Theor Appl Genet (2000) 100:160–166 

**Y.M. Charters · M.J. Wilkinson**

# The use of self-pollinated progenies as 'in-groups' for the genetic characterization of cocoa germplasm

Received: 20 January 1999 / Accepted: 25 May 1999

The value of molecular biology for the curating of large germplasm collections is limited by several practical considerations. The molecular method employed should be quick, inexpensive and technically simple, but must also be sufficiently informative to distinguish between most of the stocks held. The overall strategy for the genetic evaluation of a collection needs to identify mislabelled accessions and, at the same time, enable priorities to be established for the maintenance or removal of material. In the present study, the potential of ISSR-PCR analysis is assessed for the maintenance and genetic characterisation of an international collection of cocoa (*Theobroma cacao*) genotypes. Six primers were sufficient to distinguish all but three pairs of the 62 accessions examined. A UPGMA dendrogram was used to provide a measure of the genetic variability between genotypes. The scale was supplied by the inclusion of *Theobroma grandiflora* as an 'out group' and also by the use of two contrasting progenies as 'in groups'. The 'in groups' were obtained from the self-pollination of one plant (SPEC 54.1) known to be highly homozygous and also of a second, highly heterozygous, clone (P 19B). These reference points allowed several documentation errors to be resolved and provided a basis for identifying unwanted or low-priority material. Implications of the work for the routine maintenance of large germplasm collections are briefly discussed.

**Key words** *Theobroma cacao* · ISSR PCR · Microsatellites · Germplasm · Genebank maintenance

Communicated by P. Langridge

Y.M. Charters · M.J. Wilkinson (✉)
Department of Agricultural Botany, School of Plant Sciences, Whiteknights, P.O. Box 221, The University of Reading, RG6 6AS, UK

## Introduction

*Theobroma cacao* is a tropical crop of major significance to the economies of producer and consumer countries throughout the world. The total value of the crop has been estimated at 3.6 billion US dollars annually (Anonymous 1995). Long-term sustainability of supply relies partly on the ability of breeders to generate new cultivars that are resistant to the many pests and diseases of the crop. Several international cocoa genebanks have been established to provide breeders with ready access to source material with desirable resistance and quality traits. These repositories typically contain many hundreds or even thousands of clones originating from collecting expeditions or donations from other genebanks. The large number of accessions held and the diversity of their origins creates operational difficulties for the management of these collections. Errors of documentation can occur during the collection, transportation, propagation or maintenance of material. Curators also face the problem of limited space or manpower resources. Most collections operate at, or close to, the capacity of the institution where stocks are held. Under such circumstances, new germplasm can be accommodated only through the provision of fresh resources or, more frequently, by the removal of accessions already held. It is of vital importance, therefore, that there is an easy means of identifying areas of duplication within a collection and, equally, for the recognition of valuable, genetically distinct stocks. It is also important in collections of living material (as in cocoa) that the curator is able to identify the level of variation that is appropriate to justify the continued storage of related clones. It is doubtful, for example, that many collection managers would chose to maintain large numbers of siblings that originate from a single, highly homozygous parent. Such information is frequently lacking for many accessions.

A number of strategies have been used to assess genetic variability in cocoa collections. Morphological descriptors are the simplest of these but these are sensitive to environmental influences. Many workers have there-

fore employed molecular and biochemical approaches including RFLP (Lerceteau et al. 1997), RAPD (Cruz et al. 1995) and isozyme analysis (Warren 1994) to reveal genetic variability. These studies generate important information relating to the genetic diversity of the crop and can enable curators to identify mislabelled stocks. To-date, however, these works have not provided an internal scale against which the variation between related clones can be measured.

The selection of an appropriate molecular technique for the routine screening of material in large germplasm collections is also subject to practical considerations. Factors of particular importance include speed, simplicity, cost and reproducibility of a protocol, as well as its ability to distinguish between large numbers of closely related individuals. PCRs which target ISSRs (inter-simple sequence repeats) utilise primers that are complementary to SSRs and that contain a 1–3 base 'anchor' at either the 3′ or 5′ end (Zietkiewicz et al. 1994). This anchor ensures that the primer anneals only to one end of the repeat motif. The complex band profiles generated by these primers are reproducible and highly polymorphic (Charters et al. 1999) and have been used in several species for cultivar identification and genetic distance analysis (e.g. Charters et al. 1996; Fang and Roose 1997; Nagoaka and Ogihara 1997).

