ISSN: 0975-833X

Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 9, Issue, 05, pp.51006-51013, May, 2017

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

RESEARCH ARTICLE

DATA QUALITY ASSESSMENT AND SPATIAL ANALYSIS FOR CONSERVATION AND SUSTAINABLE USE OF ARABICA COFFEE (*Coffea arabica* L.) GENETIC RESOURCES IN ETHIOPIA

[1,*]Tamene Yohannes and [2]Sebsebe Demissew

[1]Ethiopian Biodiversity Institute, PO Box 121523, Addis Ababa, Ethiopia
[2]Department of Plant Biology and Biodiversity Management, Addis Ababa University, P.O. Box 3434, Addis Ababa, Ethiopia

ARTICLE INFO

Article History:
Received 19th February, 2017
Received in revised form
21st March, 2017
Accepted 24th April, 2017
Published online 31st May, 2017

Key words:
Biodiversity,
Spatial data,
Data quality assessment,
Coffea arabica L.

ABSTRACT

Poor data quality in database can result from various causes, such as errors in site descriptions, imprecise coordinates or even mistakes. Hence, exploratory data analysis and cleaning are among several range of approaches and methods of data auditing and cleaning, which have been suggested for maintaining or improving data quality. Since its establishment, 1976 the Ethiopian Biodiversity Institute (EBI) previously known as Plant Genetic Resource Centre of Ethiopia/PGRCE (1976), Institute of Biodiversity Conservation/IBC (1994), has been collecting crop germplasm and associated traditional knowledge with special emphasis on major crop species. In 1999 EBI has established a database and documented several biodiversity related information. As of December 2013, the national collection has reached 73,438 accessions from 423 plant species belonging to 196 genera in 54 families. *Coffea arabica* L. is among these crop species, which Ethiopia has been regarded as the center of origin and center of genetic diversity. The wild coffee populations of Ethiopia are gene pool for several important traits which has national and international importance. Data quality assessment and spatial analysis were done on 5779 Coffee accessions deposited at the EBI, which were collected all over the coffee growing parts of the country and conserved at two field gene banks. The result has showed that, among the 5779 accessions, 2251 (38 %) were entered into the database. Among these 2251 accessions, 977 (43.40%) were georeferenced. The result of visualization map of the 977 C. arabica accessions and quality control based on the comparison of the country administrative unit data in the passport, with that of the country administrative unit information extracted from thematic layers using DIVA-GIS software has showed that, a number of accessions fall out of the geographical region of Ethiopia. Moreover, there are also 11 % erroneous points, which are 107 out of the 977 accessions. Three groups of causes of errors which could be committed either during field data collecting or while data encoding or at both stages were found. All the 107 erroneous points were corrected using different methods and recommendations were given for future use.

Citation: Tamene Yohannes and Sebsebe Demissew, 2017. "Data quality assessment and spatial analysis for conservation and sustainable use of Arabica coffee (*Coffea Arabica* L.) genetic resources in Ethiopia", *International Journal of Current Research*, 9, (05), 51006-51013.

INTRODUCTION

Ethiopia has been regarded as the center of origin and center of genetic diversity of Arabica coffee, *Coffea arabica* L. (Vavilov, 1951; Meyer, 1965). The centre of origin for Arabica coffee is in the Southwestern and Southeastern parts of Ethiopia, where it occurs naturally in the undergrowth of the Afromontane rainforests between 1,000 and 2,000 m.a.s.l. (Schmitt and Grote, 2006). The most suitable range being between 1,500 – 1,800 m.a.s.l (Schmitt, 2006). These wild coffee populations are the gene pool for several important traits which has paramount national and international importance, because it has high potential for the breeding of new coffee varieties (Kassahun Tesfaye, 2006). The Ethiopian Biodiversity Institute (EBI) previously known as Plant Genetic Resource Centre of Ethiopia (PGRCE) (1976), the Institute of Biodiversity Conservation (IBC) (1994), has been collecting crop germplasm and associated traditional knowledge with special emphasis on major crop species. In 1999 EBI has established a database and documented several biodiversity related information. As of December 2013, the national collection has reached 73,438 accessions from 423 plant species belongs to 196 genera in 54 families. In order to formulate appropriate management and conservation strategies, it is critical that datasets are of high quality and are precise (Chapman, 2005a & 2005b). The use of incorrect or low-quality information may have significant consequences on the

Corresponding author: **Tamene Yohannes,**
Ethiopian Biodiversity Institute, PO Box 121523, Addis Ababa, Ethiopia

relevance and appropriateness of subsequent recommendations, decisions and utilization. Hence, the main objective of undertaking the spatial analysis of germplasm collection passport data is to provide accurate and clear information that will assist in effective decision-making processes in natural resource conservation and use. Poor data quality in database could be due to various factors; such as errors in site descriptions, imprecise coordinates, data entry errors, wrong taxonomic identification and changes in nomenclature.

In general, many of the sources of error in databases fall into one or more of the following categories: data entry errors, measurement errors and data integration errors (Chapman, 2005a). Before data are entered a database, it usually passes through several steps involving both human interaction and computation. Thus, knowing the sources of errors can be useful in designing a better data collection and storage techniques that reduce the introduction of errors.

