EVOLUTIONARY BIOGEOGRAPHY OF *MANIHOT* (EUPHORBIACEAE), A RAPIDLY RADIATING NEOTROPICAL GENUS RESTRICTED TO DRY ENVIRONMENTS

Anne Duputié [1, 2], Jan Salick [3], Doyle McKey [1]

**Aim** The aims of this study were to reconstruct the phylogeny of *Manihot*, a Neotropical genus restricted to seasonally dry areas, to yield insight into its biogeographic history and to identify the closest wild relatives of a widely grown, yet poorly known, crop: cassava (*Manihot esculenta*).

**Location** Dry and seasonally dry regions of Meso- and South America.

**Methods** We collected 101 samples of *Manihot*, representing 52 species, mostly from herbaria, and two outgroups (*Jatropha gossypiifolia* and *Cnidoscolus urens*). More than half of the currently accepted *Manihot* species were included in our study; our sampling covered the whole native range of the genus, and most of its phenotypic and ecological variation. We reconstructed phylogenetic relationships among *Manihot* species using sequences for two nuclear genes and a noncoding chloroplast region. We then reconstructed the history of traits related to growth form, dispersal ecology, and regeneration ability.

**Results** *Manihot* species from Mesoamerica form a grade basal to South American species. The latter species show a strong biogeographic clustering: species from the cerrado form well-defined clades, species from the caatinga of northeastern Brazil form another, and so do species restricted to forest gaps along the rim of the Amazon basin. Vine and tree growth habits evolved repeatedly in the genus, as did fruit indehiscence and loss of ant-mediated seed dispersal.

**Main conclusions** The genus *Manihot* probably originated in Mesoamerica, where it diversified prior to colonizing South America. Within South America, several groups then radiated within all kinds of seasonally dry lowland habitats southwards and eastwards. Some species also adapted to more humid environments, such as forest gaps, around the rim of the Amazon basin. Given the limited dispersal abilities of Manihot species, we propose that this radiation is most likely to have occurred only after, or shortly before, the completion of the Isthmus of Panama, around 3.5 Ma. Our results are in agreement with the past existence of a corridor of dry vegetation through Amazonia or along the eastern South American coast. In addition, our phylogeny allows identification of cassava's closest wild relatives.

**Keywords**: Adaptive radiation, biogeography, cassava, cerrado, dispersal, Neotropics, phylogeny, regeneration, seasonally dry environments, seedling morphology.

---

[1]  CEFE UMR5175, 1919 Route de Mende, 34293 Montpellier Cedex 5, France

[2] Present address: Section of Integrative Biology, University of Texas, Austin, TX 78712, USA.

[3]  Missouri Botanical Garden, P.O. Box 299, St. Louis, MO 63166-0299, USA

## Introduction

The Neotropical genus *Manihot* Mill. (Euphorbiaceae) has been surprisingly neglected by phylogenetic studies. The most recent taxonomic revision of the genus, performed over 35 years ago, recognizes 98 species (Rogers & Appan, 1973). One of them is domesticated cassava, *Manihot esculenta* Crantz ssp. *esculenta*, one of the most important tropical crops, and the eighth most harvested crop in the world (FAOSTAT, 2009). Several additional *Manihot* species have since been described by Allem (1978; 1979; 1999a) and others have been reduced to synonymy, based on morphological grounds (Allem, 1978; 1979; 1994; Second *et al.*, 1997). The genus is distributed from south-western North America to northern Argentina, with two centres of endemism (Rogers & Appan, 1973, p. 7): in south-western Mexico (with 18 species recognized in Mesoamerica) and in the *cerrado* of central Brazil (with 41 species described in the Brazilian states of Goiás and Distrito Federal). Only one species is distributed on both continents, North and South America: *M. brachyloba* Muell. Arg., which occurs north-east of the Amazon basin in South America, and also in a very small area of Costa Rica. It is the only species of *Manihot* that occurs in the Caribbean islands, where it is found in a restricted part of Hispaniola. Figure 1 shows a distribution map of *Manihot* (excluding the cultivated species).

Phylogenetic and phylogeographic studies have so far focused on the origin of cassava (Roa *et al.*, 1997, 2000; Olsen & Schaal, 1999; Olsen, 2004). Only one attempt to establish a phylogeny has included a large number of species unrelated to cassava (Chacón *et al.*, 2008). That study showed *Manihot* to be monophyletic. However, the resolution was poor, and the phylogenetic reconstruction resulted in a polytomy. The phylogeny of *Manihot* thus remains unclear. This is unfortunate, because its resolution could cast light not only on the origin of cassava but also on broader issues concerning the history of Neotropical environments.

Most species of *Manihot* are restricted to relatively dry environments, such as the *cerrado* of central Brazil, the *caatinga* of north-eastern Brazil, or forest-savanna ecotones. While several Neotropical genera of forests (Richardson *et al.*, 2001; Erkens *et al.*, 2007) or high-elevation areas (Hughes & Eastwood, 2006) have been thoroughly studied, the recent history of taxa restricted to dry environments – especially savannas – in this part of the world has been overlooked (Pennington *et al.*, 2009), even though some insights are currently arising from the study of trees from Neotropical seasonally dry forests (Caetano *et al.*, 2008; Pennington *et al.*, 2009; Ireland *et al.*, 2010) and even from *cerrado* plants (Simon *et al.*, 2009). The study of taxa restricted to dry environments has much to tell us about the history of the vegetation cover of the Neotropics for the past few million years, a currently hotly debated topic (Bush, 1994; Colinvaux *et al.*, 2000; Pennington *et al.*, 2000, 2004; Haffer & Prance, 2001; Bonaccorso *et al.*, 2006).

*Manihot* also shows a wide range of ecologies. Although most species are restricted to open, relatively dry environments, a few are typically found in disturbed environments in forest. A number of species possess starchy tuberous roots that facilitate resprouting after disturbance, fire probably being the most important type of disturbance over the history of the genus. Many *Manihot* species possess adaptations for myrmechochory, associated with prolonged dormancy of buried seeds. In these species, seeds are first projected by explosive dehiscence of the dry mature capsule. Each seed bears a lipid-rich caruncle, which functions as an elaiosome, attracting ants that carry diaspores to their nests (Elias & McKey, 2000), consuming the caruncle and then burying the seed, either in the nest or in nearby refuse piles (Renard *et al.*, 2010). Thus stored in the soil, seeds can remain dormant for years, their dormancy being broken by high soil temperatures that signal removal of vegetation cover by a disturbance (Pujol *et al.*, 2002). Dispersal-related traits show substantial variation in the genus. Some species have indehiscent fruits with seeds lacking a caruncle and too large to be carried by ants. Growth form also shows great variation within the genus, with shrubs, trees, and vines all being represented. As this genus is probably of recent origin (about 6.3 Myr old, according to Chacón *et al.*, 2008), it offers the opportunity to study patterns of diversification in a recent radiation.