The Intermediate Cocoa Quarantine Facility of The University of Reading/BCCCA (Biscuit, Cake, Chocolate and Confectionary Alliance) operates a service to provide cocoa material, which has been certified free of pests and disease, to breeders throughout the world. In this study, we use ISSR-PCR to evaluate the genetic status of 62 clones held in the collection. *Theobroma grandiflora* is used as an 'out group'. Two contrasting progenies from the self-pollination of a highly homozygous and a highly heterozygous parent are also used as 'in groups' to provide internal scale for the variability revealed by UPGMA.

## Materials and methods

### Plant material

Mature, but only recently fully expanded, leaves were collected at ambient temperature from accessions grown in heated polythene tunnels (18°C minimum) at The University of Reading Intermediate Cocoa Quarantine Facility (Table 1). A sample of *T. grandiflora* was kindly provided by A. Daymond of the Almirante Cacau Centre for Cocoa Research, Brazil. Samples for analysis were swabbed with 70% ethanol, rinsed in distilled water, wiped dry and stored at –20°C prior to DNA extraction.

### Seedling leaf material

Two self-pollinated progenies were generated to provide an internal scale for the genetic variability revealed by UPGMA. One progeny was derived from the self-pollination of SPEC 54.1. This plant has been found previously to be >95% homozygous (Charters and Wilkinson, unpublished data). The second progeny originated from the self-pollination of P 19B. This clone was collected by Pound from the river Nanay, Loreto, Peru and was originally described as self-incompatible. Repeated pollination with its own pollen gave rise to the progeny used in the current study.

Ten seedlings from one pod of accession P 19B and a further ten from one pod of accession SPEC 54.1 were raised in pots under glass. Cocoa beans from the same pod were grown together in a single 30-cm pot. For each seedling, one leaf from the first pair of true leaves was harvested and stored as above.

### DNA extraction and quantification

DNA extraction was performed using a commercially available DNA extraction kit (DNeasy Plant Mini Kit, Qiagen) according to the manufacturers' instructions. This required the disruption of 0.1 g of frozen leaf material by grinding in liquid nitrogen prior to cell-lysis and an RNase-A digest in lysis buffer (supplied) at 65°C. An optional centrifugation step was included prior to the removal of debris and salt-precipitated polysaccharides and proteins by column filtration. The lysates were passed through a second column matrix to which DNA binds and then washed twice with an ethanol-based buffer (supplied). DNA was removed from the column in two rinses, each using 100 µl of the elution buffer provided. DNA was quantified using a DyNA Quant 200 fluorimeter (Hoefer) and diluted to 10 ng/µl with sterile nano-pure water.

### Polymerase chain reaction (PCR)

Each reaction mixture contained 15.0 µl of sterile distilled water, 0.2 µl (1 unit) of *Taq* polymerase (Boehringer Mannheim), 2.0 µl of 15 mM $MgCl_2$ reaction buffer (Boehringer Mannheim, as supplied with the enzyme), 0.4 µl (0.2 µM) of dNTPs (equimolar solution of ATP, CTP, GTP and TTP, all Boehringer Mannheim), 0.4 µl (0.3 µM) of primer (Table 2) and 2.0 µl (20 ng) of template DNA (total volume 20.0 µl per reaction). The reaction mixture was over-laid with sterile mineral oil (Sigma) prior to PCR.

ISSR primer sequences, found to be useful in previous screenings (data not shown), were selected from UBC (University of British Columbia) set #9. These are listed in Table 2.

PCRs were performed on a Omnigene (Hybaid) or a PTC-100 (MJ Research, Inc.) thermocycler using the following programme: (1 min at 94°C, 2 min at 55°C, 30 s at 72°C) for 30 cycles followed by a 5-min extension period at 72°C. The annealing temperature was therefore 2–5°C above the mean-Tm of each primer used.

### ISSR-PCR analyses

PCRs were conducted on DNA samples derived from 83 individual plants using primers 812, 827, 834, 841, 857, 888 and 890. In addition, ten individuals from one pod of accession SPEC 54.1 and ten individuals from a single pod of accession P 19B were analysed using primers 812, 827, 834, 841, 857, 888, 889, 890 and 891.

### Electrophoresis and silver staining

Electrophoreses of pre-cast polyacrylamide gels (48 S Cleangels, Amersham Pharmacia Biotech) using the Multiphore-II flatbed system (Amersham Pharmacia Biotech) were conducted as previously described (Charters et al. 1996) with the exception that gels were re-hydrated in 75 mM Trizma acetate (Sigma) pH 6.4. Silver staining was conducted according to the method of Bassam et al. (1991) as modified by Charters et al. (1996).