Two key aspects of spatial data quality include the accuracy and precision of geographic coordinates (Scheldeman and Zonneveld, 2010). The accuracy of coordinates determines the ability to correctly represent the site of collection/observation point. Precision refers to the level of detail of the coordinates necessary to represent the described site effectively. A lack of accuracy in the analyzed data will inevitably lead to errors in the results of the analysis, while a lack of precision will often result in conclusions of limited use. Data can be very precise but inaccurate and can also be accurate but highly imprecise. Hence, exploratory data analysis and data cleaning are among several range of approaches and methods of data auditing and cleaning, which have been suggested for maintaining or improving data quality (Hellerstein, 2008). The objective of this paper is therefore, to evaluate the spatial data quality and conduct some spatial analysis of conserved *C. arabica* accessions' passport data in the EBI database, to identify possible erroneous presence points using different tools and to take corrective measures and to ensure high levels of spatial data quality for both domestic and international data exchanges, which in turn helps for the conservation and sustainable utilization of the crop.

MATERIALS AND METHODS

As described by Tukey (Tukey, 1977), in many if not most instances, data can only be cleaned effectively with some human involvement. Therefore, exploratory data analysis or exploratory data mining was employed for the study, which typically involves human in the process of understanding properties of a dataset, including the identification and possible rectification of errors. In addition, DIVA-GIS software (a GIS programme specifically designed to undertake spatial diversity analysis) version 7.5.0.0 developed by Robert J. Hijmans (DIVA-GIS, 2005) was used to conduct data visualization and quality control based on administrative unit information as described by Scheldeman and Zonneveld (Scheldeman and Van Zonneveld, 2010).

In this study, 5779 Coffee accessions conserved at the EBI were used. Among these collections, 4906 accessions were conserved at "Choche" (in southwestern part of Ethiopia) and 873 at "Bedessa" (in eastern part of the country) field gene banks respectively, as shown in Figure 1.

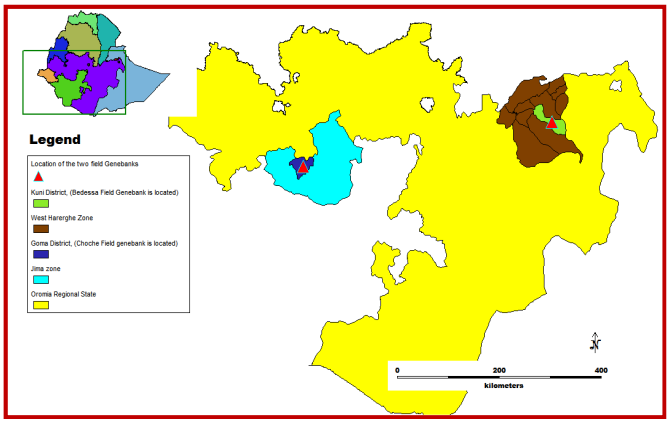

Figure 1. Map of Oromia Regional State with the relative locations of "Choche" and "Bedessa" field Genebanks shown by red triangles

RESULTS AND DISCUSSION

Among the 5779 Coffee accessions of the EBI, only 2251 (38%) were entered into a database. Thus, the analysis was carried out based on the 2251 accessions. Figure 2 shows the regional distribution of *C. arabica* accessions throughout Ethiopia. Among these 2251 accessions, only 977 accessions (43.40%) are georeferenced.

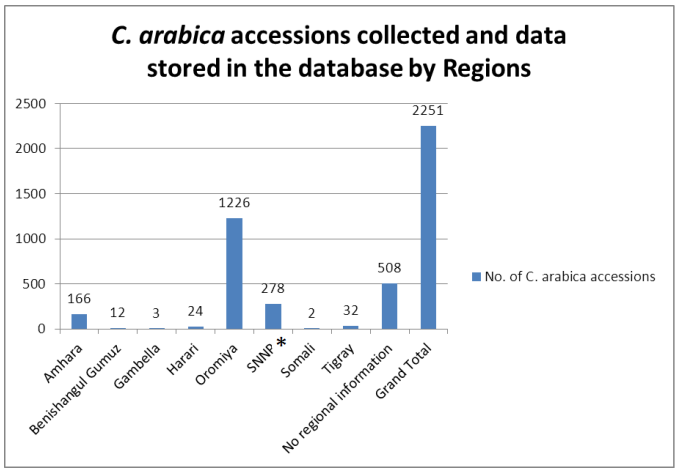

Figure 2. Collected *C. arabica* accessions by Regional States of Ethiopia. *SNNP= South Nations, Nationalities and Peoples Regional State

As figure 2 shows, the majority of the collections are from Oromia and Southern Nations, Nationalities and Peoples (SNNP) Regional States. Oromia Regional State contributed about 54.5 % of the collection; as it is the major *C. arabica* gene pool of the country as well as one of the major coffee growing regions of the country. Though, SNNP Regional State is second with the number of collected accessions; it seems under represented compared to its potential and known coffee growing localities as well as to the known specialty export quality coffee. Even though, all the 2251 accessions were collected from known geographical regions, 508 of them didn't have regional information in the database. Moreover, as the quality control result below shows, those with regional information and with geographical coordinates also have errors in spatial data which were created during the data collecting and/or data entry stages.

