ORIGINAL ARTICLE On the origin of sweet potato (*Ipomoea batatas* (L.) Lam.) genetic diversity in New Guinea, a secondary centre of diversity

C Roullier^{1,2}, R Kambouo³, J Paofa³, D McKey^{2,4} and V Lebot¹

New Guinea is considered the most important secondary centre of diversity for sweet potato (*Ipomoea batatas*). We analysed nuclear and chloroplast genetic diversity of 417 New Guinea sweet potato landraces, representing agro-morphological diversity collected throughout the island, and compared this diversity with that in tropical America. The molecular data reveal moderate diversity across all accessions analysed, lower than that found in tropical America. Nuclear data confirm previous results, suggesting that New Guinea landraces are principally derived from the Northern neotropical genepool (*Camote* and *Batata* lines, from the Caribbean and Central America). However, chloroplast data suggest that South American clones (early *Kumara* line clones or, more probably, later reintroductions) were also introduced into New Guinea and then recombined with existing genotypes. The frequency distribution of pairwise distances between New Guinea landraces suggests that sexual reproduction, rather than somaclonal variation, has played a predominant role in the diversification of sweet potato. The frequent incorporation of plants issued from true seed by farmers, and the geographical and cultural barriers constraining crop diffusion in this topographically and linguistically heterogeneous island, has led to the accumulation of an impressive number of variants. As the diversification of sweet potato in New Guinea is primarily the result of farmers' management of the reproductive biology of their crop, we argue that on-farm conservation programmes that implement distribution of core samples (clones representing the useful diversity of the species) and promote on-farm selection of locally adapted variants may allow local communities to fashion relatively autonomous strategies for coping with ongoing global change.

Heredity (2013) 110, 594-604; doi:10.1038/hdy.2013.14; published online 27 March 2013

Keywords: chloroplast and nuclear SSRs; germplasm conservation; Papua New Guinea; secondary centre of diversity; sweet potato

INTRODUCTION

Crop genetic diversity is shaped by complex interactions between natural processes and those driven by humans. Diversity tends to be concentrated in particular parts of a crop's range (Pickersgill, 1998). Primary centres of diversity occur in areas where the plant was domesticated. Here, diversity is generated by long periods of interaction between humans and their plants, often with continued gene flow with the crop's wild relatives. Secondary centres of diversity occur in areas where the crop was introduced. In these areas, despite bottlenecks associated with introduction, the shorter period of cultivation and the usual absence of wild relatives, crops have produced huge diversity in a short time.

Our study focuses on the case of sweet potato (*Ipomoea batatas* (L.) Lam.) diversification in New Guinea. This large island is considered to be the most important secondary centre of genetic diversity for sweet potato, particularly the highlands region, where the total number of cultivars grown has been roughly estimated to be about 5000 (Yen, 1974; Bourke, 2009). Sweet potato is by far the most important crop in New Guinea. It strongly dominates agricultural production in the highland areas, and this was already the case when Europeans discovered the Central New Guinea highlands during the early 1900s. The dense human population encountered appeared to be

heavily dependent on *I. batatas* for food and as fodder for pigs, and the crop played a central role in cultural rituals such as ceremonial exchange systems (Ballard, 2005). It has been argued that sweet potato triggered an 'Ipomoean revolution' in this isolated region. Its adoption, largely replacing traditional crops such as taro (*Colocasia esculenta*), led to rapid demographic growth and transformed societies (Barrau, 1957; Yen, 1974; Ballard, 2005).

Sweet potato originated from tropical America. The nature of its introduction into New Guinea and the timing of its adoption—key points for understanding how its great diversity developed—are both unclear. According to archaeological, linguistic and historical data, sweet potato could have reached the Pacific by three different main paths (Barrau, 1957; Yen, 1974), corresponding to the introduction of clones from different geographical origins at different times. The *Kumara* line is based on the hypothesis of prehistoric introductions of clones by Polynesian sailors from coastal areas of South America to eastern Polynesia (Cook, Society and Marquesas Islands), around 1100–1200 AD, according to archaeological and linguistic evidence (Green, 2005). Sweet potato could have also reached Hawaii, New Zealand and Easter Island before early European contacts according archaeological evidences (Ladefoged *et al.*, 2005). The *Batata* line assumes that later, around 1550 AD, Portuguese explorers could have

E-mail: carolineroullier@hotmail.com

¹CIRAD-BIOS, AGAP, Montpellier, France; ²CEFE, UMR 5175, Montpellier, France; ³NARI, Southern Regional Centre, Central Province, Port Moresby, Papua New Guinea and ⁴Institut Universitaire de France, Montpellier, France

Correspondence: Dr C Roullier, CIRAD-BIOS, UMR AGAP, Avenue Agropolis, 34398 Montpellier Cédex 5, France.

Received 17 March 2012; revised 27 January 2013; accepted 28 January 2013; published online 27 March 2013

595

transferred West Indian cultivars to Africa, India and their colonies in the Moluccas in eastern Indonesia. Finally, the *Camote* line supposes that Spanish 'Acapulco-Manila' galleons could have spread sweet potato clones from Mexico to the Philippines very soon after Spanish conquest of these islands in 1522.

A recent molecular study (Roullier et al., 2011) analysed 329 landraces collected from Mexico to Peru by nuclear and chloroplast microsatellite markers. Both kinds of markers supported the existence of two geographically restricted genepools, corresponding to accessions from the Peru-Ecuador region of South America (hereafter called the Southern genepool) and accessions from the Caribbean and Central America region (hereafter called the Northern genepool). According to the tripartite hypothesis of sweet potato introduction into Oceania (Oceania corresponds here to the entire insular region between Asia and the Americas, including Polynesia, Micronesia and Melanesia.), the three lines introduced different genepools into different areas. On the basis of palynological and soil depositional evidence (In the absence of direct evidence for sweet potato in pollen records, palaeoecological studies have used sediment and pollen indicators of increased landscape clearance and degradation to infer its presence in New Guinea.), it has been suggested that sweet potato could have arrived in the New Guinea highlands in pre-Magellanic time (Golson, 1977; Gorecki, 1986; Haberle, 1998; Haberle and Atkin, 2005) via the Kumara line. Linguistic data have also been considered to support early introductions (Scaglion and Soto, 1994). In contrast, a recent review of the evidence suggests that sweet potato reached New Guinea only in the post-Magellanic period (around 1650), through local trade relations from the Moluccas to West New Guinea (Names for West New Guinea (the Indonesian part of New Guinea) are multiple and contested. Irian Java was generally used, but this name has recently been replaced by Papua. However, Papuans favouring independence refer instead to West Papua. We have chosen to call this region West New Guinea in this article.) (Ballard et al., 2005). On the basis of surveys of oral history (Wiessner, 2005; Bourke and Harwood, 2009) and on archaeological data (Bayliss-Smith et al., 2005), it has been proposed that sweet potato was introduced into the New Guinea highlands around 1700, where the plant was then relatively rapidly adopted and diffused. In the lowlands, the situation is different. It seems that although sweet potato may have been cultivated early in some parts of the northern coastal lowlands, introduction in most areas occurred after 1870 and its adoption as a major component of lowland agricultural systems has taken place since 1940, giving rise to a second 'Ipomoean revolution' (Bourke, 2005). New Guinea thus appears to be a potential convergence point of these three lines, where both neotropical genepools, Southern and Northern, could have been introduced at different times.

In the past decade, molecular markers have shown their potential to investigate origins of Pacific crops (Lebot, 1999; Kreike *et al.*, 2004; Clarke *et al.*, 2006, Hinkle, 2007; Gunn *et al.*, 2011; Perrier *et al.*, 2011), complementing archaeological and anthropological data. Molecular markers have already been used to analyse the genetic base of sweet potato from Oceania and New Guinea (Zhang *et al.*, 1998; Gichuki *et al.*, 2003; Zhang *et al.*, 2004), but studies thus far have used relatively small samples (fewer than 150 accessions). Their results show that varieties from Oceania are closely related to the Mexican accessions but only weakly related to those from Peru-Ecuador, suggesting that varieties from Oceania may be only of Mesoamerican origin.

