References

Allem, A.C. 1999. The closest wild relatives of cassava (*Manihot esculenta* Crantz). *Euphytica* 107: 123-133.

Avise, J.C. 2004. *Molecular markers, natural history, and evolution*, 2nd Ed. Sinauer Associates, Sunderland.

Bernal, R. 1989. Proposal to conserve *Bactris gasipaes* Kunth against *Bactris ciliata* (R & P) C. Martius (Palmae). *Taxon* 38: 520-522.

Billotte, N. et al. 2004. A new set of microsatellite markers for the peach palm (*Bactris gasipaes* Kunth): characterization and across-taxa utility within the tribe Cocoeae. *Molecular Ecology Notes* 4: 580–582.

Blumler, M.A. 1992. Independent inventionism and recent genetic evidence on plant domestication. *Economic Botany* 46: 98-111.

Buckler, E.S.; Stevens, N.M. 2006. Maize origins, domestication, and selection. *In*: Motley *et al*., pp. 67-90.

Cavalli-Sforza, L.L.; Eduards, A.W.F. 1967. Phylogenetic analysis: models and estimation procedures. *American Journal Human Genetics* 19:233-257.

Clement, C.R. 1995. Pejibaye (*Bactris gasipaes*). *In*: Smartt, J.; Simmonds, N.W. (Eds). *Evolution of crop plants*, 2nd Ed. Longman, London. pp. 383-388.

Clement, C.R. *et al*. 1989. Pupunha brava (*Bactris dahlgreniana* Glassman): progenitora da pupunha (*B. gasipaes* H.B.K.)? *Boletin do Museu Paraense Emílio Goeldi, série Botânica* 5(1): 39-55.

Clement, C.R. *et al*. 1997. Allozyme variation in spineless pejibaye (*Bactris gasipaes* Palmae). *Economic Botany* 51(2): 149-157.

Clement, C.R. *et al*. 1999. A pupunha brava (*Bactris dahlgreniana*) no Estado do Amazonas, Brasil. *Acta Botanica Venezuelica* 22(1): 29-44.

Couvreur, T.L.P. *et al*. 2005. Close genetic proximity between cultivated and wild *Bactris gasipaes* Kunth revealed by microsatellite markers in Western Ecuador. *Genetic Resources and Crop Evolution* - in press

Ellstrand, N.C. 2003. *Dangerous liaisons? When cultivated plants mate with their wild relatives.* (Syntheses in Ecology and Evolution) John Hopkins Univ. Press, Baltimore.

Falush, D. *et al*. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164: 1567-1587.

Ferreira, E. 1999. The phylogeny of pupunha (*Bactris gasipaes* Kunth, Palmae) and allied species. *In*: Henderson, A.; Borchsenius, F. (Eds.). *Evolution, variation, and classification of palms*. Memoirs of The New York Botanical Garden 83: 225-236.

Gómez-Pompa, A. 2004. The role of biodiversity scientists in a troubled world. *BioScience* 54: 217-225.

Harter, A.V. *et al*. 2004. Single origin of extant domesticated sunflower in eastern North America. *Nature* 430: 201-205.

