South Sudan Arabica Coffee Land Race Survey in Boma
Germplasm Assessment and Conservation
Project Report
Dr. Sarada Krishnan
Dr. Aaron P. Davis

1. Introduction and Background:

Coffee is an extremely important agricultural commodity (Vega et al. 2003) produced in about 80 tropical countries, with an annual production of nearly seven million tons of green beans (Musoli et al. 2009). It is the second most valuable commodity exported by developing countries after oil, with over 75 million people depending on it for their livelihood (Vega et al. 2003; Pendergrast 2009). It is thought that coffee was introduced to Yemen from its origins in Ethiopia around the sixth century (Pendergrast 1999). From Yemen, two genetic bases spread giving rise to most of the present commercial cultivars of Arabica coffee grown worldwide (Anthony et al. 2002). The two sub-populations of wild coffee introduced from Ethiopia to Yemen underwent successive reductions in genetic diversity with the first reduction occurring with the introduction of coffee to Yemen 1,500 to 300 years ago (Anthony et al. 2002). Introduction of coffee to Java, Amsterdam, and La Réunion at the beginning of the 18th century led to further reductions in genetic diversity (Anthony et al. 2002).

In addition to Ethiopia, wild plants of *C. arabica* were observed in the Boma Plateau of South Sudan (Thomas 1942; Meyer 1965) and Mount Marsabit in northern Kenya (Meyer 1965). A consortium led by Texas A&M University's Norman Borlaug Institute for International Agriculture has been commissioned to support the John Garang University of Science and Technology (JG-MUST) of South Sudan. The goal of this United States Agency for International Development (USAID) funded project is to improve applied agriculture and technical education, create research capacity and institutional linkages to increase productivity, conservation, and resource management of coffee.

As part of the JG-MUST project and because the Borlaug Institute is also headquarters for the global collaborative research program known as World Coffee Research (WCR), a sub-project was proposed and accepted under JG-MUST to determine if wild *C. arabica* was still growing indigenously within the Boma forest of South Sudan as earlier reported by Thomas (1942). If confirmed, then WCR would fund a genetic diversity study to ascertain if the genetic diversity found was unique to South Sudan which is the subject matter of this report. If deemed unique or useful to coffee improvement activities of WCR, then WCR would work with the South Sudanese government to collect and evaluate its germplasm for the benefit of South Sudan and the coffee sector at large. Finally, because South Sudan is a new country with very few trained scientists, this JG-MUST-WCR sub-project was designed to train students and professors from the JG-MUST, as well as pertinent South Sudanese government agricultural officials, in the wild coffee germplasm collection, evaluation and eventually, sector development, if and when Arabica coffee was determined to be a valuable income generator for the populations.

The specific research objectives of this first mission to South Sudan were to:

1. Determine the current status of wild and cultivated Arabica coffee in the Boma Plateau.
2. Assessment of genetic diversity of wild populations of *Coffea arabica* in the Boma Plateau.

3. Based on genetic diversity studies, establish a *Coffea* field genebank in South Sudan using specimens representing the greatest genetic diversity.

2. Field Work:

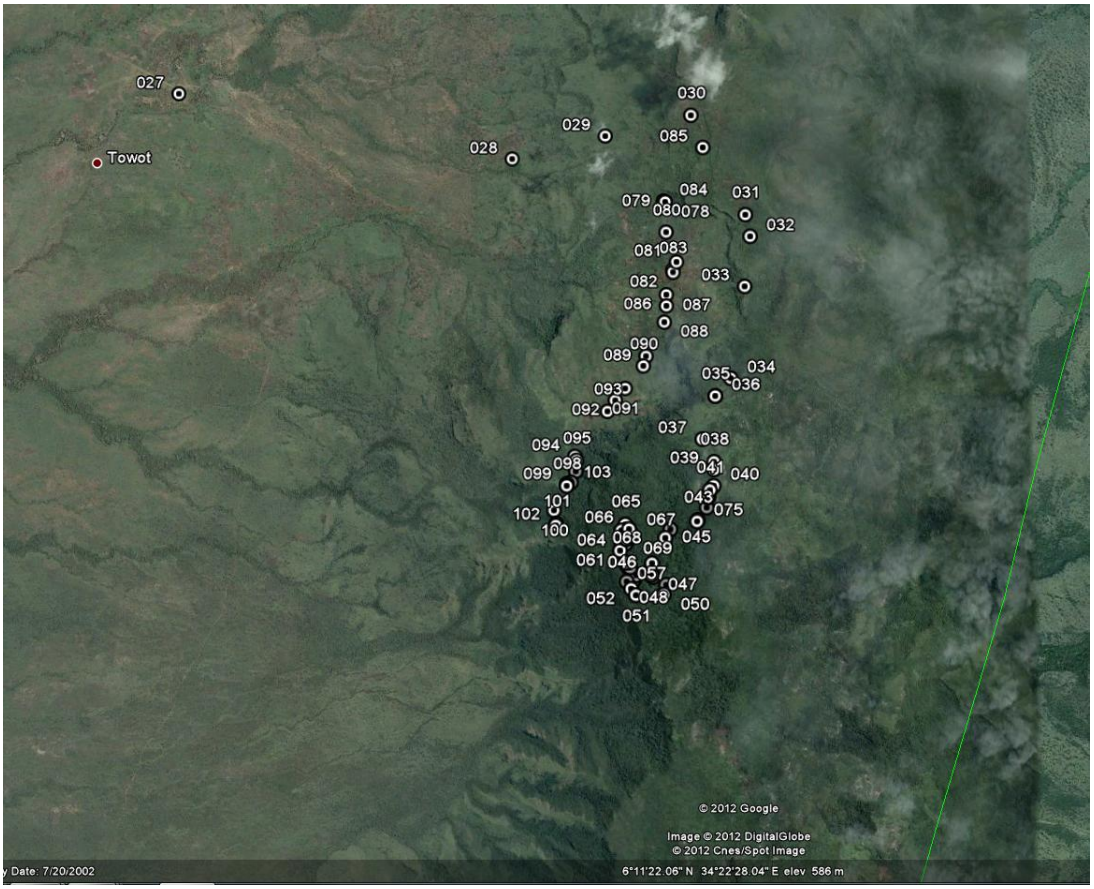
Field work in Upper Boma in the Jonglei State of South Sudan was conducted April 9 – 12, 2012. The expedition consisted of the following personnel:

