

ANNEX 8

**Example for a comparative analysis of the blind test data on
geographic claims of the ITTO-Africa Project**

Example for a comparative analysis of the blind test data on geographic claims of the ITTO-Africa Project

Bernd Degen

(02/11/2015)

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

Background

We discussed among all involved partners the results of the blind test during a project meeting in Großhansdorf on the 23.06.2015. The meeting have shown that the different groups used quite variable approaches to analyse the data. The range of methods varied from a pure expert evaluation with no statistics up to advanced statistical approaches. Also the used criteria to reject a claim on origin and to include or exclude individuals of reference and blind tests sets in the analysis were not unique among the groups. Thus the proportion of correct results in the blind test is influenced by two different factors:

- a) by the performance of the method and the quality of the reference data
- b) by the statistical approach and the criteria to include or exclude incomplete data as well as the thresholds to reject or accept a declaration of origin

I analysed the provided reference data and blind test data of the ITTO project using a unique approach and using the same criteria to judge on claims in order to have a more objective basis to compare the results of the different groups. The raw data were shared with all involved laboratories on the 14/09/2015. The presented approach and the results are thought as a science based contribution to the ongoing discussion of the performance of the methods.

Method

I computed pairwise distance measures between the individual reference data and the test data. For the metric isotope data I made a z-transformation of the different elements and calculated then the city-block-distance among the isotope profile of the different individuals (Deichsel & Trampisch 1985; page 22 ff.). For the SNP-data in genetics I computed the genetic distance of Gregorius among the multilocus genotypes of the different individuals (Gregorius 1978). Then I computed exclusion probabilities to judge on the claim “correct country of origin”. In all cases with at least 10 reference data in the declared country, the

exclusion probability is equal the proportion of cases where the distance among a test sample and a reference sample was higher than the distance of this reference sample and all other reference samples that fulfil the requirements of the claim (e.g. from the same claimed country). The pairs of individuals were included in the analysis only if at least 80% of the data of the two compared individual were measured. Then the reference samples are ordered from the smallest to the largest distance after the distances among a test individual and all reference samples have been computed. A small distance means similar genetic composition or similar stable isotope profile of reference sample and test individual. If the declared country of origin is correct than we would expect a significant higher proportion of reference samples from the declared country of origin among the most similar reference samples. The index RND is computed for all present countries in the reference samples. The Index RND varies between -1 and +1. If the Index RND is equal 0 then the proportion of similar reference samples is exactly as high as randomly expected according to the sample intensity in a country. A positive Index RND indicates an excess and a negative index stands for a deficit of similar reference samples compared to the sample proportion. The departure between observed and expected numbers of similar individuals has been checked with an Exact Fisher's test.

Thresholds and rules to judge on claims

The below presented results were computed based on the following thresholds:

- a) test individuals and reference data with less than 80% of the gene markers or stable isotopes completed were not included in the analysis. => "unsolved"
- b) for test individuals with an exclusion probability $\geq 80\%$ the declaration was "rejected"
- c) for test individuals with an exclusion probability $< 80\%$ but an RND-Index below 0.1 and an alternative country with a significant positive and higher RND-Index => "rejected"
- d) all test individuals that could not be rejected by the rules b and c are classified as "accepted_a"
- e) all test individuals that could not be rejected by the rules b and c and got a significant positive RND-Index > 0.1 for the declared country are classified as "accepted_confirmed".

Results

The detailed results are given in the attached three tables. For each species there is one table with the direct comparison of isotope and genetic results.