### Gel analysis

Forty five selected band positions generated by the six most-informative primers were scored as present (1) or absent (0) from the profiles generated by 83 plants (Table 3). These data were then assembled into a data matrix. A UPGMA cluster analysis was performed using NTSYS-pc software (Jaccard co-efficient, phenetic-tree style).

**Table 1** Geographic origins of *Theobroma* accessions sampled for DNA extraction

| Group 1 | Origin |
|---|---|
| AM 1.8 | Ecuador, Hacienda Amalia |
| AM 3.9 (AM 1.39?)[a] | Ecuador, Hacienda Amalia |
| AM 2.43 | Ecuador, Hacienda Amalia |
| B 13.7 | Ecuador, Hacienda Balao |
| SLC 12 | Ecuador, H. Santa Lucia |
| EET 95 | Ecuador |
| JA 119 | Ecuador, Hacienda Javilla |
| JA 10.12 | Ecuador, Hacienda Javilla |
| JA 62 | Ecuador, Hacienda Javilla |
| **Group 2** | |
| B 5.7 | Ecuador, Hacienda Balao |
| B 13.1 | Ecuador, Hacienda Balao |
| **Group 3** | |
| P 19B parent | Peru, R. Nanay |
| P 19B progeny 6 | Peru, R. Nanay |
| P 19B progeny 8 | Peru, R. Nanay |
| P 19B progeny 10 | Peru, R. Nanay |
| P 19B progeny 1 | Peru, R. Nanay |
| P 19B progeny 2 | Peru, R. Nanay |
| P 19B progeny 5 | Peru, R. Nanay |
| P 19B progeny 9 | Peru, R. Nanay |
| P 19B progeny 3 | Peru, R. Nanay |
| P 19B progeny 4 | Peru, R. Nanay |
| P 19B progeny 7 | Peru, R. Nanay |
| **Group 4** | |
| AMAZ 12 | Peru, R. Amazonas, Iquitos |
| LCTEEN 86 | Ecuador, R. Yacuambi |
| CC 38 | Costa Rica (Amazon type) |
| **Group 5** | |
| CLM 90 | Ecuador, Hacienda Clementina |
| DOM 25 | Dominica |
| RIM 189 | Mexico |
| UF 221 | Costa Rica (Ecuador hybrid) |
| UF 677 | Costa Rica (Ecuador hybrid) |
| UF 705 | Costa Rica (Ecuador hybrid) |
| CUMAO 177 | Venezuela |
| R [10] MEX | Mexico/Pacific coast |
| **Group 6** | |
| DOM 30 | Dominica |
| MAR 9 | Martinique |
| MAR 13 | Martinique |
| GS 6 | Grenada |
| RED 127 | Amazon Valley |
| SIC 5 | Brazil, Bahia |
| SIAL 339 | Brazil, Bahia |
| Playa Alta 2 | Venezuela |
| **Group 7** | |
| SJ 119 | Ecuador, San Juan |
| **Group 8** | **Origin** |
| RB 29 | Brazil, Rio Branco |
| RB 41 | Brazil, Rio Branco |
| **Group 9** | |
| SPEC 54.1 parent | Columbian Amazon |
| SPEC 54.1 progeny 1 | Colombian Amazon |
| SPEC 54.1 progeny 2 | Colombian Amazon |
| SPEC 54.1 progeny 3 | Colombian Amazon |
| SPEC 54.1 progeny 4 | Colombian Amazon |
| SPEC 54.1 progeny 5 | Colombian Amazon |
| SPEC 54.1 progeny 6 | Colombian Amazon |
| SPEC 54.1 progeny 7 | Colombian Amazon |
| SPEC 54.1 progeny 8 | Colombian Amazon |
| SPEC 54.1 progeny 9 | Colombian Amazon |
| SPEC 54.1 progeny 10 | Colombian Amazon |
| **Group 10** | |
| AMAZ 6–3 | Peru, R. Amazonas, Iquitos |
| AMAZ 6–3 progeny | Peru, R. Amazonas, Iquitos |
| AMAZ 15–15 | Peru, R. Amazonas, Iquitos |
| EET 272 | Peru, Nanay Region[a] |
| SPEC 194–75 | Hybrid (Colombia×?)[a] |
| NA 756 | Peru, R. Nanay |
| NA 776 | Peru, R. Nanay |
| **Group 11** | |
| BORNE 7 A2 | French Guiana., R. Kerindioutou |
| GU 154 C | French Guiana, R. Camopi |
| GU 290 H | French Guiana, R. Camopi |
| GU 221 C | French Guiana, R. Camopi |
| PINA | French Guiana, Oyapok |
| KER 2/E | French Guiana, R. Kerindioutou |
| KER 9 | French Guiana, R. Kerindioutou |
| KER 3 | French Guiana, R. Kerindioutou |
| **Group 12** | |
| EET 162 | Hybrid (Ecuador×?)[a] |
| SIAL 93 | Hybrid (IMC 67×PA 218) |
| PA 56 | Peru, Parinari |
| PA 88 | Peru, Parinari |
| PA 184 | Peru, Parinari |
| **Group 13** | |
| CRUZ 7–11 | Hybrid ? (Brazil×?)[a] |
| ICS 70 | Trinidad |
| MO 20 ? | Peru ? |
| LCTEEN 31? | Hybrid ? (Ecuador[b]) |
| SCA 6 | Peru, Rio Ucayali |
| SCA 19 | Peru, Rio Ucayali |
| SCA 24 | Peru, Rio Ucayali |
| **Group 14** | |
| *T. grandiflora* | Outgroup |