- Dr. Timothy Schilling, World Coffee Research
- Dr. Aaron Davis, Royal Botanic Gardens, Kew
- Dr. Sarada Krishnan, Denver Botanic Gardens
- Emma Bladyka, Specialty Coffee Association of America
- Lindsey Bolger, Green Mountain Coffee
- Solomon Kuol Deng, State Ministry of Agriculture and Forestry, Jonglei/Bor
- Richard Yona, John Garang Memorial University of Science and Technology
- Majok Ayuen, John Garang Memorial University of Science and Technology
- Thon Nyok Dor, John Garang Memorial University of Science and Technology
- Solomon Moore, Wall Street Journal

3. Geographic Information:

Figure 1 shows geographic distance covered by the expedition during four days of field work.

Figure 1: Map showing area covered during the expedition. Way point 027 corresponds to landing strip at Lower Boma. Way point 030 corresponds to the expedition's camp site in Jonglei Village in Upper Boma.



4. Collections:

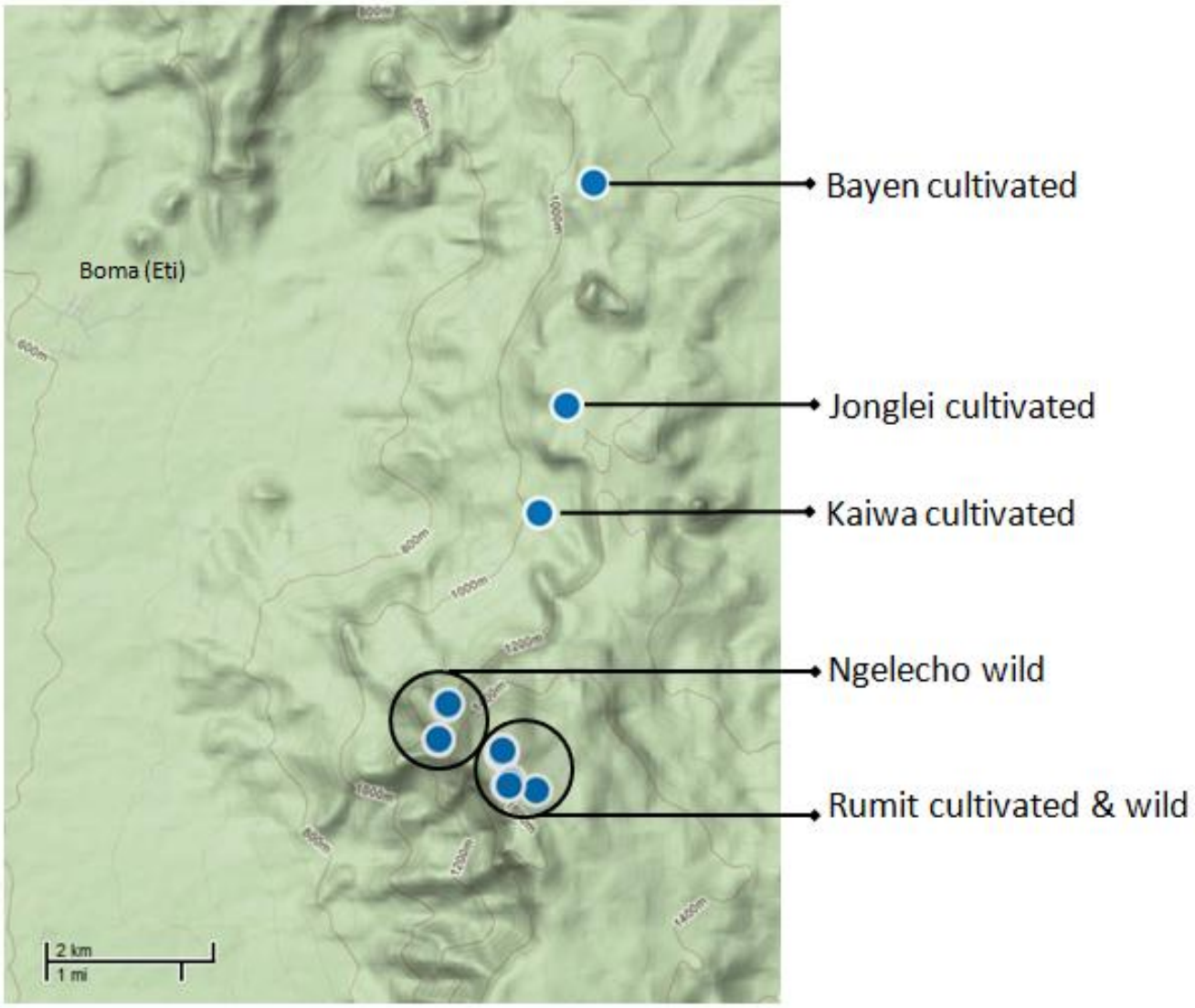
Over four days, collections were made from cultivated plants in local villages (Rumit/Zoch, Kaiwa, Jonglei and Bayen) and two forest locations (Rumit and Ngelecho). Table 1 gives a summary of the different populations sampled. An accession list with location coordinates is provided in Appendix 1. Figure 2 shows the geographic locations of the individual populations.

