Czech University of Life Sciences Prague

Faculty of Tropical AgriSciences

Molecular Characterization of *Plukenetia volubilis* L. and Analysis of Seed Storage Protein Pattern and Protein Fractions

Dissertation Thesis

Department of Crop Sciences and Agroforestry

Author: Ing. Martin Ocelák

Supervisor: doc. Ing. Bohdan Lojka, Ph.D.

Co-supervisors: Ing. Petra Hlásná Čepková, Ph.D.

Ing. Iva Viehmannová, Ph.D.

In Prague, September, 2016

Acknowledgment

I would like to express my gratitude to my supervisor doc. Ing. Bohdan Lojka, Ph.D. and co-supervisors Ing. Petra Hlásná Čepková, Ph.D. and Ing. Iva Viehmannová, Ph.D. for their guidance, advices, help and also patience during the studies, laboratory works and mainly during the writings.

My thanks also belong to IIAP represented by Ing. Danter Cachique Huansi and Lucas Garcia Chujutalli for their cooperation in samples collection, to Ing. Anna Prohasková for her guidance during analysis of proteins in Crop Research Institute in Prague - Ruzyně, to Ing. Eva Beoni, Ph.D. and Ing. Lenka Havlíčková, Ph.D. for their help in learning how to work in the laboratory; to Ing. Zdislava Dvořáková, Ph.D. for her help, teaching and encouragement and to Ing. Blanka Křivánková, Ph.D. for providing some useful materials.

Also my family contributed with their support in all means. So great thanks belong to my parents Jan and Jaroslava Ocelákovi and my boyfriend Ioannis Nikolakis for their love and support in all possible means.

This research was supported financially by an Internal Grant Agency of the University of Life Science Prague, CIGA (Project No. 20135004), by an Internal Grant Agency of the Faculty of Tropical AgriSciences - University of Life Science Prague, IGA (Project No. 20145020).

Declaration

I, Martin Ocelák, declare that this thesis, submitted in partial fulfilment of the requirements for the degree of Ph.D., in the Faculty of Tropical AgriSciences of the Czech University of Life Sciences Prague, is wholly my own work unless otherwise referenced or acknowledged.

……

Ing. Martin Ocelák

Abstract

Plukenetia volubilis is a higly nutritious oilcrop from the Southern America. This study aimed to assess the diversity and genetic relations in 173 sacha inchi samples from 10 locations in San Martín region in Peruvian Amazon. Genetic analysis was elaborated using ISSR markers, and the total protein and protein fractions content in its seeds were measured using Osborne´s and Kjeldahl´s methods. Results of Kjeldahl method were used for calibration model of FT NIRS. Eight primers showed variability in tested samples. ISSR fragments ranged from 200 to 2,500 bp. The obtained level of genetic variability was 36% among tested populations and 64% within populations. The mean of Nei's genetic diversity index H was 0.18, the mean of Shannon's information index was 0.28, the Nei´s gene differentiation index G_{st} was calculated to be 0.290 and gene flow index N_m estimated at 1.227. That is why the cluster analysis well revealed eight clusters containing mainly samples belonging to individual populations. PCoA clearly distinguished Chumbaquihui, Pucallpa, Dos de Mayo, and Aguas de Oro populations, the others were intermixed, and there were nine clearly distigished clusters in the NJ dendrogram. The obtained results indicated the level of genetic diversity is present in this location of Peru although it is influenced by anthropological aspects and independent on the geographical distances. The crude protein content of seeds in sacha inchi populations was detected at a range between 15.95% and 23.83% by the Kjeldahl method with mean at 20.61%. The protein fractions estimated by modified Osborne's method ranged between 10.2 and 18.4% for albumins-globulins; 0.1 and 0.5% for gliadins; and 3.6% and 8% for glutenins. The calibration model for FT NIRS had $R^2 = 0.88$ and RMSEP = 0.46%. Protein bands of total seed protein were detected in the wide range of molecular weight 10-75 kDa with seven bands detected on polyacrylamide gel. This study mapped current genetic diversity in selected populations in San Martín region of Peruvian Amazon. The significant differences were discovered both in genetic and protein contents. FT NIRS was successfully tested for protein content prediction.

