ANNEX 4

Technical reports on genetics

Thünen Institute of Forest Genetics	Dr Céline Blanc-Jolivet
	PD Dr Bernd Degen
Université Libre de Bruxelles	Dr Kasso Dainou
	Dr Olivier Hardy2
University od Adelaide	Mr. Duncan Jardine
	Mr. Joey Gerlach
	Dr Elly Dormontt
	Dr Andrew Lowe
Plant Genetic Diagnostics Ltd	BSc. Maike Paulini
	Dr. Aki Michael Höltken

Development of a genetic reference map for Sapelli

Céline Blanc-Jolivet, Bernd Degen

Thünen Institute of Forest Genetics, Sieker Landstrasse 2, 22927 Grosshansdorf, Germany,
E-mail: bernd.degen@ti.bund.de

Material and methods

A total of 1192 *Entandrophragma* spp. samples were sampled in Ghana, Ivory Coast, Cameroon, Congo Brazzaville, Democratic Republic of Congo (DRC) and Gabon. Although indication of species identity was provided, we analyzed all samples and not only *Entandrophragma cylindricum* (Sapelli). One individual from Cameroon and one individual from DRC were selected for molecular marker (SNP) discovery through Restriction Site-Associated DNA sequencing (RADSeq). A set of putative SNPs located in the nuclear genome were selected for screening with a MassARRAY technology (Mc Kernan et al. 2002) conducted by the INRA Genome-Transcriptome Facility (GTF). Genotypes from a total of 190 individuals representing the whole distribution range were used to select the final set of SNP loci. All individuals were then genotyped at the selected loci to build the genetic reference data.

To detect genetic structure, we conducted a Bayesian clustering analysis (Pritchard et al. 2000). This method allows the grouping of individuals in genetic entities regardless of geographical origin. We proceeded to grouping of reference individuals based on the results of the cluster analysis and on the country of origin. This data was further used to control the geographical origin of blind samples.

Based on the results of the cluster analysis, we selected 10 individuals for a genome skimming approach using the MiSeq platform. The goal here was to analyze plastid genome (mitochondrial and chloroplastic) to find species-specific markers and to increase the resolution of the reference data based on nuclear SNP discovered with the RADseq method. Putative chloroplastic SNPs were also screened on a MassARRAY platform by the GTF.

Results

RADseq yielded more than 1,000 putative SNP loci. Among those, 131 were organized in four multiplexes for a MassARRAY genotyping. Three groupings were applied to find the loci with the strongest geographical signal: per bloc, per country, per country excluding Western African countries. For each loci and dataset, we estimated the correlation between geographical distance and genetic distance and geographical distance and differentiation index. We selected in priority loci, which had a high correlation at several datasets. Genetic assignment using a Bayesian approach (Rannala & Mountain 1997) by grouping reference samples by country was conducted for all loci and for the reduced set of loci to compare the accuracy of the reduced set of loci compared to results with all loci. A final set of 74 loci (two multiplexes) was defined for the screening of all individuals.

Nine genetic clusters could be identified. Among those, two could be attributed to other species, *Entandrophragma angolese* and *Entandrophragma utile*. A dendrogramm (Figure 1) illustrates the

genetic relationship among the identified clusters and also suggests that cluster 2 might represent another *Entandrophragma* species.

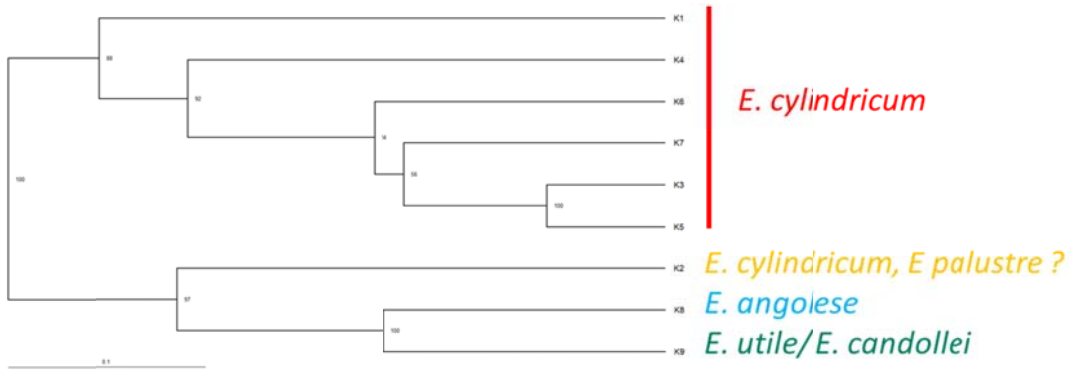

Figure 1: Dendrogramm representing the genetic relationships among the identified genetic clusters

The results highlight species misidentification problems in the field, especially in DRC where a lot of samples declared as *E. cylindricum* were attributed to *E. angolese* by the genetic data. The geographical distribution of the genetic clusters in presented in Figure 2.