Table 1: Summary of *Coffea arabica* populations sampled

Population Name	**Collection Numbers***	**# Samples**
Rumit Cultivated 1	B1 – B32	32
Rumit Cultivated 2	B47 – B56	10
Kaiwa & Jonglei Cultivated	B57; B59	2
Bayen Cultivated	B61 – B63	3
Rumit Wild	B33 – B46	14
Ngelecho Wild	B64 – B76	13

*Detailed collections list with location coordinates and herbarium sample information is provided in Appendix 1.

Figure 2: Map showing locations of individual populations.



5. Genetic Analysis:

a. Materials & Methods:

Plant material

Table 1 lists the cultivated and wild populations of *Coffea arabica* sampled. In addition, sampling was also done from two wild plants of *C. neoleroyi*, a wild relative of cultivated coffee, though genetic study results for these two samples are not included in this report. Location coordinates were recorded using a Garmin eTrex Vista NCx and a Holux M-241. Several leaves of each individual plant were collected and placed in plastic bags with silica gel (Chase and Hillis, 1991). Voucher specimens of selected samples were collected in replicates of three and will be housed at the herbaria at the Royal Botanic Gardens, Kew (K), Missouri Botanical Garden (MO), and at a suitable location in South Sudan (herbaria not yet established). As collections were being made, individual trees were tagged with numbered aluminum tags nailed to the ground, and with orange tree tape with the corresponding number tied to the truck of the tree. Tagging was done for ease of location and plant identification during subsequent seed collecting expeditions.

As there were only one sample each from Kaiwa and Jonglei villages, these were pooled together as one population for the genetic diversity analysis.

DNA extraction and molecular markers

Genomic DNA was extracted from 10 mg of silica-dried leaf material using GenCatch™ Plant Genomic DNA Purification kits (Epoch Biolabs) at the Conservation Genetics lab at Denver Botanic Gardens. Slight modifications were made to the extraction protocols. A detailed account of the extraction procedure is described in Krishnan et al. (2012). Extracted DNA was sent to Nevada Genomics, Reno, Nevada, USA, for quantification, optimization, fragment analysis, and scoring using microsatellite markers. Initially 20 microsatellite markers were selected based on Combes et al. (2000), Poncet et al. (2004), and Cubry et al. (2008). Of these 20 markers, one did not amplify (M253), as so this marker was dropped from the study. The remaining 19 markers were used (Table 2).

The DNA was quantified and normalized to 5 ng/µl. PCR amplifications were conducted using an MJ thermocycler. Each 10 µl PCR amplification reaction contained 4 µl of 5 ng/µl genomic DNA, 1 µl Primer Panel mix, and 5 µl Qiagen Multiplex PCR Mix. The nineteen microsatellite markers were multiplexed in four panels with a different annealing temperature for each panel. The PCR cycling conditions were as follows: 95°C for 15 min; 95°C for 30 sec; specific panel annealing temperature for 45 sec; 40 cycles at 72°C for 45 sec; 72°C for 30 min. Panel 1 consisted of M780, M790, M259, M254, and M257 with an annealing temperature of 60°C. Panel 2 consisted of M258, M746, M782, M837, and M784 with an annealing temperature of 61°C. Panel 3 consisted of M256, M838, M255, M809, and M883 with an annealing temperature of 63°C. Panel 4 consisted of M260, M779, M774, and M764 with an annealing temperature of 62°C. The samples were run on an Applied Biosystems Prism 3730 DNA Analyzer. The filter set used was G5,

which detects the fluorescent dyes 6-FAM, VIC, NED, and PET. The samples were run with the 500 MW size standards labeled with LIZ.

The fragment analysis results were scored using GeneMarker 2.2.0 Software by SoftGenetics®.

Genetic Diversity Analysis

Data analyses to assess genetic diversity were performed using GenAlEx Version 6.4 software (Peakall and Smouse 2006). Parameters estimated included mean number of alleles per locus (Na), mean number of effective alleles per locus (Ne), the mean observed (H_o) and the mean expected (H_e) heterozygosities based on Hardy-Weinberg assumptions, the allelic fixation index (F_{is}), the number of observed genotypes per population (G), the number of private alleles per population (Npr), and the percentage of polymorphic loci per population (P). Nei's genetic distance (Nei 1978) was calculated between pairs of populations to estimate genetic distance among populations. The distinctiveness among populations was investigated using population assignment tests based on allele frequencies for populations 1 and 2 (Rumit cultivated 1 and 2 respectively). Hierarchical genetic structure was examined through an analysis of molecular variance (AMOVA) (Excoffier et al. 1992) with 999 permutations using GenAlEx V6.4 (Peakall and Smouse 2006). AMOVA was applied to estimate the components of variance among and within populations based on F_{st} for each population.