Spatial Data Visualization and Quality Control

The result of visualization map of the 977 *C. arabica* accessions has shown that, a number of accessions fall out of the geographical ranges of Ethiopia. Some fall in the Indian Ocean, others in the Sudan, Somalia and Nigeria (Figure 3).
The result of quality control based on the comparison of the country administrative unit data included in the passport, with that of the country administrative unit information extracted from thematic layers (map) showed that; there are 11% erroneous points, which are 107 out of the 977 accessions. This means that, for the 107 accessions the passport country information does not coincide with the country lo
cation of the collecting point on the map (Figure 3). Therefore, identifying the sources of error and then taking corrective measures are crucial to upgrade the quality of the data and increase the usability and acceptability both for internal and external uses. Internal uses such as GIS based conservation planning, collection gap analysis and recollection of dead accessions. External uses such as, international data exchanges with international organizations such as Global Biodiversity Information Facility (GBIF), which works on huge global biodiversity data which needs quality spatial data.

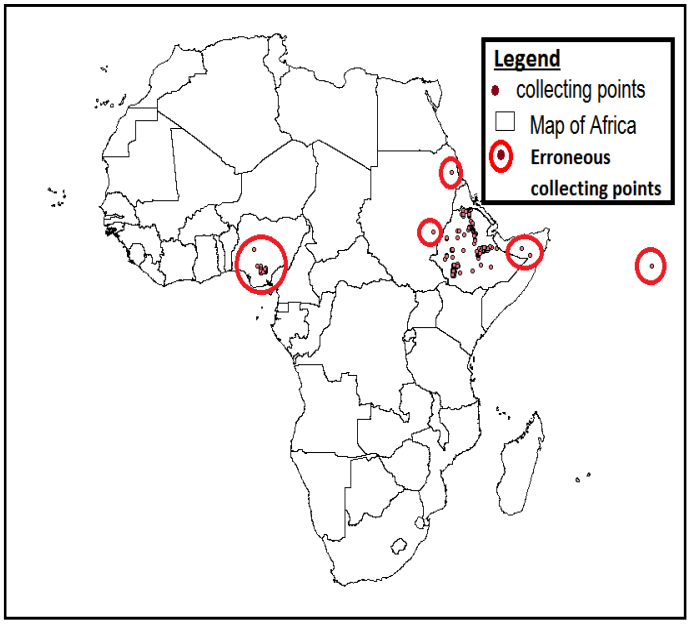

Figure 3. Visualization of the spatial data of 977 *C. arabica* accessions at the EBI, overlaid on the Africa map

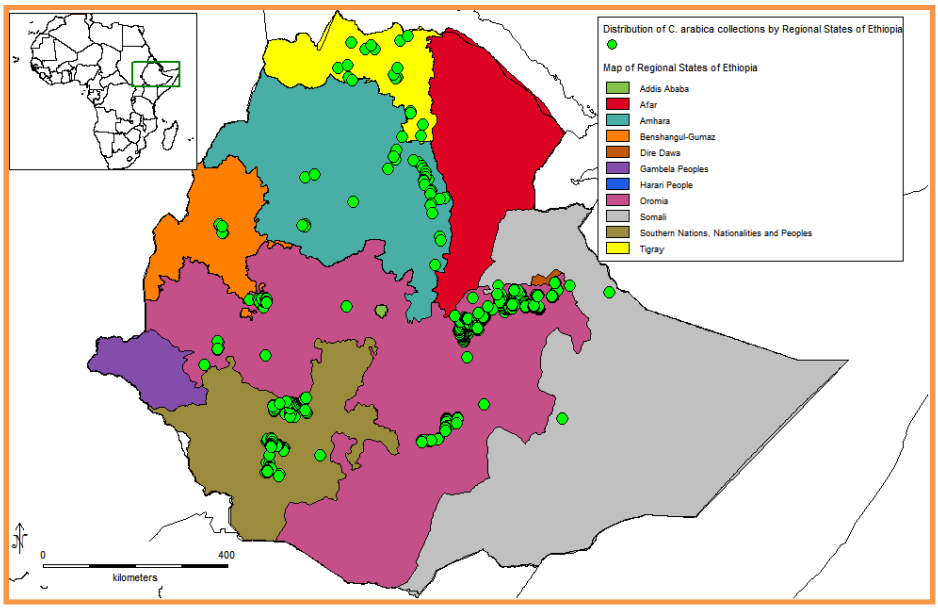

Figure 4. Corrected collecting points of *Coffea arabica* L. (total 997, among which 107 are corrected erroneous points) overlaid on the Ethiopia map