Several sources of variation may contribute to the diversification of clonally propagated crops by farmers (McKey *et al.*, 2010). Somatic mutation is an important source of phenotypic variation in clonally

propagated crops, especially in the Pacific (Lebot, 1992; Sardos et al., 2008), and epigenetic inheritance may also contribute (McKey et al., 2010). In some clonally propagated crops, sexual reproduction is certainly an important mechanism in the generation of diversity (McKey et al., 2010). Variants selected by farmers are then maintained and multiplied by clonal propagation. Increasing evidence shows that many clonally propagated crops exhibit a mixed reproductive system in which evolutionary dynamics result from the interaction of clonal and sexual reproduction (Elias et al., 2001; Caillon et al., 2006; Scarcelli et al., 2006; Sardos et al., 2008; Delêtre, 2010). It has been reported that New Guinea farmers traditionally adopt sweet potato plants issued from true seed, resulting from the spontaneous germination of the numerous self-sown seeds the crop produces in New Guinean farming systems (Yen, 1974; Schneider, 1995). Incorporation of volunteer seedlings, selected by farmers across a wide array of ecological and cultural conditions throughout this large island, could be a major mechanism for creating diversity (Yen, 1974; Fajardo et al., 2002).

In Papua New Guinea, about 1000 accessions of sweet potato are maintained in ex situ collections at the Highlands Agricultural Experiment Station at Aiyura in the Eastern Highlands Province and at the Lowlands Agricultural Experiment Station at Keravat in East New Britain. Maintaining and evaluating such collections is laborious and expensive (Lebot et al., 2005). A comprehensive study of the genetic diversity of New Guinea sweet potato is needed to eliminate duplicates and to identify the diverse sources of germplasm in the core collections (Brown, 1989). We analysed a subsample of accessions from these collections, representative of the agro-morphological diversity they contain. Using both chloroplast and nuclear microsatellites, we analysed the extent of diversity existing in subsets of both the Highlands Agricultural Experiment Station and Lowlands Agricultural Experiment Station collections to characterize the distribution of allelic diversity throughout Papua New Guinea (samples available from West New Guinea were also included) and compare this diversity with that existing in tropical America. In particular, we analysed the origin of New Guinea sweet potato landraces in both highland and lowland agroecosystems. Finally, we attempted to explain the factors that contributed to the development of a secondary centre of genetic diversity in New Guinea.

MATERIALS AND METHODS

Plant materials and areas sampled

A total of 369 sweet potato landraces were collected from the Highlands Agricultural Experiment Station and Lowlands Agricultural Experiment Station collections of the National Agricultural Research Institute of Papua New Guinea. Accessions were selected to represent agro-morphological variation across the range of sweet potato cultivation in highlands and lowlands of the country (Figure 1 and Supplementary Material 1). We also analysed 48 accessions from the 'Yen collection' maintained in The National Institute for Agrobiological Sciences (NIAS) (Tsukuba, Japan), including 30 accessions originating from Papua New Guinea, and 18 from West New Guinea (8 from lowland and 10 from highland regions). Healthy young leaves were collected from accessions maintained in greenhouses or fields and dried in an oven at 37 °C for 2 days, and then conserved in silica gel. All DNA extractions (for both chloroplast and nuclear DNA) were conducted using the Qiagen 96 Plant kit for lyophilized tissues (Qiagen, Hilden, Germany), which proved suitable for tissues dried by our methods.

We recognized *a priori* two levels of hierarchical subdivisions for genetic analysis according to agro-geographical criteria (Figure 1). At the regional scale within Papua New Guinea, we distinguished the northern coast, the southern coast, the islands region and the highlands region, and within West New Guinea, the highlands and lowlands regions. Finally, we grouped these regions



Figure 1 Geographic provenience of the 417 New Guinean sweet potato landraces from the National Agricultural Research Institute and National Institute for Agrobiological Sciences collections, and definition of the regional and agroecosystem groups used in our study. The island of New Guinea is divided into West New Guinea and Papua New Guinea, the latter being divided into 20 provinces, each abbreviated by two letters on this map. Our subdivision for genetic analysis distinguishes sampled areas from the highlands (in black) (Wissel Lakes and Baliem localities for West New Guinea highlands and EH, SH, SI and WH provinces for Papua New Guinea highlands) to the lowlands (in grey) (Biak and Merauke localities for West New Guinea). Lowlands of Papua New Guinea were also subdivided into three regional groups: the south coasts of the mainland (WE, GU, CE and ND provinces), the north coast of the mainland (OR, MO, MA, ES and WS provinces) and the islands region (MB, BO, WB, EB, NI, and MN provinces).

into New Guinea lowland and highland agroecosystem types (Figure 1), following a conventional subdivision in New Guinea agriculture (Bourke, 2009).

Chloroplast and nuclear genotyping and coding

All individuals were genotyped with six chloroplast microsatellites (ccmp2, NTCP 26, NTCP 28, Ibcp 5, Icbp 8 and Ibcp 10) and 11 nuclear microsatellites (J263, J315E, J522A, J116a, Ib297, J206A, J1809E, IbR16, IbC5, J544b and IbS11) described in a previous study (Roullier et al., 2011). All loci were amplified independently using Multiplex PCR Taq (Qiagen) in a final volume of 10 µl, using 30 ng of DNA per reaction. The following programme was conducted using a PTC-100 Thermocycler (MJ Research, Waltham, MA, USA): 15 min at 95 °C, 35 cycles of 30 s at 94 °C, 1 min 30 s at 57 °C, 1 min at 72 °C and finally 30 min at 72 °C. The 96-well plates were used for amplification and three wells per plate were devoted to negative (one well) and migration controls (two wells). The reproducibility of reactions was checked using replicated samples (25% of the total amount of reactions) on distinct plates; error rate was <0.05 for the nuclear data set and non-replicated samples were counted as missing data. For the chloroplast data set, we only kept accessions with complete data. Amplification products were analysed with an ABI 3130 XL 16 capillary sequencer (ABI Prism; Applied Biosystems, Foster City, CA, USA) and allele scoring was eye-checked by two investigators using Genemapper (Applied Biosystems).

Geographical patterns of nuclear and chloroplast genetic diversity in New Guinea landraces

Spatial patterns of genetic diversity were inferred by computing different diversity estimators for both chloroplast and nuclear simple sequence repeats (SSRs) at different predefined geographical scales (regional and agroecosystem types (highlands or lowlands)). A rarefaction procedure was applied in some cases to account for unequal sampling among geographical areas (Petit *et al.*, 1998).

Nuclear SSR. As sweet potato varieties are clonally propagated, we estimated first the number of unique multilocus genotypes (MLGs) and then identified sets of clones or genets (pairwise genetic distances equal to 0) and clonal lineages (multilocus lineages) by screening each MLG pair presenting extremely low genetic distances (here individuals differing in fewer than six variables). These pairs accounted for a small peak (at very low values) in the frequency distribution of distances, making this distribution bimodal rather than unimodal (Arnaud-Haond *et al.*, 2007). Two indices of clonal diversity, G_c and G_1 (number of MLGs or of multilocus lineages, respectively, divided by the total number of accessions), were calculated. The R software was used to

obtain the frequency distribution of pairwise distances (Manhattan distance) (R Development Core Team, 2009) within New Guinea lowland and highland groups separately. We then computed the mean number of alleles (N_A) , the allelic richness $(A_p$ the rarefied mean number of alleles per locus) and the rarity index, *R*. *R* is a rarefied index that reflects the proportion of rare alleles in a given data set, being inversely proportional to the allele frequency in the global dataset. A_r and *R* were averaged from 1000 resamplings of eight individuals for the regional scale and 150 for the agroecosystem-type scale. Number of individuals found in one of the six regions and in one of the two agroecosystem types, respectively. Computations were made using custom R scripts.

The partitioning of genetic variation among agroecosystem types and regional groups was quantified with an analysis of molecular variance (AMOVA) using the Arlequin software 3.1 (Excoffier *et al.*, 2005). We also computed a principal coordinate analysis (PCoA) based on a 'Lynch' distance matrix between individuals ('polysat' R CRAN package; Clark and Jasieniuk, 2011). We then performed a one-way analysis of variance between PCoA first axis coordinates and agroecosystem-type groups, followed by a Tukey's test to assess the significance of the differentiation following this axis (the most discriminating for the data set) between the New Guinea highlands and lowlands groups, using the R software (aov and TukeyHSD functions; R Development Core Team, 2009).