Keywords: FT NIRS; genetic diversity; *Plukenetia volubilis*; oil crop, protein content; protein fractions content; Sacha inchi, SDS-PAGE

Abstrakt

Plukenetia volubilis je vysoce nutriční olejnina z Jižní Ameriky. Tato studie se zaměřila na zjištění genetické diverzity mezi 173 genotypy sacha inchi celkem z 10 lokalit v regionu San Martín v Peruánske Amazónii. K tomu bylo využito ISSR markerů a obsah bílkovin a bílkovinných frakcí v jejich semenech byl zjištěn Osbornovou a Kjeldahlovou metodou. Výsledky Kjeldahlovy metody byly využity pro sestavení kalibračního modelu spektroskopie v blízké infračervené oblasti. 8 ISSR primerů amplifikovalo jasné a reprodukovatelné polymorfní bandy. Amplifikované fragmenty měly délky v rozmezí od 200 do 2500 pb. Zjištěná míra genetické variability byla rozdělena na 36% mezi populacemi a 64% uvnitř populací. Průměrná hodnota indexu Neiovy genetické diverzity H byla 0,18, průměr Shannonova informačního indexu I 0,28. Neiův diferenciační index G_{st} nabyl hodnoty 0,290 a index genetického toku N_m 1,227. Díky tomu byla detekována jasná genetická struktura, díky níž PCoA jasně odlišila populace Chumbaquihui, Pucallpa, Dos de Mayo a Aguas de Oro, ale ostatní populace nebylo možné odlišit, zatímco v dendrogramu bylo zřetelně diferenciováno 9 klastrů převážně s jedinci té které populace. Zjištěné hodnoty naznačují, že genetická diverzita, ačkoli je ovlivněna antropologickými aspekty, zde existuje, a je nezávislá na geografické vzdálenosti. V těchto populacích byl v rámci této studie stanoven i průměrný obsah bílkovin 20,61%, s rozmezím 15,95% až 23,83% Kjeldahlovou metodou. Bílkovinové frakce byly stanoveny modifikovanou Osbornovou metodou mezi 10,2% a 18,4% pro albuminoglobulinovou frakci; 0,1 and 0,5% pro gliadiny; a gluteninů mezi 3,6% a 8%. Kalibrační model FT NIRS operoval s hodnotami koeficientu determinace $R^2 = 0.88$ a RMSEP = 0,46%, což jsou hodnoty odpovídající kvalitnímu modelu. Celkem 7 bandů celkových bílkovin detekovaných na polyakrylamidovém gelu se pohybovalo mezi 10 – 75 kDa molekulární hmotnosti. Tato studie zmapovala současnou genetickou diverzitu ve vybraných populacích regionu San Martín v peruánské Amazonii. Statisticky významné rozdíly byly mezi populacemi detekovány jak v oblasti genetiky, tak v oblasti obsahů bílkovin. Využití metody FT NIRS bylo úspěšně otestováno pro predikci obsahu celkových bílkovin.

Klíčová slova: FT NIRS; genetická diverzita; *Plukenetia volubilis*; obsah bílkovin; obsah bílkovinových frakcí; olejnina, Sacha inchi; SDS-PAGE

Resumen

Plukenetia volubis es una planta muy nutritiva de Sudamérica. El presente trabajo estudió la variabilidad genética de 173 muestras de sacha inchi de 10 localidades en la Región San Martín en la Amazonía peruana. La variabilidad genética fue analisada con marcadores ISSR y el contenido de las proteínas y sus facciones en sus semillas col el métodos de Osborne y Kjeldahl. Los resultados de la análisis de Kjeldahl fueron utilizados para compilar el modelo de calibración de la espectroscopía infraroja cercana. 8 marcadores ISSR amplificarón claras y reproducibles bandas polimorfas. Fragmentos amplificados fueron de la longitud entre 200 y 2500 pares de bases. La cantidad de la varibilidad genética detectada fue dividida a 36% entre las poblaciones y 64% dentro de las poblaciones. El promedio del índice de diversidad genética de Nei I fue 0.18, el promedio del índice de información de Shannon H fue 0.28. El índice de diferenciación de Nei G_{st} fue 0.290 y el índice de flujo genético N_m fue 1.228. Gracias a estas valores la estructuralización genética fue detectada claramente, por eso PcoA podría distinguir claramente las poblaciones Chumbaquihui, Pucallpa, Dos de Mayo y Aguas de Oro y la dendrograma de NJ contenió 9 racimos con mayormente individualidades de las mismas poblaciones. Los valores calculados indican que la diversidad genética, aunque afectada con aspectos antropológicos, existe en la area de investigación y es independiente de la distancia geográfica. En estas poblaciones hemos investigado el contenido de proteínas a ser 20.61% en promedio entre mínimo 15.95% y máximo 23.83% con el método de Kjeldahl. Las facciones fueran investigadas con el método de Osborne cuál encontró albúminas-globulinas a ser entre 10.2% y 18.4%; gliadinas entre 0.1 y 0.5%; y gluteninas entre 3.6% y 8%. Modelo de calibración de FT NIRS trabajó con coeficiente de determinación $R^2 = 0.88$ y RMSEP = 0,46%, cuáles son los valores apropiadas para el modelo de la alta calidad. En total 7 bandas de proteínas detectadas en gel de polyacrylamida osciló entre 10 – 75 kDa de peso molecular. Este estudio ha investigado la diversidad genética recien en poblaciones selectadas en la Región San Martín en la Amazonía peruana. Las diferencias importantes entre las poblaciones fueran detectadas ambos en la genetica y los contenidos de proteínas. Uso del método de FT NIRS fue probado con éxito para la predicción del contenido de protínas crudas.