Table 2: GenBank EMBL accession number, locus code, primer sequences, repeat motif structures, product size, and reference of the 19 primer pairs used

EMBL accession #	Locus code	Primer sequence	Repeat motif	Product size (bp)	Reference
AJ250254	M254	F: GGCTCGAGATATCTGTTTAG R: TTTAATGGGCATAGGGTCC	(CA)15(CG)4CA	132	Combes et al. 2000
AJ250255	M255	F: CCCTCCCTGCCAGAAGAAGC R: AACCACCGTCCTTTTCCTCG	(GT)5/CT(GT)2/(GT)12	160	Combes et al. 2000
AJ250256	M256	F: AGGAGGGAGGTGTGGGTGAAG R: AGGGGAGTGGATAAGAAGG	(GT)11	118	Combes et al. 2000
AJ250257	M257	F: GACCATTACATTTCACACAC R: GCATTTTGTTGCACACTGTA	(CTCACA)4/(CA)9	103	Combes et al. 2000
AJ250258	M258	F: AACTCTCCATTCCCGCATTC R: CTGGGTTTTCTGTGTTCTCG	(CA)3/(CA)3/(CA)18	100-132	Combes et al. 2000
AJ250259	M259	F: ATCCGTCATAATCCAGCGTC R: AGGCCAGGAAGCATGAAAGG	(GT)3/(GT)7	72	Combes et al. 2000
AJ250260	M260	F: TGATGGACAGGAGTTGATGG R: TGCCAATCTACCTACCCCTT	(CT)9/(CA)8/(CT)4/(CA)5	100	Combes et al. 2000
AJ308746	M746	F: GGCCTTCATCTCAAAAACCT R: TCTTCCAAACACACGGAGACT	(CT)12/(CA)11	378	Rovelli et al. 2000
AJ308764	M764	F: CTGGCATTAGAAAGCACCTTG R: GCTTGGCTCACTGTAGGACTG	(CA)10	158	Rovelli et al. 2000
AJ308774	M774	F: GCCACAAGTTTCGTGCTTTT R: GGGTGTCGGTGTAGGTGTATG	(CA)7	228	Rovelli et al. 2000
AJ308779	M779	F: TCCCCCATCTTTTTCTTTCC R: GGGAGTGTTTTTGTGTTGCTT	(GT)17	116	Rovelli et al. 2000
AJ308780	M780	F: ATTCTCTCCCCCTCTCTG R: GTTAGTATGTGATTTGGTGTGG	(CA)6	95	Rovelli et al. 2000
AJ308782	M782	F: AAAGGAAAATTGTTGGCTCTGA R: TCCACATACATTTCCCAGCA	(GT)8	114	Rovelli et al. 2000
AJ308784	M784	F: TTGCTTGCTTGTTCTGTTAT R: TGACACGAGAGTTAGAAATGA	(GT)7/(GC)7/(GT)7	126	Rovelli et al. 2000
AJ308790	M790	F: TTTTCTGGGTTTTCTGTGTTCTC R: TAACTCTCCATTCCCGCATT	(GT)21	134	Rovelli et al. 2000
AJ308809	M809	F: AGCAAGTGGAGCAGAAGAAG R: CGGTGAATAAGTCGCAGTC	(ATG)11	144	Rovelli et al. 2000
AJ308837	M837	F: CTCGCTTTCACGCTCTCTCT R: CGGTATGTTCCTCGTTCCTC	(GT)16(GA)11	102	Rovelli et al. 2000
AJ308838	M838	F: CCCGTTGCCATCCTTACTTA R: ATACCCGATACATTTGGATACTCG	(CA)9	100	Rovelli et al. 2000
AJ308883	M883	F: CGTCTCGTTTCACGCTCTCT R: GATCTGCATGTACTGGTGCTTC	(GT)15	237	Rovelli et al. 2000

b. Results:

Four loci were monomorphic across all populations (M774, M780, M809, M838). Table 3 summarizes the mean population level genetic variability across the 19 loci tested. Genetic diversity was highest for the Rumit Cultivated 1 population. The mean number of alleles (Na) for this population was 2.63, which was significantly higher than all other populations (two-tailed *t* test, $P < 0.05$) and had nine private alleles (Npr). The Rumit Wild population had the second highest Na with an average of 2.16, which was significantly different from all other populations except Rumit Cultivated 2, which had an Na of 1.90 (two-tailed *t* test, $P < 0.05$). The Rumit Wild population had one private allele. The Bayen Cultivated population had seven private alleles. The observed heterozygosity was higher than expected heterozygosity for all populations, leading to negative F_{is} values, suggesting heterozygosity excess.

Table 4 gives a summary of private alleles for the three populations with private alleles with the locus, allele size and frequency information. Locus M254 had five private alleles across all three populations and locus M258 had four private alleles, two each in Rumit Cultivated 1 and Bayen Cultivated populations.

Table 5 gives the Nei's Genetic Distance, showing the pairwise matrix of genetic distance between populations. Population 4 (Bayen Cultivated) is the most distant from all other populations with a distance ranging from 0.224 between Population 2 (Rumit Cultivated 2) to 0.432 between Population 6 (Ngelecho Wild). The distance among all other populations range from 0.012 (Pop 1 and Pop 2) to 0.157 (Pop 2 and 6). Population assignment test also showed the genetic distinctiveness of the Bayen Cultivated population (Pop 4), which clustered separately than all the other populations (Figure 3).

Since the Bayen Cultivated population was genetically distinct from all other populations as seen from the Nei's Genetic Distance matrix (Table 5), AMOVA was performed with all six populations and again with the Bayen Cultivated population removed to see if this would reduce the among-population variation for the rest of the five populations. The AMOVA results are presented in Table 6. Most of the variation was within populations. When genetic variation was partitioned among all six populations, the among-population variation was 10% compared to 7% among-population variation when the Bayen Cultivated population was removed.