Chloroplast SSRs. We identified distinct haplotypes and plotted their geographic distributions. We computed $H_{\rm p}$ the rarefied haplotypic diversity, for regional groups using the program CONTRIB (Petit *et al.*, 1998) (results averaged from resamplings of eight individuals).

Assessing the genetic relationships between New Guinean and tropical American genepools

For purposes of comparison, we added to the New Guinea data set nuclear and chloroplast data sets already published in a previous study (Roullier *et al.*, 2011). The tropical America data set comprises 130 individuals (76 accessions from the Northern genepool and 54 from the Southern genepool) for nuclear data, and 329 individuals (192 from the Northern genepool and 137 from the Southern genepool) for chloroplast data.

Comparison of chloroplast data. We first compared haplotypes identified in New Guinea and tropical America. Our previous study showed that tropical American sweet potato landraces are partitioned in two geographically restricted cp lineages, cp 1 and 2, corresponding to most of the accessions from Southern and Northern genepools, respectively. We estimated the

proportions of haplotypes from each lineage, in each province of Papua New Guinea and in collecting sites in West New Guinea.

Comparison of nuclear data. We aimed to identify the origin of New Guinean landraces and the relative contribution of Northern and Southern genepools from the neotropics to New Guinea genetic diversity. To do this, we attempted to 'assign' New Guinean individuals to their potential sources, the Northern and Southern genepools, using two complementary kinds of assignment methods. First, we used a 'general method', the discriminant analysis of principal components (DAPC), a multivariate analysis recently developed and implemented in the adegenet R packages (Jombart, 2008). DAPC provides an efficient description of genetic clusters using a few synthetic variables (called the discriminant functions). This method seeks linear combinations of the original variables (alleles), which show differences between groups as best as possible while minimizing variation within clusters. On the basis of the retained discriminant functions, the analysis derives probabilities for each individual of membership in each of the different groups. This coefficient can be interpreted as 'genetic proximity' of individuals to the different clusters and might provide an 'assignment measure' of individuals to predefined groups.

In our case, we first constructed the linear model and obtained synthetic variables on the tropical America data set. Then, we used New Guinean landraces as supplementary individuals (that were not used in constructing the model). We did this by transforming the New Guinea data set using the 'centring' and 'scaling' of the Tropical America data set, and then using the same discriminant coefficients as for the individuals that contributed to model construction to predict the position of the new individuals on the discriminant functions. The analysis provides probabilities for each New Guinea accession that it originated from each of a number of predefined neotropical groups. DAPC in itself requires construction of prior groups. Then, we first ran a sequential K-means clustering algorithm for K=2-10, after transforming the data using a PCA (notably to reduce the number of variables and thereby speed the clustering algorithm). This allowed identifying an optimal number of genetic clusters to describe the data, by comparing the different clustering solutions using the Bayesian Information Criterion. Following this analysis, three genetic clusters were considered optimal to summarize the data. However, we consider grouping obtained for K=2, an accurate and simpler summary of the tropical American genepools (see Supplementary Material 2 for details of the DAPC analysis). We then performed DAPC for K=2, retaining five PCA components (29.3% of the total variance) for prior data transformation, corresponding to the 'optimal' value following the a-score optimization procedure proposed in the adegenet R packages (Jombart et al., 2010). A landrace is considered to be well assigned to its source group if the associated membership probability is >0.8.

We also compared non-model-based DAPC assignment with that obtained by a Bayesian model-based method implemented in the software Structure 2.3.3 (Falush *et al.*, 2007). By prespecifying source genepools in tropical America (as defined by the *K*-means clustering algorithm for K=2), the algorithm implemented in Structure estimates ancestry for additional individuals (from New Guinea), updating allele frequencies using only those from tropical America. We ran the admixture model with K=2, correlated allele frequencies, 50 000 burn-in iterations and 150 000 Markov chain–Monte Carlo steps and data coding for handling genotype ambiguity for codominant markers in polyploids (Falush *et al.*, 2007).

Finally, we performed a PCoA computed on a 'Lynch' distance matrix between tropical American and New Guinean accessions to describe genetic variation between New Guinean landraces and tropical American genepools and to quantify the partitioning of genetic variation among tropical American (total, neotropical Southern and Northern genepools) and New Guinean accessions with an AMOVA using the Arlequin software. We also measured differentiation between groups, calculating intra- and intergroup mean pairwise distances (Lynch distance, R package polysat) between genotypes and a measure of $F_{\rm ST}$ based on the estimation of allelic frequencies ('simpleFreq' and 'calcFst' functions of the R package polysat).

To compare allelic diversity between the neotropics and New Guinea, we computed A_p , R, H_e (expected heterozygosity, derived from allelic frequencies estimated using the 'simplefreq' function of the R package polysat), H_o (observed heterozygosity), R_a (the allelic size range), G_c and G_l (indices of clonal diversity) and H (total number of haplotypes) using custom R scripts. All indices were estimated following a rarefaction procedure (except for N_A , G_c , G_l and H), where results were averaged from the resampling of 50 individuals for the comparison of Northern and Southern neotropical genepools and New Guinea highlands and lowlands groups, and 130 for the global comparison of New Guinean and tropical American accessions. To explain observed neutral diversity calculated in terms of allelic richness, we also established a curve characterizing the increase in the total number of captured alleles for a given sample size (based on a 1000 resamplings procedure) for each group. This curve allowed us to estimate the number of clones that have been introduced from tropical America to New Guinea.

RESULTS

Geographical patterns of genetic diversity in New Guinea

Chloroplast markers. We identified a total of six different haplotypes in our sample of New Guinea landraces (H1_2, H5, H6_9, H12, H13 and H14). Haplotype richness varied from 0.473 (\pm 0.162) in the southern coastal region to 0.687 (\pm 0.009) in the highlands of Papua New Guinea (Figure 1). H5 was found only once, in the highlands of Papua New Guinea. H12 was found in both New Guinea highland and lowland sites but was very rare (freq < 0.01). Another uncommon haplotype, H6_9 (freq = 0.091), was also found in both agroecosystem groups, while haplotype H13 (freq = 0.027) was present mostly in the New Guinea highlands. H14 and H1_2 were clearly dominant (freq = 0.33 and 0.5, respectively) and were distributed over all areas sampled (Table 1).

Regions	Ν	Н	H <i>1_2</i>	H <i>6_9</i>	H14	H <i>13</i>	H <i>12</i>	H <i>5</i>	D_H	N _A	Ar	R
Northern coast of Papua New Guinea	40	4	21	3	13		1		0.552 (0.057)	7.273	5.933 (0.660)	0.230 (0.050)
Islands of Papua New Guinea	118	4	73	14	28	3			0.542 (0.041)	8.273	5.942 (0.780)	0.260 (0.053)
Southern coast of Papua New Guinea	12	3	8	1	2				0.473 (0.162)	5.910	5.655 (0.326)	0.179 (0.012)
West New Guinea lowlands	8	2	2		6				0.476	5.022	5.022	0.235
Total lowlands	179	5	104	18	49	3	1		0.510 (0.086)	8.546	8.500 (0.158)	0.255 (0.004)
Highlands of Papua New Guinea	229	6	109	22	83	9	3	1	0.687 (0.009)	8.728	5.701 (0.790)	0.215 (0.053)
West New Guinea highlands	10	2	6		2				0.500 (0.128)	4.900	4.845 (0.160)	0.260 (0.019)
Total highlands	238	6	115	22	85	9	3	1	0.595 (0.068)	8.73	8.349 (0.448)	0.211 (0.006)

Abbreviations: A_n allelic richness; D_H , rarefied haplotype diversity; H, number of distinct haplotypes; N, total number of individuals sampled; N_A , mean number of alleles; R, rarity index. All diversity indices were computed following a rarefaction procedure (by re-sampling 8 and 150 individuals (1000 times each) for comparison of regional groups and agroecosystem groups, respectively). Standard deviations are indicated within parentheses.



Figure 2 Patterns of genetic differentiation of sweet potato landraces in New Guinea. (a) Frequency distribution of genetic dissimilarity based on the calculation of all pairwise Manhattan distances between genotypes within landraces of New Guinea lowlands (in grey) and New Guinea highlands (in black). The first peak of the bimodal curve represents clones (pairwise dissimilarities equal to 0), possible variants by somatic mutations and/or scoring errors. (b) Principal coordinate analysis based on the Lynch distance between samples. Highland and lowland genotypes appear in black and grey, respectively.