Henderson, A. 1995. *The palms of the Amazon*. Oxford Univ. Press, Oxford.
Henderson, A. 2000. *Bactris* (Palmae). *Flora Neotropica* 79: 1-181.
Henderson, A. *et al*. 1995. *Field guide to the palms of the Americas*. Princeton Univ. Press, Princeton.
Hernández, J.A. 2005. Descripción de la diversidad y estructura genética de las poblaciones silvestres y cultivadas de pejibaye (*Bactris gasipaes* Kunth), y sus relaciones de parentesco, utilizando marcadores microsatélites. Tesis de Maestria, Univ. Costa Rica, San José. (in two chapters at the following URLs) http://www.pejibaye.ucr.ac.cr/Diversidad/Diversidad9.htm; http://www.pejibaye.ucr.ac.cr/Diversidad/Diversidad10.htm
Huber, J. 1904. A origem da pupunha. *Bol. do Museu Paraense Emilio Goeldi, Botanica* 4: 474-475.
Johnson, R.C.; Hodgkin, T. (Eds.). 1999. *Core collections for today and tomorrow*. IPGRI, Roma.
Martínez, A.K. et al. 2002. Primer Note: Microsatellite loci in *Bactris gasipaes* (Arecaceae): their isolation and characterization. *Molecular Ecology Notes* 2: 408-410.
Martins, P.S. 2001. Dinâmica evolutiva em roças de caboclos amazônicos. *In*: Vieira, I.C.G. *et al*. (Eds.). *Diversidade biológica e cultural da Amazônia*. Museu Paraense Emílio Goeldi, Belém. pp.369-384.
Miller, A.; Schaal, B.A. 2005. Domestication of a Mesoamerican cultivated fruit tree, *Spondias purpurea*. *PNAS* 102: 12801-12806.
Mora-Urpí, J. 1993. Diversidad genética en pejibaye: II. Origen y evolución. *In*: Mora Urpí, J. et al. (Eds). *IV Congresso Internacional sobre Biologia, Agronomia e Industrialización del Pijuayo*. Editorial Univ. Costa Rica, San José. pp. 21-29.
Mora-Urpí, J. 1999. Origen y domesticación. *In*: Mora-Urpí, J.; Gainza E., J. (Eds.). *Palmito de pejibaye (Bactris gasipaes Kunth): su cultivo e industrialización*. Editorial de la Universidad de Costa Rica, San José. pp. 17-24.
Mora-Urpí, J. *et al*. 1997. *Peach palm.* Bactris gasipaes *Kunth.* Promoting the conservation and use of underutilized and neglected crops. 20. Institute of Plant Genetics and Crop Plant Research - IPK, Gatersleben/International Plant Genetic Resources Institute - IPGRI, Rome.
Morcote-Rios, G.; Bernal, R. 2001. Remains of palms (Palmae) at archaeological sites in the New World: A review. *Botanical Review* 67(3): 309–350.
Motley, T.J. *et al*. (Eds.). *Darwin's harvest: New approaches to the origins, evolution, and conservation of crops*. Columbia Univ. Press, New York.
Nei, M. *et al*. 1983. Accuracy of estimated phylogenetic trees from molecular data. *J. Molecular Evolution* 19: 153-170.
Olsen, K.M. 2004. SNPs, SSRs and inferences on cassava's origin. *Plant Molecular Biology* 56: 517-526.
Posada, D.; Crandall, K.A. 2001. Intraspecific gene genealogies: trees grafting into networks. *Trends Ecol. Evol.* 16: 37-45.
Pritchard, J.K. *et al*. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959.
Reiseberg, L.H.; Harter, A.V. 2006. Molecular evidence and the evolutionary history of the domesticated sunflower. *In*: Motley *et al*., pp. 31-48.
Rodrigues, D.P. *et al*. 2004. Molecular marker-mediated validation of morphologically defined landraces of pejibaye (*Bactris gasipaes*) and their phylogenetic relationships. *Genetic Resources and Crop Evolution* 51(8):871-882.
Rodrigues, D.P. *et al*. 2004. Novel microsatellite markers for *Bactris gasipaes* (Palmae). *Molecular Ecology Notes* 4(4): 575-576.
Rojas-Vargas, S. *et al*. 1999. Polimorfismo isoenzimático en cuatro razas y un híbrido de *Bactris gasipaes* (Palmae). *Rev. Biología Tropical* 47(4): 755-761.
Schaal, B.A. *et al*. 2006. Evolution, domestication, and agrobiodiversity in the tropical crop cassava. *In*: Motley *et al*., pp. 269-284.
Saldías-Paz, M. 1993. La chonta de castilla (*Bactris gasipaes* H.B.K.): taxonomía y algunos datos económicos en Santa Cruz y su distribución en Bolivia. *In*: Mora Urpí, J. et al. (Eds). *IV Congresso Internacional sobre Biologia, Agronomia e Industrialización del Pijuayo*. Editorial Univ. Costa Rica, San José. pp. 115-126.
Sanders, R.W. 1991. Cladistics of *Bactris* (Palmae): survey of characters and refutation of Burret's classification. *Salbyana* 12: 105-133.
Santos, R.P.; Farias Neto, J.T. 2005. Primeira expedição de prospecção e coleta de material genético de pupunha na calha do Rio Madeira. Reunião Técnico do sub-projeto ProBio Pupunha: raças primitivas e parentes silvestres, INPA, Manaus, AM. http://www.inpa.gov.br/pupunha/probio/expedicao_i_madeira.pdf

Silva, C.C. 2004. Análise molecular e validação de raças primitivas de pupunha (*Bactris gasipaes*) por meio de marcadores RAPD. Dissertação de Mestrado, Univ. Fed. São Carlos/Univ. Fed. Amazonas, Manaus.
Silva, J.B.F.; Clement, C.R. 2005. Wild pejibaye (*Bactris gasipaes* var. *chichagui*) in southeastern Amazonia. *Acta Botanica Brasilica* 19(2): 283-286.
Ward, S.M. 2006. Appendix 1. Molecular marker and sequencing methods and related terms. *In*: Motley, T.J. *et al*. pp. 347-369.