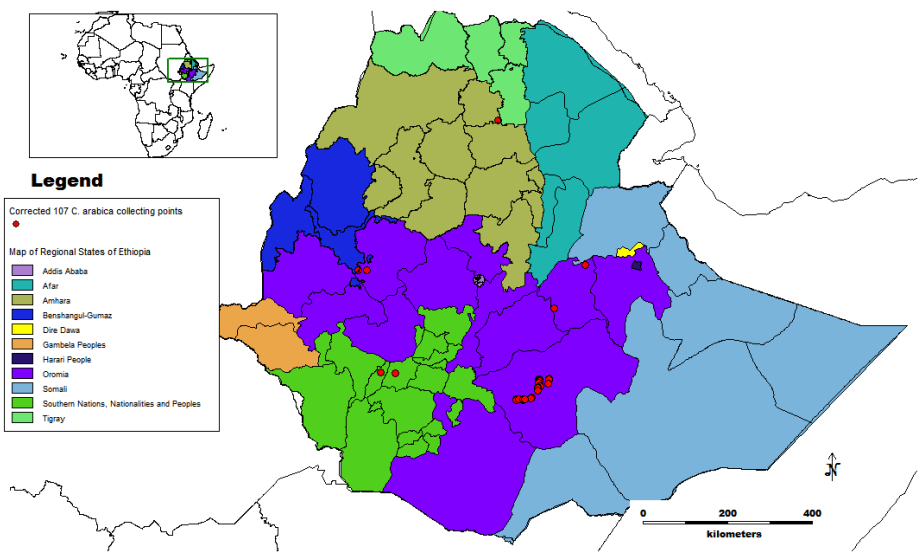

Figure 5. Corrected collecting points of *Coffea arabica* L. for the 107 erroneous points, overlaid on the Ethiopia map

Annex 1. Original and corrected coordinates for the 107 *C. arabica* accessions of the EBI collections