[a] For hybrid or suspected hybrid (?) plants, the origins of parental clone(s) are given where possible. MO 20 exhibited atypical morphology for this accession and may be a hybrid or mislabelled clone

[b] LCTEEN 31 is known to originate from Ecuador but on the basis of its morphology the plant held at Reading was suspected of being a hybrid by Dr. B. Bartley

## Reproducibility

Reproducibility of the ISSR protocol was assessed in three ways. First, independent DNA extractions were performed from leaves taken from three separate branches of clone SIAL 339. ISSR-PCR was conducted according to the protocol described above using four different primers (812, 841, 890 and 891).

Second, DNA extracted from AMAZ 15–15 was subjected to replicate ISSR-PCR with primers 812, 834 and 841 using a Omnigene (Hybaid) or a PTC-100 (MJ Research, Inc.) thermocycler. Products were fractionated and stained as described above.

Third, independent PCRs from the following clones were duplicated over different gels for all primers and used as reference pro-

**Table 2** Primers used in ISSR-PCR analysis

| I.D. No. | Sequence[a] | Source | I.D. No. | Sequence[a] | Source |
|---|---|---|---|---|---|
| 812 | $[GA]_8$–A | Gibco BRL | 888 | BDB-$[CA]_7$ | Gibco BRL |
| 827 | $[AC]_8$–G | Pharmacia | 889 | DBD-$[AC]_7$ | Gibco BRL |
| 834 | $[AG]_8$–YT | Pharmacia | 890 | VHV-$[GT]_7$ | Gibco BRL |
| 841 | $[GA]_8$–YC | Gibco BRL | 891 | HVH-$[TG]_7$ | Gibco BRL |
| 857 | $[AC]_8$–YG | Pharmacia | | | |

[a] Where Y=C or T; B=C, G or T (i.e. not A), and similarly D=not C, H=not G, V=not T

**Table 3** Bands selected for statistical analysis

| Primer | No. of bands selected | Primer | No. of bands selected |
|---|---|---|---|
| 812 | 10 | 841 | 8 |
| 827 | 5 | 888 | 10 |
| 834 | 2 | 890 | 10 |

files for the genetic-variation experiments: CC 38; EET 272; JA 119; RED 127 and SPEC 54.1.

## Results

Technique reproducibility: band profiles generated by different leaves of the clone SIAL 339 were identical for all four primers employed and no variation was detected in the band profile of AMAZ 15–15 generated by different PCR thermocyclers using any of the primers examined. No variation was observed between reference profiles in any of the gels used in the analysis of genetic variation (see below).

Genetic variation revealed: ISSR-PCR profiles varied between accessions for all of the primers screened (Fig. 1). Fingerprints produced using primers 812, 827, 834, 841, 888 and 890 yielded the greatest variability between accessions. Profiles generated by these primers were used to compile a data matrix for 82 cocoa accessions on the basis of 45 band positions.