Palabras clave: contenido de proteínas; contenido de facciones de proteínas; diversidad genética; FT NIRS; oleaginosa; *Plukenetia volubilis*; Sacha inchi, SDS-PAGE

List of used abbreviations

^{1}H NMR	Proton Nuclear Magnetic Resonance Spectroscopy
2DM	Dos de Mayo
AD	Anno Domini
ADO	Aguas de Oro
AFLP	Amplified Fragment Length Polymorphism
ANOVA	Analysis of Variance
AUC	Aucaloma
bp	Base Pair
BSA	Bovine Serum Albumin
cDNA	Complementary DNA
cpDNA	Chloroplast DNA
cpSSR	Chloroplast simple sequence repeat
CRI	Crop Research Institute
CTAB	Cetyl trimethylammonium bromide
CULS	Czech University of Life Sciences Prague
DALP	Direct Amplification of Length Polymorphism
DART	Diversity Arrays Technology
DNA	Deoxyribonucleic Acid
EST	Expressed Sequence Tag
FAO	Food and Agriculture Organization
FISH	Fluorescent *in situ* hybridisation
FTA	Faculty of Tropical AgriSciences
FT NIR	Fourier Transformation Near Infrared (spectroscopy)
CHU	Chumbaquihui
ILP	Intron length polymorphism
INIA	Instituto Nacional de las Inovaciones Agrarias (National Institute of Agrarian Innovations)
IPA	Inca Peanut Albumin
IR	Infrared
IRTAP	Inter-retrotransposon-amplified polymorphism
ISSR	Inter Simple Sequence Repeat
ITS	Internal Transcribed Spacers
kDa	Kilodalton
MIS	Mishquiyacu
msAFLP	Methyl sensitive AFLP
mtDNA	Mitochondrial DNA
NJ	Neighbour Joining
nrDNA	Nuclear Ribosomal DNA
QTL	Quantitative trait locus
PAC	Pacchilla
PCR	Polymerase Chain Reaction
PCoA	Principal Coordinates Analysis

PLS	Partial least squares regression
PUC	Pucallpa
RAC	Ramón Castillo
RAPD	Random Amplification of Polymorphic DNA
RBIP	Retrotransposon-based Insertion Polymorphism
RFLP	Random Fragment Length Polymorphism
SCAR	Sequence characterized Amplified Region
SDS-PAGE	Sodium dodecyl sulfate - Polyacrylamide gel electrophoresis
SCR	Santa Cruz
SLU	Santa Lucía
SNP	Single nucleotide polymorphism
SSRs	Simple Sequence Repeats
TRAP	Target Region Amplification Polymorphism
UBC	University of British Columbia
USDA	United States Department of Agriculture
WHS	Weight of Hundred Seeds

List of Figures

Figure 1. Distribution of *Plukenetia volubilis* in Southern America (Křivánková, 2012) 4
Figure 2. Map of regions of Peru, where sacha inchi is cultivated (Netmaps, 2016) 5
Figure 3. *Plukenetia volubilis* L. plant - its habit and fruits. (picture taken by the author of thesis) 6
Figure 4. *Plukenetia volubilis* is a liana which under cultivation needs support to avoid putrefaction of fruits. Plantations are usually intercropped (*Musa*, *Zea mays*, *Ananas comosus*, *Erythrina* spp., etc.) (picture taken by the author of thesis) 6
Figure 5. *Plukenetia volubilis* (Mouré, 1967): A – habit (ca x ½); B – male flower (ca x 8); c = female flower (ca x 2) 8
Figure 6. The map of Peru pointing San Martín Region (Perry-Castañeda Library, 2016) 20
Figure 7. The samples collection sites in San Martín Region, Peru.(Google Earth, 2015) 21
Figure 8. Neighbor-joining (NJ) based analysis of 173 individuals of *P. volubilis* using Jaccard's dissimilarity coefficient 33
Figure 9. Three-dimensional principal coordinate analysis (PCoA) of the genetic data based on Nei's genetic coefficients for tested 173 individuals of *Plukenetia volubilis* marking four distinguished compact populations 34
Figure 10. Distribution of crude protein content in sample set. 38
Figure 11. PLS calibration model for prediction protein content in collection of sacha inchi seeds 39
Figure 12. Electroforeograms of classical sodium dodecyl sulphate – polyacrylamide gel electrophoresis patterns of total seed protein (bulked samples) extracted from sacha inchi seeds 40
Figure 13. Electroforeograms of classical sodium dodecyl sulphate – polyacrylamide gel electrophoresis patterns of albumins+globulins (bulked samples) extracted from sacha inchi seeds 40
Figure 14. Electroforeograms of classical sodium dodecyl sulphate – polyacrylamide gel electrophoresis patterns of prolamins (bulked samples) extracted from sacha inchi seeds 41