Table 3: Summary of mean population level genetic variability within *Coffea arabica* at 19 microsatellite loci. N = sample size, Na = mean number of alleles per locus, Ne = mean number of effective alleles per locus, H_o = mean observed heterozygosity, H_e = mean expected heterozygosity, Fis = mean allelic fixation index, G = mean number of genotypes, Npr = number of private alleles per population, and P = percentage of polymorphic loci.

Population	**N**	**Na**	**Ne**	**H_o**	**H_e**	**F_{is}**	**G**	**Npr**	**P**
Rumit Cultivated 1	32	2.63	1.78	0.561	0.361	-0.401*	2.21	9	78.95
Rumit Cultivated 2	10	1.90	1.72	0.595	0.347	-0.706	1.42	0	68.42
Kaiwa & Jonglei Cultivated	2	1.84	1.75	0.579	0.349	-0.651	1.32	0	68.42
Bayen Cultivated	3	1.63	1.54	0.439	0.260	-0.647	1.21	7	52.63
Rumit Wild	14	2.16	1.71	0.579	0.352	-0.546	1.63	1	73.68
Ngelecho Wild	13	1.84	1.67	0.603	0.332	-0.745	1.37	0	68.42

*$P<0.001$ (significantly deviates from Hardy-Weinberg equilibrium)

Table 4: Summary of private alleles by population

Population	Locus	Allele size	Frequency
Rumit Cultivated 1	M254	154	0.016
Rumit Cultivated 1	M254	173	0.047
Rumit Cultivated 1	M257	121	0.031
Rumit Cultivated 1	M258	131	0.016
Rumit Cultivated 1	M258	133	0.031
Rumit Cultivated 1	M260	116	0.125
Rumit Cultivated 1	M784	130	0.016
Rumit Cultivated 1	M837	117	0.031
Rumit Cultivated 1	M837	124	0.047
Bayen Cultivated	M254	161	0.667
Bayen Cultivated	M254	163	0.333
Bayen Cultivated	M258	125	0.333
Bayen Cultivated	M258	129	0.167
Bayen Cultivated	M746	365	0.333
Bayen Cultivated	M837	110	0.500
Bayen Cultivated	M883	247	0.500
Rumit Wild	M254	170	0.036

Table 5: Nei's Genetic Distance among populations

	Pop 1	**Pop2**	**Pop3**	**Pop4**	**Pop5**	**Pop6**
Pop 1	0.000					
Pop 2	0.012	0.000				
Pop 3	0.087	0.103	0.000			
Pop 4	0.228	0.224	0.341	0.000		
Pop 5	0.019	0.008	0.083	0.227	0.000	
Pop 6	0.124	0.157	0.026	0.432	0.151	0.000

Pop 1 – Rumit Cultivated 1
Pop 2 – Rumit Cultivated 2
Pop 3 – Kaiwa & Jonglei Cultivated
Pop 4 – Bayen Cultivated
Pop 5 – Rumit Wild
Pop 6 – Ngelecho Wild