The plant material compared included two self-pollinated progenies derived from a single pod on parental trees thought to differ in their levels of heterozygosity. These progenies were used to provide an internal scale for the degree of genetic variability detected between clonal accessions. *T. grandiflora* was employed as an 'out group' and was used to provide a scale for the more-distant clusters generated. The progeny of P 19B could all be distinguished from each other and also from the parental clone. All bands detected in each of the offspring were present in the profile of the parental tree

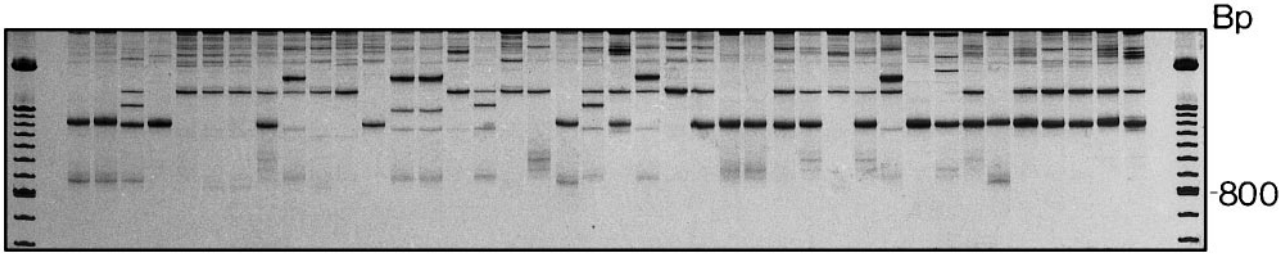


**Fig. 1** Banding profiles of 40 *T. cacao* individuals generated using primer 827. *Lanes* from left to right: (*1*) 100 bp ladder; (*2*) negative control (no template DNA); (*3–42*) banding profiles of 40 individual cocoa accessions; (*43*) negative control; (*44*) 100 bp ladder. The gel section shown is approximately 0.75× life size

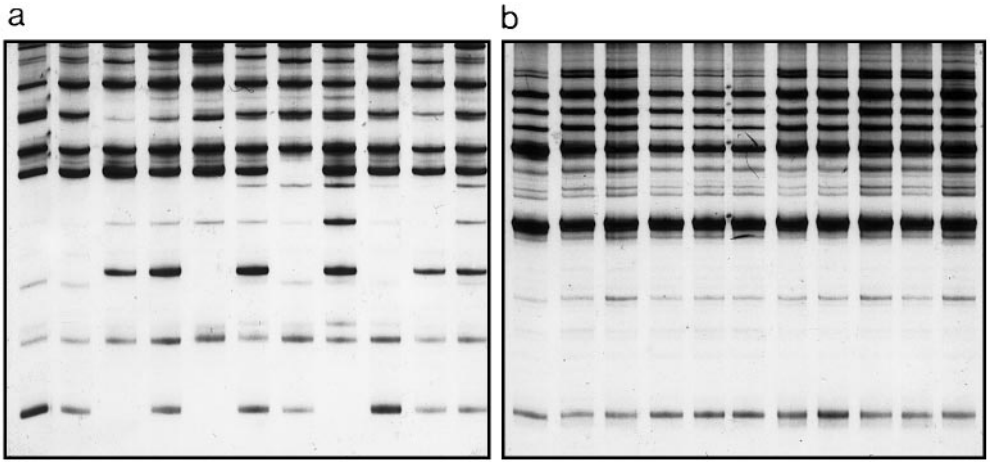


**Fig. 2a, b** Banding profiles of two progenies generated using primer 888. **a** *Lanes* from left to right show the profiles of ten full siblings and the heterozygous parental accession SPEC 54.1. The gel section shown is approximately 1.3× life size. **b** *Lanes* from left to right show the profiles of ten full siblings and the homozygous parental accession P 19B. The gel section is shown approximately 1.3× life size