The origin of cassava is uncertain, because few molecular studies that address this question have included a sufficiently wide range of species. Current studies support the hypothesis that cassava is derived from populations of *Manihot esculenta* ssp. *flabellifolia* (Pohl) Ciferri from the south-western rim of the Amazon basin (Olsen & Schaal, 1999; Olsen, 2004; Léotard *et al.*, 2009), but in fact few other candidate species have been studied in any detail. Groups of species considered to comprise the "primary gene pool" and "secondary gene pool" of cassava have been proposed on morphological grounds (Allem *et al.*, 2001), but the phylogenetic relationships among these species, and to species supposedly more distant from cassava, are unknown.

In this study, we present a phylogeny for *Manihot*, based on two nuclear genes and one noncoding chloroplast region, using a wide sample of species across the distributional range of the genus. Three issues are addressed: (1) What is the biogeographic history of the genus? (2) How did *Manihot* species diverge and colonize new habitats? (3) What are the species most closely related to domesticated cassava?
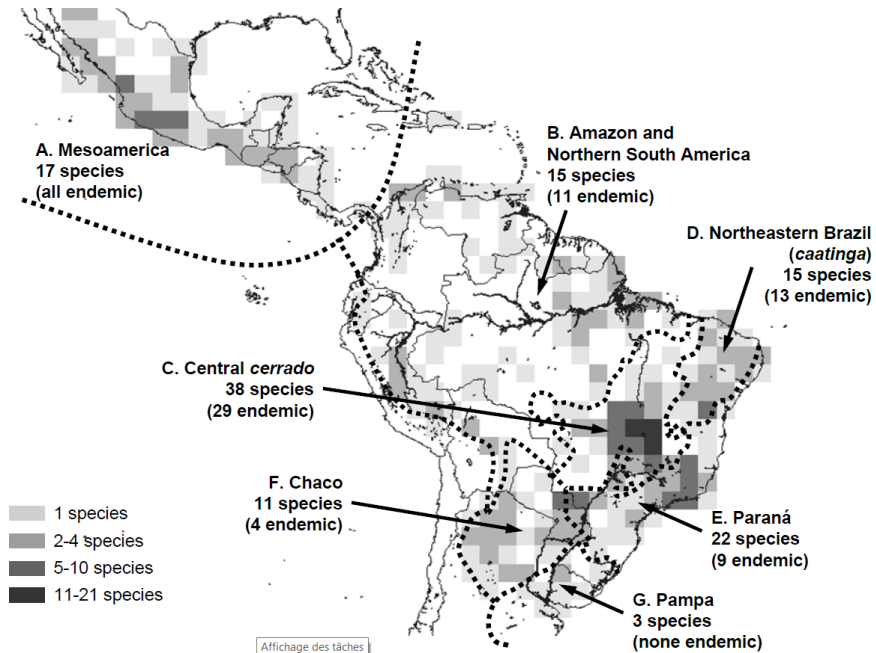
**Figure 1 :** Map of the natural range of the genus *Manihot* (*i.e.*, not including *M. esculenta* or introduced *M. glaziovii*) showing species richness on a 2° × 2° grid. This map was compiled from the observations of Rogers & Appan (1973), which are the most complete and accurate available so far. The map reflects their sampling effort, and is likely to underestimate the number of *Manihot* species present in some areas. Dashed lines indicate the boundaries considered in the DIVA and LAGRANGE analyses.

## Material & Methods

### Plant material

Leaf samples of 222 *Manihot* specimens were collected from herbaria (UC Davis herbarium, CA, USA; Missouri Botanical Garden Herbarium, MO, USA; New York Botanical Garden Herbarium, NY, USA; Royal Botanic Gardens, Kew, UK; Herbier du Muséum d'Histoire Naturelle de Paris, France) and nine samples were collected in the field, in French Guiana. Obtaining high-quality DNA from herbarium material that has not been collected, dried and stored under appropriate conditions  is not always successful (e.g. Rogers & Bendich, 1985). As a result, only 101 specimens (52 species and all three subspecies of *M. esculenta*) for which at least the two nuclear genes could be sequenced were included in our analysis. Fresh leaves were collected for two outgroups: *Cnidoscolus urens* (L.) Arthur and *Jatropha gossypiifolia* L. Detailed information on vouchers is provided in Table S1 in the Supporting Information.

Whenever possible, we used herbarium samples determined by the late D. J. Rogers (University of Colorado) and his former graduate student S. G. Appan or by A. C. Allem (formerly of EMBRAPA-Cenargen, Brasilia). Rogers and Appan observed and measured a large number of *Manihot* specimens in herbaria throughout the world and collected *Manihot* samples all over the Neotropics, prior to writing their monograph on *Manihot* (Rogers & Appan, 1973). Allem collected numerous samples of *Manihot* species in Brazil and proposed modifications of the taxonomy established by Rogers & Appan (e.g. Allem, 1978, 1979, 1999a). Despite these precautions, species identification sometimes was dubious, as discussed later.

After initial tests, three published primer pairs were retained for further analysis: the nuclear genes glyceraldehyde-3-phosphate dehydrogenase (G3*pdh*, 4 introns and 5 exons covered; Strand *et al.*, 1997) and nitrate reductase (*NIA*-i3, covering one intron, Howarth & Baum, 2002), and the noncoding chloroplast region *trn*C-D (Taberlet *et al.*, 1991). G3*pdh* and *NIA* could be sequenced for all 101 *Manihot* specimens included in this study; *trn*C–D was sequenced for 57 specimens (39 species). All three fragments were amplified for *J. gossypiifolia*, but we were unable to retrieve *trn*C–D and *NIA* sequences from *C. urens*.

### DNA amplification and sequencing

DNA extraction was conducted under sterile conditions, using Qiagen DNEasy Plant mini kits (Qiagen, Foster City, CA, USA) in a final volume of 50 μL. As DNA yield was very low with most herbarium samples, we tested different brands of Taq polymerase and used a range of amplification conditions. For all samples, amplification took place in 20 μL final volume, using Qiagen Multiplex Taq Polymerase following the manufacturer's recommendations, and using 1 to 5 μL undiluted DNA (5 μL of 1/50 diluted DNA was enough for fresh samples). Amplification cycles were as follows: 95°C for 15 min, 35 to 40 cycles of denaturation (94°C, 30 s), annealing (57 to 60°C, 90 s), elongation (72°C, 90 s), and a final elongation step of 10 min at 72°C. Amplification quality was assessed by migrating polymerase chain reaction (PCR) products on 1 % agarose gels. Samples were then purified using BigDye and sequenced on an ABi3730 sequencer (Life Technologies, Carlsbad, CA, USA). Because numerous PCR cycles were prone to lead to the amplification of contaminants, each sequencing reaction was run at least twice from independent amplification reactions.