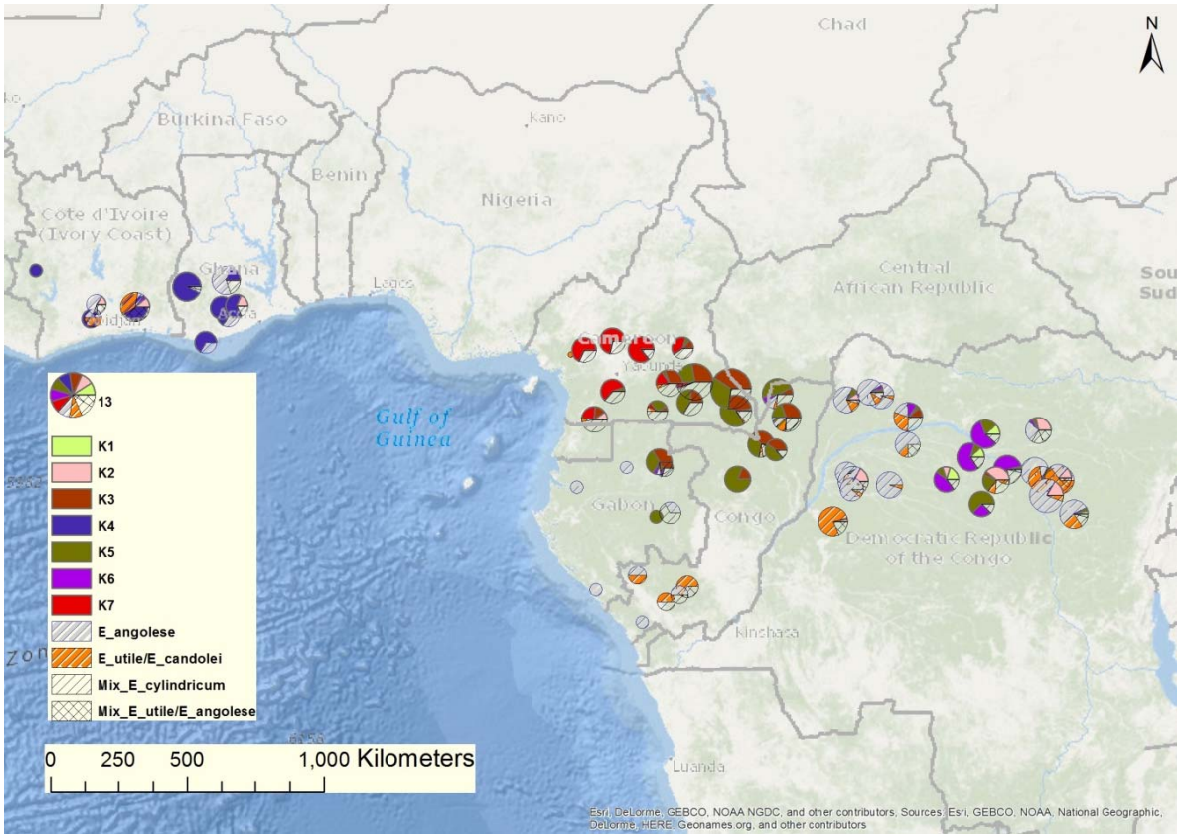

Figure 1: Distribution map of the *Entandrophragma spp.* genetic clusters

The map show that cluster 4 (K4) is restricted to Western Africa, clusters 1 (K1) and 6 (K6) to DRC, cluster 7 (K7) to Cameroon, while clusters 3 (K3) and 5 (K5) are common in all Central African countries. This means if clusters 1, 4, 6 and 7 are observed in a timber sample, then controlling the country of origin is straightforward. However, for clusters 3 and 5, more precise statistical analysis is required. Genome skimming and screening of putative SNPs on 190 individuals allowed the identification of a set of 13 loci from the chloroplast genome for species identification. Unfortunately, no further geographical structure could be identified on *E. cylindricum* samples.

Precision of the reference data

Individual assignment tests were performed using the Bayesian multilocus-approach (Rannala & Mountain 1997) in GDA_NT (Degen, unpublished) and a new approach based genetic distances of individuals (GeoAssign). All individuals from the reference data were self-classified to the country of origin using the leave-one-out approach (self-assignment, Efron 1983). The below table gives the results for the different countries:

Population	**Sample Size**	**Tested ind/**	**Bayesian approach**	**Distance approach**	
			% correct assigned	**% correct assigned**	**% claim accepted**
Cameroon	434	431	61	68	90
Congo_Braz	141	140	51	57	97
DRC	487	420	80	72	92
Gabon	36	36	22	3	84
Ghana	69	66	80	95	89
Ivory Coast	25	23	17	0	64
Total/Mean			**66**	**65**	**90**

The precision measured by the % of correct assigned individuals varies among the countries and statistical approaches from 0% to 95%. Most of the wrong assignment in West Africa is mixing up Ghana and Ivory Coast.

Blind test

For the blind tests we used the reference data and the Bayesian approach based on allele frequencies (Rannala & Mountain 1997). We tested different approaches of classification of the reference data (only classified by country, classified by country and genetic cluster). The results of this approach are given in the blind test reports (Annex 7.1 and Annex 7.2). In addition, we applied the approach based on genetic distances among individuals (Gregorius 1978) and higher thresholds for data completeness and the criteria to reject a claim (Annex 8). The different ways of analyzing the data, treating missing data and the different thresholds for the blind test lead to an overall performance from 50% to 83% correct results for claims on the country of origin.

Discussion

Species identity

The newly developed markers could efficiently assign individuals to species. However, the low sample size for *E. candollei* (six individuals) and potential identification mistakes in the field do not allow the genetic characterization of this species. Indeed, *E. candollei* and *E. utile* individuals were both belonging to the genetic cluster 9 based on nSNPs. However, chloroplast SNPs showed a subdivision of the cluster, but we could not define whether one group corresponds to *E. candollei* and the other to *E. utile*, or whether the second was another species. Further sampling with reliable taxonomic identification is thus needed to clarify the genetic differentiation between *E. candollei* and *E utile*.

The genotypes of individuals from cluster 2 raised a lot of questions. First, cluster 2 was genetically distant to the *E. cylindricum* group (Figure 1), which suggests that those individuals belong to another *Entandrophragma* species. This cluster mostly occurs in central DRC, where the species *E. palustre* is present. Again, sampling of well-taxonomically identified *E. palustre* should help to clarify this question. Second, cluster 2 was found to be very polymorphic at the chloroplast SNPs. It is therefore possible that this group is not genetically homogenous. Low amplification success at the nSNPs, probably due to low DNA quality, and low sampling size probably also artificially led to the grouping of these individuals in one single cluster. This could probably explain the occurrence of cluster 2 in Western Africa.

Genetic structure of Sapelli

Genetic data at the nSNPs clearly identified a strong differentiation among Western Africa and Central Africa. This pattern is also observed in many other species and results from the recolonization from different glacial refugias after the last glaciations. The data show that a contact zone between the two lineages exists in Western Cameroon (hybrid individuals cluster 4 (Western Africa) and 7 (Cameroon). We therefore expect that populations in Nigeria have intermediate genotypes.

The lack of genetic structure in Central Africa has also been reported in other species. Populations found in Cameroon, Congo Brazaville and Gabon probably originated from the same glacial refugia located on the coast. The relative climate homogeneity and the absence of geographical barriers further prevented differentiation in the Congo Bassin region. Populations from center DRC differ from the Congo Basin populations, probably also due to presence of another glacial refugia. Unfortunately, species identification mistakes in the field were very common in the Western DRC, Southern Gabon and Southern Congo Brazaville samples, thereby strongly reducing the number of *E. cylindricum* samples. Therefore we judge the reference data in this area for this species as not sufficient for timber tracking.