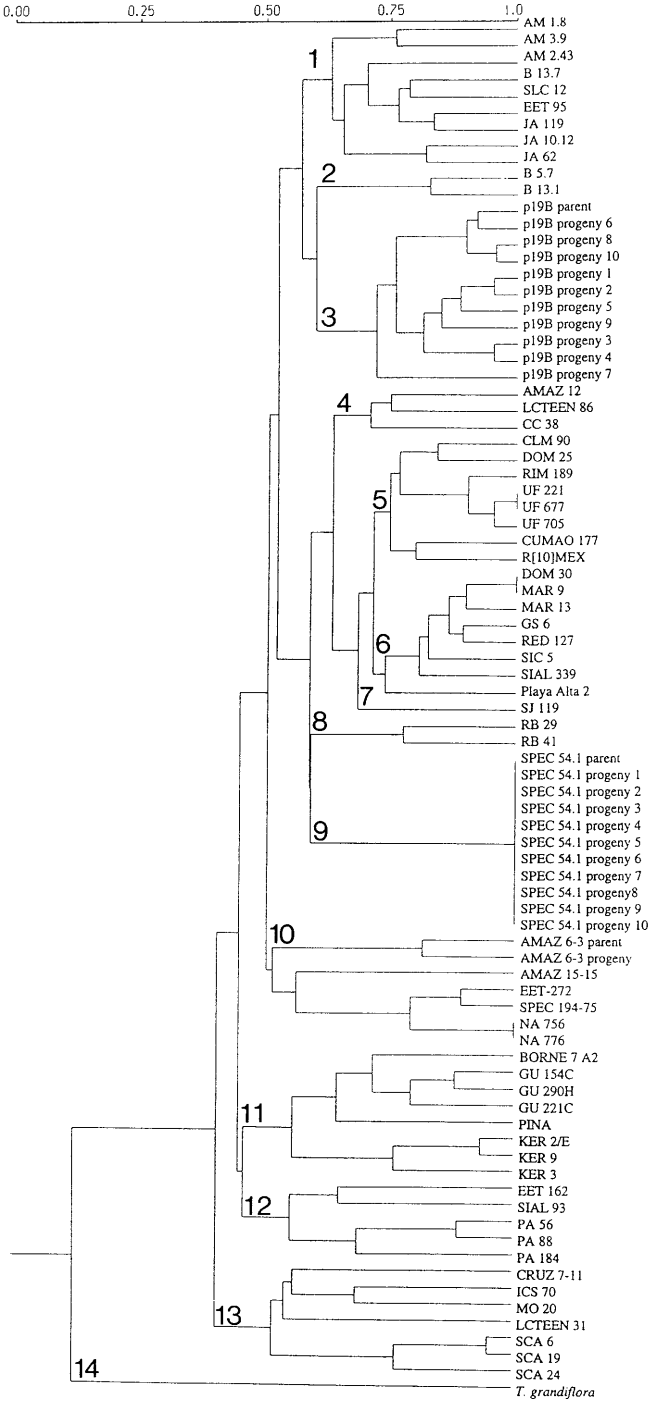


**Fig. 3** Dendrogram comparing 82 *T. cacao* individuals and one representative of *T. grandiflora* (outgroup) on the basis of the profiles generated using six ISSR primers

(Fig. 2a), as would be expected from a progeny produced by self-pollination. The parental profile contained 25 bands of which 13 (52%) segregated in the progeny, thereby confirming the highly heterozygous nature of P 19B. In contrast, the ten self-pollinated progeny from SPEC 54.1 generated identical band profiles that were indistinguishable from the parental plant for all of the 45 band positions scored (Fig. 2b).

Overall, 56 of the 62 accessions could be distinguished on the basis of their ISSR band profiles. Three pairs of genotypes (DOM 30/MAR 9, NA 756/NA 776 and UF 221/UF 677) produced identical profiles for all bands scored.

There was a tendency for clones collected from the same area to cluster together on the UPGMA dendrogram (Fig.3, and see Table 1). The 14 main clusters identified (Fig. 3) included two containing the 'in group' progenies (clusters 3 and 9), eight containing clones from a single region or country plus hybrids (clusters 1,2,7,8,10,11,12,14), and only four containing material from several regions and hybrids (clusters 4,5,6,13). Interestingly, material collected from the Caribbean islands was restricted to these mixed groups.

The clusters formed by the two self-pollinated progenies differed in the level of similarity revealed between siblings. Those derived from SPEC 54.1 were identical and so formed a uniform cluster of individuals with a similarity of 1.0. The progeny of the highly heterozygous P 19B, in contrast, formed a more heterogeneous group that divided from other clones at 0.72. The major clusters identified all diverged from other clones at this (clusters 5 and 6) or lower levels of similarity.