## Sequence analysis

Sequences were checked by eye under SEQUENCHER v.4.5 (Gene Codes, Ann Arbor, MI, USA). Length-variant heterozygotes were identified and one of the two haplotypes was retained when they could easily be distinguished; otherwise the specimen was not included in the analysis. Alignment was conducted with MUSCLE (Edgar, 2004) on the EMBL-EBI Computational Platform (http://www.ebi.ac.uk/Tools/), after removal of a minisatellite included in a G3*pdh* intron, which provided no phylogenetic information (Olsen, 1999). Alignments were then edited under BIOEDIT (Hall, 1999). Sequences are deposited in the EMBL database (accessions FN551734–FN552000).

Two datasets were considered: one included all 101 specimens and sequences for G3*pdh* and *NIA*; the other included the 57 specimens sequenced for all three genes. The first dataset was partitioned into three sets: G3*pdh* introns, G3*pdh* exons, and *NIA*, while the second set contained a fourth partition for *trn*C–D.

Maximum likelihood tree reconstruction was conducted using PHYML online interface (Guindon & Gascuel, 2003), with the GTR + Γ substitution model, and 500 bootstrap replicates, and the approximate likelihood-ratio test (aLRT, Anisimova & Gascuel, 2006). Bayesian analyses were conducted using MRBAYES v3.1.2 (Huelsenbeck & Ronquist, 2001) on the CBSU computation cluster (http://cbsuapps.tc.cornell.edu/mrbayes.aspx) with default parameters and $10^7$ Markov chain Monte Carlo (MCMC) iterations (half were burn-in) for the three-gene dataset, and $2 \times 10^6$ iterations (800,000 burn-in) for the two-gene dataset. In both cases, convergence was achieved by the end of the burn-in.

## Ancestral character reconstruction

Seven geographic zones were defined, corresponding to biogeographic units defined by Morrone (2006; see Fig. 1): Mesoamerica (A); Amazon region and northern South America (B); the *cerrado* region (C); the *caatinga* region, in north-eastern South America (D); the Paraná region (E; synonym to "Atlantic Forest" in Simon *et al.,* 2009); the chaco (F); and the pampa (G).

A dispersal–vicariance analysis was carried out using DIVA version 1.1 (Ronquist, 1996), without setting a maximum number of ancestral areas or setting the maximal number of ancestral areas to different values. No difference was observed among results of these different parameterizations, except at the basal node of the tree. The analysis was run several times, varying the parameters for extinction, dispersal, vicariance and duplication, with no change in the results.

As DIVA analysis is prone to errors, notably when the biogeographic history includes numerous dispersal events,

and as it is prone to finding very widespread ancestors towards the base of the tree (Ronquist, 1996; Clark *et al.*, 2008; Kodandaramaiah, 2009), we also carried out an analysis of ancestral ranges using LAGRANGE, a reconstruction method of the ancestral areas inherited at each node based on a dispersal–extinction–cladogenesis model, and allowing model comparison (Ree & Smith, 2008). We used the tree generated by MRBAYES (which did not differ in topology from the maximum likelihood tree, with polytomies reduced to 0-length branches using package APE in R, Paradis, 2004), and set equal coefficients (=1.0) in the adjacency and in the dispersal matrices. This allows species to be co-distributed freely among regions, and to disperse with equal probabilities among them. This is probably an overestimate for Mesoamerica, and therefore our analysis is likely to be biased toward overestimating the probabilities of branches being present in Mesoamerica. The maximum number of ancestral areas was set to four. For each branch and each of the areas except the pampa (which was never found to be ancestral to any branch), the relative probabilities of the branch being present in the area of interest were summed among the most likely models (those within four points of log-likelihood of the best one). Therefore, the probabilities shown on the rectangles on the branches of the tree in Figure 2 may sum to more than 1, if the most likely models inferred the branch to belong to several areas.

Data on life history traits were collected from descriptions found in Rogers & Appan (1973) and on herbarium sheets. The traits considered were linked to dispersal (fruit dehiscence, seed size, presence of a caruncle on the seed), capacity to resprout after disturbance (presence of a tuberous root), growth form and habitat. Reconstruction of ancestral characters was carried out using MESQUITE (Maddison & Maddison, 2010). Characters coded as binary variables were reconstructed using the maximum likelihood Mk1 model; other characters were analysed using maximum parsimony.
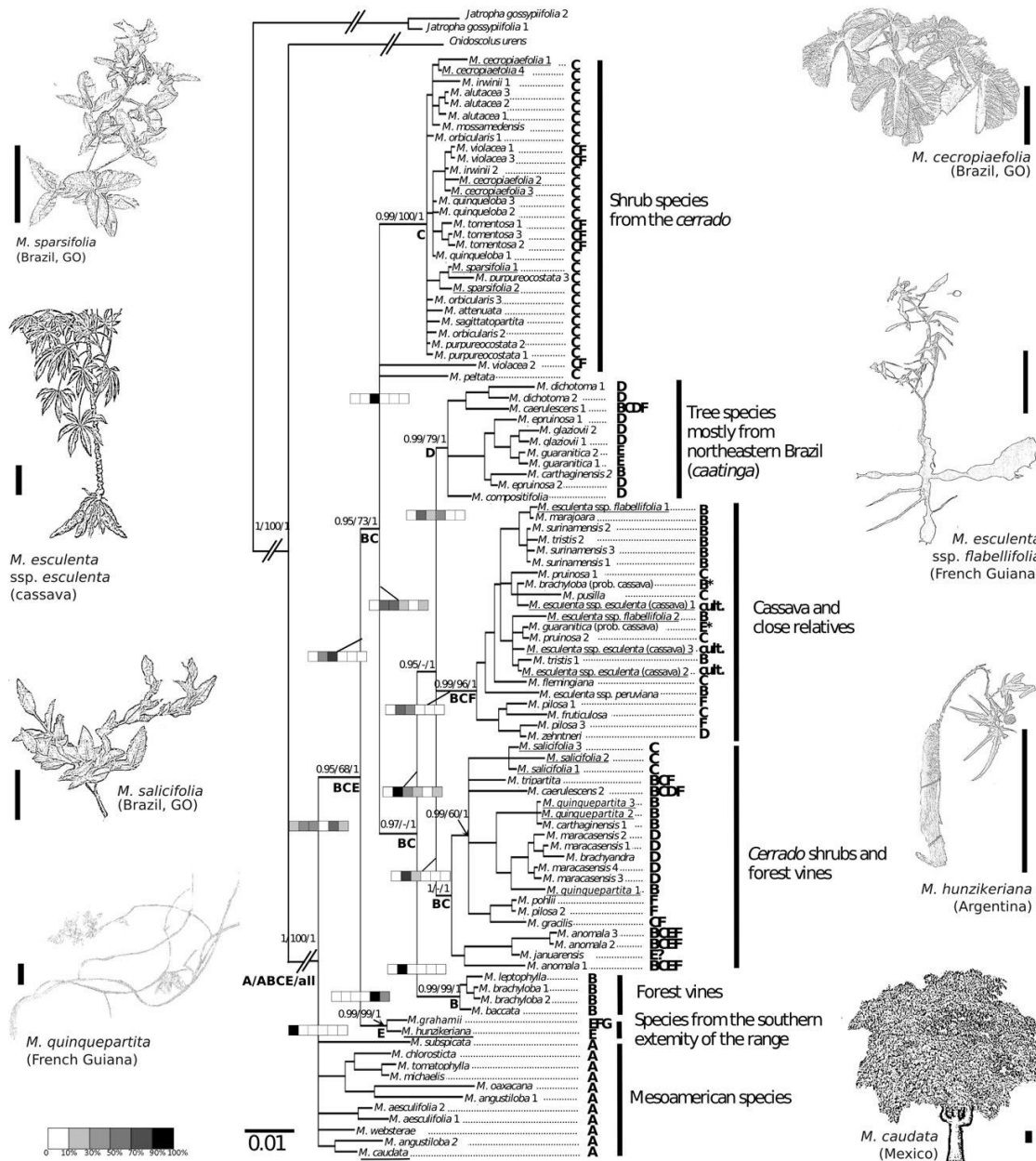
**Figure 2.** Phylogeny of *Manihot*, based on sequences of the two nuclear genes (*NIA* and G3*pdh*). Bayesian tree in which the branches supported by posterior probabilities (PP) < 0.95 were collapsed. Posterior probability / bootstrap percentages (when > 50 %) / approximate likelihood-ratio test (aLRT) score of the maximum likelihood tree are indicated next to the nodes. Dispersal–vicariance analysis (DIVA) results are shown below nodes (area codes correspond to those on Figure 1), and LAGRANGE results are presented as rectangles on the branches. The relative probabilities of each branch being present in each area, summed over the most likely models (within four units of log-likelihood of the best model), are shown on a grey scale (legend at the bottom left), with areas arranged from left to right as ABCDEF (Mesoamerica, Amazon and northern South America, *cerrado*, *caatinga*, Paraná, chaco). Line drawings illustrate some of the life forms encountered in the genus (species represented by the drawings are underlined in the tree; the specimens pictured here do not necessarily correspond to individuals included in our analysis). Scale bars: 20 cm. These drawings were prepared following photographs: *M. caudata* Greenm.: photo by Mark Olson, http://www.explorelifeonearth.org/manihot.html; *M. hunzikeriana*: herbarium sample at P (Krapovickas *et al.*, 16742); *M. quinquepartita*: photo courtesy of Léa Ménard; other photographs used were taken by the authors.