Literature

Efron B (1983) Estimating the error rate of a prediction rule - improvement on cross-validation, J. Am. Stat. Assoc. 78 , 316-331.

Gregorius H-R (1978). The concept of genetic diversity and its formal relationship to heterozygosity and genetic distance. Mathematical Bioscience 41: 253-271.

McKernan K, Fujii C, Ziauddin J, Malek J, McEwan P (2002). A high throughput and accurate method for SNP genotyping using Sequenom MassARRAY (TM) system. Am J Hum Genet 71(4): 454-454.

Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. Genetics 155(2): 945-959.

Rannala B, Mountain JL (1997). Detecting immigration by using multilocus genotypes. Proceedings of the National Academy of Sciences of the United States of America 94(17): 9197-9201.

Development of a genetic reference map for Iroko

Céline Blanc-Jolivet[1], Kasso Dainou[2], Olivier Hardy[2], Bernd Degen[1]

1) Thünen Institute of Forest Genetics, Sieker Landstrasse 2, 22927 Grosshansdorf, Germany, Email: bernd.degen@ti.bund.de
2) Université Libre de Bruxelles, 50 Av. F. Roosevelt; 1050 Brussels/Belgium

Material and methods

A total of 1920 *Milicia excelsa* and *Milicia regia* samples were sampled in Ghana, Ivory Coast, Cameroon, Congo Brazzaville, Democratic Republic of Congo (DRC), Gabon and Kenya. One individual from Benin and one individual from Kenya, available before sampling activities, were selected for molecular marker (SNP) discovery through Restriction Site-Associated DNA sequencing (RADSeq). A set of putative SNPs located in the nuclear genome were selected for screening with a MassARRAY technology (Mc Kernan et al. 2002) conducted by the INRA Genome-Transcriptome Facility (GTF). Genotypes from a total of 95 individuals representing the whole distribution range were used to select the final set of SNP loci. 1833 individuals were then genotyped at the selected loci to build the genetic reference data.

To detect genetic structure, we conducted a Bayesian clustering analysis (Pritchard et al. 2000). This method allows the grouping of individuals in genetic entities regardless of geographical origin. We proceeded to grouping of reference individuals based on the results of the cluster analysis and on the country of origin. This data was further used to control the geographical origin of blind samples.

Results

RADseq yielded more than 1,000 putative SNP loci. Among those, 138 were organized in four multiplexes for a MassARRAY genotyping. Three groupings were applied to find the loci with the strongest geographical signal: per bloc, per country, per country excluding Western African countries. For each loci and dataset, we estimated the correlation between geographical distance and genetic distance and geographical distance and differentiation index. We selected in priority loci, which had a high correlation at several datasets. Genetic assignment by grouping reference samples by country was conducted for all loci and for the reduced set of loci to compare the accuracy of the reduced set of loci compared to results with all loci (Rannala & Mountain 1997). A final set of 79 loci (two multiplexes) was defined for the screening of all individuals, which included two chloroplastic loci formerly identified (Dainou et al., 2010).

Five genetic clusters could be identified, among those one represented the species *M. regia*, commonly observed in Western Africa. One cluster was only present in Western Africa (K2), while three clusters (K3, K4 and K5) were present in central Africa.K3 was dominant in the Congo Basin, while K4 was restricted to the South-Western area and K5 on the Eastern Coast (Kenya and Eastern DRC) (Figure 1).

Further work was conducted to allow genetic screening of SNPs with PCR-RFLP, which can be easily applied in African laboratories. Among the 79 loci used for the final screening, 23 were selected for the development of a PCR-RFLP approach. Assignment tests showed that the use of these 23 loci reduced self-assignment success of 10% compared to the set of 79 loci.