List of Tables

Table 1. Amino acid profile of sacha inchi compared to protein[a,b] of other oilcrops cultivated in Southern America (Hamaker et al., 1992) 9
Table 2. Chemical composition of the sacha inchi seeds and physiochemical properties of their crude oil (Gutiérrez et al., 2011). 10
Table 3. Fatty Acid profile of sacha inchi in comparison with other oil crops cultivated in Southern America (Hamaker et al., 1992). 11
Table 4. Comparison of some molecular marker systems using for plant genome analysis (modified from Farooq and Azam, 2002; Harris, 2003; Semagn et al., 2006; Park et al., 2009; Abdel-Mawgood, 2012) in thesis by Dvořáková (2014). 14
Table 5. Molecular markers used in Euphorbiaceae family 15
Table 6. Sampled location in the Peruvian Amazon with basic specifications 22
Table 7. ISSR primers selected for optimization of ISSR screening 25
Table 8. Results for ISSR markers used for screening of *Plukenetia volubilis* 30
Table 9. Measures of genetic diversity in the 10 populations of *P. volubilis.* 31
Table 10. Genetic distances among investigated populations 31
Table 11. Cochran, Hartley, Bartlett test of the homogeneity of variances for total protein content, albumins and globulins, gliadin, and glutenins 35
Table 12. Overall results of the protein and protein fractions analysis (%). 36
Table 13. Quality evaluation of calibration model 38

List of Annexes

Annex 1. The primary data from the laboratory measurements of the total protein and protein fractions contents in the dry matter of the seeds. 66

TABLE OF CONTENT

1. Introduction 1
2. Literature review 3
2.1 Genus *Plukenetia*'s origin, distribution and taxonomy 3
2.2 Botanical description of *P. volubilis* species 5
2.3 Nutritional properties of *P. volubilis* and its uses 8
2.4 Cultivation of *Plukenetia volubilis* 11
2.5 Genetic diversity 12
3. Hypotheses and objectives 19
4. Materials and methods 20
4.1 Study area 20
4.2 Plant material collection 21
4.3 DNA analysis 23
4.3.1 DNA extraction 23
4.3.2 ISSR analysis 24
4.3.3 Statistical analysis of ISSR analysis 26
4.4 Protein and protein fractions analysis 26
4.4.1 Total protein analyses – Kjeldahl method and FT NIRS 26
4.4.2 Protein fractions analyses – Osborne's method 27
4.4.3 Statistical analysis of total proteins and protein fractions contents 28
4.4.4 Protein solubilization 28
4.4.5 Protein separation by SDS-PAGE 28
5. Results 30
5.1 Molecular analysis 30
5.1.1 ISSR profile and analysis 30
5.1.2 Cluster analyses based on the ISSR genotyping profile 32
5.1.3 Principal coordinate analysis 34
5.2 Protein and protein fractions analysis 35
5.2.1 Total protein content analysis 36
5.2.2 Albumins-globulins fraction analysis 36
5.2.3 Gliadin fraction analysis 37
5.2.4 Glutenin fraction analysis 37
5.3 Near Infrared Spectroscopy 37
5.4 SDS PAGE analysis 39
6. Discussion 42

6.1 Genetic analysis 42
6.2 Protein and protein fraction analysis 45
7. Conclusion 49
8. References 50
9. Annexes 66
10. List of author's publications 71

1. Introduction

It has been claimed that plant breeding reduces genetic diversity in elite germplasm, which could seriously jeopardize the continued ability to improve crops (Reif et al. 2005). The genetic diversity is a key component for evolution and adaptation. It also plays an important role in the population fitness. The loss of genetic diversity is related to inbreeding, and inbreeding reduces reproductive fitness (Reed and Frankham, 2003). That is why we need to maintain populations with high genetic diversity, mainly of cultivated crops, as it provides possibilities for plant improvements.