Table 2 AMOVA of nuclear SSRs within New Guinea^a and among tropical America^b and New Guinea

Source of variation	Sum of squares	Variance components	Variation percentage	Fixation indices (P-value)
Within New Guinea				
Among agroecosystem groups	97.600	0.119	1	F _{CT} =0.010 (0.04)
Among regional groups	90.259	0.424	3.590	$F_{SC} = 0.036 \ (< 0.0001)$
Within regional groups	4640.420	11.290	95.410	F _{ST} =0.046 (<0.0001)
Among tropical America and New Guinea				
Among groups 'New Guinea-neotropical Northern'	176.139	1.27855	9.810	$F_{ST} = 0.098 \ (< 0.0001)$
Within groups	5773.581	11.759	90.190	
Among groups 'New Guinea-neotropical Southern'	236.245	2.350	16.930	$F_{ST} = 0.169 \ (< 0.0001)$
Within groups	5407.241	11.52930	83.07	

Abbreviations: AMOVA, analysis of molecular variance; SSR, simple sequence repeats. ^aAmong agroecosystem groups, among regional groups and within regional groups. ^bSouthern and Northern genepools.

Nuclear markers. We found a total of 378 MLGs and 329 putative clonal lineages (Manhattan distance between genotypes within a lineage <6) ($G_c = 0.9$ and $G_l = 0.79$). Duplicates were sometimes not only locally restricted but were also shared between distant regions (shared notably between New Guinea highland and lowland regions). Frequency distribution of pairwise Manhattan distances for New Guinea highland and lowland genotypes showed a bimodal curve, attesting to the existence, in both groups, of clones (identical MLGs) and of apparent clonal lineages (or sampling errors) differing by only few somatic mutations. Both occurred at low frequency $(G_c = 0.9 \text{ and } 0.94 \text{ and } G_l = 0.72 \text{ and } 0.87 \text{ in New Guinea highlands}$ and lowlands, respectively) (Figure 2a).

The mean number of alleles per locus ranged from 4.9 in the highlands of West New Guinea to 8.721 in the highlands of Papua New Guinea, with a mean value for all six regions of 6.68 (Table 1). Values of allelic richness, an estimator that allows comparison of samples of unequal size, ranged from 4.845 (± 0.160) in the highlands of West New Guinea to 5.942 (± 0.78) in the islands region of Papua New Guinea. Similar values for allelic richness were found in New Guinea highland and lowland landraces (8.349 ± 0.448) and 8.5 ± 0.158), respectively). Values for the rarity index ranged from $0.179 (\pm 0.012)$ in the southern coastal region of Papua New Guinea to 0.26 for the islands region of Papua New Guinea (±0.053) and highlands of West New Guinea (± 0.019). Values of the rarity index in New Guinea highland and lowland agroecosystems were 0.211 (± 0.006) and 0.255 (± 0.004) , respectively.

Heredity

Results of the AMOVA underlined the weak genetic differentiation between New Guinea highland and lowland agroecosystems and among regions (Table 2). Most of the variation occurred within regional groups (95.5%), with little variation between highland and lowland systems and among regions (1% and 3.5%, respectively). Results of the PCoA confirmed the weak differentiation among regions (Supplementary Material 3) and between ecosystem types within New Guinea sweet potato landraces. However, the first axis, which accounted for 12.6% of the variance, still showed a slight but significant (P < 0.01) differentiation between New Guinea lowland and highland accessions (Figure 2b).

Comparing genetic diversity between tropical American and New Guinean landraces

Chloroplast markers. Chloroplast diversity found in New Guinea was a subset of that present in tropical America (Table 3). Out of 21 haplotypes found in tropical America, 6 were found in New Guinea (H1_2, H5, H6_9, H12, H13 and H14). New Guinea haplotypes included representatives of both cp groups 1 and 2, defined in a previous study (Roullier et al., 2011): H14 and H13 were the most frequent haplotypes of cp group1. Widespread throughout the Southern region in the neotropics, they are also present in the Northern region, although H13 is rare there. H6_9 and H1_2 are the most frequent haplotypes of cp group 2, found throughout the Northern region, but are also present in the neotropical Southern

Table 3 Comparison of nuclear and chloroplast diversity between tropical America and New Guinea.

Landrace groups	Regions	N _{nuc}	N _A	A _r	H _e	H _o	R _a	R	K1	Private alleles (>0.05)	G _c	G,	N _{cp}	cpDNA 1	Н
Tropical America	Total Northern Southern	130 76 54	10.91 9.727 8.455	10.9 9.35 8.372	0.769 0.768 0.733	0.367 0.385 0.343	25.455 24.57 24.247	0.354 0.371 0.332	 0.276 0.724	24 (7) 9 (3) 8 (2)	0.94 0.99 1	0.83 0.8 0.87	329 192 137		21 13 15
New Guinea	Total Highlands Lowlands	417 239 178	9.545 8.818 8.636	8.3 7.381 7.63	0.747 0.734 0.758	0.387 0.381 0.396	23.074 22.153 23.863	0.169 0.158 0.186	0.716 0.264 0.310	16 (3) 6 (0) 4 (0)	0.9 0.9 0.95	0.79 0.727 0.872	407 236 171	36.27 41.94 28.48	6 6 6

Abbreviations: A_n allelic richness; cpDNA 1, frequency of cpDNA lineage 1 haplotypes and within parentheses the same index considering only the nuclear subsampling; G_c and G_i, indices of clonality (giving the ratio of the number of unique multilocus genotypes and multilocus clonal lineages, respectively, identified from nuclear data to the number of samples); H_n number of distinct haplotype; H_e, expected heterozygosity; H_o, observed heterozygosity; K1, mean K1 ancestry value given by the Bayesian clustering analysis; N_A, mean number of alleles; N_{cp}, total number of accessions of the chloroplast data set; N_{nuc}, total number of accessions for the nuclear data set; R, rarity index; R_a, allelic size range. Five indices (averaged over all loci) were computed following a rarefaction procedure based on 1000 re-samplings of 130 accessions for New Guinea and tropical America total data sets and 50

Five indices (averaged over all loci) were computed following a rarefaction procedure based on 1000 re-samplings of 130 accessions for New Guinea and tropical America total data sets and 50 accessions each for neotropical Southern, neotropical Northern, New Guinea highland and New Guinea lowland data sets.



Figure 3 Geographical origin of New Guinea and Tropical America sweet potato landraces (as distinguished by chloroplast data, and nuclear data with DAPC analysis at K=2) and possible paths of sweet potato introduction into New Guinea. The size of circles is proportional to the total number of individuals by country, province and sampled area, for tropical America, Papua New Guinea and West New Guinea, respectively. In the pie charts, the top half of the circle represents the proportions of individuals belonging to each chloroplast group (cp group 1 and cp group 2), while the bottom half represents the proportions belonging to nuclear clusters K1 and K2. A means 'Admixed' for individuals with a membership probability (DAPC) or ancestry value (Bayesian clustering) inferior to 0.8 for both clusters. Arrows indicate proposed routes and dates of introduction by humans.

genepool. H5 and H12 are very rare cp group1 haplotypes, restricted to Peru and Bolivia, respectively (Supplementary Material 4).

Both cp groups 1 and 2 were represented throughout New Guinea, but cp group 2 dominated (63% on average), except for the highlands of Papua New Guinea, where the proportion of cp group 1 haplotypes was 42% (Table 1 and Figure 3). Minor cp group 1 haplotypes (H5, H12 and H13) were found more frequently in New Guinea highland than in New Guinea lowland sites.

Nuclear markers. For almost all of the estimators (N_A , A_p , R_a , R and H_e), nuclear genetic diversity tended to be greater in tropical America than in New Guinea (Table 3), particularly comparing that of

the Northern genepool with that of New Guinea group. Only H_0 was slightly higher in New Guinea (0.387), close to the value for this index (0.385) for the Northern group. Allelic richness of the Southern genepool (8.372) was equivalent to that found in New Guinea (8.3), but the rarity index was higher in the Southern genepool (R = 0.32) than in New Guinea (R = 0.169). The tropical America group had 24 private alleles and the New Guinea group 16 (all the latter were rare, with frequencies <0.1). Neotropical Northern and Southern genepools, and New Guinea lowland and highland groups, had 9, 8, 6 and 4 private alleles, respectively. Clonality indices G_c and G_l were, respectively, 0.90 and 0.79 for New Guinea accessions and 0.94 and 0.83, respectively, for those from tropical America.