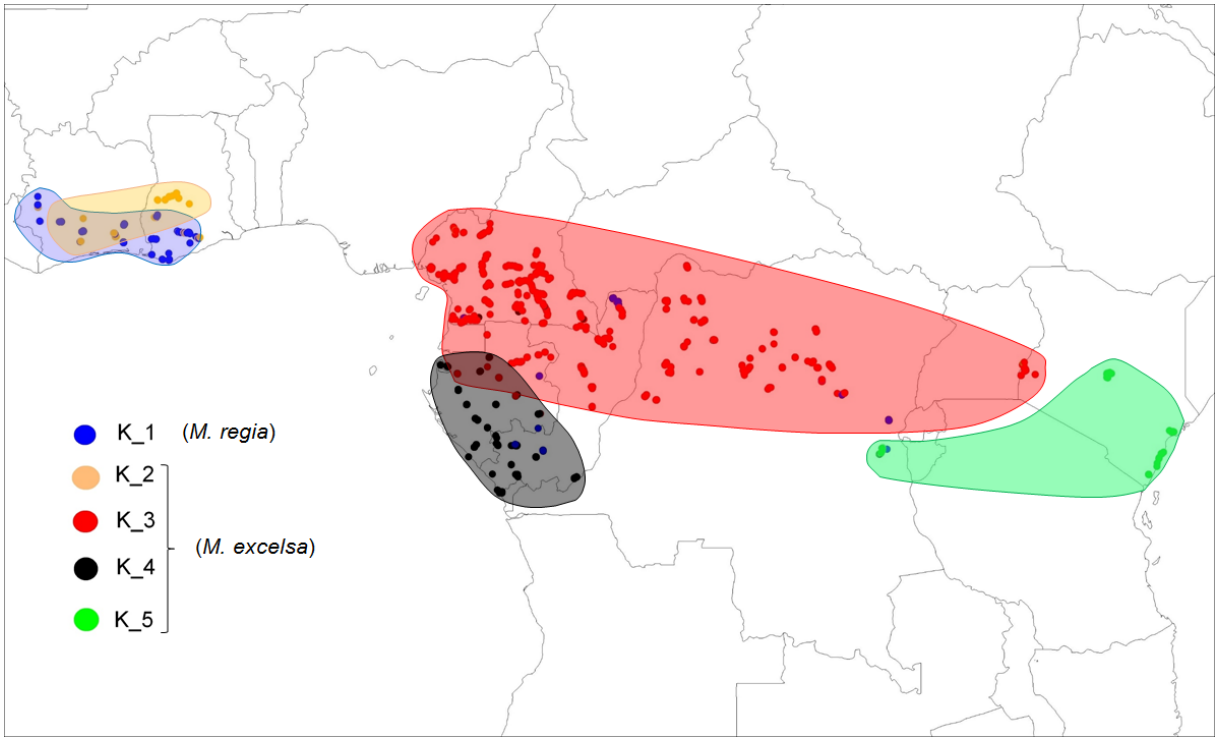

Figure 1: Distribution map of the *Milicia spp.* genetic clusters

Precision of the reference data

Individual assignment tests were performed using the Bayesian multilocus-approach (Rannala & Mountain 1997) in GDA_NT (Degen, unpublished) and a new approach based genetic distances of individuals (GeoAssign). All individuals from the reference data were self-classified to the country of origin using the leave-one-out approach (self-assignment, Efron 1983). The below table gives the results for the different countries:

Population	**Sample Size**	**Tested ind/**	**Bayesian approach**	**Distance approach**	
			% correct assigned	**% correct assigned**	**% claim accepted**
Cameroon	306	305	66	43	83
Congo_Braz	260	259	49	42	87
DRC	412	411	53	45	80
Gabon	252	251	62	71	91
Ghana	46	46	72	42	89
Ivory Coast	101	101	50	94	100
Kenya	103	103	81	77	95
Total/Mean			**59**	**54**	**86**

Blind test

For the blind tests we used the reference data and the Bayesian approach based on allele frequencies (Rannala & Mountain 1997). We tested different approaches of classification of the reference data (only classified by country, classified by country and genetic cluster). The results of this approach are given in the blind test reports (Annex 7.1 and Annex 7.2). In addition, we applied the approach based on genetic distances among individuals (Gregorius 1978) and higher thresholds for data completeness and the criteria to reject a claim (Annex 8). The different ways of analyzing the data and the different thresholds for the blind test lead to an overall performance from 40% to 60% correct results for claims on the country of origin.

Discussion

Species identity

The set of 79 SNP loci could efficiently differentiate the two *Milicia* species. Interestingly, a few individuals from Central Africa were assigned to the *M. regia* cluster and their leaves showed similar morphology. Further analysis could determine that these individuals genetically diverge from *M. regia* samples occurring in Western Africa. This indicates that *M. regia* was formerly widespread in Central Africa.

Genetic structure Iroko

Genetic data at the nSNPs clearly identified a strong differentiation among Western Africa and Central Africa. This pattern is also observed in many other species and results from the recolonization from different glacial refugias after the last glaciations. The presence of distinct genetic group in Kenya indicates the presence of a glacial refugia on the Eastern coast with restricted subsequent dispersal, probably due to geographical barriers (rift valley). The lack of genetic structure in Central Africa has also been reported in other species. Populations found in Cameroon, Northern Congo Brazaville and Northern Gabon (K3) probably originated from the same glacial refugia located on the coast, while the presence of K4 suggests that a second refugia existed more South. The relative climate homogeneity and the absence of geographical barriers further prevented differentiation in the Congo Bassin region.

Reference

Dainou K, Bizoux JP, Doucet JL, Mahy G, Hardy OJ, Heuertz M. (2010). Forest refugia revisited: nSSRs and cpDNA sequences support historical isolation in a wide-spread African tree with high colonization capacity, *Milicia excelsa* (Moraceae). Mol Ecol 19: 4462–4477.

Efron B (1983) Estimating the error rate of a prediction rule - improvement on cross-validation, J. Am. Stat. Assoc. 78 , 316-331.

Gregorius H-R (1978). The concept of genetic diversity and its formal relationship to heterozygosity and genetic distance. Mathematical Bioscience 41: 253-271.

McKernan K, Fujii C, Ziauddin J, Malek J, McEwan P (2002). A high throughput and accurate method for SNP genotyping using Sequenom MassARRAY (TM) system. Am J Hum Genet 71(4): 454-454.

Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. Genetics 155(2): 945-959.

Rannala B, Mountain JL (1997). Detecting immigration by using multilocus genotypes. Proceedings of the National Academy of Sciences of the United States of America 94(17): 9197-9201.

Development of Genographic map for Ayous (*Triplochiton* scleroxylon) using SNP markers

Duncan Jardine, Joey Gerlach, Elly Dormontt, Andrew Lowe
University of Adelaide, Australia,
Email : andrew.lowe@adelaide.edu.au

Objective

Develop single nucleotide polymorphic (SNP) markers for ayous, to allow genetic identification and verification of the geographic source of origin of timber.

Material and Methods

Genographic map

SNP marker discovery

Partial genome sequencing for initial single nucleotide polymorphism (SNP) discovery was performed using an in-house reduced representation genome library preparation method modified for double restriction enzyme digest
method (after Vos et al. 1995; van Orsouw et al. 2007) (see Jardine et al. 2015 for more information on the protocol). A total of 48 individuals from 10 populations and 3 countries, representing the geographical distribution of ayous were used in this marker discovery stage. Genomic DNA of these samples was extracted at the Thünen Institute-Forest Genetics (TIFG), with the extractions sent to Adelaide for molecular analysis. The samples were sequenced using an Ion Torrent PGM™ system (Life Technologies), with the subsequent data analysed using the CLC-bio workbench (Qiagen) and Geneious (Biomatters) platforms. A shortlist of 127 potential loci was identified, and a final panel of 117 loci designed into three multiplexes were prepared for genotyping on the Sequenom® MassARRAY® iPLEX™ GOLD platform.

DNA from 90 samples, consisting of the original 48 samples used in the marker discovery, as well as an extra 48 samples were used to screen the potential SNP loci for suitability. Any loci that failed to amplify or found to be uninformative were excluded from further use.

A second set of SNP discovery and development was undertaken by TIFG using three of the reference samples on a Rad-Seq platform (details available from TIFG). A final panel of SNP markers was developed using a combination of the most suitable markers from both the Adelaide and TIFG marker sets. This final panel consisted of 235 markers across six multiplexes and was used to screen all individuals in the subsequent analysis.

