Genetic Structure in Wild and Cultivated Populations of *Inga edulis* Mart. (Fabaceae) in Peruvian Amazon

Alexandr Rollo[1, 2], Bohdan Lojka[1], Marie Kalousová[1, 2], Bohumil Mandák[3], Maria Margarida Ribeiro[4]

[1] Czech University of Life Sciences Prague, Faculty of Tropical AgriSciences, Dept. of Crop Sciences and Agroforestry, Czech Republic
[2] Students for the Living Amazon o.p.s., Prague, Czech Republic
[3] Czech University of Life Sciences Prague, Faculty of Environmental Sciences, Czech Republic
[4] Instituto Politécnico de Castelo Branco, Escola Superior Agrária, Department of Natural Resources and Sustainable Development, Castelo Branco, Portugal

The purpose of this study was to compare wild and cultivated *Inga edulis* Mart. populations' genetic structure spanning the Peruvian Amazon and to identify the level of differentiation among studied populations. Does the domesticated form of the species have local origion?

Materials and methods

We evaluated a total of 259 trees, 197 cultivated and 62 wild, sampled in 27 geographically different populations (Tab. 1). The sexually mature trees were selected randomly and the minimum average distance between two sampled individuals was 200 m. Young leaves were collected and preserved in silica gel. The total genomic DNA was extracted for each sample and genotyped with four microsatellite primers: (Pel5, Inga03, Inga08 and Inga33). Diversity parameters were calculated and cultivated populations were compared with the wild populations. Hierarchical analysis of molecular variance (AMOVA) was done. To assess the population structure, we estimated the number of genetic clusters (*K*). Due to the weak population structure, we used a "locprior" model, which incorporated a priori sampling information. Two groups (cultivated/wild) of populations were used as priors. The number of clusters (*K*) was set at each value from one through twenty-eight, and the simulation was run ten times at each *K* value to confirm the repeatability of the results.