The loss of genetic diversity was encountered in many cultivated crops. However in crops, whose breeding is just at the beginning, we can avoid the loss by preserving their natural and diverse populations. Such crop may also be *Plukenetia volubilis* L., a highly nutritious traditional food crop of the Peruvian Amazon. It has gained world's attention since the oil derived from it's seeds won the gold medal at the "World Edible Oil" competition in Paris in 2004 (Agroindustrias Amazónicas, 2006). Since then, it became subject of research in various institutions. This "peanut of the Incas", or sacha inchi is a native plant whose origins lays in the Peruvian Amazon and its potential revenue from cultivation could aid poor indigenous and mestizo communities to move out of poverty and improve the diets in the same time (Hamaker et al., 1992; Hofmeijer, 2010). Sacha inchi was noticed already in 1992 by Hamaker (1992), who described its nutritional properties.

Sacha inchi has seeds of a lenticular shape, which are rich in oil and proteins and contain heat-labile substances with a bitter taste, which can be removed by roasting. It has traditionally been consumed by the Indians of Peru. It was probably cultivated by the pre-Incas and the Incas because representations of this plant and of its fruits have been found in vessels in Inca tombs; although it has been in danger of extinction, projects are now underway-developed by several universities, industries, local institutions and farmers – to recover its cultivation (Guillén et al. 2003). Recently, the cultivation of this crop spreads not just across Peru, but abroad too. It is now cultivated in Cambodia, Thailand or China. The Amazon natives obtain flour and oil from sacha inchi seeds. These products are used in the preparation of various meals and beverages; roasted seeds either mixed with corn meal and peppers or alone, and cooked tender leaves are also consumed. However, this plant has rarely been studied, and its importance from the nutritional and functional point of view is still a subject of research (Guillén et al. 2003; Sathe et al., 2002).

Sacha inchi is not valued only for its importance in alimentation, culture or history, but also in its economical rentability. It is a potential economically efficient crop with great possibilities of industrialization and is going to be grown intensively but a very high genetic, morphological and phytochemical variation was observed (Arévalo, 1995), which may lead to need of further investigations. At this moment, agronomy assessment is in progress and genetic improvement is just at the beginning. Only a few publications about sacha inchi's genetics is available today (Corazon-Guivin et al., 2008; Corazon-Guivin et al., 2009; Rodríguez et al., 2010; Křivánková et al., 2012; Rodrigues et al., 2013 – Table 5) Therefore this study attempts to provide some additional knowledge about existing genetic diversity of *P. volubilis* in the country of its origin - Peru, because the knowledge of genetic relationship and variability is a useful tool to improve selection of appropriate genotypes for breeding.

The aim of this thesis was to assess the genetic variability and to analyse seed protein patterns and fractions in selected populations of *P. volubilis*, a species for which few genetic experiments were elaborated, in Peruvian Amazon, the genetic centre of that species. Mapping of genetic diversity and description protein polymorphism can help to recover the genetic relationship and variability important for future researches and improvements.