Genotyping

A total of 911 individuals, representing 45 populations from five countries (Figure 1), were genotyped using the final SNP panel. Samples were available as either leaf tissue or cambium plugs, and DNA was extracted at the Australian Genome Research Facility (AGRF). Samples and loci that failed to meet a strict 95% sequencing success threshold were removed from further analysis. The remaining individuals and loci (792 and 190 respectively) were analysed for linkage disequalibrium (LD) and Hardy Weinberg (HW) equilibrium (removing outlier loci), heterozygosity and F_{ST}. When a pair of loci was identified as being linked, the locus with the lowest heterozygisity and/or lowest HW ratio was removed. The final dataset consisted of 753 individuals and 105 loci. A STRUCTURE analysis to identify the most appropriate number of genetic clusters. This was done using standard parameters and incorporating a burnin length of 3000000mcmc's and 700000mcmc run length.

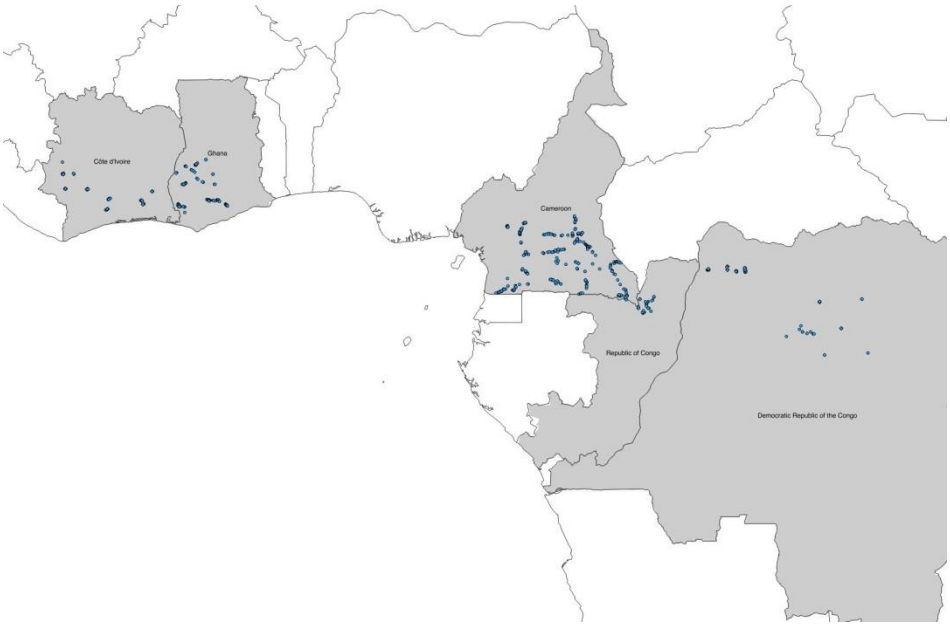

Figure 1. Location of all the samples used in the second round of genotyping

Blind samples

Twenty blind test samples were sent to Adelaide, 10 via WWF and 10 via G2S. DNA was extracted from all samples, using either a modified Analatik Jena innuPREP Plant DNA extraction kit or standard BoTAB for timber protocol. DNA was genotyped using the final SNP marker set on the MassARRAY platform at AGRF. The results from the genotyping of blind samples were screened and filtered to remove samples that had less than a 95% sequencing success rate. The geographic source of samples was estimated using the GeoAssign program, which provides a likely position of test samples as the centroid point of the 10 most genetically similar individuals from the reference data set.

Results and discussion

The results of the STRUCTURE analysis identified that a clustering of K=2 (Figure 2) was the most applicable for the dataset, followed by K=4. The K=4 clustering (Figure 3) was considered to be the most useful for geographic structuring and assignment as the following genetic clusters could be identified; Ghana/Ivory Coast; most of Cameroon; SE Cameroon, northern Republic of Congo and NW DRC; and central DRC.

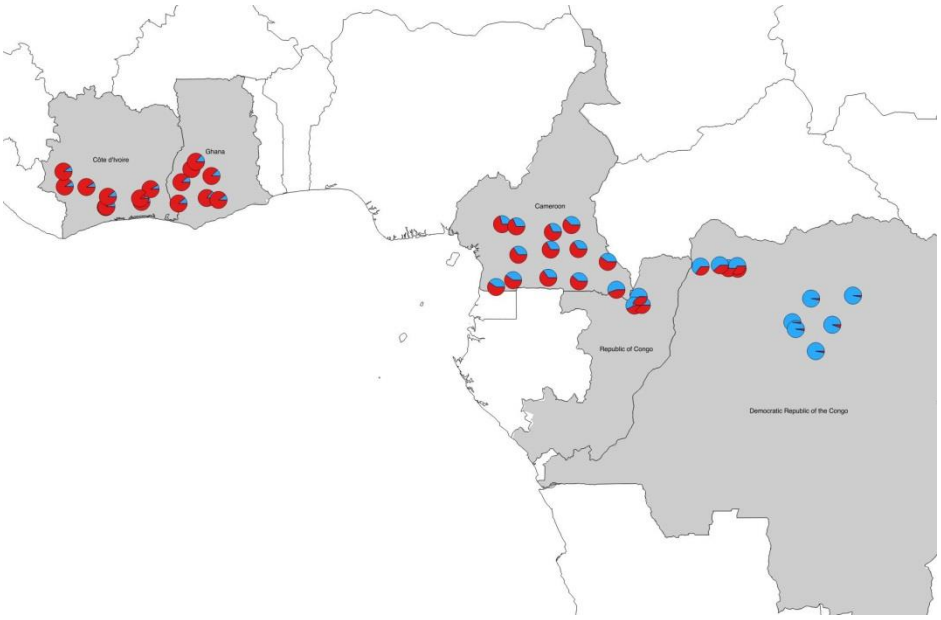

Figure 2. Proportional membership of populations and individuals under K=2 STRUCTURE clustering.