Accessions sharing a common code prefix are frequently related or else known to have been collected from a common population or area (see Table 1). The precise nature of the relationship (e.g. full or half siblings) is rarely documented. For the most part, 'related' clones with the same code prefix clustered closely together and showed a higher level of similarity than the P 19B 'in group' progeny. For example, three clones with the prefix KER were all collected from a single population along the Oyapock tributary of the river Kerindioutou in French Guiana. These plants formed a discrete grouping with a similarity of 0.76 on the UPGMA dendrogram. Likewise, the clones with the GU prefix formed a cluster with a similarity of 0.81. These plants originate from the progenies of a few fruits collected along the river Camopi, French Guiana. There were some cases, however, where clones with the same prefix did not cluster together. In the majority of these examples, plants with the same code prefix do not share a common origin or else included both hybrids and selections. One notable exception concerns the separation of B13.7 (cluster 1) and B13.2 (cluster 2). These plants were documented as originating from a single progeny collected by Pound in Ecuador.

## Discussion

ISSR-PCR generated band profiles that could distinguish between 56 of the 62 cocoa accessions studied using only six primers. This level of informativeness, coupled with the speed, simplicity and reproducibility (Charters et al. 1999) of the protocol, demonstrates its potential for the large-scale evaluation of cocoa germplasm collections.

The UPGMA dendrogram generated from the ISSR band profiles was largely able to separate accessions into clusters containing plants from a single geographic area (plus hybrids). Four clusters, however, contained mixtures of material originating from several regions. There are several possible explanations for this. It could be argued that more data are required to resolve these clusters into smaller, more genetically meaningful, groupings. This possibility cannot be dismissed. The fact that all representatives of the more heterogeneous 'in group' clustered together, as did the vast majority of clones from a common source, suggests, however, that this may not be the entire answer. Perhaps a more plausible explanation lies in the widespread transportation and distribution of clones during the history of the crop. It is interesting to note, for example, that all four clusters containing accessions from several regions included one or more accessions from the Caribbean islands. Wood (1985) suggested that cocoa cultivation spread to the Caribbean Islands after the Spanish conquest of Mexico and that new stocks were repeatedly re-introduced into the area from both Central and South America after cocoa crops suffered periodic devastation by hurricane or disease. It follows that the area currently contains a complex mixture of clones and hybrids from a range of geographic origins. This would account for the failure of material from this area to cluster together. Mislabelling and other forms of documentation errors may also account for at least some of the observed clustering of material from diverse sources. For example, B. Bartley (personal communication) noted that MO 20, represented in cluster 13, has a distinctly different morphology to a tree with the same code in the USDA (United States Department of Agriculture) collection at Mayaguez, Puerto Rico. It is possible, therefore, that this tree is mislabelled.

The inclusion of contrasting selfed progenies as 'in groups' and *T. grandiflora* as an 'out group' in the analysis provides three reference points against which genetic similarity of the cocoa accessions could be measured. The finding that band profiles from the (homozygous) SPEC 54.1 progeny were all indistinguishable illustrates that identical band patterns are not proof of genetic fidelity. Rather, it provides evidence that such plants are either one clone (e.g. cuttings) or else are genetically extremely similar (e.g. full siblings from a homozygous parent). In the context of the collection as a whole, however, such accessions should be viewed as being 'effectively' genetic duplicates. It follows that one of the two duplicate accessions should be considered of very low priority for continued storage if fresh material needs to be accommodated.

The high proportion of segregating bands observed in the progeny of P 19B confirms the heterozygous nature of this plant. The progeny generated from this plant therefore represents the opposite extreme of variability expected by self-pollination. This progeny formed a distinct cluster that separated from the other clones at 0.72. It is perhaps interesting to compare this level of variation with that seen between AMAZ 6–3 and the self-pollinated progeny of this plant, AMAZ 6–3 progeny. These two clones were more-closely clustered than the P 19B progeny. In contrast, EET 162 is the product of a cross between disparate accessions (IMC 67×PA 218). The UPGMA included this plant and three close relatives of the paternal parent (PA 56, PA 88, PA 184). In the resultant dendrogram, EET 162 clustered with the PA clones, but at lower levels of similarity than observed between the PA clones themselves or between members of the P 19B 'in group'. These data are consistent with the level of similarity shown by the P 19B 'in group' providing an approximate limit to that expected from a self-pollinated progeny. This information can be used to assess the level of genetic diversity between groups of accessions and can also be used to resolve documentation anomalies. For instance, there were three clones possessing the AM prefix in the study: AM 1.8, AM 3.9 and AM 2.43. These plants all derive from the progenies of two trees from Ecuador, labelled AM 1 and AM 2. There was not parental plant labelled AM 3. Thus, AM 3.9 was suspected to be a mislabelled progeny from an original AM 1.39 or AM 2.39. This plant clustered more-closely to AM 1 than AM 2. Furthermore, it showed a greater level of similarity to AM 1 and a lower level of similarity to AM 2 than the P 19B 'in group' plants showed to each other. These results are therefore consistent with the mislabelled plant being a progeny of AM 1 and this plant should probably be re-labelled AM 1.39. Another example concerns two accessions labelled as siblings from fruit collected by Pound in Ecuador B 13.1 and B 13.7. Clone B 13.1 clustered closely with a third accession taken from the same collection (B 5.7) but was more dissimilar to B 13.7 than the P 19B progeny were to each other. There are two trees of B 13.7 held in Reading. These are morphologically distinct from each other, and the individual not used here (Plant 2) possesses a phenotype typical of the 'B' code prefix (Bartley, personal communication). These results therefore support the morphological evidence suggesting that B 13.7 plant 1 may have been mislabelled.