## Results

The three-gene alignment resulted in 3065 bp (2629 bases when only *Manihot* was included), 419 (15.9 %) of which were variable and 207 (7.9 %) of which were parsimony-informative within *Manihot*. Tree topology was similar based on the two-gene dataset (Fig. 2), including only both nuclear genes) and three-gene (Fig. S1) datasets. Maximum Likelihood and Bayesian reconstructions led to the same topology when poorly supported branches [Bootstrap (BP) < 50 % and posterior probabilities (PP) < 0.95] were collapsed (Fig. 2).

As found in earlier studies (Chacón *et al.*, 2008), the genus is monophyletic. Species from Mesoamerica form a grade (Fig. S1) with low support, but appearing in both topologies (ML and Bayesian). The low-support nodes have been reduced to polytomies in Fig. 2, but can be seen in Fig. S1. South American species form a clade nested within this grade. DIVA analysis shows that the genus most likely originated in Mesoamerica, or from a widespread ancestor. Due to the lack of data on a very closely related outgroup, LAGRANGE estimates are not available for the basal node of *Manihot*. Species from the southern extremity of the range (*M. grahamii* Hook. and *M. hunzikeriana* Mart. Crov.) form a supported clade, sister to all other South American samples.

This latter group contains several well-supported clades (Fig. 2):

1. a well-supported clade (BP = 100 %, PP = 0.99, aLRT = 1) of shrubby species endemic to the *cerrado* of central Brazil, originating in the *cerrado*;

2. a clade grouping *cerrado* shrubs and forest vines (BP = 44 %, PP = 1, aLRT = 1), originating in the *cerrado* and/or the Amazon or northern South America;

3. a group of vines (BP = 99 %, PP = 0.99, aLRT = 1) found in forest gaps;

4. a clade of mostly tree species endemic to the *caatinga* (northeastern Brazil; BP = 79 %; PP = 0.99, aLRT = 1);

5. and a well-supported clade (BP = 96 %, PP = 0.99, aLRT = 1) that includes cassava and its closest relatives, originating in the cerrado and/or the Amazon or northern South America.

Reconstruction of ancestral characters revealed that most show convergent evolution (Fig. S2). While the genus was ancestrally of shrubby habit, vines and trees evolved independently several times, both in Mesoamerica and in South America, sometimes followed by a reversion towards a shrubby habit (Fig. S2a). Fruits were ancestrally dehiscent, but a small number of species independently evolved towards indehiscent fruits (Fig. S2b). Most Mesoamerican species have large seeds with a relatively inconspicuous elaiosome, which probably reflects the ancestral character state. A large number of South American species, notably those from the *cerrado* area, evolved smaller seeds, usually with large elaiosomes, which probably facilitated myrmechochory (Figs S2c, S2d). Convergent evolution is most striking in several lineages that have independently colonized forest-gap habitats and evolved a suite of adaptations to them: pronounced viny habit, large seeds lacking an elaiosome, and indehiscent fruits (Figs S2a-S2d). Some *Cnidoscolus* and *Jatropha* species are known to have tuberous roots, a character that has been lost several times in *Manihot*, notably among forest species (Figs S2e, S2f).

## Discussion

The genus *Manihot* is monophyletic and may be of Mesoamerican origin. Using DNA markers, the phylogeny was not fully resolved, but clades of species that partly correspond to biogeographic units and ecological strategies were supported. We identified the closest relatives of cassava, among which are *M. tristis* Muell. Arg., *M. surinamensis* Rogers & Appan, *M. marajoara* Huber, *M. pruinosa* Pohl, *M. flemingiana* Rogers & Appan, *M. pilosa* Pohl and the wild subspecies of *M. esculenta*.

### Taxonomy of the genus

For several species, we were able to obtain sequences from two or more specimens, and often found either that species were not monophyletic for these gene phylogenies (*e.g.*, *cerrado* species), or that they fell into two different clades (e.g., *M. quinquepartita* Huber ex Rogers & Appan). These observations underline the plasticity of *Manihot* species. For example, the wild progenitor of cassava, *Manihot esculenta* ssp. *flabellifolia* (Olsen & Schaal, 1999; Léotard *et al.*, 2009) can grow either as a vine or as a shrub, depending on its environment (A. Duputié, D. McKey, pers. obs.), and even its seedling morphology shows plasticity (Pujol *et al.*, 2005). This can make species identification difficult, notably when determining herbarium specimens. Although a thorough study of herbarium material is necessary for revision of the genus, it alone is unlikely to resolve species delimitations and biosystematic studies are needed. The genus appears to have diversified rapidly, and equivocal morphological delimitation of related species (e.g. in the clade of *cerrado* species) may reflect recurrent hybridization events, ongoing speciation, or ancestral polymorphism. The same complexities may also explain cases in which species with very distinctive morphology (such as *M. purpureocostata* Pohl, which has simple entire leaves) are not discriminated from others by the phylogenetic analyses (Fig. 2).