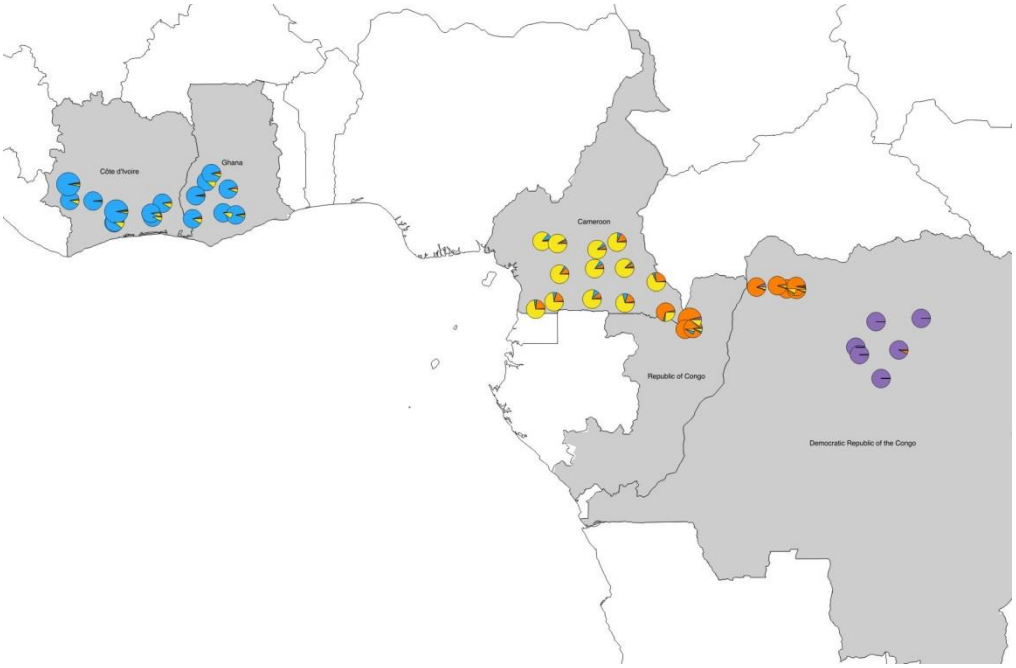

Figure 3. Proportional membership of populations and individuals under K=4 STRUCTURE clustering

For the blind samples (Table 1), 15 out of 20 samples produced high quality DNA that could be reliably genotyped, therefore 75% of tests could be processed. Using the GeoAssign analytical routine to identify the most likely country of origin, the claimed origin was accepted in eight cases and rejected in six cases. Scored against the true origin 11 out of 15 tests (73.3.%) were correctly accepted/rejected.

Sample ID	Claimed origin	GeoAssign origin	Real origin	GeoAssign Reject claim?	Correct?
BT_2014_533	Ghana	Ghana/Ivory Coast	Ghana	No	Yes
BT-2014_543	Cameroon	Cameroon	Gabon	Yes	No
BT_2014_547	DRC	Ghana	Ghana	Yes	Yes
BT_2014_551	CIV	Ivory Coast	Ghana	No	No
BT_2014_563	CIV	Ivory Coast	Ghana	No	No
BT_2014_567	DRC	Ghana	Ghana	Yes	Yes
BT_2014_568	Cameroon	Cameroon	Cameroon	No	Yes
BT_2014_578	Cameroon	Ghana	Ghana	Yes	Yes
BT_2014_594	Ghana	Ghana/Ivory Coast	Ghana	No	Yes
BT_2014_598	Ghana	Ghana	Ghana	No	Yes
G2S_O_T1	Ghana	DNA quality too poor			
G2S_O_T3	Ghana	Ghana	Ghana	No	Yes
G2S_O_T6	Cameroon	DNA quality too poor			
G2S_O_T7	Congo Braz	DNA quality too poor			
G2S_O_T9	DRC	Cameroon	Cameroon	Yes	Yes
G2S_O_T11	Cameroon	Cameroon	Cameroon	No	Yes
G2S_O_T13	Cameroon	Cameroon	Cameroon	No	Yes
G2S_O_T15	Cameroon	DNA quality too poor			
G2S_O_T16	Cameroon	DNA quality too poor			
G2S_O_T18	Cameroon	Ghana	Cameroon	Yes	No
Overall DNA analysis success = 75%			**Correct claim accept/reject = 73.3%**		

Table 1. Blind test results, indicating claimed origin, genetically identified origin and true origin, claim rejection/acceptance and correct determination.

The results provide an insight into where further sampling would be useful and would improve future blind test results, in particular the following locations; Ghana/Ivory Coast, Gabon and southern Congo Braz and DRC and western Cameroon/Nigeria.

ITTO project "Development and implementation of a species identification and timber tracking system in Africa with DNA fingerprints and stable isotopes"

Report: DNA barcoding

Dr. Aki Michael Höltken & BSc. Maike Paulini (Plant Genetic Diagnostics Ltd., Thünen Institute of Forest Genetics)

Background

A genetic barcoding method for the identification of 21 tropical tree species based on timber samples requires molecular markers (a) with low intraspecific but sufficient interspecific variability, (b) a high-copy number of the target DNA fragments because of the low yield of DNA following extraction and (c) short in length due to high degradation of the DNA. The DNA of chloroplasts (cpDNA) combines all these features. It is present in multiple copies per cell, the ring structure of the cp-genome gives higher stability to the DNA molecule and, because cpDNA is maternally inherited, there is no recombination of the cp genome in contrast to nuclear DNA. Various protocols have been published to extract DNA from wood, from recently logged (almost fresh) up to processed timber and timber products from different steps in the chain-of-custody. Most of the approaches are developed to mitigate the effects of contamination of the samples with external DNA and to minimise further demolition of already highly degraded DNA sequences (DE FILIPPIS & MAGEL 1998, DEGUILLOUX et al. 2002, RACHMAYANTI et al. 2006, ASIF & CANNON 2007).

Methodological approach

The selected tree species of this project belong to 9 botanical families and 8 orders. This circumstance poses a huge challenge for the development of a universal DNA barcoding system. Although universal fragments are available for plant species identification, these are mostly not suitable for timber species identification and require major modifications:

1. Official barcoding fragments are too long for amplification from timber samples. Universal primers have to be designed within these DNA sequences to reduce the length of the amplification products.
2. Intergenic sequences are too variable in size as well as in the order of nucleotides. An alignment of the sequences as well as the detection of overlapping DNA sections for the design of universal primers is not possible.
3. Coding sequences seem to display the only option of finding universal barcoding sequences for this large botanical variety of tropical timber species.