- The isotope data provided results for all test samples data. For the genetics between 17% and 47% of the samples did not have enough SNPs amplified. This is due to the degradation of DNA in timber.
- Based on all test individuals that had sufficient data the success rate of the isotopes varied between 55% and 75%. For the genetic data the proportion of correct results varied between 60% and 83%.
- For the isotopes overall all species 10 out of 60 samples have been wrongly classified as “rejected” (17 %) and for the genetics 4 out of 43 samples (9 %). These are very critical errors in face of a later practical application because it indicates the risk that a correct declaration gets rejected. Further work is needed to see if this error could be minimised with other thresholds during the data analysis. But it also clearly indicates problems with the reference data itself (insufficient number and distribution of reference samples, errors on geographic co-ordinates, too weak spatial structure of the measured variables).
- Over all species there were 28 cases for which both methods got a result and either one (82%) or both methods (18 %) made a wrong decision on the claim. This low rate of overlap in the errors indicates that a combination of methods has the potential to minimise the proportion of errors. The critical point is then the best way to judge which methods gets for a particular case the best and “correct” decision. This requires further work.

References

- Deichsel G, Trampisch HJ (1985) Clusteranalyse und Diskriminanzanalyse. Gustav Fischer Verlag: Stuttgart.
- Gregorius H-R (1978). The concept of genetic diversity and its formal relationship to heterozygosity and genetic distance. *Mathematical Bioscience* 41: 253-271.

Results blind test *Triplochiton scleroxylon* (Ayous)

ID_Sample	Claim	Solution	Isotopes								Genetics							
	Group	Group	Decision	N Cluster	Exclusion Probability	Index Cameroon	Index DRC	Index Ghana	Index Con_Braz	Index CIV	Decision	N Cluster	Exclusion Probability	Index Cameroon	Index Con_Braz	Index DRC	Index Ghana	Index CIV
G2S_O_T1	Ghana	Ghana	accept_a	30	0.69	0.35*	-0.36**	0.25ns	-0.07ns	-0.12ns								
G2S_O_T3	Ghana	Ghana	accept_a	30	0.71	0.30ns	-0.32*	-0.04ns	0.07ns	0.00ns	accept_a	38	0.51	-0.28**	-0.06ns	-0.23**	0.16ns	0.46**