Figure 4 Tropical American origin of the genetic diversity of New Guinean landraces as assessed by nuclear markers. (a) Two bar plots showing for each individual the probabilities of membership in nuclear clusters *K*1 and *K*2 as determined by DAPC (top graph), the ancestry values in nuclear clusters *K*1 and *K*2 as determined by DAPC (top graph), the ancestry values in nuclear clusters *K*1 and *K*2 as determined by the Bayesian clustering method implemented in Structure (bottom graph). Each individual is represented as a vertical bar, with colours corresponding to membership probabilities (or ancestry value) in clusters *K*1 (pale grey) and *K*2 (grey). (b) Principal coordinate analysis based on the Lynch distance between genotypes from New Guinea and tropical America. Accessions are labelled following Southern (S) and Northern genepools (N), both in pale grey, and New Guinea highland (H) and lowland (L) agroecosystem groups, in black and dark grey, respectively.

We used grouping obtained with K-means clustering for K=2 to perform the DAPC and the Bayesian assignation analysis (Supplementary Material 2): the cluster K1 mainly contained samples from the Southern genepool (79%) and cluster K2 mostly those from the Northern genepool (91%) (Figure 4a). DAPC results showed that most of the New Guinea accessions (370 accessions), from both lowland and highland agroecosystem types, were clearly assigned (with a membership probability > 0.8) to the cluster K2 (mean K2 membership probability of 0.92 ± 0.20) (Figures 4a and b). Only few accessions (10) were found to be assigned to the cluster K1 (mean K1 membership probability of 0.08). Some others were not well assigned to a given cluster (membership probability <0.8) and are considered to be 'admixed'. These included 37 accessions of uncertain assignment to genetic clusters K1 or K2. Bayesian 'assignation' provided congruent results (Figure 4a). New Guinea accessions exhibited a mean K2 ancestry value of 0.716 (\pm 0.116), a value quite similar to that found in the Northern genepool (0.724 ± 0.145) , while the Southern genepool exhibited a higher mean K1 ancestry value (0.710 ± 0.185) (Table 3).

The PCoA analysis differentiated neotropical Northern and Southern genepools and the New Guinea samples, principally on the first axis (16.3% of variance explained) (Figure 4b). The Southern genepool was the most divergent from the New Guinea genepool, as also attested by highest values of both intergroup mean pairwise distance and F_{st} (Table 4), confirming DAPC results. Results of AMOVA also supported this tendency: 9.81% of the total variance was between Northern and New Guinea genepools and 16.93% was between Southern and New Guinea genepools (Table 2). Moreover, PCoA showed that lowland accessions appear less divergent from the Northern genepool than highland accessions (Figure 4b).

DISCUSSION

History of sweet potato introduction into New Guinea: the relative contributions of Kumara, Batata and Camote lines

The timing and nature of sweet potato introduction into New Guinea are still controversial. New Guinea forms the point at which the three lines, *Kumara*, *Batata* and *Camote*, could have converged and their relative contributions to New Guinea sweet potato diversity over time have been widely debated among archaeologists and anthropologists of Oceania. Some consider that the high degree of dependence between sweet potato and highlands societies is best explained as the

Table 4 Distance and differentiation between accessions from New Guinea (highland and lowland groups) and the neotropics (Northern and Southern genepools)^a

		New Guinea	Neotropical			
	Total Lowlands		Highlands	Northern	Southern	
New Guinea						
Total	0.341	_	_	0.403	0.434	
Lowlands		0.343	0.349	0.405	0.433	
Highlands	—	0.004 0.339		0.402	0.433	
Neotropical						
Northern	0.006	0.008	0.010	0.386	0.433	
Southern	0.012	0.020	0.020	0.020	0.356	

 $^{\rm a}\text{Measured}$ by intra- and intergroup pairwise distances (Lynch distances) above the diagonal and Fst values below the diagonal.

result of a long period of evolution (implying ancient introduction via the *Kumara* line). Opponents argue that widespread adoption and diversification could have taken place within 300 to 400 years, a hypothesis compatible with sixteenth-century transfers to New Guinea via the *Batata* and *Camote* lines. According to the proponents of this hypothesis, the intrinsic properties of sweet potato (its broad tolerance of environmental conditions and high productivity, its relative ease of cultivation), combined with innovations developed by highland societies in a constraining and heterogeneous environment, are sufficient to explain the intensity and relative rapidity of the adoption of sweet potato and its diversification in New Guinea (Ballard, 2005).

All results on nuclear markers show that New Guinean landraces are clearly divergent from the genotypes of the Southern genepool, while they are only weakly differentiated from those of the Northern genepool. These results confirm previous results (Gichuki *et al.*, 2003; Zhang *et al.*, 2004), suggesting a main Central and Caribbean American origin for New Guinean landraces via the *Camote* and *Batata* lines (Figure 3).

However, both cp groups (1 and 2) typical of the Southern and Northern genepools, respectively, were identified in New Guinea landraces. Two hypotheses could explain the presence of haplotypes typical of the Southern genepool in New Guinea (Figure 3). First, they may already have been present in clones introduced from the Northern genepool. Recombination between migrant clones of Northern and Southern genepools has occurred since ancient times in tropical America (Roullier et al., 2011), blurring the phylogeographic pattern in chloroplast DNA diversity and complicating assignation of clones introduced into New Guinea to areas of origin in the Neotropics. In fact, the most common haplotypes of cp group 1 are also found, albeit at lower frequency (14.7%), in the Northern genepool. Second, clones of Southern origin may have been introduced into New Guinea, either directly from South America or from other regions in Oceania into which they had been introduced previously. Interestingly, three rare haplotypes of cp group 1, mostly restricted to the Southern genepool, were found in New Guinea, which could be seen as an indication of an early introduction through the Kumara line.

On the basis of palaeoecological data, some archaeologists have associated the intensification of land use, assessed from Casuarina pollen records at about 1200 BP, to an early introduction of sweet potato from Polynesia (Golson, 1977; Gorecki, 1986; Haberle, 1998). Following the same reasoning, linguists have argued that the presence of terms related to the Polynesian word 'Kumara' could support a hypothesis of early introduction (Scaglion and Soto, 1994). Although our chloroplast data are compatible with the existence of such introductions, our nuclear data indicate that they cannot account for the main genetic background of New Guinean sweet potato diversity, either in the New Guinea highlands or lowlands. Such introductions may rather have been occasional, resulting more probably from post-Magellanic introductions of clones by whalers, missionaries and Polynesians, as well as through twentiethcentury movements across the Pacific (Bourke, 2005) (Figure 3). Alternatively, primary Kumara line introductions could have been widely reshuffled by later European reintroductions from the Northern genepool.

Our nuclear data also suggest a slight differentiation between landraces from New Guinea highlands and lowlands. Moreover, the highlands group is more differentiated from the ancestral neotropical Northern genepool than is the lowland group. Two complementary scenarios could explain this pattern. The observed differentiation could be largely drift-driven, resulting from a sequence in which sweet potato first colonized the island region, then mainland coastal areas and finally the highlands. The slight decrease of genetic diversity from lowland to highland areas might be an indication of this humanmediated dispersal. On the basis of surveys of oral history, it has been proposed that sweet potato was introduced around 1700 into the New Guinea highlands, where the plant was rapidly adopted and diffused (Bourke, 2009). It is likely that sweet potato was available on the coasts of New Guinea as early as AD1521 and certainly at least by AD1633. Therefore, sweet potato probably reached the highlands from the Sepik area along one of the corridors that link coastal regions to the Border Mountains. Furthermore, it seems that even though sweet potato was cultivated early in some lowland parts of New Guinea (north coastal areas of the mainland), its adoption in lowland regions dates mostly to a (re)introduction after 1870 by Europeans, Polynesians and other outsiders. Widespread adoption of sweet potato into lowland agricultural systems has occurred since 1940 (Bourke, 2005). Thus, the weak differentiation between the islands region and the neotropical Northern genepool could be associated with repeated recent introductions, and thus a shorter time during which drift could have occurred (Figure 3).