### Biogeography of *Manihot*

Our analysis shows that *Manihot* is probably of Mesoamerican origin: Mesoamerican species form a grade within which is nested the clade of all South American species

(Fig. S1). However, the methods of ancestral area reconstruction could not definitely confirm this result: when restricted to one ancestral area, DIVA analysis inferred Mesoamerica to be the ancestral area of *Manihot*, but when multiple ancestral areas were allowed, it inferred a more widespread ancestral area. The dispersal–extinction–cladogensis (DEC) approach led to interesting results within the tree, but without inclusion of an outgroup close enough to *Manihot*, the origin of the genus could not be inferred.

Several observations suggest that *Manihot* colonized South America only recently, perhaps only just before the formation of the Isthmus of Panama, 3.5 Ma. First, reproductive isolation does not seem to be complete, even between relatively distantly related South American species, such as *M. glaziovii* Muell. Arg. and cassava (Lefèvre & Charrier, 1993). Even though reproductive isolation is not always correlated with time since divergence (e.g. Parks & Wendel, 1990; Moyle *et al.*, 2004), this observation seems consistent with a relatively young age for the genus. Furthermore, genetic variability for chloroplast markers was very low within the South American clade (Duputié, 2008). In addition, *Manihot* species are not present in Caribbean islands (except for *M. brachyloba* on Hispaniola, where it may have been introduced by the Amerindians), suggesting that *Manihot* species cannot disperse over large bodies of water (of the order of 100 km). *Manihot* should thus have been able to cross the strait of Panama only when it was less than a few dozen km wide (later than about 7 Ma, see Coates & Obando, 1996; Kirby *et al.*, 2008), or even after the completion of the landbridge (about 3.5 Ma, Coates & Obando, 1996; Burnham & Graham, 1999). The hypothesis of a relatively recent colonization of South America is also consistent with the young age recently inferred for the genus by Chacón *et al.* (2008) and with the proposed age of 4-8 Ma for diverse *cerrado* clades, which could have diversified *in situ* from species adapted to wetter environments (Pennington *et al.*, 2006; Simon *et al.*, 2009). None of these observations is by itself sufficient to support our claim, but their conjunction strongly suggests a recent origin of the genus, and an even more recent colonization of South America.

The completion of the Isthmus of Panama landbridge was a long process, extending over 12 Myr and ending with the complete closure of the Isthmus around 3.5 Ma (Coates & Obando, 1996). Whether the Mesoamerican peninsula existed over a long period, or consisted of islands until the completion of the isthmus, remains a matter of debate (Kirby *et al.*, 2008). In any case, the completion of the land bridge provided new habitats for taxa originating both in North and South America. This exchange of taxa has been well documented for mammal species, termed the "Great American Interchange" (reviewed in MacFadden, 2006; Webb, 2006), but the dispersal of plant species was also facilitated (e.g. Burnham & Graham, 1999; Weir *et al.*, 2009). Although migrations occurred in both directions (Burnham & Graham, 1999), most plant genera studied so far colonized

South America from the north: *Lupinus* L. species (Fabaceae) colonized the Andes from western North America (Hughes & Eastwood, 2006), and the species-rich forest genus *Guatteria* Ruiz & Pav. (Annonaceae) was found to be of Mesoamerican origin, and to have started diversifying in South America about 6 Ma (Erkens *et al.*, 2007). The same scenario seems to apply to *Manihot* (Chacón *et al.*, 2008; this study). This estimated date precedes the closure of the Isthmus of Panama. This discrepancy could reflect imprecision in date estimation based on molecular data. An alternative hypothesis is that *Guatteria*, *Manihot* and other plants were able to cross the narrowing gap between Mesoamerica and South America before complete closure of the Isthmus of Panama, as has been documented for some passerine species (Weir *et al.*, 2009) and also suggested for plants (Burnham & Graham, 1999; Pennington & Dick, 2004; Pennington *et al.*, 2004; Saslis-Lagoudakis *et al.*, 2008; Ireland *et al.*, 2010). It should be noted that some of these genera (e.g. *Ateleia* (DC.) Benth., Fabaceae) may be wind-pollinated and wind-dispersed (Ireland *et al.*, 2010), so they may have greater ability for long-distance dispersal than does *Manihot*. The pattern for *Manihot* resembles that observed for *Ruprechtia* C.A.Mey (Polygonaceae; Pennington *et al.*, 2004), which seems to have crossed the seaway around 4.1 Ma. In other genera, the pattern of colonization is more complex, including several independent dispersals and/or vicariant events between the two continents, or secondary recolonization of the continent of origin (*Krameria* L., Krameriaceae, *Hoffmannseggia* Cav. and *Pomaria* Cav., both Fabaceae, Simpson *et al.*, 2004, 2005, 2006).

A notable result of our analysis is that *Manihot* species are phylogenetically clustered according to geography and ecology: shrub species from the *cerrado* of central Brazil form a well-supported clade, vines from the Amazon form another, and trees endemic to north-eastern Brazil form a third clade. From our data, it is not possible to analyse phylogenetic community structure finely (Pennington *et al.*, 2009), but *Manihot* species clearly show phylogenetic clustering in similar habitats. This can be due either to habitat filtering or to *in situ* diversification (Vamosi *et al.*, 2009).

The genus *Manihot* seems to have reached rapidly the southern part of its range, radiating in the *cerrado* of central Brazil, as shrubs restricted to seasonally dry environments, and in all open habitats of lowland tropical South America (eastward to the *caatinga*, southward to dry areas, and northwards and westwards into forest gaps). If the northern rim of the Amazon basin had been colonized very early, there would have been an opportunity for speciation to take place north of the Amazon, as happened in the rodent genus *Calomys* Waterhouse (Muridae), which is restricted to dry Neotropical environments (Almeida *et al.*, 2007). This does not seem to have occurred in the genus *Manihot*, as haplotypes of species found north and south of the Amazon show low intra-specific but large interspecific divergence. This suggests either that *Manihot* species colonized the

northern rim of the Amazon basin only recently, or that they were, until recently, connected to populations located further south, via corridors of dry forest or savanna (Pennington *et al.*, 2004), either along the Atlantic coast or across central Amazonia (Quijada-Mascareñas *et al.*, 2007).