Accession Number	Species name	Country	Region/State/ Province	Zone	Woreda/ District	Original Lat. (hddd°mm′.ss″)	Original Long. (hddd°mm′.ss″)	Corrected Lat. (hddd.ddddd°)	Corrected Long. (hddd.ddddd°)
8727	*Coffea arabica*	Ethiopia	Oromiya	MISRAK WELLEGA	SASIGA	09-11-56-N	06-11-11-E	9.19889	36.18639
8730	*Coffea arabica*	Ethiopia	Oromiya	MISRAK WELLEGA	SASIGA	19-12-10-N	36-22-07-E	9.20278	36.36861
8780	*Coffea arabica*	Ethiopia	SNNP	SEMEN OMO	ISARA TOCHA	06-58-59-N	66-59-07-E	6.98306	36.98528
8825	*Coffea arabica*	Ethiopia	SNNP	SEMEN OMO	KUCHA	06-59-53-N	06-40-06-E	6.99806	36.66833
8833	*Coffeaarabica*	Ethiopia	Oromiya	BALE	MENNANA HARENA BULU	05-64-36-N	07-08-54-E	6.41308	39.58415
8834	*Coffea arabica*	Ethiopia	Oromiya	BALE	MENNANA HARENA BULU	05-64-36-N	07-08-54-E	6.41308	39.58415
8835	*Coffea arabica*	Ethiopia	Oromiya	BALE	MENNANA HARENA BULU	05-64-36-N	07-08-54-E	6.41308	39.58415
8836	*Coffea arabica*	Ethiopia	Oromiya	BALE	MENNANA HARENA BULU	05-64-36-N	07-08-54-E	6.41308	39.58415
8837	*Coffea arabica*	Ethiopia	Oromiya	BALE	MENNANA HARENA BULU	05-64-36-N	07-08-54-E	6.41308	39.58415
8838	*Coffea arabica*	Ethiopia	Oromiya	BALE	MENNANA HARENA BULU	05-64-35-N	07-08-54-E	6.41308	39.58407
8839	*Coffea arabica*	Ethiopia	Oromiya	BALE	MENNANA HARENA BULU	05-64-17-N	07-09-10-E	6.41552	39.58129
8840	*Coffea arabica*	Ethiopia	Oromiya	BALE	MENNANA HARENA BULU	05-64-17-N	07-09-10-E	6.41552	39.58129
8841	*Coffea arabica*	Ethiopia	Oromiya	BALE	MENNANA HARENA BULU	05-64-17-N	07-09-10-E	6.41552	39.58129
8842	*Coffea arabica*	Ethiopia	Oromiya	BALE	MENNANA HARENA BULU	05-64-17-N	07-09-10-E	6.41552	39.58129
8843	*Coffea arabica*	Ethiopia	Oromiya	BALE	MENNANA HARENA BULU	05-64-17-N	07-09-10-E	6.41552	39.58129
8844	*Coffea arabica*	Ethiopia	Oromiya	BALE	MENNANA HARENA BULU	05-64-17-N	07-09-10-E	6.41552	39.58129
8845	*Coffea arabica*	Ethiopia	Oromiya	BALE	MENNANA HARENA BULU	05-64-47-N	07-08-34-E	6.41004	39.58582
8846	*Coffea arabica*	Ethiopia	Oromiya	BALE	MENNANA HARENA BULU	05-64-47-N	07-08-34-E	6.41004	39.58582
8847	*Coffea arabica*	Ethiopia	Oromiya	BALE	MENNANA HARENA BULU	05-64-47-N	07-08-34-E	6.41004	39.58582
8848	*Coffea arabica*	Ethiopia	Oromiya	BALE	MENNANA HARENA BULU	05-64-47-N	07-08-31-E	6.40963	39.58574
8849	*Coffea arabica*	Ethiopia	Oromiya	BALE	MENNANA HARENA BULU	05-64-56-N	07-08-17-E	6.40750	39.58717
8850	*Coffea arabica*	Ethiopia	Oromiya	BALE	MENNANA HARENA BULU	05-64-54-N	07-08-17-E	6.40746	39.58692
8851	*Coffea arabica*	Ethiopia	Oromiya	BALE	MENNANA HARENA BULU	05-65-22-N	07-08-15-E	6.40716	39.59108
8852	*Coffea arabica*	Ethiopia	Oromiya	BALE	MENNANA HARENA BULU	05-65-22-N	07-08-15-E	6.40716	39.59108
8853	*Coffea arabica*	Ethiopia	Oromiya	BALE	MENNANA HARENA BULU	05-65-26-N	07-08-13-E	6.40683	39.59164
8854	*Coffea arabica*	Ethiopia	Oromiya	BALE	MENNANA HARENA BULU	06-63-54-N	07-09-55-E	6.42231	39.57791
8855	*Coffea arabica*	Ethiopia	Oromiya	BALE	MENNANA HARENA BULU	05-63-54-N	07-09-55-E	6.42231	39.57791
8856	*Coffea arabica*	Ethiopia	Oromiya	BALE	MENNANA HARENA BULU	05-63-54-N	07-09-55-E	6.42231	39.57791
8857	*Coffea arabica*	Ethiopia	Oromiya	BALE	MENNANA HARENA BULU	05-63-54-N	07-09-55-E	6.42231	39.57791
8858	*Coffea arabica*	Ethiopia	Oromiya	BALE	MENNANA HARENA BULU	05-65-33-N	07-08-8 -E	6.40614	39.59281
8859	*Coffea arabica*	Ethiopia	Oromiya	BALE	MENNANA HARENA BULU	05-65-33-N	07-08- 8-E	6.40614	39.59281
8860	*Coffea arabica*	Ethiopia	Oromiya	BALE	MENNANA HARENA BULU	05-71-45-N	07-10-21-E	6.42607	39.64884
8861	*Coffea arabica*	Ethiopia	Oromiya	BALE	MENNANA HARENA BULU	05-71-30-N	07-11-12-E	6.43371	39.64659
8862	*Coffea arabica*	Ethiopia	Oromiya	BALE	MENNANA HARENA BULU	05-71-30-N	07-11-12-E	6.43371	39.64659
8863	*Coffea arabica*	Ethiopia	Oromiya	BALE	MENNANA HARENA BULU	06-11-25-N	07-15-35-E	6.47297	39.91712
8864	*Coffea arabica*	Ethiopia	Oromiya	BALE	MENNANA HARENA BULU	06-04-29-N	07-15-36-E	6.47310	39.91785
8865	*Coffea arabica*	Ethiopia	Oromiya	BALE	MENNANA HARENA BULU	05-98-52-N	07-13-41-E	6.45580	39.89415
8866	*Coffea arabica*	Ethiopia	Oromiya	BALE	MENNANA HARENA BULU	05-98-52-N	07-13-41-E	6.45580	39.89415
8867	*Coffea arabica*	Ethiopia	Oromiya	BALE	MENNANA HARENA BULU	05-98-48-N	07-13-41-E	6.45582	39.89346
8868	*Coffea arabica*	Ethiopia	Oromiya	BALE	MENNANA HARENA BULU	05-81-31-N	07-09-13-E	6.41567	39.73293
8869	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-17-53-N	07-55- 8-E	6.83036	40.06678
8870	*Coffea arabica*	Ethiopia	Oromiya	BALE	MENNANA HARENA BULU	05-81-9 -N	07-09-19-E	6.41662	39.73377
8871	*Coffea arabica*	Ethiopia	Oromiya	BALE	MENNANA HARENA BULU	05-81-18-N	07-09-41-E	6.41993	39.73523
8872	*Coffea arabica*	Ethiopia	Oromiya	BALE	MENNANA HARENA BULU	05-81-47-N	07-09-49-E	6.42116	39.73951
8873	*Coffea arabica*	Ethiopia	Oromiya	BALE	MENNANA HARENA BULU	05-84-27-N	07-10-44-E	6.42943	39.76350