G2S_O_T6	Cameroon	Con_Braz	accept_a	30	0.45	0.30ns	-0.32*	0.25ns	-0.11ns	-0.08ns								
G2S_O_T7	Con_Braz	Con_Braz	reject	30	0.72	0.35*	-0.32*	0.17ns	-0.07ns	-0.08ns								
G2S_O_T9	DRC	Cameroon	reject	30	0.89	0.30ns	-0.36**	0.17ns	-0.07ns	0.00ns	reject	26	0.99	-0.05ns	-0.08ns	-0.24*	0.24ns	0.17ns
G2S_O_T11	Cameroon	Cameroon	accept_a	30	0.71	0.30ns	-0.36**	0.08ns	-0.07ns	0.08ns	accept_a	32	0.43	0.13ns	0.03ns	0.12ns	-0.12ns	-0.18ns
G2S_O_T13	Cameroon	Cameroon	accept_a	30	0.39	0.17ns	-0.36**	0.38*	-0.11ns	-0.04ns	accept_a	19	0.19	0.07ns	0.11ns	0.13ns	-0.20ns	-0.15ns
G2S_O_T15	Cameroon	Cameroon	accept_a	30	0.6	0.26ns	-0.36**	0.17ns	-0.07ns	0.04ns								
G2S_O_T16	Cameroon	Cameroon	accept_a	30	0.63	0.30ns	-0.36**	0.21ns	-0.07ns	-0.04ns								
G2S_O_T18	Cameroon	Cameroon	accept_a	30	0.55	0.30ns	-0.36**	0.17ns	-0.07ns	0.00ns	reject	26	0.89	-0.05ns	-0.08ns	-0.24*	0.24ns	0.17ns
BT_2014_533	Ghana	Ghana	accept_a	30	0.37	0.13ns	-0.32*	0.25ns	-0.11ns	0.08ns	reject	23	0.82	0.06ns	-0.10ns	-0.21ns	0.05ns	0.31ns
BT_2014_543	Cameroon	Gabun	accept_confirmed	30	0.5	0.35*	-0.36**	0.25ns	-0.07ns	-0.12ns	accept_confirmed	28	0.76	0.38*	-0.04ns	-0.09ns	-0.13ns	-0.11ns
BT_2014_547	DRC	Ghana	accept_a	30	0.57	-0.04ns	-0.14ns	0.21ns	-0.11ns	0.12ns	reject	23	0.92	-0.28*	-0.10ns	-0.21ns	0.32ns	0.38ns
BT_2014_551	Ivory_Coast	Ghana	accept_a	30	0.39	-0.13ns	-0.18ns	0.21ns	-0.11ns	0.24ns	accept_a	39	0.76	-0.23*	-0.06ns	-0.19*	0.25*	0.30ns
BT_2014_563	Ivory_Coast	Ghana	reject	30	0.57	0.13ns	-0.27ns	0.33*	-0.11ns	-0.04ns	accept_a	21	0.66	-0.31*	-0.05ns	-0.24ns	0.35ns	0.29ns
BT_2014_567	DRC	Ghana	accept_a	30	0.47	0.04ns	0.05ns	0.00ns	-0.07ns	0.04ns	reject	17	0.85	-0.15ns	-0.06ns	-0.21ns	0.00ns	0.58*
BT_2014_568	Cameroon	Cameroon	accept_confirmed	30	0.47	0.35*	-0.36**	0.08ns	-0.07ns	0.04ns	accept_a	20	0.62	0.07ns	-0.05ns	0.13ns	0.06ns	-0.21ns
BT_2014_578	Cameroon	Ghana	accept_a	30	0.3	0.22ns	-0.32*	0.29ns	-0.11ns	-0.04ns	reject	31	0.88	-0.29**	-0.07ns	-0.24*	0.40**	0.24ns
BT_2014_594	Ghana	Ghana	accept_confirmed	30	0.35	0.00ns	-0.32*	0.38*	-0.11ns	0.08ns	accept_confirmed	24	0.66	-0.28ns	-0.09ns	-0.26*	0.40*	0.19ns
BT_2014_598	Ghana	Ghana	accept_a	30	0.46	0.00ns	-0.14ns	0.25ns	-0.07ns	0.00ns	accept_a	31	0.79	-0.21ns	-0.03ns	-0.24*	0.16ns	0.38*
		Correct		13	65%							10	67%					
		Wrong		7	35%							5	33%					
		Unsolved		0	0%							5	25%					