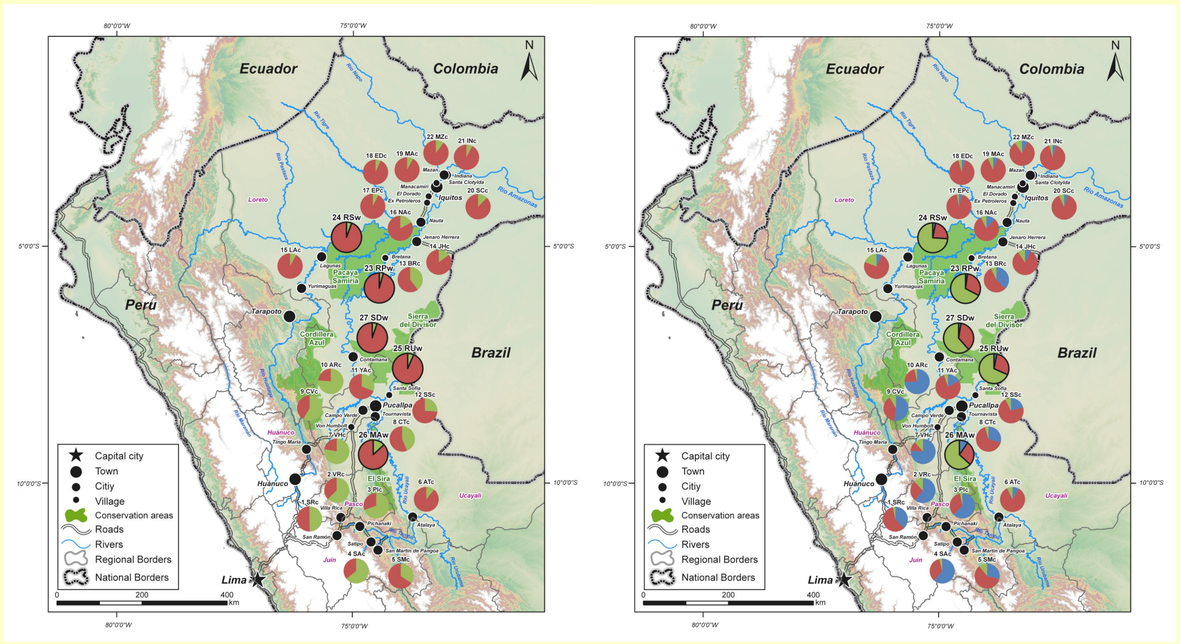

Figure 1. Results of Bayesian clustering for $K = 2$, plotted in the map of Peru. Populations assigned to two clusters (green and red) corresponding to the *I. edulis* wild (bigger black-lined pie charts) and cultivated populations (smaller no-lined pie charts) in the Peruvian Amazon.

Figure 2. Results of Bayesian clustering for $K = 3$, plotted in the map of Peru. Populations assigned to three clusters (visualized as blue. green and red) corresponding to the *I. edulis* wild (bigger black-lined pie charts) and cultivated populations (smaller no-lined pie charts) in the Peruvian Amazon.

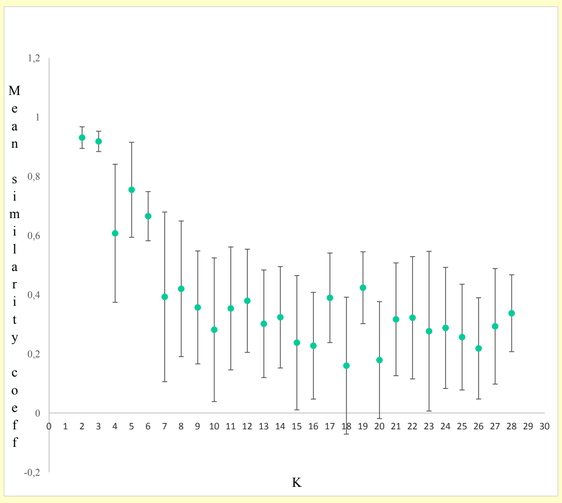

Figure 3 Number of clusters (K)

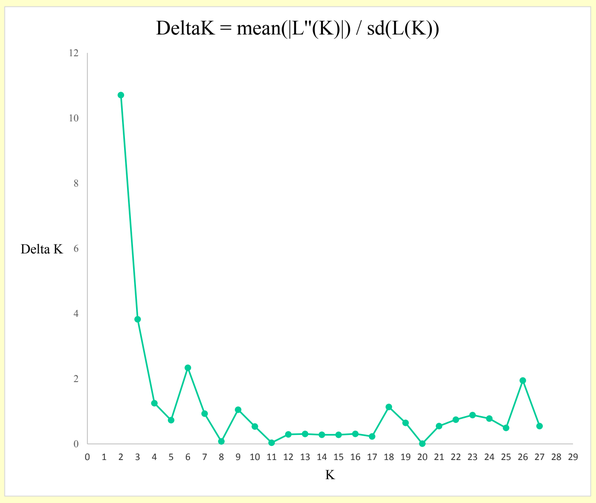

Figure 4. DeltaK = mean(|L''(K)|) / sd(L(K)).

Table 1. Type of populations sampled (Cultivated or Wild), study site (Site), population code (Pop.), sample size (N), geographic location - GPS coordinates in WGS84 (Latitude S and Longitude W) and altitude in meters above sea level.