Processes of diversification: on the origin of a secondary centre of diversity

Several examples of secondary diversification have been well documented: *Citrus* spp. in Europe (Ollitrault and Luro, 2001); watermelon, *Citrullus lanatus* (Thunb.) Mats. & Nakai in Brazil (Romão, 2000); bean, *Phaseolus vulgaris* L., in Africa (Pickersgill, 1998; Asfaw *et al.*, 2009); barley, *Hordeum vulgare* L., in Africa (Pickersgill, 1998); and cassava, *Manihot esculenta* Crantz subsp. *esculenta*, in Africa (Pickersgill, 1998; Delêtre, 2010) and Vanuatu (Sardos *et al.*, 2008). Causes of secondary diversification have been discussed (Pickersgill, 1998; Romão, 2000; Sardos *et al.*, 2008; Delêtre, 2010). Multiple introductions of a crop can mitigate any bottleneck effects, especially if introductions have led to the mixing of different genepools from the original range. Subsequent recombination between genepools, associated with new selective pressures, both natural and cultural, could lead to the rapid selection of a large number of new variants.

Chloroplast markers indicate a reduction of diversity in New Guinea compared with tropical America (from 21 haplotypes to 6). The most frequent haplotypes from tropical America were found (H14, H1_2 and H6_9) in New Guinea. Nuclear data also revealed that New Guinea accessions exhibited lower diversity than that observed for the original Northern genepool, and a level of diversity similar to that of the Southern genepool, except for rare alleles. These results highlight the introduction bottleneck. However, as the reduction of genetic diversity is relatively moderate, it suggests a quite limited bottleneck. Multiple independent introductions can reduce bottleneck effects by increasing the genetic diversity introduced into the new area. Such scenarios have been invoked to explain the impressive diversity of beans in East Africa (Pickersgill, 1998; Asfaw et al., 2009) and of watermelons in Brazil (Romão, 2000). Beans were domesticated independently in Mesoamerica and in the Andean region, giving rise to two differentiated genepools. Both were introduced into Africa, leading to a concentration of diversity at the scale of individual fields (Asfaw et al., 2009). Bean evolutionary history shows parallels with that of sweet potato. However, as bean genotypes from both American genepools rarely recombined (Asfaw et al., 2009; Burle et al., 2010), it still remains easy to identify original sources in introduced areas.

This identification is less evident in our case, because we are dealing with a situation of initial admixture in tropical America before introduction into New Guinea, as detected by both nuclear and chloroplast data (Figure 3 and Supplementary Material 2; Roullier et al., 2011). We propose therefore that independent introductions of 'pure' individuals from the Northern neotropical genepool (accessions with cp group 2), introgressed individuals (Northern genepool accessions with cp group 1) and probably a few individuals directly from the Southern neotropical genepool (Southern genepool accessions with cp group 1) could have led to the accumulation of a substantial subset of the genetic diversity present in tropical America (Figure 4). We estimated that the introduction of 63 clones from the Northern genepool (or of 45 clones if these came from throughout tropical America) would have been sufficient to explain the current New Guinea diversity in terms of allelic richness (Supplementary Material 5).

Our data show, moreover, that few clonemates were identified among landraces through the island of New Guinea. The 417 landraces considered included 329 genotypes that were most likely derived from distinct sexual reproduction events (208 over 230 highland landraces, and 170 over 179 lowland landraces). The greatest proportion of the diversity among cultivars in New Guinea thus appears to have resulted from intensive recombination among introductions, both in highland and in lowland conditions. This situation of secondary admixture may explain the difference observed between chloroplast and nuclear genetic signals: introductions from the Southern genepool recombined with local material, most of which was of Central American and Caribbean origin (that is, Northern genepool), blurring their original nuclear signal.

Sweet potato is vegetatively propagated by stem cuttings. Various studies have reported the widespread occurrence and use by farmers of plants issued from true seed in different New Guinea highland farming systems (Bulmer, 1965; Yen, 1974; Schneider, 1995). Cultivation practices favour the appearance and subsequent adoption of such variants, similarly to what has been described in detail and reviewed for cassava in the Neotropics (McKey et al., 2010) and for taro in Oceania (Caillon et al., 2006; Sardos et al., 2011). In the highlands, sweet potato plants are harvested over a long period of time and then the parcel is left as fallow, giving the remaining plants ample time to reach flowering and fruiting stages. Moreover, flowering of sweet potato is recognized to be particularly abundant under highland conditions (P Van Wijmeersch, personal communication, 2011). Furthermore, mixtures of different varieties (landraces) are a ubiquitous characteristic of highland gardens. The number of varieties grown by different communities varies from 6 to 71, with a mean of 33 varieties per garden (Bourke, 2009), favouring even more genetic mixing. As sweet potato is an obligatorily outcrossing plant, with a sporophytic incompatibility system, seeds are produced only if different incompatibility groups are planted in the same garden (or in gardens close enough to permit pollen flow). In all, 17 autoincompatibility groups have been described in tropical America (Nakanishi and Kobayashi, 1979). Multiple introductions may have facilitated the arrival of several incompatibility groups, leading to high probability of effective crossings. The practice of clearing and burning after the fallow period, widespread in the New Guinea Highlands, also creates favourable conditions for seed germination. Finally, it has been observed that farmers recognize the value and take care of plants issued from true seed. They name those that are selected and progressively add some to their stock of landraces (Schneider, 1995).

The highlands, a mountainous environment, exhibit a diversity of environmental, ecological and cultural conditions (Hays, 1993). In the Baliem Valley, for example, there are two major agroecologies; alluvial with wetter conditions in the centre of the valley and dryer on the surrounding slopes. Farmers clearly distinguish their landraces in respect to their broad adaptation to each of these environments (Schneider, 1995). Moreover, mountainous terrain and tribal hostilities may have limited the movement of people and therefore also that of plant germplasm (Lutulele, 2001). Geographical isolation in this heterogeneous environment must have been important for generating new, locally adapted genotypes.

In lowland areas, no such precise observations have been reported, and flowering is probably a rare event in these wet, almost equatorial, conditions, although it likely does occur in drier areas or periods (P Van Wijmeersch, personal communication, 2011). Our results demonstrate that pairwise dissimilarities between lowland landraces are high, even higher than between different highland landraces. Recombination must also have been a major mechanism of landrace diversification in lowland agroecosystems, suggesting that here as well, seedlings appear and are selected. Alternatively, repeated introductions from tropical America, where sexuality is also a major contributor to evolutionary processes (Roullier *et al.*, 2011), could also contribute to the pattern of diversity in lowland accessions.

Patterns of diversity in New Guinea: some insights into the conservation of sweet potato

Our data showed that there is weak genetic differentiation within the New Guinea genepool. Allelic diversity in New Guinea does not appear to show strong geographical structure. Moreover, the level of diversity in terms of allelic richness and rarity seems to be relatively homogeneous across geographical regions. Relatively recent (around 300–400 years ago) introduction, predominantly from one American genepool, combined with movements of clones across the island, and very active sexual recombination, are factors that may explain the near absence of geographical structure assessed with neutral markers.

These results highlight the fact that neutral nuclear markers are not the most efficient for selecting varieties to constitute core subsets. There is certainly a differentiation between lowland and highland varieties that is based on ecological gradients. A total of 678 highland varieties, representing approximately 60% of the highland collections, have already been evaluated in lowland environments. Most of them gave only very low yields or even no yield at all (Van Wijmeersch, 2001). However, neutral markers do not reflect very well patterns of variation related to environmental gradients, including divergence in adaptive traits (McKay and Latta., 2002; Hoffmann and Willi, 2008; Gebremedhin *et al.*, 2009). For New Guinea sweet potato landraces, conventional evaluation of morphological, physiological or other attributes (for example, organoleptic traits) remains an appropriate approach for identifying useful genotypes and defining core subsets for *ex situ* conservation.