Sequencing: After sequencing diverse coding cpDNA regions using reference material provided by project partners (cambium, herbarium dried leaf material) and comparing the sequencing results with (partly) available data from the NCBI data bank, the *rbcL* gene showed the best features for primer design in order to develop barcoding sequences. This fragment is 575 basepairs long and contains, besides some indels of only 1 bp., 95 SNPs (single nucleotide poplymorphisms), a part of them species or genera specific, others family or order specific (see figue 1).

Primer design: When comparing all sequence sites, we found that the alignment revealed no region with sufficient consensus to accommodate a unique single oligonucleotide for use as primer although this fragment is one of the most conservative regions within the cpDNA genome. In all cases, one to three nucleotides were mismatched. When designing primers for this region, we introduced a degenerate site, or "wobble", to compensate for the variability in the target sequence (see figure 1).

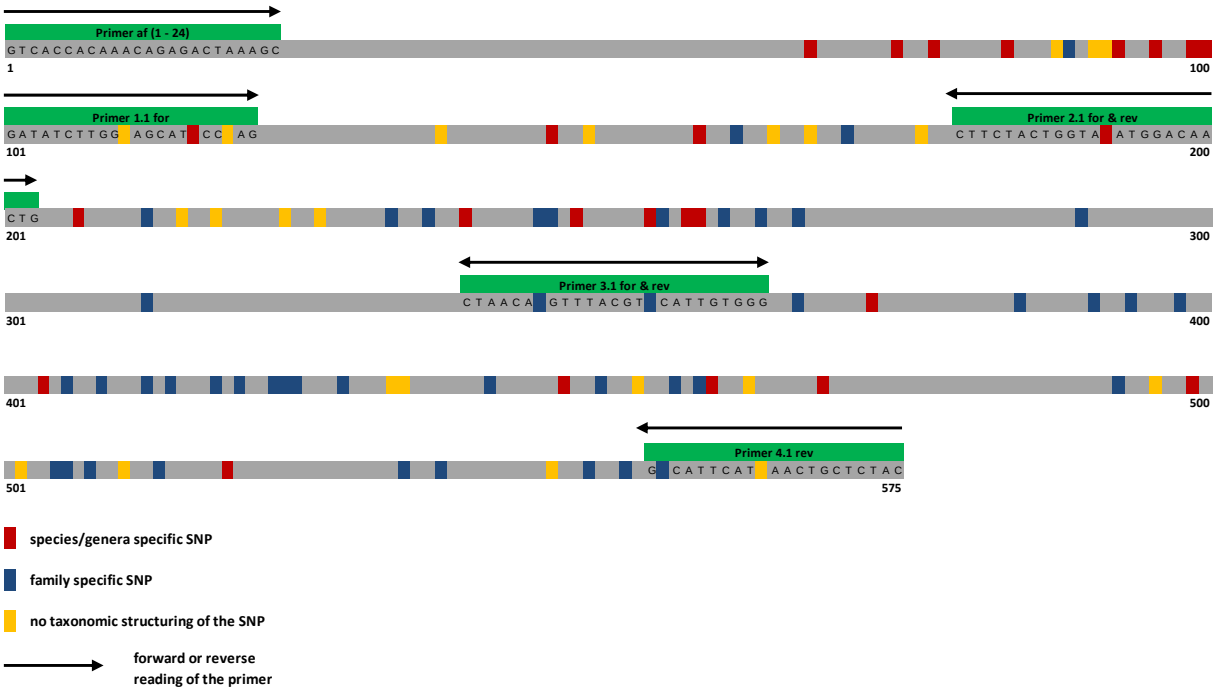

Figure 1: Used cpDNA barcoding sequence (*rbc*L) for differentiating the 22 tropical timber tree species, the forward and reverse primers and the specifity of the detectes SNPs (species/genera or family specific SNPs or SNPs with no taxonomic structuring)

Results and conclusions from the timber sample blind test

95 SNPs (single nucleotide polymorphisms) have been detected within this cpDNA fragment between the different taxonomic species, genera, families and orders (see figure 1). The results of applying this fragment in a blind test on 50 timber samples are shown in table 1.

As positive cases we declined the blind test timber samples for which we could identify the right species or genera. In most of these cases we stopped our identification on the genera level, because most of the listed genera consist of many species world wide for which we have no reference material or which could not been differentiated by the chosen cpDNA marker (see G2S_S_1.0, for example). In many cases there were also no or too bad PCR amplification products for interpretation (no results from the lab) or even wrong interpretation of the outcome of the genetic analysis. Altogether, it turned out to be difficult to apply DNA techniques for the differentiation between taxonomically very distant species, e.g. species belonging to different families or even different taxonomical orders. Due to this high variability of the DNA between taxonomically distant species, the design of well functioning consensus primers requires the introduction of “wobbles”. These might have caused the problems in amplifying the identified barcoding fragments, particularly of the highly degraded DNA. Further, there was no chance to analyse further cpDNA sequences because there was no overlapping information making the outcome of alignment procedures impossible to interpret. For taxonomically distant families and orders it is much more effective to apply morphological methods as used by Gerald Koch.

But a very interesting and useful application of DNA techniques should be considered in cases, in which morphological methods are not sufficient to differentiate between different species. This is the case for very closely related species and has already successfully tested in previous genetic projects (see DNA-based identification of mahogany species [*Swietenia*] Höltken et al. 2011). Here we can circumvent the above mentioned “alignment” problems of the DNA sequences. A claim from the morphological survey/expertise for the genera or the family in the first instance would be and then we can detect the species name with genera or family specific molecular methods. In the case of differentiating closely related tree species we are able to use intergenic sequences *(psb*A-*trn*H, *mat*K-

*trn*H etc.). Further, for species groups that have been studies by next generation sequencing techniques (see RAD-sequencing of *Entandrophragma*) and for which thousands of SNPs have been developed, the MassARRAY Techonology (Agena Bioscience)can be applied for a much more reliable species determination. Further, this new technique requires low DNA quantities after extraction, so that we could differentiate even between species of the *Entandrophragma* genus in this study (see samples G2S_S_5.0, RM_2014_48). Further candidates would be species of the genera *Afzelia*, *Erythrophleum*, *Khaya*, *Milicia* etc. requiring studies on species differentiation.