The radiation of *Manihot* in South America implies changes in habitat, in growth habit, and in several traits such as dispersal mode and perhaps the development of tuberous roots (although this trait might be plastic, at least in some species; D.M., pers. obs., in Brazil and French Guiana). Adaptation to disturbance might be the key ecological feature driving *Manihot* evolution, with most species having a starchy tuberous root, allowing for regrowth after disturbance. According to Simon *et al.* (2009), such adaptation to frequent fires may have occurred quite rapidly in diverse genera that assembled a few million years ago to form the *cerrado* flora.

From our data, the ancestor of *Manihot* is inferred to have been characterized by small seeds with inconspicuous elaiosomes, but this conclusion has to be tempered, because most species of the closely related genus *Cnidoscolus* have large seeds bearing a large elaiosome (Leal *et al.*, 2007). A large number of *Manihot* species seem to have evolved larger elaiosomes and smaller seeds, facilitating dispersal and burial by ants, thereby protecting them from fire (as well as from granivores). Dormancy is often broken by elevated soil temperature (Pujol *et al.*, 2002; Duputié, 2008) due to canopy opening or to fires. In forest species, a different combination of traits, probably associated with dispersal by frugivorous vertebrates, evolved independently several times: indehiscent fruits and larger seeds lacking an elaiosome.

### Cassava's closest wild relatives

Our analysis confirms that several species are close relatives of cassava: the two wild subspecies of *M. esculenta* (Roa *et al.*, 1997, 2000; Olsen & Schaal, 1999; Allem *et al.*, 2001; Olsen, 2004), *M. pruinosa* (Olsen & Schaal, 1999), *M. tristis* (Roa *et al.*, 2000), *M. surinamensis* and *M. marajoara*. Allem (1994) considered the latter three species to be synonyms of *M. esculenta* ssp. *flabellifolia*, and our data are consistent with this treatment. According to Allem *et al.* (2001), *Manihot pruinosa* is the closest relative of *M. esculenta*. Our results certainly support the close relationship of these species; indeed, they do not differentiate them.

Two specimens identified as other species (*M. guaranitica* Chodat & Hassl. and *M. brachyloba*) also grouped within the clade containing cassava and its closest wild relatives. Both specimens were from plants found in cultivation or in sympatry with cassava. As other accessions of the same taxa, collected from the wild, were in different clades, both these specimens are misidentified cassava accessions.

Two of the three accessions of *M. pilosa* and one of *M. zehntneri* Ule (all four of them initially determined by A. C. Allem as *M. pilosa*) also figured among cassava's closest relatives, consistent with Allem (1999b). The fourth accession (*M. pilosa* 2) strongly differs in morphology from the other three, and was in a different clade. Either specimens labelled *M. pilosa* include two distinct species, one of them being very close to cassava, or this "species" is genetically heterogeneous, perhaps due to hybridization.

Unexpectedly, the only sequenced specimens of *M. pusilla* Pohl, *M. fruticulosa* (Pax) Rogers & Appan and *M. flemingiana* also appeared among cassava's closest relatives. Sequencing of other accessions of these taxa will help determine whether their placement reflects misidentifications, laboratory errors, or biological reality.

Finally, Allem (1999b, 2001) considered some species to be only somewhat more distantly related to cassava, belonging to its "secondary gene pool": *M. anomala* Pohl, *M. brachyloba*, *M. dichotoma* Ule, *M. epruinosa* Pax & Hoffm., *M. glaziovii* Muell. Arg., *M. gracilis* Pohl, *M. leptophylla* Pax, *M. pohlii* Wavra, *M. tripartita* (Sprengel) Muell. Arg., *M. triphylla* Pohl, and even two Mesoamerican species: *M. aesculifolia* (H.B.K.) Pohl and *M. chlorosticta* Standley & Goldman. Our results show that they are no more closely related to cassava than are other *Manihot* species outside the clade of the crop's closest relatives.

In conclusion, the genus *Manihot* is likely to be of Mesoamerican origin, and to have secondarily diversified in South America, colonizing all available types of lowland seasonally dry environments. Even though some uncertainties remain, we identified the closest wild relatives of cassava. Determining whether the observed lack of species-level resolution in several clades reveals ancestral polymorphism not yet sorted by coalescent processes or hybridization between closely related species, or whether it simply reflects inaccurate species determinations or inaccurate assessment of species boundaries, stemming from phenotypic plasticity, will require not only further phylogenetic reconstructions, but also extensive field biosystematic studies.

### Acknowledgements

## References

Allem, A.C. (1978) Notas taxonômicas e novos sinônimos em espécies de *Manihot* – II (Euphorbiaceae). *Revista Brasileira de Biológia*, **38**, 721-726.

Allem, A.C. (1979) Notas taxonômicas e novos sinônimos em espécies de *Manihot* – V (Euphorbiaceae). *Revista Brasileira de Biológia*, **39**, 891-896.

Allem, A.C. (1994) The origin of *Manihot esculenta* Crantz (Euphorbiaceae). *Genetic Resources and Crop Evolution*, **41**, 133-150.

Allem, A.C. (1999a) A new species of *Manihot* (Euphorbiaceae) from the Brazilian Amazon. *International Journal of Plant Sciences*, **160**, 181-187.

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

Allem, A.C., Mendes, R.A., Salomão, A.N. & Burle, M.L. (2001) The primary gene pool of cassava (*Manihot esculenta* Crantz subspecies *esculenta*, Euphorbiaceae). *Euphytica*, **120**, 127-132.

Almeida, F.C., Bonvicino, C.R. & Cordeiro-Estrela, P. (2007) Phylogeny and temporal diversification of *Calomys* (Rodentia, Sigmodontinae): implications for the biogeography of an endemic genus of the open/dry biomes of South America. *Molecular Phylogenetics and Evolution*, **42**, 449-466.

Anisimova, M. & Gascuel, O. (2006) Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Systematic Biology*, **55**, 539-52.

Bonaccorso, E., Koch, I. & Peterson, A.T. (2006) Pleistocene fragmentation of Amazon species' ranges. *Diversity and Distributions*, **12**, 157-164.

Burnham, R.J. & Graham, A. (1999) The history of Neotropical vegetation: new developments and status. *Annals of the Missouri Botanical Garden*, **86**, 546-589.

Bush, M.B. (1994) Amazonian speciation: a necessarily complex model. *Journal of Biogeography*, **21**, 5-17.

Caetano, S., Prado, D., Pennington, R.T., Beck, S., Oliveira, A., Spichiger, R. & Naciri, Y. (2008) The history of seasonally dry tropical forests in eastern South America: inferences from the genetic structure of the tree *Astronium urundeuva* (Anacardiaceae). *Molecular Ecology*, **17**, 3147-3159.

Chacón, J., Madriñán, S., Debouck, D., Rodriguez, F. & Tohme, J. (2008) Phylogenetic patterns in the genus *Manihot* (Euphorbiaceae) inferred from analyses of nuclear and chloroplast DNA regions. *Molecular Phylogenetics and Evolution*, **49**, 260-267.