…… *Continue*

8874	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-19-13-N	07-56-57-E	6.84680	40.07897
8875	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-19-13-N	07-56-57-E	6.84682	40.07897
8876	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-19-14-N	07-56-50-E	6.84574	40.07907
8877	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-19-15-N	07-56-47-E	6.84535	40.07927
8878	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-19-13-N	07-57-49-E	6.85461	40.07892
8879	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-19-13-N	07-57-49-E	6.85461	40.07892
8880	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-19-11-N	07-57-56-E	6.85577	40.07871
8881	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-19-11-N	07-57-56-E	6.85575	40.07869
8882	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-19-11-N	07-57-56-E	6.85576	40.07871
8883	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-17-47-N	07-57-16-E	6.84903	40.07503
8884	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-18-46-N	07-57-11-E	6.84901	40.07482
8885	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-18-48-N	07-57-4 -E	6.84797	40.07524
8886	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-17-29-N	07-55-7 -E	6.83032	40.06317
8887	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-17-25-N	07-55-11-E	6.83081	40.06263
8888	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-17-24-N	07-55-5 -E	6.83001	40.06252
8889	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-17-21-N	07-55-9 -E	6.83062	40.06206
8890	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-17-37-N	07-55-9 -E	6.83058	40.06447
8891	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-17-46-N	07-55-12-E	6.83100	40.06577
8892	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-17-47-N	07-55-12-E	6.83104	40.06597
8893	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-18-5 -N	07-54-30-E	6.82465	40.06873
8894	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-18-22-N	07-54-24-E	6.82371	40.07572
8895	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-17-19-N	07-49-2 -E	6.77521	40.06165
8896	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-17-27-N	07-49-8 -E	6.77619	40.06274
8897	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-17-37-N	07-49-10-E	6.77645	40.06437
8898	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-17-42-N	07-49-10-E	6.77644	40.06504
8899	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-17-47-N	07-49-7 -E	6.77605	40.06584
8900	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-17-51-N	07-49-6 -E	6.77586	40.06642
8901	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-17-44-N	07-49-13-E	6.77690	40.06531
8902	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-18-17-N	07-49-13-E	6.77689	40.07034
8903	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-20-42-N	07-41-49-E	6.70988	40.09203
8904	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-20-39-N	07-41-49-E	6.70991	40.09161
8905	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-20-50-N	07-41-48-E	6.70983	40.09330
8906	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-20-55-N	07-41-48-E	6.70974	40.09407
8907	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-20-58-N	07-41-56-E	6.71101	40.09448
8908	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-21-4 -N	07-41-56-E	6.71090	40.09545
8909	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-21-31-N	07-41-31-E	6.70721	40.09948
8910	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-21-26-N	07-41-9 -E	6.70381	40.09867
8911	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-21-17-N	07-41-3 -E	6.70297	40.09733
8912	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-21-17-N	07-41-3 -E	6.70297	40.09733
8913	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-15-42-N	07-38-9 -E	6.67679	40.04680
8914	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-15-42-N	07-38-11-E	6.67710	40.04669
8915	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-15-28-N	07-38-30-E	6.68451	40.04466
8916	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-15-25-N	07-38-30-E	6.68000	40.04416
8917	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-15-23-N	07-38-30-E	6.68003	40.04396
8918	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-15-19-N	07-38-31-E	6.68015	40.04337
8919	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-15-26-N	07-38-21-E	6.67868	40.04434
8920	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-15-35-N	07-38-13-E	6.67750	40.04573
8921	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-16-19-N	07-37-18-E	6.66912	40.05238
8922	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-16-28-N	07-37-18-E	6.66916	40.05372
8923	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-16- 9-N	07-32-19-E	6.62404	40.05074

…… *Continue*

8924	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-16-42-N	07-31-52-E	6.62003	40.05573
8925	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-16-42-N	07-31-52-E	6.62003	40.05573
8926	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-42-31-N	07-55-8 -E	6.82995	40.28968
8927	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-42-39-N	07-59-7 -E	6.86591	40.29104
8928	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-42-37-N	07-56-18-E	6.86749	40.29072
8929	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-42-40-N	07-58-40-E	6.86185	40.29113
8930	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-42-41-N	07-58-28-E	6.86001	40.29132
8931	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-42-36-N	07-58-7 -E	6.85687	40.29061
8932	*Coffea arabica*	Ethiopia	Oromiya	BALE	BERBERE	06-40-46-N	07-48-22-E	6.76868	40.27375
243918	*Coffea arabica*	Ethiopia	Oromiya	MIRAB HARERGE	DAROLEBU	08-22-70-N	48-24-11-E	8.38611	40.40306
244135	*Coffea arabica*	Ethiopia	Oromiya	MIRAB HARERGE	DOBA	09-17-62-N	47-03-98-E	9.30056	41.07722
244251	*Coffea arabica*	Ethiopia	Amara			11-23-93-N	33-32-87-E	12.40917	39.19083