Figure 3: Population assignment based on allele frequency

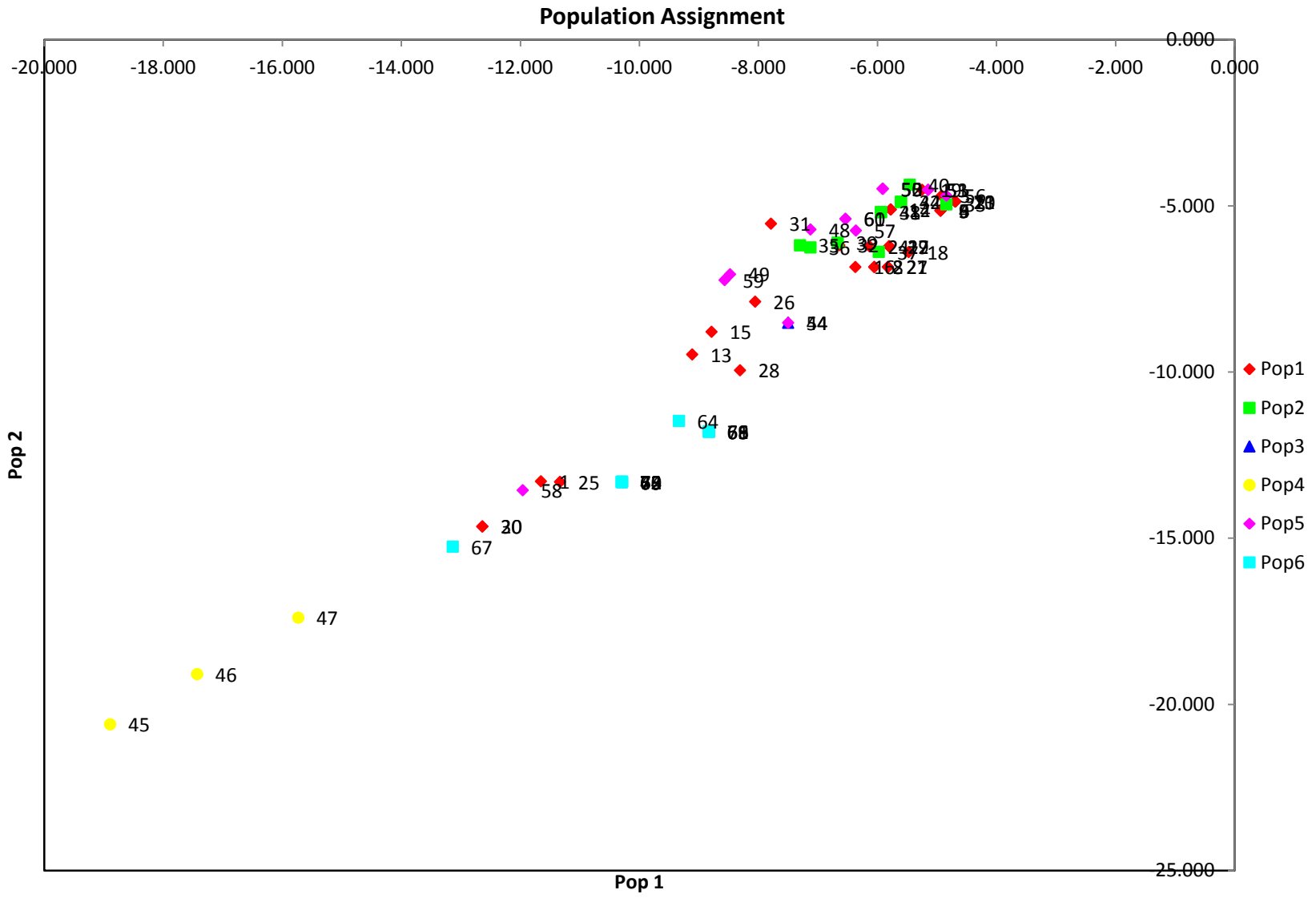
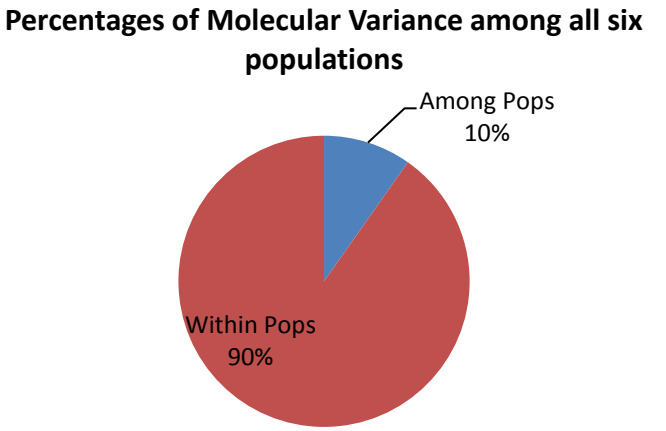
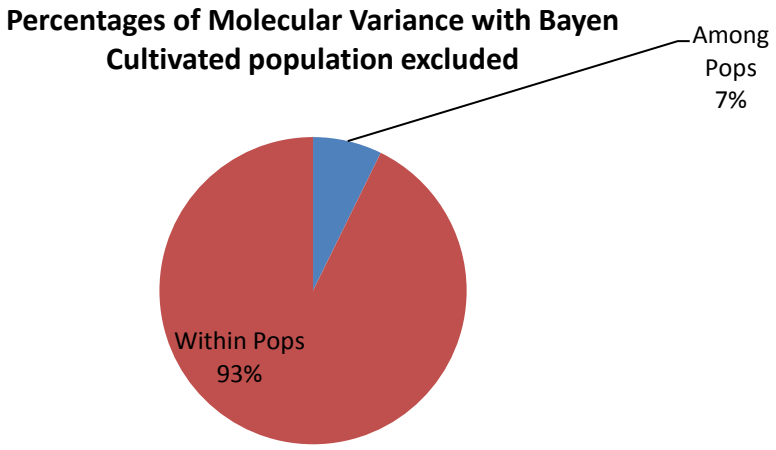


Table 6: Summary of partitioning of genetic variation among populations of *Coffea arabica* using AMOVA

Source of variation	***df***	**Sum of squares**	**Est. Variance**	**Percentage of variation**
Variation among all six populations:				
Among populations	5	57.292	0.373	10
Within populations	142	489.451	3.447	90
Variation among five populations after removing the Bayen cultivated population:				
Among populations	4	40.974	0.271	7
Within populations	137	474.618	3.464	93





c. Discussion:

The overall genetic diversity of the *C. arabica* populations from South Sudan (Boma Plateau) is comparable to those from other regions within the indigenous range of this species (i.e. Ethiopia). The Na ranged from 1.63 (Bayen Cultivated) to 2.63 (Rumit Cultivated 1) and the H_o ranged from 0.439 (Bayen Cultivated) to 0.603 (Ngelecho Wild) in this study (Table 3). The lower diversity of the Bayen Cultivated population could be due to the fact that there were only three plants in that population. In comparison, Cubry et al. (2007) report a mean Na of 2.0 and mean H_o of 0.49 using 60 SSR markers for both cultivated and wild Ethiopian accessions of *C. arabica*. Maluf et al. (2005) report an average Ne of 2.87 for cultivated *C. arabica* inbred lines and an average H_e of 0.33. The H_e in this study ranged from 0.260 (Bayen Cultivated) to 0.361 (Rumit Cultivated 1). When a comparison of cultivated and wild coffee from Ethiopia, Yemen and Brazil was performed (Silvestrini et al. 2007), the commercial cultivars had a Ne of 2.0 and for the wild coffee ranged from 2.0 in Eritrea, 2.2 in Yemen and in Ethiopia from 2.3 (Harar region) to 3.2 (Kaffa region).