Clark, J.R., Ree, R.H., Alfaro, M.E., King, M.G., Wagner, W.L. & Roalson, E.H. (2008) A comparative study in ancestral range reconstruction methods: retracing the uncertain histories of insular lineages. *Systematic Biology*, **57**, 693-707.

Coates, A.G. & Obando, J.A. (1996) The geologic evolution of the Central American isthmus. *Evolution and environment in tropical America* (ed. by J.B.C. Jackson, A.F. Budd and A.G. Coates), pp. 21-56. The University of Chicago Press, Chicago, IL.

Colinvaux, P.A., de Oliveira, P.E. & Bush, M.B. (2000) Amazonian and neotropical plant communities on glacial time-scales: the failure of the aridity and refuge hypotheses. *Quaternary Science Reviews*, **19**, 141-169.

Duputié, A. (2008) *From the radiation of* Manihot *within a diversity hotspot to the domestication of cassava*. PhD Dissertation, University Montpellier II, Montpellier, France.

Edgar, R.C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, **5**, 113.

Elias, M. & McKey, D. (2000) The unmanaged reproductive ecology of domesticated plants in traditional agroecosystems: an example involving cassava and a call for data. *Acta Oecologica*, **21**, 223-230.

Erkens, R.H.J., Chatrou, L.W., Maas, J.W., van der Niet, T. & Savolainen, V. (2007) A rapid diversification of rainforest trees (*Guatteria*; Annonaceae) following dispersal from Central into South America. *Molecular Phylogenetics and Evolution*, **44**, 399-411.

FAOSTAT (2009) *Agricultural production data*. FAO, Rome, Italy. Available at: http://faostat.fao.org/site/339/default.aspx.

Guindon, S. & Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, **52**, 696-704.

Haffer, J. & Prance, G.T. (2001) Climatic forcing of evolution in Amazonia during the Cenozoic: on the refuge theory of biotic differentiation. *Amazoniana*, **16**, 579-605.

Hall, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95-98.

Howarth, D.G. & Baum, D.A. (2002) Phylogenetic utility of a nuclear intron from nitrate reductase for the study of closely related plant species. *Molecular Phylogenetics and Evolution*, **23**, 525-528.

Huelsenbeck, J.P. & Ronquist, F. (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, **17**, 754-755.

Hughes, C. & Eastwood, R. (2006) Island radiation on a continental scale: exceptional rates of plant diversification after uplift of the Andes. *Proceedings of the National Academy of Sciences USA*, **103**, 10334-10339.

Ireland, H.E., Kite, G.C., Veitch, N.C., Chase, M.W., Schrire, B., Lavin, M., Linares, J. & Pennington, R.T. (2010) Biogeographical, ecological and morphological structure in a phylogenetic analysis of *Ateleia* (Swartzieae, Fabaceae) derived from combined molecular, morphological and chemical data. *Botanical Journal of the Linnean Society*, **162**, 39-53.

Kirby, M.X., Jones, D.S. & MacFadden, B.J. (2008) Lower Miocene stratigraphy along the Panama Canal and its bearing on the Central American peninsula. *PLoS ONE*, **3**, e2791.

Kodandaramaiah, U. (2009) Use of dispersal–vicariance analysis in biogeography – a critique. *Journal of Biogeography*, **37**, 3-11.

Leal, I.R., Wirth, R. & Tabarelli, M. (2007) Seed dispersal by ants in the semi-arid *caatinga* of North-East Brazil. *Annals of Botany*, **99**, 885-894.

Lefèvre, F. & Charrier, A. (1993) Isozyme diversity within African *Manihot* germplasm. *Euphytica*, **66**, 73-80.

Léotard, G., Duputié, A., Kjellberg, F., Douzery, E.J.P., Debain, C., de Granville, J.J. & McKey, D. (2009) Phylogeography and the origin of cassava: new insights from the northern rim of the Amazon basin. *Molecular Phylogenetics and Evolution*, **53**, 329-334.

MacFadden, B.J. (2006) Extinct mammalian biodiversity of the ancient New World tropics. *Trends in Ecology and Evolution*, **21**, 157-165.

Maddison, W.P. & Maddison, D.R. (2010) Mesquite: a modular system for evolutionary analysis. Version 2.73. Available at: http://mesquiteproject.org..

Morrone, J.J. (2006) Biogeographic areas and transition zones of Latin America and the Caribbean islands based on panbiogeographic and cladistic analyses of the entomofauna. *Annual Review of Entomology*, **51**, 467-94.

Moyle, L.C., Olson, M.S. & Tiffin, P. (2004) Patterns of reproductive isolation in three Angiosperm genera. *Evolution*, **58**, 1195-1208.

Olsen, K.M. (1999) Minisatellite variation in a single-copy nuclear gene: phylogenetic assessment of repeat length homoplasy and mutational mechanism. *Molecular Biology and Evolution*, **16**, 1406-1409.

Olsen, K.M. (2004) SNPs, SSRs and inferences on cassava's origin. *Plant Molecular Biology*, **56**, 517-526.

Olsen, K.M. & Schaal, B.A. (1999) Evidence on the origin of cassava: phylogeography of *Manihot esculenta*. *Proceedings of the National Academy of Sciences USA*, **96**, 5586-5591.

Paradis, E. (2004) APE: analyses of phylogenetics and evolution in R language.

*Bioinformatics*, **20**, 289-290.

Parks, C.R. & Wendel, J.F. (1990) Molecular divergence between Asian and North American species of *Liriodendron* (Magnoliaceae) with implications for interpretation of fossil floras. *American Journal of Botany*, **77**, 1243-1256.

Pennington, R.T. & Dick, C.W. (2004) The role of immigrants in the assembly of the South American rainforest tree flora. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **359**, 1611-1622.

Pennington, R.T., Prado, D.E. & Pendry, C.A. (2000) Neotropical seasonally dry forests and Quaternary vegetation changes. *Journal of Biogeography*, **27**, 261-273.

Pennington, R.T., Lewis, G.P. & Ratter, J.A. (2006) Neotropical savannas and seasonally dry forests: plant diversity, biogeography and conservation. *Neotropical savannas and seasonally dry forests: plant diversity, biogeography and conservation* (ed. by R.T. Pennington, G.P. Lewis & J.A. Ratter), pp. 1-29. CRC Press, Boca Raton, FL.

Pennington, R.T., Lavin, M. & Oliveira-Filho, A. (2009) Woody plant diversity, evolution and ecology in the tropics: perspectives from seasonally dry tropical forests. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 437-57.

Pennington, R.T., Lavin, M., Prado, D.E., Pendry, C.A., Pell, S.K. & Butterworth, C.A. (2004) Historical climate change and speciation: neotropical seasonally dry forest plants show patterns of both Tertiary and Quaternary diversification. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **359**, 515-537.