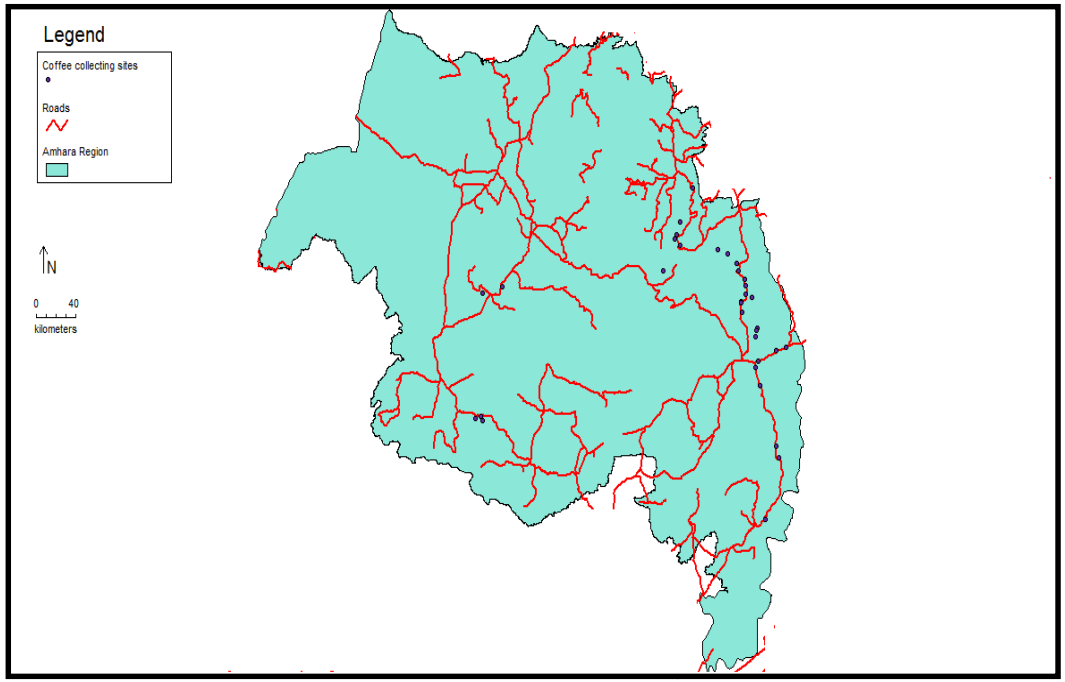

Figure 6. Map Amhara Region coffee collecting sites overlaid on Ethiopian road map

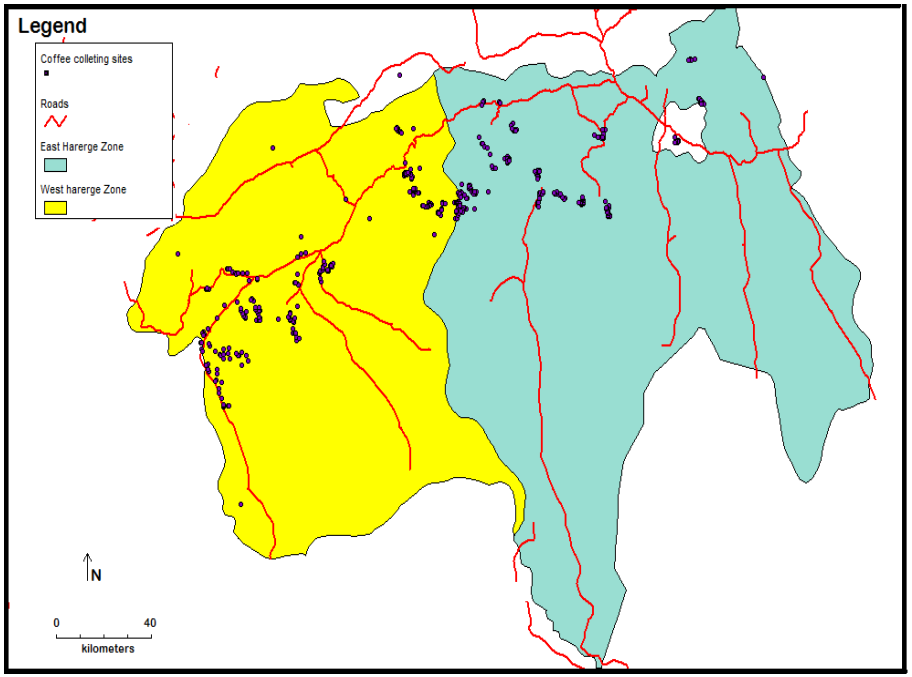

Figure 7. Map of East and West Harerge zones coffee collecting sites overlaid on Ethiopian road map

Correction of erroneous points based on quality control results

When error points are recognized, the first step is to go back and check the original collection sheet (field book) to verify the origin of the point; then fix the error immediately if the error is caused during data entry. That is why it is so important to keep the field book or collection form for further checkup even after the data were entered to the database. When the original collection sheet is not available, the next step is to identify the possible causes of the error. For the present study three groups of causes of errors were identified which might be committed either during field data collecting or while data encoding or at both stages:

Case 1: Longitude errors: Accessions 8727, 8780, 8825, 243918,244135 and 244251. In this group either the latitude value itself is entered in the longitude column or error values such as 66 instead of 36 were entered.

Case 2: Latitude errors: Only accession 8730 was found in this group. Latitude value 19 degree which is out of the range of Ethiopian latitude value was entered instead of possibly nine degree latitude. For the above two error groups, the errors were corrected, visualized, checked for their accuracy and crosschecked with the locality information.