Table 1: Results of the DNA-barcoding blind test on tropical timber samples

Sample code	Claims on species names	True species names	Lab results		Comment
G2S_S_1.0	*Guibourtia ehie.*	*Afzelia africana*	Afzelia		identified genus right
G2S_S_1.5	*Baillonella toxisperma*	*Afzelia spp*		Pterocarpus	wrong
G2S_S_2.0	*Khaya anthotheca*	*Khaya ivorensis*		Guibourtia	wrong
G2S_S_3.5	*Baillonella toxisperma*	*Baillonella toxisperma*		no reference	wrong
G2S_S_5.0	*Entandrophragma*	*Entandrophragma angolense*	E. angolense		identified species right
G2S_S_8.0	*Entandrophragma candollei*	*Entandrophragma utile*		Guibourtsia	wrong
G2S_S_8.5	*Entandrophragma*	*Entandrophragma utile*	X	X	no results from the lab
G2S_S_10.0	*Milicia excelsa*	*Erythrophleum suaveolens*		Guibourtsia	wrong
G2S_S_11.5	*Guibourtia spp.*	*Guibourtia spp.*	Guibourtsia		identified genus right
G2S_S_13.0	*Lophira alata*	*Lophira alata*		no reference	wrong
G2S_S_13.5	*Lophira alata*	*Lophira alata*	X	X	no results from the lab
G2S_S_14.0	*Erythrophleum suaveolens*	*Milicia excelsa*	X	X	no results from the lab
G2S_S_15.0	*Milicia regia*	*Milicia regia*	Milicia		identified genus right
G2S_S_16.5	*Millettia laurentii*	*Millettia laurentii*	X	X	no results from the lab
G2S_S_18.5	*Khaya spp*	*Pericopsis elata*	X	X	no results from the lab
G2S_S_20.0	*Terminalia superba*	*Terminalia superba*	Terminalia		identified genus right
G2S_S_21.5	*Pterocarpus soyauxii*	*Pterocarpus soyauxii*		Malvaeceae family	wrong
G2S_S_24.0	*Triplochiton scleroxylon*	*Triplochiton scleroxylon*	*Triplochiton*		identified genus right
G2S_S_25.5	*Entandrophragma utile*	*Mansonia altissima*	no reference		exclusion right
G2S_S_30.5	*Triplochiton scleroxylon*	*Sterculia rhinopetala*		*Triplochiton*	wrong
G2S_S_33.5	*Erythrophleum ivorense*	*Lovoa trichiloides*		Pterocarpus	wrong
G2S_S_35.5	*Afzelia spp*	*Afzelia spp*		Erythrophleum ssp.	wrong
G2S_S_38.5	*Nauclea diderrichii*	*Nauclea diderrichii*	X	X	no results from the lab
G2S_S_41.5	*Aningeria robusta*	*Mansonia altissima*		Cyclodiscus	wrong
G2S_S_47.5	*Cylicodiscus gabunensis*	*Cylicodiscus gabunensis*	*Cylicodiscus*		identified species right
RM_2014_03	*Milicia excelsa*	*Millettia laurentii*	X	X	no results from the lab
RM_2014_04	*Erythrophleum ivorense*	*Erythrophleum suaveolens*	Erythrophleum ssp.		identified genus right
RM_2014_13	*Khaya ivorensis*	*Khaya anthotheca*	X	X	no results from the lab
RM_2014_37	*Erythrophleum suoveolens*	*Erythrophleum ivorense*	X	X	no results from the lab
RM_2014_39	*Entandrophragma utile*	*Aucoumea kleineana*	Aucoumea		identified genus right
RM_2014_42	*Entandrophragma angolense*	*Nauclea diderrichii*	Nauclea		identified genus right
RM_2014_45	*Afzelia pachyloba*	*Afzelia bipindensis*	Afzelia		identified genus right
RM_2014_48	*Entandrophragma cylindricum*	*Entandrophragma angolense*	E. angolense		identified species right
RM_2014_49	*Aningeria robusta*	*Baillonella toxisperma*	Sapotaceae family		identified family right
RM_2014_59	*Aucoumea klaineana*	*Afzelia pachyloba*	Afzelia		identified genus right
RM_2014_60	*Cyclodiscus gabunensis*	*Entandrophragma utile*	X	X	no results from the lab
X2-57	*Ptercarpus soyauxii*	*Pericopsis elata*	X	X	no results from the lab
X2-58	*Baillonella toxisperma*	*Aningeria robusta*	X	X	no results from the lab
X2-59	*Afzelia bipindensis*	*Afzelia africana*	X	X	no results from the lab
X2-65	*Guibourtia ehie.*	*Guibourtia tessmanii*		Nauclea	wrong
X2-66	*Millettia laurentii*	*Millicia excelsa*	X	X	no results from the lab
X2-67	*Khaya grandiflora*	*Khaya ivorensis*	X	X	no results from the lab
X2-68	*Milicia regia*	*Milicia excelsa*	X	X	no results from the lab
X2-69	*Terminalia superba*	*Terminalia superba*	Terminalia superba		identified species right
X2-74	*Pericopsis elata*	*Pterocarpus soyauxii*	X	X	no results from the lab
X2-75	*Nauclea diderrichii*	*Aucoumea kleineana*	X	X	no results from the lab
X2-76	*Khaya ivorensis*	*Entandrophragma cylindricum*	X	X	no results from the lab
X2-78	*Triplochyton scleroxylon*	*Triplochyton scleroxylon*	X	X	no results from the lab
X2-79	*Pericopsis elata*	*Cylicodiscus gabunensis*	X	X	no results from the lab
X2-81	*Lophira alata*	*Lophira alata*	X	X	no results from the lab

Nevertheless, on fresh material (leaf, cambium, bud tissues) this chosen *rbc*L fragment turned out to differentiate many tropical timber species very effectively. It can be used as a quick-check procedure improving the true species´ identity of tropical tree sample collections (quality management system).

Literature

HÖLTKEN A.M., SCHRÖDER H., WISCHNEWSKI N., MAGEL E. & FLADUNG M. (2012): Development of DNA-based methods to identify CITES-protected timber species: A case study in the Meliaceae family. *Holzforschung* 66: 97-104.