Pujol, B., Gigot, G., Laurent, G., Pinheiro-Kluppel, M., Elias, M., Hossaert-McKey, M. & McKey, D. (2002) Germination ecology of cassava (*Manihot esculenta* Crantz, Euphorbiaceae) in traditional agroecosystems: seed and seedling biology of a vegetatively propagated domesticated plant. *Economic Botany*, **56**, 366-379.

Pujol, B., Mühlen, G.S., Garwood, N., Horoszowski, Y., Douzery, E.J.P. & McKey, D. (2005) Evolution under domestication: contrasting functional morphology of seedlings in domesticated cassava and its closest wild relatives. *New Phytologist*, **166**, 305-318.

Quijada-Mascareñas, J.A., Ferguson, J.E., Pook, C.E., da Graça Salomão, M., Thorpe, R.S. & Wüster, W. (2007) Phylogeographic patterns of trans-Amazonian vicariants and Amazonian biogeography: the Neotropical rattlesnake (*Crotalus durissus* complex) as an example. *Journal of Biogeography*, **34**, 1296-1312.

Ree, R.H. & Smith, S.A. (2008) Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology*, **57**, 4-14.

Renard, D., Schatz, B. & McKey, D. (2010) Ant nest architecture and seed burial depth: implications for seed fate and germination success in a myrmecochorous savanna shrub. *Écoscience*, **17**, 194-202.

Richardson, J.E., Pennington, R.T., Pennington, T.D. & Hollingsworth, P.M. (2001) Rapid diversification of a species-rich genus of neotropical rain forest trees. *Science*, **293**, 2242-2245.

Roa, A.C., Maya, M.M., Duque, M.C., Tohme, J., Allem, A.C. & Bonierbale, M.W. (1997) AFLP analysis of relationships among cassava and other *Manihot* species. *Theoretical and Applied Genetics*, **95**, 741-750.

Roa, A.C., Chavarriaga-Aguirre, P., Duque, M.C., Maya, M.M., Bonierbale, M.W., Iglesias, C. & Tohme, J. (2000) Cross-species amplification of cassava (*Manihot esculenta*) (Euphorbiaceae) microsatellites: allelic polymorphism and degree of relationship. *American Journal of Botany*, **87**, 1647-1655.

Rogers, D.J. & Appan, S.G. (1973) Manihot, *Manihotoides (Euphorbiaceae)*. *Flora Neotropica. Monograph 13*. Haffer Press, New York.

Rogers, S.O. & Bendich, A.J. (1985) Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant tissues. *Plant Molecular Biology*, **5**, 69-76.

Ronquist, F. (1996) *DIVA version 1.1.* Computer program and manual available by anonymous FTP from Uppsala University Available at: (http://www.ebc.uu.se/systzoo/research/diva/diva.html).

Saslis-Lagoudakis, C., Chase, M.W., Robinson, D.N., Russell, S.J. & Klitgaard, B.B. (2008) Phylogenetics of Neotropical *Platymiscium* (Leguminosae: Dalbergieae): systematics, divergence times, and

biogeography inferred from nuclear ribosomal and plastid DNA sequence data. *American Journal of Botany*, **95**, 1270-1286.

Second, G., Allem, A.C., Mendes, R.A., Carvalho, L.J.C.B., Emperaire, L., Ingram, C. & Colombo, C. (1997) Molecular marker (AFLP)-based *Manihot* and cassava numerical taxonomy and genetic structure analysis in progress: implications for their dynamic conservation and genetic mapping. *African Journal of Root and Tuber Crops*, **2**, 140-147.

Simon, M.F., Grether, R., de Queiroz, L.P., Skema, C., Pennington, R.T. & Hughes, C.E. (2009) Recent assembly of the Cerrado, a neotropical plant diversity hotspot, by in situ evolution of adaptations to fire. *Proceedings of the National Academy of Sciences USA*, **106**, 20359-64.

Simpson, B., Larkin, L., Weeks, A. & McDill, J. (2006) Phylogeny and biogeography of *Pomaria* (Caesalpinioideae: Leguminosae). *Systematic Botany*, **31**, 792-804.

Simpson, B.B., Tate, J.A. & Weeks, A. (2005) The biogeography of *Hoffmannseggia* (Leguminosae, Caesalpinioideae, Caesalpinieae): a tale of many travels. *Journal of Biogeography*, **32**, 15-27.

Simpson, B.B., Weeks, A., Helfgott, M. & Larkin, L.L. (2004) Species relationships in *Krameria* (Krameriaceae) based on ITS sequences and morphology: implications for character utility and biogeography. *Systematic Botany*, **29**, 97-108.

Strand, A.E., Leebens-Mack, J. & Milligan, B.G. (1997) Nuclear DNA-based markers for plant evolutionary biology. *Molecular Ecology*, **6**, 113-118.

Taberlet, P., Gielly, L., Pautou, G. & Bouvet, J. (1991) Universal primers for amplification of three noncoding regions of chloroplast DNA. *Plant Molecular Biology*, **17**, 1105-1109.

Vamosi, S.M., Heard, S.B., Vamosi, J.C. & Webb, C.O. (2009) Emerging patterns in the comparative analysis of phylogenetic community structure. *Molecular Ecology*, **18**, 572-592.

Webb, S.D. (2006) The Great American Biotic Interchange: patterns and processes. *Annals of the Missouri Botanical Garden*, **93**, 245-257.

Weir, J.T., Bermingham, E. & Schluter, D. (2009) The Great American Biotic Interchange in birds. *Proceedings of the National Academy of Sciences USA*, **106**, 21737-21742.

## SUPPORTING INFORMATION

*Additional Supporting Information may be found in the online version of this article:*

**Table S1** List of the samples used for the molecular phylogeny.

**Figure S1** Phylogeny of *Manihot*, based on the sequences of all three genes, obtained by Bayesian reconstruction (the maximum likelihood reconstruction led to the same topology).

**Figure S2** Reconstruction of character history: (a) growth habit, (b) fruit dehiscence, (c) size of the elaiosome, (d) seed size, (e) presence of a starchy root or tuber, (f) habitat.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

*BIOSKETCHES*

**Anne Duputié** is a postdoctoral fellow studying the evolution of tree distribution ranges in a changing environment. The work presented here is part of her PhD, dealing with the genetic and ecological aspects of the domestication of cassava.

**Jan Salick** is a professor at Washington University and University of Missouri in Saint Louis, and the curator of ethnobotany at Missouri Botanical Garden. She is interested in ethnobotany, agroecology, conservation biology and tropical ecology.

**Doyle McKey** is a senior researcher in tropical ecology. He is particularly interested in coevolution between plants and insects, and between humans and their crops.

Author contributions: All authors designed the work and collected plant material; J.S. coordinated collection of plant material from herbaria in the USA (DAV, MO, NY); A.D. performed the laboratory work and the analyses; A.D. and D.M. wrote the article.

Editor: Pauline Ladiges