The use of progenies of contrasting provenance as 'in groups' has thus provided valuable information that has enabled genetic variability revealed by ISSR-PCR to be placed into context. The inclusion of such groups should allow the effective evaluation of new material being introduced to a collection and also to identify areas of effective genetic duplication. Accessions that are no more different to existing stocks than is expected from the selfed progeny of a heterozygous parent have low priority for maintenance compared with those that are significantly more diverse. In the material studied here, 47 of the 62 accessions differed from others by less than this level of similarity. This information could be used to identify superfluous clones should the removal of stocks be required to accommodate new material at a future date. Accessions that cannot be distinguished by their combined profiles would be replaced first, followed by those that are only minor variants of existing stocks (less variable than full sibs). The point of divergence of the

'out group' (*T. grandiflora*) sets the outer limit of genetic diversity expected in the collection. This helps to confirm the species identity of all stocks examined and provides a measure of the genetic diversity contained in the collection as a whole.

We conclude that the use of genetically characterised progenies as 'in groups' as well as a related species as an 'out group,' provides valuable reference points that have practical value for the genetic evaluation of clonal germplasm collections.

**Acknowledgements** We thank Cocoa Research (UK) Ltd. for funding this work. We also thank Dr. Basil Bartley, Ms. Caroline Ford and Dr. Michelle End for their help and advice on the origins of material used in the study. Our thanks also to Dr. Alistair Culham for his assistance with the UPGMA software and to Professor Paul Hadley of The University of Reading/BCCCA Intermediate Cocoa Quarantine Facility for plant material.

## References

Anonymous (1995) Quarterly bulletin of cocoa statistics 21. ICCO (International Cocoa Organisation), London, UK

Bassam BJ, Caetano-Anolles G, Gresshoff PM (1991) Fast and sensitive silver staining of DNA in polyacrylamide gels. Ann Biochem 196:80–83

Charters YM, Robertson A, Wilkinson MJ, Ramsay G (1996) PCR analysis of oilseed rape (*Brassica napus* L. ssp. *oleifera*) using 5′-anchored simple sequence repeat (SSR) primers. Theor Appl Genet 92:442–447

Charters YM, Culham A, End M, Hadley P, Wilkinson MJ (1999) The potential of anchored microsatellite analysis for cocoa germplasm characterization. Proc 12th Int cocoa Res Conf 17–23 Nov 1996, Salvador-Bahia, Brazil

Cruz de la M, Whitkus R, Gomez-Pompa A, Mota-Bravo L (1995) Origins of cacao cultivation. Nature 375:542–543

Fang DQ, Roose ML (1997) Identification of closely related citrus cultivars with inter-simple sequence repeat markers. Theor Appl Genet 95:408–417

Lerceteau E, Robert T, Petiard V, Crouzillat D (1997) Evaluation of the extent of genetic variability among *Theobroma cacao* accessions using RFLP and RAPD markers. Theor Appl Genet 95:10–19

Nagaoka T, Ogihara Y (1997) Applicability of inter-simple sequence repeat polymorphisms in wheat for use as DNA markers in comparison to RFLP and RAPD markers. Theor Appl Genet 94:597–602

Warren JM (1994) Isozyme variation in a number of populations of *Theobroma cacoa* L. obtained through various sampling regimes. Euphytica 72:121–126

Wood GAR (1985) History and development. In: Wood GAR, Lass RA (eds) Cocoa (4th edn). Longman, London New York, pp 1–10

Zietkiewicz E, Rafalski A, Labuda D (1994) Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. Genomics 20:176–183