Farmers and their practices appear to be keystone actors in determining the evolutionary and adaptive dynamics of sweet potato, because of their ability to manage the crop's sexual reproductive biology. Throughout the island, subsistence farmers do not readily adopt improved varieties developed away from farming systems and local environments (Van Wijmeersch, 2001). It thus seems more appropriate to develop new cultivars in the different agroecological zones. The geographical distribution of allelic diversity, integrating both ex situ conservation and participatory breeding (Lebot et al., 2005), has been proposed as a practical alternative to avoid the costs and laborious maintenance of ex situ collections in developing countries, and to help farmers to fashion relatively autonomous strategies for coping with ongoing global change. The distribution of allelic diversity approach focuses on the broadening of local genetic basis, by the assemblage and evaluation in research stations of genotypes representative of the useful diversity of the species. Then, successive recombination-selection cycles on farm should be able to give rise quite rapidly to locally adapted variants. Success of such an approach depends largely on how frequently seedlings are incorporated by farmers. Generally, farmers do not really take care of seedlings for utilitarian purposes; their management of crops over many years cannot be characterized as consciously having the purpose of plant breeding. They have been taking advantage of a natural breeding/crossing process that resulted in the continual production of spontaneously germinating true sweet potato seeds. The number of seedlings produced will depend largely on features of the agroecosystem, such as the length of the fallow period and the planting (or not) of varieties belonging to different self-incompatibility groups in the same field. Once produced, the extent to which plants issued from sexually produced true seed actually contribute to diversity and adaptation depends on farmers' knowledge about them and attitudes towards them. On-farm approaches to conserving the dynamics of sweet potato diversification should thus promote not only farmers' knowledge about volunteer seedlings but also agroecological conditions and cultivation practices that favour their abundant production.

602

CONCLUSION

Combining nuclear and chloroplast data, we showed that New Guinea landraces are principally derived from the Northern neotropical genepool (Caribbean and Central America), but that some South American clones may also have been introduced, either early by Polynesians themselves or (more likely) later by whalers and missionaries and through twentieth-century movements across the Pacific. Subsequent recombination between these multiple introductions, the frequent incorporation of plants issued from true seed by farmers and the geographical and cultural barriers constraining crop diffusion in this topographically and linguistically heterogeneous area have generated quite rapidly an impressive number of variants, adapted to a wide range of agroecological zones through the island. Integrating such evolutionary processes in conservation strategies could provide plant scientists with a powerful tool in supporting local communities to adapt their farming systems to ongoing environmental and societal changes.

DATA ARCHIVING

Data deposited in the Dryad repository: doi:10.5061/dryad.bd6v0.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This study was funded by grants from the 'Pacific Funds' distributed by the French Ministry of European and Foreign Affairs. The first author, CR, was supported by a doctoral grant from the French Ministry of Higher Education and Research. We are very grateful to National Agricultural Research Institute field helpers for collecting plant material. We also thank Paul Van Wijmeersch (formerly of the National Agricultural Research Institute) for useful comments on previous versions of the manuscript and for stimulating discussions. Data used in this work were produced through the technical facilities of the Centre Méditerranéen Environnement Biodiversité.

- Asfaw A, Blair MW, Almekinders C (2009). Genetic diversity and population structure of common bean (*Phaseolus vulgaris* L.) landraces from the East African highlands. *Theor Appl Genet* **120**: 1–12.
- Arnaud-Haond S, Duarte CM, Alberto F, Serrão EA (2007). Standardizing methods to address clonality in population studies. *Mol Ecol* 16: 5115–5139.
- Ballard C (2005). Still good to think with: sweet potato in Oceania. In: Ballard C, Brown P, Bourke RM, Harwood T (eds.) *The Sweet Potato in Oceania: A Reappraisal.* Ethnology Monographs 19 Department of Anthropology, University of Pittsburgh: Pittsburgh, PA, USA, Oceania Monograph 56, The University of Sydney: Sydney, Australia, pp 3–13.
- Barrau J (1957). L'énigme de la patate douce en Océanie. Etudes d'Outre-Mer 40: 83–87. Bayliss-Smith T, Golson J, Hughes P, Blong R, Ambrose W (2005). Archaeological evidence for the Ipomoea Revolution at Kuk swamp, upper Waghi Valley, Papua New Guinea. In: Ballard C, Brown P, Bourke RM, Harwood T (eds.). The Sweet Potato in Oceania: A Reappraisal. Ethnology Monographs 19 Department of Anthropology, University of Pittsburgh: Pittsburgh, PA, USA, Oceania Monograph 56, The University of Svdnev. Svdnev. Australia. pp 3–13.
- Bourke RM (2005). The continuing Ipomoea revolution in Papua New Guinea. In: Ballard C, Brown P, Bourke RM, Harwood T (eds.) *The Sweet Potato in Oceania: A Reappraisal.* Ethnology Monographs 19 Department of Anthropology, University of Pittsburgh: Pittsburgh, PA, USA, Oceania Monograph 56 The University of Sydney: Sydney, Australia, pp 171–180.
- Bourke M (2009). Sweet potato in Oceania. In: Loebenstein G, Thottappilly G (eds.) *The Sweet Potato.* Springer: Berlin.
- Bourke RM, Harwood T (2009). Food and Agriculture in Papua New Guinea. ANU E Press, The Australian National University: Canberra, Australia.
- Brown AHD (1989). Core collections—a practical approach to genetic resources management. Genome 31: 818–824.
- Bulmer R (1965). Beliefs concerning the propagation of new varieties of sweet potato in two New Guinea Highland societies. *J Polynesian Soc* **74**: 237–239.
- Burle ML, Fonseca JR, Kami JA, Gepts P (2010). Microsatellite diversity and genetic structure among common bean (*Phaseolus vulgaris* L.) landraces in Brazil, a secondary centre of diversity. *Theor Appl Genet* **121**: 801–813.

