# DNA Barcoding of Plants at Shaw Nature Reserve Using *matK* and *rbcL* Genes.

Livingstone Nganga

DNA barcoding is the use of short DNA sequences to identify different species of plants, animals and fungi. The DNA sequences are usually around 700 bp long and are conserved within a species but variable among species. The difference in nucleotide sequences between species makes it possible to tell them apart. Barcode can be used for easy identification of newly discovered species or of specimens such as fossils whose morphology has been greatly degraded. It can also be used in detection of illegal trades of restricted products such as across borders and by taxonomists for groups with similar morphology that makes it difficult to tell them apart. *Cytochrome oxidase subunit 1 (CO1)* gene is used for barcoding animals. It has all the qualities of a good DNA barcode i.e short, easily PCR amplified and sequenced, conserved intraspecifically but variable interpsecifically and has universal primers. The plant barcode however has not been fully developed. Several genes have been proposed by different researchers for use as a plant barcode. However, none of them works perfectly due to low amplification, low discrimination and/or lack of universality. Having a two or three locus barcode has been seen as the best solution. The Consortium for Barcode of Life (CBOL) plant working group adopted a combination of *matK* and *rbcL* genes as the standard plant barcode. Some researchers have said that a supplemental barcode maybe required for it to be effective. In our research, we set out to test how effective the *rbcL+matK* barcode was as a plant barcode. Sixtytwo plant species were collected at Shaw Nature Reserve (SNR). SNR is a collection of more than 1000 plant species growing in upland and bottomland forest, praire, wetland, abandoned graze and farmlands. It is about 30 miles West of Saint Louis, MO and a part of Missouri Botanical Garden. DNA was extracted from each specimen and *rbcL* gene PCR amplified using primers *rbcLa_F* and *rbcLa_R*. *matK* was PCR amplified using primers *KIM_F* and *KIM_R*. Successful amplicons, as determined by 1% gel electrophoresis, were sequenced at Yale University DNA sequencing facility. Sequences over 500 bp long for both *matK* and *rbcL* were used to identify the species on BOLD, Barcode of Life Database (http://www.barcodinglife.com/). The highest sequence match of the species was then compared to the identity of the species as determined using *Steyermark's Flora of Missouri, Vol 1* and *2* (Yatskievych). Fortyfive of the 62 species successfully amplified for *rbcL* and 38 for *matK*. Of the 45 *rbcL* amplifications, 21 were successfully sequenced (had 500 bp or longer). For *matK* 26 0f the 38 amplifications were successful (500 bp or longer). 20 species had successful amplifications for both *matK* and *rbcL*. 6 of the 20 species with successful amplifications were successfully identified to their species level on BOLD with 100% sequence match. 10 of them identified the genus of the species with 98-99% sequence match with the species. 4 of them identified the family of the species as well as other families. The other six could identify the genus level with the members of the genus having sequence match of above 97%. However, the individual species did not show up on the database. In conlusion, the *rbcL+matK* barcode works for some species. For others, it can only identify the species mostly to the genus level and the family level for a few of them. A supplemental barcode maybe required for the barcode to work perfectly. Better primers that work with all the land plants should be developed and if not possible, primers universal to each land group should be developed. It was also clear that sequences for some species were not available on BOLD. For barcoding to be successful, barcode sequences for all species will have to be uploaded on Genbank/BOLD.