Genotypically, most of the allelic variation is represented in the Rumit Cultivated 1 population, which had nine private alleles (Table 4). Other than one private allele in the Rumit Wild population, it was similar to the Rumit Cultivated 1. The Ngelecho Wild population was also similar to Rumit Cultivated 1. The similarity between the three cultivated (Rumit Cultivated 1 & 2 and Kaiwa/Jonglei Cultivated) and two wild populations (Rumit Wild and Ngelecho Wild) indicate that the plants in the villages were originally collected from the neighboring forests and established around village dwellings. The higher genetic diversity of the Rumit Cultivated 1 population could be indicative of the diversity that was present in the original forests when these plants were brought to the villages for cultivation and collected from multiple surrounding locations. The natural forests, which are in close proximity or contiguous with cultivated areas, have a long history of human disturbance. The forested area north of Rumit, which Thomas (1942) referred to as Barbuk, could not be located during the field expedition of 2012 and appears to have been converted to arable land. There is also evidence that the South Sudan populations have been negatively impacted by climate change (Davis et al. 2012, in review). The net result has been a reduction in the range, density and health of the *C. arabica* populations on the Boma Plateau, no doubt resulting in a loss of the genetic diversity in the populations. The Bayen Cultivated population was different from all the other populations with seven private alleles, and is genetically the most distant (Table 5; Figure 3). All the other cultivated and wild populations cluster together, which would infer that the Bayen Cultivated population is of a different origin, possibly from neighboring Ethiopia. Discussion with several individuals from the local population in Upper Boma revealed that in the late 1940s, the British had attempted coffee cultivation in this area and may have introduced foreign germplasm at that time. It is also possible that these are remnants of the Barbuk population seen by Thomas (1942). Geographically, they are the most distant from Rumit and Ngelecho (Figure 2). When an AMOVA was performed, the among-population variation with all six populations was 10% and when the Bayen Cultivated population was removed and the other five populations were compared, the among-population genetic variation was reduced to 7% (Table 6).

Evidence from this study, in combination with geographical proximity to populations in Ethiopia (less than 100 km) and bioclimatic suitability (Davis et al. 2012, in review), strongly supports the assertion that *C. arabica* is indigenous in the humid forests of the Boma Plateau in South Sudan. A comparative study between the populations in South Sudan and Ethiopia should be undertaken to confirm our genetic analysis.

6. Germplasm Conservation Recommendations:

The genetic diversity study results show that the Rumit Cultivated 1 population is the most genetically diverse of all populations tested. The allelic combinations in all other populations, except Bayen Cultivated and a single plant from the Rumit Wild population, are represented in this population. Table 7 lists all the individual plants with private alleles with their collection numbers.

Table 7: List of samples with one or more private alleles

Population	**Sample #**	**Loci w/ private alleles**
Rumit cultivated 1	B1	M254; M257; M784; M837
Rumit cultivated 1	B2	M837
Rumit cultivated 1	B8	M837
Rumit cultivated 1	B15	M260; M837
Rumit cultivated 1	B16	M257
Rumit cultivated 1	B17	M260
Rumit cultivated 1	B18	M260
Rumit cultivated 1	B20	M254; M258
Rumit cultivated 1	B22	M260
Rumit cultivated 1	B26	M260; M837
Rumit cultivated 1	B28	M258; M260
Rumit cultivated 1	B29	M260
Rumit cultivated 1	B30	M254; M258
Rumit cultivated 1	B32	M260
Bayen cultivated	B61	M254; M258; M746; M837; M883
Bayen cultivated	B62	M254; M258; M746; M837; M883
Bayen cultivated	B63	M254; M258; M837; M883
Rumit wild	B44	M254

Based on this, plants from which seed collections should be made are listed in Appendix 2. This list shows all the collections with their collection numbers and locations, size, phonological condition, and whether tagged or not. The individual plants recommended for collecting are highlighted in yellow based on private alleles and unique allele combinations.

Even though Appendix 2 provides a comprehensive list for collections acquisition, based on practical field conditions, these are the recommendations for establishing a field genebank:

- Rumit Cultivated 1 – Collect seeds from the following 19 samples: B01-02, B08, B12 – 18, B20-22, B25-26, B28-30, B32. Of these B16, 18, 29 & 32 did not have any flowers or fruit set during the April visit. Hence, these may have to be collected during the following or subsequent season(s).
- Rumit Cultivated 2 – These are genotypically similar to Rumit Cultivated 1 and hence any collections would only lead to redundant genotypes. No collecting is needed from this population at this time.
- Kaiwa/Jonglei Cultivated – one each plant specimen is currently being grown in these villages. Recommendation is to collect seeds from both plants. The plant in Kaiwa village (B57) was not flowering or fruiting. May have to wait until the following season to collect seeds. The plant at Jongeli (B59) was in flower in April and so should have fruit production.
- Bayen Cultivated – All three plants are unique with seven private alleles. Seeds should be collected from all three plants (B61, 62, 63). The plants were not blooming or fruiting in April and hence may have to wait until the following season.
- Rumit Wild – one plant (B44) had one private allele. Unfortunately this plant was not tagged nor GPS coordinates noted. Since a majority of genotypes of this population are represented in Rumit Cultivated 1, collecting from this population is optional, though a few genotypes could be important in representing any local adaptations of this population. None of the trees were flowering or fruiting, which may necessitate propagating by cuttings since more of the plants are in juvenile growth stage. If the decision is made to collect from this population, in addition to B44, recommendation is to collect B42 and B43 as well.
- Ngelecho Wild – this population did not have any unique genotypes (using the 19 markers) that weren't already represented in Rumit Cultivated 1. Hence collecting from this forest is not necessary, but for the sake of representation of any local adaptations, it would be advisable to collect from a few plants in the future.
- *Coffea neoleroyi* – collect from both plants (B 58, 60), though both were not tagged. GPS coordinates for both have been noted. Even though not used in cultivation, it will important to conserve these plants in the genebank.