2. Literature review

2.1 Genus *Plukenetia*'s origin, distribution and taxonomy

The tribe *Plukenetieae* belongs to the subfamily *Acalyphoideae*, the largest and the least understood of the five euphorbiaceous subfamilies. It includes 13 genera distributed worldwide in the tropical and warm temperate regions. *Plukenetia* species in tropical America varies from 7 to 12 (Standley and Steyemark, 1949; Hutchinson, 1969) and to recent date there were reported 12 in Americas, 3 in Africa, 3 in Madagascar and 1 in Asia (Gillespie, 1993; Gillespie 2007). The non-american species belonging to genus *Plukenetia* are *P. africana* Sond., *P. conophora* Müll. Arg. and *P. procumbens* Prain. from Africa, *P. corniculata* Sm., from Southern Asia, *P. decidua* L.J.Gillespie, *P. ankaranensis* L.J.Gillespie and *P. madagascarariensis* Leandri. from Madagascar (Gillespie, 2007; Bisby et al., 2010). Many species are twining vines or lianas, which are unusual habits in the family, other species are erect herbs, shrubs, or rarely small trees. Although flowers are small and apetalous, floral morphology is diverse, particularly the style and adroecium. Another uncommon feature is the presence of stinting hairs in many species. Gillespie (1993) in her synopsis compared eleven Neotropical species, such as *P. serrata* (Vell.) L.J.Gillespie, *P. lehmanniana* (Pax & K.Hoffm.) Huft & Gillespie, *P. polyadenia* Müll. Arg., *P. brachybotrya* Müll. Arg., *P. multiglandulosa* Jabl., *P. loretensis* Ule, *P. penninervia* Müll. Arg., *P. verrucosa* Sm., *P. volubilis* L. and newly described *P. supraglandulosa* L.J.Gillespie and *P. stipellata* L.J.Gillespie. There were some new Neotropical species distinguished recently, such as *Plukenetia carolis-vegae* (Bussman et al., 2013) and *Plukenetia huayllabambana* Bussmann, C.Téllez & A.Glenn which has very large seeds with a high content of fatty acids (Bussman et al. 2009). Jiménez (2000) identified *P. carabiasiae* J.Jiménez Ram. endemic to Oaxaca, Mexico. Based on ISSR analysis, Rodríguez et al. (2010) suggested a new species originally identified as *P. volubilis*. This suggestion may be supported by Gillespie (1993) who pointed on *P. volubilis* accession from Cuzco at elevation 1600 – 2000 m.a.s.l. not matching entirely any other *Plukenetia* species.

Full taxonomic classification of *Plukenetia volubilis* according to nomenclature database of Missouri Botanical Garden and Macbride (1990) is following:

Kingdom: Plantae
Phyllum: Magnoliophyta
Class: Magnoliopsida
Order: Euphorbiales
Family: Euphorbiaceae
Subfamily: Acalyphoideae
Tribe: Plukenetieae
Genus: *Plukenetia*
Species: *Plukenetia volubilis* L.

The presence of *Plukenetia volubilis* in America was recorded in Bolivia, Brazil, Colombia, Costa Rica, Ecuador, French Guiana, Mexico, Panama, Peru, Suriname, Venezuela, Lesser Antilles and in West Indies (Macbride, 1951; Correa y Bernal 1992; Gillespie, 1993) as indicated in Figure 1.

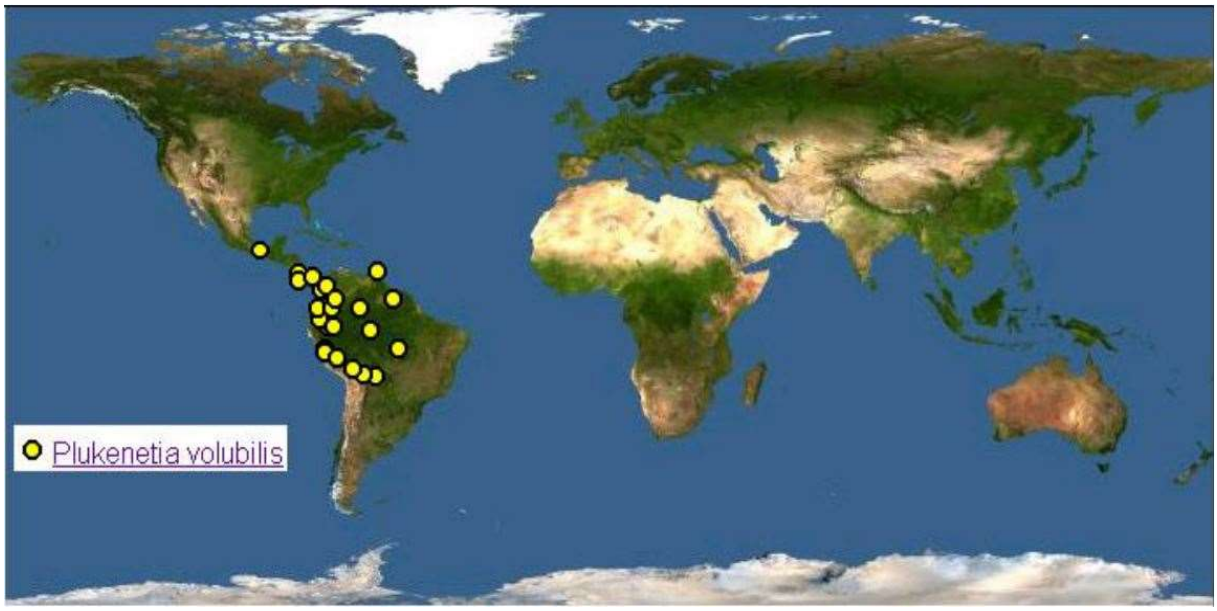

Figure 1. Distribution of *Plukenetia volubilis* in Southern America (Křivánková, 2012)

Its growth within Peru was recorded in Regions San Martín, Ucayali, Huanuco, Cuzco, Amazonas, Loreto y Madre de Dios (Figure 2). In San Martín it was recorded in nearly entire basin of river Huallaga, in the province of Lamas, Valle de Sisa, Alto Mayo and Bajo Mayo. It grows in elevations between 100 and 2 000 m. a.s. l. (Manco, 2006).