Case 3: Both Latitude and Longitude errors: A total of 100 continuous accessions, (8833 to 8932) were found in this error group. These accessions are the ones which fell in the geographical region of the Nigeria in the visualization map (Figure 3). The problem of this group was very complicated, since no accession can fall to the correct (the passport locality information) geographical region of Ethiopia by correcting or dragging either longitude or latitude or both. Therefore, the collection sheets of each accession were scrutinized for the possible error. The result showed that, just six digit numbers were recorded in the latitude and longitude fields without any units and other spatial data such as Datum and UTM zone. Moreover, these numbers were again changed to degree, minute and second while the data entered into the database without knowing the original units, so that the final result ended up with such erroneous points. For this group the final result showed that, the data were collected in Universal Transverse Mercator **(**UTM) unit, though it was not indicated in the collection sheets. When these UTM units are converted to decimal degree and projected in the map, they exactly corresponded with passport locality information (Figures 4 and 5).

The total set of the corrected spatial data of *C. arabica* is presented in Annex 1 and submitted to the database unit of the EBI for further scrutiny and use.

Conclusion

For the last four decades, the Ethiopian Biodiversity Institute has been striving to collect and conserve genetic resources of local and global importance. The quest of safeguarding these invaluable resources was not simple and it is indeed a great achievement. In addition to such effort there are some gaps to be filled. The spatial data are among them which need a great attention while data collection, entry and quality control.

Recommendations

- The Ethiopian Biodiversity Institute needs to have standards for recording Geographical coordinates (to use either UTM, Decimal degree, degree-minute-second or other format and also the level of its precision) and to use similar GPS settings such as Datum (as recommended by Chapman (6)), all over the data to avoid or reduce errors which can arise from conversion and other related issues.
- It is important to give formal training for data collators about effective use of GPS receivers, the accuracy and precision of spatial data collection, importance of accurate and precise field notes, and proper description on collecting localities, geo-referencing of accessions without coordinate data by using field notes, maps, gazettes, geomancers and other available methods and to make the database more clean and usable.
- By taking this preliminary study as a model, it is important to conduct similar data quality assessment and analysis for other data types and other species in the database.
- For future collecting missions, it is strongly recommended that to use the spatial data for gap analysis, since there are many known coffee growing areas which were not adequately covered. Moreover, most of the collections lie down near the main roads (for example Annex 2 compares the distribution of the collections in the Amhara Region found near the main roads with that of Eastern and Western Hareregе collections which are relatively collected away from the main roads). Thus, the gap analysis will clearly show which regions, localities and specific coffee growing areas to be addressed in future collecting missions.
- It is also important to correct and complete the incomplete spatial data in the database, based on the information in the data sheet and using the latest geo-referencing software to make the data more useful for conservation planning.

Acknowledgement

The Authors are thankful to Ethiopian Biodiversity Institute and its staff for their unreserved assistance and providing us the required information and passport data.

REFERENCES

Chapman, A. D. 2005a. Principles of Data Quality, version 1.0. Report for the Global Biodiversity Information Facility, Copenhagen.

Chapman, A. D. 2005b. Principles and Methods of Data Cleaning – Primary Species and Species-Occurrence Data, version 1.0. Report for the Global Biodiversity Information Facility, Copenhagen.

DIVA-GIS, 2005. User Manual, version 5.2 [on line]. Available from: http://www.diva-gis.org/docs/DIVA-IS5_manual.pdf. (Date accessed: May 2014).

Hellerstein, J. 2008. Quantitative Data Cleaning for Large Databases. EECS Computer Science Division, UC Berkeley. http://db.cs.berkeley.edu/jmh. (Date accessed: May 2014).

Kassahun Tesfaye, 2006. Genetic diversity of wild *Coffea arabica* populations in Ethiopia as a contribution to

conservation and use planning. Ecology and Development Series, No. 44. Cuvillier Verlag, GÖttingen.

Meyer, F.G 1965. Notes on wild *Coffea arabica* from southwestern Ethiopia, with some historical considerations. *Economic Botany*, 19(2): 136-151.

Scheldeman, X., van Zonneveld, M. 2010.Training Manual on Spatial Analysis of Plant Diversity and Distribution. Bioversity International, Rome, Italy.

Schmitt, C. 2006. Montane rainforest with wild *Coffea arabica* in the Bonga region (SW Ethiopia): plant diversity, wild coffee management and implications for conservation. Ecology and Development Series, No. 47. Cuvillier Verlag, Göttingen.

Schmitt, C. and Grote, U. 2006. Wild coffee production in Ethiopia: the role of coffee certification for forest conservation. (http://web.fuberlin.de/ffu/akumwelt/bc2006/ papers/ Schmitt_Grote _ Coffee.pdf) (Date accessed: May 2014).

Tukey, J. 1977. Exploratory Data Analysis. Addison-Wesley.

Vavilov N.I 1951. The origin, variation, immunity and breeding of cultivated plants. Chronica Botanica 13. Ronald Press, New York.

****