If all collections are made as specified in Appendix 2, the total collections will be 29 plants (19 Rumit Cultivated 1; 1 Kaiwa; 1 Jongeli; 3 Bayen; 3 Rumit Wild; 2 *Coffea neoleroyi*).

7. Next Steps:

- Identify location of field genebank.
- Establish nursery for propagation of seeds collected for establishment of field genebank. As part of nursery development, need to identify and establish reliable water source to irrigate plants.
- Seed collecting expedition timed to coincide with seed maturation.

- Collect seeds from Rumit Cultivated 1, Kaiwa Cultivated (if seeds have been set), Jonglei Cultivated and Bayen Cultivated (if seeds are available). Collect at least 10 seeds per plant.
- For plants that do not have seeds, propagation by cuttings is an option. Vegetative propagation experiments are being conducted at Denver Botanic Gardens. The results are not yet available to report.
- Develop an accessioning system and database to keep track of collections.
- Collaborative research with Ethiopian scientists to compare South Sudanese *C. arabica* populations with Ethiopian populations.

8. References:

Anthony, F., M. C. Combes, C. Astorga, B. Bertrand, G. Graziosi, and P. Lashermes. 2002. The origin of cultivated *Coffea arabica* L. varieties revealed by AFLP and SSR markers. Theoretical and Applied Genetics. 104:894-900.

Chase, M. W. and H.H Hillis. 1991. Silica gel: an ideal material for field preservation of leaf samples for *DNA* studies. Taxon 40: 215-220.

Combes M. C., S. Andrzejewski, F. Anthony, B. Bertrand, P. Rovelli, G. Graziosi, and P. Lashermes. 2000. Characterization of microsatellite loci in *Coffea arabica* and related coffee species. Molecular Ecology. 9: 1178-1180

Cubry, P., P. Musoli, H. Legnate, D. Pot, F. de Bellis, V. Poncet, F. Anthony, M. Dufour, and T. Leroy. 2008. Diversity in coffee assessed with SSR markers: structure of the genus *Coffea* and perspectives for breeding. Genome. 51: 50-63.

Excoffier, L., P. E. Smouse, and J. M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics. 131: 479-491.

Krishnan, S, T. A. Ranker, A. P. Davis, and J. J. Rakotomalala. 2012. An assessment of the genetic integrity of *ex situ* germplasm collections of three endangered species of *Coffea* from Madagascar: implications for the management of field germplasm collections. Genetic Resources and Crop Evolution. DOI 10.1007/s10722-012-9898-3.

Maluf, M. P., M. Silvestrini, L. M. de C. Ruggiero, O. G. Filho, and C. A. Colombo. 2005. Genetic diversity of cultivated *Coffea arabica* inbred lines assessed by RAPD, AFLP and SSR marker systems. Scientia Agricola. 62(4): 366-373.

Meyer, F. G. 1965. Notes on wild *Coffea arabica* from southwestern Ethiopia, with some historical considerations. Economic Botany. 19: 136-151.

Musloi, P., P. Cubry, P. Aluka, C. Billot, M. Dufour, T. De Bellis, D. Pot, D. Bieysse, A. Charrier, and T. Leroy. 2009. Genetic differentiation of wild and cultivated populations: diversity of *Coffea canephora* Pierre in Uganda. Genome. 52: 634-646.

Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics. 89: 583-590.

Peakall, R, and P. E. Smouse. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes. 6: 288-295.

Pendergrast, M. 1999. Uncommon Grounds: The History of Coffee and How It Transformed Our World. New York: Basic Books.

Pendergrast, M., 2009. Coffee second only to oil? Tea & Coffee Trade Journal. April:38-41.

Poncet, V., P. Hamon, J. Minier, C. Carasco, S. Hamon, and M. Noirot. 2004. SSR cross-amplification and variation within coffee trees (*Coffea* spp.). Genome. 47: 1071-1081.

Rovelli, P., R. Mettulio, F. Anthony, F. Anzueto, P. Lashermes, and G. Graziosi. 2000. Microsatellites in *Coffea arabica* L. In: Sera, T., C. R. Soccol, A. Pandey, and S. Roussos (eds) Coffee Biotechnology and Quality. Kluwer Academic Publishers, Netherlands, pp 123-133.

Silvestrini, M., M. G. Junqueira, A. C. Favarin, O. Guerreiro-Filho, M. P. Maluf, M. B. Silvarolla, and C. A. Colombo. 2007. Genetic diversity and structure of Ethiopian, Yemen and Brazilian *Coffea arabica* L. accessions using microsatellite markers. Genetic Resources and Crop Evolution. 54: 1367-1379.

Thomas, A. S. 1942. The wild Arabica coffee on the Boma Plateau, Angle-Egyptian Sudan. Empire Journal of Experimental Agriculture. 10: 207-212.

Vega, F., E. Rosenquist, and W. Collins. 2003. Global project needed to tackle coffee crisis. Nature. 425: 343.

Acknowledgments:
Thanks to World Coffee Research and its parent organizations, the Norman Borlaug Institute for International Agriculture, Texas A & M University and Specialty Coffee Association of America for inviting us to participate in this research. We acknowledge Dr. Tim Schilling for his leadership and for coordinating all the logistics, personnel at the Norman Borlaug Institute for travel logistics and the United States Agency for International Development (USAID) for funding this project.