Figure 2. Map of regions of Peru, where sacha inchi is cultivated (Netmaps, 2016)

2.2 Botanical description of *P. volubilis* species

Sacha inchi is a vigorous semiwoody liana of indefinite length (Manco, 2006). Leaves are alternate with petiole 2.5-7.5 cm long. Blade is mebraneous, triangularovate 9-16 cm long and 6-9 cm wide. Apex is long-acuminate and base truncate to cordate, glabrescent below, 3-veined at base. The plant is basically monoecious twining vine (as can be seen in Figure 3 and 4) or slender liana with stems up to 5m long (Gillespie, 1993).

Figure 3. *Plukenetia volubilis* L. plant - its habit and fruits. (picture taken by the author of thesis)

Figure 4. *Plukenetia volubilis* is a liana which under cultivation needs support to avoid putrefaction of fruits. Plantations are usually intercropped (*Musa*, *Zea mays*, *Ananas comosus*, *Erythrina* spp., etc.) (picture taken by the author of thesis)

Inflorescence is axillary or terminal on short shoot, racemous 5-18 cm long. Female flowers are 1 or rarely 2 at base; male flowers are white, numerous in condensed cymes above (Gillespie, 1994). There are approximately 200 flower buds/male inflorescence, 23 anthers/flower bud, 4 nests/anther and 8 grains of pollen/nests. The polen grains have elongated shapes, rounded at the extremes, with a medial cross section that extends from extreme to extreme, turning from transparent color to crystalline (Noriega et al. 2010). Sacha inchi is a dichogamous plant. According to its flowers, it can be considered as pollinated by wind, but is visited by insects too (Cachique, 2006). *P. volubilis* differs from other species by a single glandular knob at the petiole apex, longer stylar column is the longest of any species of *Plukenetia* and may attain a length of 3 cm (15 – 30 mm) and staminate flowers with short (0.8 mm) slender-conical filaments and four sepals (Gillespie, 1993) (Figure 5). Cachique (2006) estimated allogamy is the most effective way of pollination in sacha inchi, but autogamy may occur too and it never forms fruit without pollination. The female flowers once mature open at 5 – 5:30 a.m. and keep open for 35 – 48 hours. It may be flowering for approximately 46 days if not pollinated sooner. Once pollinated the fruit needs approximately 105 days for maturation.

The principal defining synapomorphies of tribe *Plukenetia* are the four-carpellate ovaries and the associated character of four pistillate sepals (Gillespie, 1993). *P. volubilis* produces a tetralobular 3.5-4.5 cm large, glabrous, initially fleshy, becoming woody and dehiscent capsule with loculi that contain one seed each each with white cotyledons and a hard, nutlike seed coat (Hamaker et al., 1992), however several collections have unusually large capsules with five or six carpels, which Gillespie (1993) attributes to individuals under cultivation. Seeds are lenticular, broadly oblong in outline 1.3-2.1 cm in diameter, brown with course dark brown markings.

Figure 5. *Plukenetia volubilis* (Mouré, 1967): A – habit (ca x ½); B – male flower (ca x 8); c = female flower (ca x 2)

2.3 Nutritional properties of *P. volubilis* and its uses

The seeds collected in wild have long been a component of diets of the Chancas Indians and other tribal groups in the region. It is eaten either roasted or ground and mixed with maize meal and peppers. The protein content ranges between 24.7% and 27% (Gutiérez et al., 2011; Hamaker et al., 1992). The protein content of the defatted flour is about 53%. The sacha inchi protein, if completely digested, is deficient only in leucine and lysine. The amino acid profile was comparable to, and in some respect better than that of other oilseed in Andean region, such as peanut, soybean or cottonseed. The content of methionine, cysteine, tyrosine, threonine and tryptophan was higher; leucine and lysine was lower than other oilseed (Hamaker et al., 1992). Further amino acids comparison is available in Table 1. A water soluble storage albumin protein from sacha inchi seed accounted for 25% (w/w) of defatted seed flour weight, representing 31% of the total seed protein. The albumin fraction is mainly composed of a single storage protein that accounted for a substantial portion of total seed proteins. Sacha inchi seed has an estimated saccharide content of 4.8% ±0.92% (n=6). IPA (Inca Peanut Albumin) is a basic protein (p/ of 9.4) and contains all of the essential aminoacids in adequate amounts when compared to the FAO/WHO recommended pattern for a human adult. The trypthophan content of IPA is

unusually high (44 mg/g of protein), whereas the phenylalanine content is low (9 mg/g of protein). IPA is a highly digestible protein *in vitro* (Sathe et al. 2002).The comparison of individual oil crops in the mean of amino acids content can be seen in Table 1.