- Caillon S, Quero-Garcia J, Lescure JP, Lebot V (2006). Nature of taro (*Colocasia esculenta* (L.) Schott) genetic diversity prevalent in a Pacific Ocean island, Vanua Lava, Vanuatu. *Genet Resour Crop Evol* 53: 1273–1289.
- Clarke AC, Burtenshaw MK, McLenachan PA, Erickson DL, Penny DSMBE Tri-National Young Investigators (2006). Reconstructing the origins and dispersal of the Polynesian bottle gourd (*Lagenaria siceraria*). *Mol Biol Evol* 23: 893–900.
- Clark LV, Jasieniuk M (2011). Polysat: an R package for polyploid microsatellite analysis. Mol Ecol Resour 11: 562–566.
- Delêtre M (2010). The ins and outs of cassava diversity in Gabon, Central Africa: a pluridisciplinary approach to the dynamics of genetic diversity of Manihot esculenta Crantz (Euphorbiaceae). PhD dissertation, Trinity College, University of Dublin. http:// tel.archives-ouvertes.fr/tel-00623219/fr/
- Elias M, Penet L, Vindry P, McKey D, Panaud O, Robert T (2001). Unmanaged sexual reproduction and the dynamics of genetic diversity of a vegetatively propagated crop plant, cassava (*Manihot esculenta* Crantz), in a traditional farming system. *Mol Ecol* **10**: 1895–1907.
- Excoffier L, Laval G, Schneider S (2005). Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evol Bioinform* 1: 47–50.
- Fajardo DD, La Bonte DR, Jarret RL (2002). Identifying and selecting genetic diversity in Papua New Guinea sweet potato *Ipomoea batatas* (L.) Lam. germplasm collected as botanical seed. *Genet Resour Crop Evol* 49: 463–470.
- Falush D, Stephens M, Pritchard JK (2007). Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol Ecol Notes* 7: 574–578.
- Gichuki ST, Berenyi M, Zhang D, Hermann M, Schmidt J, Glössl J et al. (2003). Genetic diversity in sweetpotato (*Ipomoea batatas* (L.) Lam.) in relationship to geographic sources as assessed with RAPD markers. Genet Resour Crop Evol 50: 429–437.
- Golson J (1977). No room at the top: agricultural intensification in the New Guinea Highlands. In: Allen J, Golson J, Jones R (eds.) *Sunda and Sahul*. Academic Press: London, UK, pp 601–638.
- Gorecki PP (1986). Human occupation and agricultural development in the Papua New Guinea Highlands. Mountain Res Dev 6: 159–166.
- Green RC (2005). Sweet potato transfers in Polynesian prehistory. In: Ballard C, Brown P, Bourke RM, Harwood T (eds.). *The Sweet Potato in Oceania: a Reappraisal*. Ethnology Monographs 19 Department of Anthropology, University of Pittsburgh: Pittsburgh, PA, USA, Oceania Monograph 56 The University of Sydney: Sydney, Australia, pp 43–62.
- Gebremedhin B, Ficetola GF, Naderi S, Rezaei HR, Maudet C, Rioux D et al. (2009). Frontiers in identifying conservation units: from neutral markers to adaptive genetic variation. Anim Conserv 12: 107–109.
- Gunn BF, Baudouin L, Olsen KM (2011). Independent origins of cultivated coconut (Cocos nucifera L.) in the Old World tropics. PLoS One 6: e21143.
- Haberle SG (1998). Dating the evidence for prehistoric agricultural change in the highlands of New Guinea: the last 2000 years. *Arch Oceania* **47**: 1–19.
- Haberle SG, Atkin G (2005). Needles in a haystack: searching for sweet potato (*Ipomoea batatas* (L.) Lam.) in the fossil pollen record. In: Ballard C, Brown P, Bourke RM, Harwood T (eds.) *The Sweet Potato in Oceania: A Reappraisal*. Ethnology Monographs 19 Department of Anthropology, University of Pittsburgh: Pittsburgh, PA, USA, Oceania Monograph 56 The University of Sydney: Sydney, Australia, pp 25–34.
- Hays TE (1993). 'The New Guinea Highlands': region, culture area, or fuzzy set? Curr Anthropol 34: 141–164.
- Hoffmann AA, Willi Y (2008). Detecting genetic responses to environmental change. Nat Rev Genet 9: 421–432.
- Hinkle AE (2007). Population structure of Pacific *Cordyline fruticosa* (Laxmanniaceae) with implications for human settlement of Polynesia. *Am J Bot* **94**: 828–839.
- Jombart T (2008). Adegenet: R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**: 1403–1405.
- Jombart T, Devillard S, Balloux F (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet* **11**: 94.
- Kreike CM, Van Eck HJ, Lebot V (2004). Genetic diversity of taro, *Colocasia esculenta* (L.) Schott, in Southeast Asia and the Pacific. *Theor Appl Genet* 109: 761–768.
- Ladefoged TN, Graves MW, Coil JH (2005). The introduction of sweet potato in Polynesia: early remains in Hawai'i. J Polyn Soc 114: 359–373.
- Lebot V (1992). Genetic vulnerability of Oceania's traditional crops. *Exp Agric* 28: 309-323.
- Lebot V (1999). Biomolecular evidence for plant domestication in Sahul. *Genet Resour Crop Evol* **46**: 619–628.
- Lebot V, Ivancic A, Abraham K (2005). The geographical distribution of allelic diversity, a practical means of preserving and using minor root crop genetic resources. *Exp Agric* **41**: 475–489.
- Lutulele RP (2001). Sweet potato variety developments in the PNG highlands: implications for future research and extension focus. In: Bourke RM, Allen MG, Salisbury JD (eds.). Food Security for Papua New Guinea. *Proceedings of the Papua New Guinea Food and Nutrition 2000 Conference (ACIAR Proceedings 99)*. University of Technology: Lae, pp 683–688.
- McKay JK, Latta RG (2002). Adaptive population divergence: markers, QTL and traits. *Trends Ecol Evol* 17: 285–291.
- McKey D, Elias M, Pujol B, Duputie A (2010). The evolutionary ecology of clonally propagated domesticated plants. New Phytol 186: 318–332.
- Nakanishi T, Kobayashi M (1979). Geographic distribution of cross incompatibility groups in sweet potato. *Incompat News/ett* 11: 72–75.
- Ollitrault P, Luro F (2001). Citrus. In: Charrier A, Jacquot M, Hamon S, Nicolas D eds. Tropical Plant Breeding. CIRAD: Paris, France, pp 55–77.

- Perrier X, De Langhe E, Donohue M, Lentfer C, Vrydaghs L, Bakry F et al. (2011). Multidisciplinary perspectives on banana (*Musa* spp.) domestication. *Proc Natl Acad Sci USA* 108: 11311–11318.
- Petit RJ, El Mousadik A, Pons O (1998). Identifying populations for conservation on the basis of genetic markers. *Conserv Biol* **12**: 844–855.
- Pickersgill B (1998). Crop introductions and the development of secondary areas of diversity. In: Prendergast HDV, Etkin NL, Harris DR, Houghton PJ (eds.). *Plants for Food and Medicine*. Royal Botanic Gardens: Kew, London, UK, pp 93–105.
- R Development Core Team (2009). R: A Language and Environment for Statistical Computing, Reference Index Version 2.8.1. R Foundation for Statistical Computing: Vienna, Austria.
- Romão RL (2000). Northeast Brazil: a secondary centre of diversity for watermelon (*Citrullus lanatus*). Genet Resour Crop Evol **47**: 207–213.
- Roullier C, Rossel G, Tay D, McKey D, Lebot V (2011). Combining chloroplast and nuclear microsatellites to investigate origin and dispersal of New World sweet potato landraces. *Mol Ecol* **20**: 3963–3977.
- Sardos J, McKey D, Duval MF, Malapa R, Noyer JL, Lebot V (2008). Evolution of cassava (*Manihot esculenta* Crantz) after recent introduction into a South Pacific Island system: the contribution of sex to the diversification of a clonally propagated crop. *Genome* 51: 912–921.
- Sardos J, Noyer JL, Malapa R, Bouchet S, Lebot V (2011). Genetic diversity of taro (*Colocasia esculenta* (L.) Schott) in Vanuatu, Oceania: an appraisal of the distribution of allelic diversity (DAD) with SSR markers. *Genet Resour Crop Evol* 59: 805–820.
- Scaglion R, Soto K (1994). A prehistoric introduction of sweet potato in New Guinea? In Strathern AJ, Sturzenhofecker G (eds.). Migration and Transformation: Regional

Perspectives on New Guinea. ASAO Monograph No. 15 University of Pittsburgh PressPittsburgh, PA, USA, pp 257–294.

- Scarcelli N, Tostain S, Vigouroux Y, Agbangla C, Daïnou O, Pham JL (2006). Farmers' use of wild relative and sexual reproduction in a vegetatively propagated crop. The case of yam in Benin. *Mol Ecol* **15**: 2421–2431.
- Schneider J (1995). Farmer practices and sweet potato diversity in highland New Guinea. In: Schneider J (eds.)Indigenous Knowledge in Conservation of Crop Genetic Resources. *Proceedings of an International Workshop held in Cisarua*. CIP-ESEAP/ CRIFC: Bogor, Indonesia, pp 63–70.
- Van Wijmeersch P (2001). The status of sweet potato variety evaluation in PNG and recommendations for further research. In: Bourke RM, Allen MG, Salisbury JD (eds.). Food Security for Papua New Guinea. Proceedings of the Papua New Guinea Food and Nutrition 2000 Conference (ACIAR Proceedings 99). University of Technology: Lae, Papua New Guinea., pp 683–688.
- Wiessner P (2005). Social, symbolic, and ritual roles of the sweet potato in Enga, from its introduction until first contact. In: Ballard C, Brown P, Bourke RM, Harwood T (eds.). *The Sweet Potato in Oceania: a Reappraisal*. Ethnology Monographs 19 Department of Anthropology, University of Pittsburgh: Pittsburgh, PA, USA, Oceania Monograph 56 The University of Sydney: Sydney, Australia, pp 121–130.
- Yen DE (1974). The Sweet Potato in Oceania: An Essay in Ethnobotany. Bernice P. Bishop Museum 236, Bishop Museum Press: Honolulu, HA, USA.
- Zhang DP, Ghislain M, Huamán Z, Golmirzaie A, Hijmans R (1998). RAPD variation in sweet potato (*Ipomoea batatas* (L.) Lam.) cultivars from South America and Papua New Guinea. Genet Resour Crop Evol 45: 271–277.
- Zhang DP, Rossel G, Kriegner A, Hijmans R (2004). AFLP assessment of diversity in sweetpotato from Latin America and the Pacific region: its implications on the dispersal of the crop. *Genet Resour Crop Evol* **51**: 115–120.

Supplementary Information accompanies this paper on Heredity website (http://www.nature.com/hdy)

604