Inga edulis	Site	Pop.	N	Latitude S	Longitude W	Altitude (m)
Cultivated	San Ramon	1 SRc	10	11°08´	75°21´	828 - 1200
	Villa Rica	2 VRc	5	10°44´	75°16´	1467 - 1494
	Pichanaqui	3 PIc	10	10°55´	74°52´	497 - 631
	Satipo	4 SAc	10	11°16´	74°38´	550 - 677
	San Martín de Pangoa	5 SMc	10	11°26´	74°30´	788 - 949
	Atalaya	6 ATc	10	10°43´	73°45´	223 - 244
	Von Humboldt	7 VHc	8	8°51´	75°00´	210 - 243
	Campo Verde-Tournavista	8 CTc	18	8°35´	74°46´	180 - 207
	Campo Verde	9 CVc	12	8°31´	74°47´	198 - 210
	Antonio Raimondi	10 ARc	11	8°29´	74°49	147 - 158
	Yarinacocha	11 YAc	5	8°20´	74°36´	144 - 154
	Santa Sofia	12 SSc	8	8°09´	74°15´	152 - 159
	Bretaña	13 BRc	16	5°15´	74°20´	103 - 108
	Jenaro Herrera	14 JHc	5	4°54´	73°40´	100 - 127
	Lagunas	15 LAc	10	5°14´	75°37´	108 - 135
	Nauta	16 NAc	5	4°30´	73°34´	106 - 139
	Ex Petroleros	17 EPc	5	4° 5'	73° 27'	97 - 108
	El Dorado	18 EDc	12	3° 57'	73° 25'	109 - 151
	Manacamiri	19 MAc	5	3° 43'	73° 17'	95 - 97
	Santa Clotylda	20 SCc	5	3° 40'	73° 15'	93 - 128
	Indiana	21 INc	7	3°29´	73°02	92 - 108
	Mazán	22 MZc	10	3°30´	73°04´	93 - 122
Sum			197			
Wild	Pacaya river	23 RPw	12	5° 41'	74° 57'	110-131
	Samiria river	24 RSw	6	5° 14'	75° 28'	105-123
	Utiquinia river	25 RUw	12	8° 10'	74° 17'	150-160
	Macuya	26 MAw	27	8° 53'	75° 0'	216-233
	Sierra del Divisor	27 SDw	5	7° 13'	74° 57'	196-231
Sum			62			

Table 4. AMOVA between cultivated and wild population groups, among populations within cultivated and wild population groups and within *I. edulis.* populations. degrees of freedom (df), sum of squared deviation (SS) , Φ statistics = fixation indexes, level of probability of obtaining a more extreme component estimate by chance alone (P).

Source of variation	df	SS	Variance components	% of total variance	Φ statistics	P
Between groups	26	94.23	0.10	6.86	$\Phi_{CT} = 0.07$	<0.000001
Among populations within groups	232	379.91	0.22	14.33	$\Phi_{SC} = 0.15$	<0.000001
Within populations	259	311.00	1.20	78.80	$\Phi_{ST} = 0.21$	<0.000001
Total	756	785.14	1.52			

Table 2. Diversity parameters per population. sample size (N), average number of alleles per locus (*A*), allelic richness (R_S), observed heterozygosity (H_O), expected heterozygosity (H_E), inbreeding coefficient (F_{IS}), Populations deviating from Hardy-Weinberg equilibrium at $P < 0.05$ are marked with an asterisk.