Results blind test Entandrophragma cylindricum (Sapelli)																				
ID_Sample	Claim	Solution	Isotopes									Genetics								
	Group	Group	Decision	N Cluster	Exclusion Probability	Index Ghana	Index Gabon	Index DRC	Index CIV	Index Con_Braz	Index Cameroon	Decision	N Cluster	Exclusion Probability	Index Cameroon	Index Con_Braz	Index DRC	Index Gabon	Index Ghana	Index CIV
G2S_O_E4	Ghana	Ghana	reject	30	0.88	0.00ns	-0.04ns	-0.60*	0.11ns	0.00ns	0.29ns	unsolved		-	-	-	-	-	-	-
G2S_O_E6	DRC	Cameroon	reject	30	0.67	-0.07ns	-0.04ns	-0.67**	0.11ns	-0.04ns	0.46**	unsolved		-	-	-	-	-	-	-
G2S_O_E8	Cameroon	Cameroon	accept_a	30	0.58	0.11ns	-0.04ns	-0.27ns	0.11ns	0.00ns	-0.04ns	accept_a	25	0.71	0.30ns	-0.10ns	0.00ns	-0.04ns	0.00ns	0.00ns
G2S_O_E11	Con_Braz	Con_Braz	accept_a	30	0.58	-0.04ns	-0.04ns	-0.07ns	0.07ns	0.00ns	0.04ns	unsolved		-	-	-	-	-	-	-
G2S_O_E13	Con_Braz	Con_Braz	accept_a	30	0.55	0.00ns	-0.04ns	-0.27ns	0.07ns	0.00ns	0.13ns	unsolved		-	-	-	-	-	-	-
G2S_O_E14	Con_Braz	Con_Braz	accept_a	30	0.52	0.00ns	-0.04ns	-0.27ns	0.07ns	0.00ns	0.13ns	unsolved		-	-	-	-	-	-	-
G2S_O_E18	DRC	DRC	accept_a	30	0.43	0.14ns	-0.04ns	-0.40ns	0.07ns	0.00ns	0.04ns	unsolved		-	-	-	-	-	-	-
G2S_O_E20	DRC	DRC	reject	30	0.44	0.04ns	-0.04ns	-0.60*	0.11ns	0.04ns	0.21ns	accept_a	66	0.63	-0.30ns	0.15ns	0.08ns	0.02ns	-0.08*	-0.02ns
G2S_O_E22	DRC	DRC	accept_a	30	0.57	-0.04ns	-0.04ns	0.00ns	0.04ns	0.00ns	0.04ns	unsolved		-	-	-	-	-	-	-
G2S_O_E24	DRC	DRC	accept_a	30	0.58	-0.04ns	-0.04ns	-0.07ns	0.07ns	0.00ns	0.04ns	accept_confirmed	48	0.48	-0.45ns	0.08ns	0.30**	-0.04ns	-0.09ns	-0.02ns
BT_2014_510	Ghana	Ghana	reject	30	0.96	-0.04ns	0.00ns	-0.87***	0.18ns	0.00ns	0.38*	accept_confirmed	32	0.59	-1.00***	-0.19*	-0.10ns	-0.03ns	0.79***	0.19*
BT_2014_522	DRC	DRC	reject	30	0.55	0.07ns	0.00ns	-0.67**	0.11ns	0.04ns	0.17ns	accept_a	40	0.62	0.13ns	-0.03ns	-0.03ns	0.05ns	-0.03ns	-0.03ns
BT_2014_525	Ghana	Ghana	reject	30	0.85	0.00ns	0.00ns	-0.40ns	0.11ns	0.00ns	0.13ns	accept_confirmed	25	0.68	-1.00***	-0.19ns	-0.14ns	-0.04ns	0.83***	0.12ns
BT_2014_534	Con_Braz	DRC	accept_a	30	0.29	-0.04ns	0.07ns	-0.47ns	0.07ns	0.00ns	0.17ns	accept_a	45	0.6	-0.28ns	-0.05ns	0.15ns	0.05ns	-0.07ns	0.02ns
BT_2014_552	DRC	DRC	reject	30	0.48	-0.04ns	0.11ns	-0.80***	0.07ns	0.00ns	0.33*	accept_a	51	0.49	-0.19ns	0.09ns	0.15ns	-0.02ns	-0.09ns	-0.02ns
BT_2014_557	DRC	DRC	reject	30	0.45	-0.04ns	0.07ns	-0.67**	0.07ns	-0.04ns	0.33*	accept_confirmed	51	0.63	-0.38ns	0.12ns	0.24**	-0.04ns	-0.09ns	-0.02ns
BT_2014_573	DRC	DRC	accept_a	30	0.29	0.11ns	0.04ns	-0.47ns	0.11ns	-0.07ns	0.08ns	accept_confirmed	39	0.56	-0.31ns	0.03ns	0.23*	-0.03ns	-0.06ns	-0.03ns
BT_2014_577	Cameroon	Cameroon	accept_a	30	0.36	-0.04ns	0.07ns	-0.80***	0.07ns	0.11ns	0.25ns	accept_a	39	0.65	0.00ns	0.09ns	0.03ns	-0.03ns	-0.08ns	0.00ns
BT_2014_580	Cameroon	Con_Braz	accept_confirmed	30	0.59	-0.07ns	-0.04ns	-0.53*	0.11ns	-0.04ns	0.38*	unsolved		-	-	-	-	-	-	
BT_2014_592	Con_Braz	DRC	accept_a	30	0.23	0.04ns	0.00ns	-0.47ns	0.07ns	0.04ns	0.13ns	accept_a	62	0.6	0.08ns	0.06ns	0.02ns	-0.02ns	-0.09*	-0.02ns
		Correct	11	55%								10	83%							
		Wrong	9	45%								2	17%							
		Unsolved	0	0%								8	40%							