I. edulis	Pop.	N	*A*	R_S	H_O	H_E	F_{IS}
Cultivated	1 SRc	10	5.5	3.904	0.525	0.666	0.221*
	2 VRc	5	5.5	5.500	0.700	0.828	0.170
	3 PIc	10	5.0	3.983	0.550	0.670	0.187*
	4 SAc	10	5.8	4.053	0.575	0.633	0.096
	5 SMc	10	6.3	4.395	0.600	0.671	0.111
	6 ATc	10	4.8	3.664	0.525	0.613	0.151
	7 VHc	8	5.3	4.541	0.625	0.773	0.202*
	8 CTc	18	7.3	4.587	0.653	0.756	0.139*
	9 CVc	12	5.8	4.453	0.667	0.764	0.133
	10 ARc	11	5.5	4.314	0.614	0.744	0.182*
	11 YAc	5	5.0	5.000	0.650	0.661	0.019
	12 SSc	8	5.5	4.630	0.500	0.752	0.351*
	13 BRc	16	8.0	4.760	0.640	0.715	0.107
	14 JHc	5	3.8	3.750	0.500	0.661	0.266
	15 LAc	10	6.8	4.823	0.650	0.728	0.112
	16 NAc	5	4.3	4.250	0.600	0.639	0.068
	17 EPc	5	3.8	3.750	0.550	0.694	0.228
	18 EDc	12	5.0	3.768	0.458	0.624	0.274*
	19 MAc	5	3.8	3.750	0.500	0.606	0.192
	20 SCc	5	3.8	3.750	0.450	0.656	0.339*
	21 INc	7	3.8	3.324	0.429	0.500	0.153
	22 MZc	10	5.5	4.327	0.600	0.733	0.189*
Wild	23 RPw	12	8.3	5.234	0.625	0.717	0.134*
	24 RSw	6	6.5	5.818	0.750	0.792	0.058
	25 RUw	12	7.3	5.154	0.667	0.764	0.132*
	26 MAw	27	11.0	5.400	0.657	0.745	0.120*
	27 SDw	5	4.0	4.000	0.700	0.600	–0.191
Average			5.7	4.403	0.591	0.693	0.153
SD			1.7	0.644	0.084	0.073	0.104

Tabld 3. Diversity parameters comparison between cultivated and wild populations. Allelic richness (R_S), observed heterozygosity (H_O), average genetic diversity (H_S), inbreeding coefficient (F_{IS}) and among populations differentiation (F_{ST}), probability values for differences between groups for two-sided t-tests after 10.000 permutations (*P*) * = significant test ($P < 0.05$).

Diversity measures	Cultivated	Wild	*P*
R_S	4.240*	5.121*	0.0031
H_O	0.580*	0.665*	0.0275
H_S	0.700	0.740	0.2170
F_{IS}	0.172	0.101	0.0584
F_{ST}	0.076*	0.017*	0.0452

Results

We identified 71 alleles, with an average of 17.8 alleles per locus. For the 27 populations with at least 5 samples, the average number of alleles was 5.7, the average allelic richness 4.4, the observed heterozygosity 0.59, and the expected heterozygosity 0.69. The heterozygote deficit was non-significant, but the inbreeding coefficient was 0.153. Twelve populations were not in Hardy-Weinberg equilibrium. These populations were distributed randomly across the *I. edulis* sampled range and across cultivated and wild populations (Tab. 2) . We compared the wild *vs.* the cultivated populations, and significant deviation from the null expectation emerged. The allelic richness and the observed heterozygosity were lower in the group of cultivated populations. The level of differentiation among populations was significantly higher in the cultivated compared to the wild populations (Tab. 3). A hierarchical analysis of molecular variance revealed that a majority of the genetic diversity was partitioned within populations (78.8%), 14.3% was partitioned among populations within groups (wild and cultivated) and 6.9% was partitioned between cultivated and wild groups (Tab. 4). Results of Bayesian clustering were interpreted only for $K = 2$ (Fig. 1) and K = 3 (Fig. 2). Analysis of similarity coefficients as well as ΔK indicated that two and three clusters best explained the genetic structuring of *I. edulis* populations.(Fig. 3 and 4.).

Conclusion

According to the Bayesian structuring results we inferred that the cultivated material in Peruvian Amazon has different origins.

Acknowledgments

Foundation Nadace Nadání Josefa, Marie a Zdeňky Hlávkových, Czech Republic (Foundation Grant). We thank to Servicio Nacional de Areas Naturales Protegidas for the investigation authorization (N°004-2012-SERNANP-RNPS-J); National University of San Marcos (UNMSM), Veterinary Institute of Tropical and Highland Investigations, Regional Herbarium of Ucayali IVITA - Pucallpa, Peru.

Contact: Alexandr Rollo, rollo@ftz.czu.cz

CZECH UNIVERSITY OF LIFE SCIENCES PRAGUE

Faculty of Tropical AgriSciences

STUDENTS FOR THE LIVING AMAZON

agrinatura The European Alliance on Agricultural Knowledge for Development