Table 1. Amino acid profile of sacha inchi compared to protein[a,b] of other oilcrops cultivated in Southern America (Hamaker et al., 1992)

Amino Acid	**Sacha inchi**	**Soybean**	**Peanut**	**Cottonseed**	**Sunflower**	**FAO/WHO/ UNU Scoring Pattern[c]**
Total Protein (%)	27	28	23	33	24	
Essential (mg/g of N)						
His	26	25	24	27	23	19
Ile	50	45	34	33	43	28
Leu	64	78	64	59	64	66
Lys	43	64	35	44	36	58
Met	12	13	12	13	19	-
Cys	25	13	13	16	15	-
Met + Cys	37	26	25	29	34	25
Phe	24	49	50	52	45	-
Tyr	55	31	39	29	19	-
Phe + Tyr	79	80	89	81	64	64
Thr	43	39	26	33	37	34
Trp	29	13	10	13	14	11
Val	40	48	42	46	51	35
Nonessential (mg/g of N)						
Ala	36	43	39	41	42	-
Arg	55	72	112	112	80	-
Asp	111	117	114	94	93	-
Glu	133	187	183	200	218	-
Gly	118	42	56	42	54	-
Pro	48	55	44	38	45	-
Ser	64	51	48	44	43	-
TEAA[d]	411	418	349	365	366	-
TAA[e]	976	985	945	936	941	-
TEAA as percent of TAA	42	42	37	39	39	-

[a] Values for soybean, peanut, cottonseed and sunflower were taken from Bodwell and Hopkins (1985).

[b] Values shown are miligrams/gram of protein, unless otherwise noted (N × 6.25).

[c] Recommended level for children of preschool age (2 – 5 years), alhough recently recommended for evaluation of dietary protein quality for all age groups except infants (Joint FAO/WHO Expert Consultation (1990).

[d] TEAA = Total Essential Amino Acids

[e] TAA = Total Amino Acids

The nut has a high lipid content (48.5%), while the shell contains only 1.2% of lipids (Pereira de Souza et al., 2013). The lipid fractionation of the sacha inchi oil yielded mainly neutral lipids (97.2%), and lower amounts of free fatty acids (1.2%) and phospholipids (0.8%). The physicochemical properties of the oil include: saponification number 185.2; iodine value 193.1; density 0.9187 g/cm3, refractive index 1.4791 and viscosity of 35.4 $mPa.s^{-1}$ (Guttiérez et al., 2011). Sacha inchi nut is an excellent source of essential fatty acids, it contains high amounts of tocopherols and presents anti-atherogenic, anti-thrombogenic and hypocholesterolenic effects (Pereira de Souza et al., 2013). The main minerals present in sacha inchi seeds were potassium (5563.5 ppm), magnesium (3210 ppm) and calcium (2406 ppm). A fatty acid analysis revealed that α-linolenic (50.8%) and linoleic (33.4%) acids were the main fatty acids in sacha inchi oil (Guttiérez et al., 2011). Mineral content and oil features are summarized in Table 2.

Table 2. Chemical composition of the sacha inchi seeds and physiochemical properties of their crude oil (Gutiérrez et al., 2011).

Component	Value
Seeds	
Moisture (%)	3.3 ± 0.3
Fat (%)	42.0 ± 1.1
Protein (%)	24.7 ± 0.5
Ash (%)	4.0 ± 0.7
Total carbohydrate (%)	30.9 ± 0.6
Potassium (mg/kg)	5,563.5 ± 6.4
Magnesium (mg/kg)	3,210.0 ± 21.2
Calcium (mg/kg)	2,406.0 ± 7.1
Iron (mg/kg)	103.5 ± 8.9
Zinc (mg/kg)	49.0 ± 1.1
Sodium (mg/kg)	15.4 ± 0.5
Copper (mg/kg)	12.9 ± 0.3
Crude Oil	
Iodine value (g l2/100g)	193.1 ± 1.0
Saponification value (mg KOH/g)	185.2 ± 0.5
Refractive index at 25°C	1.4791 ± 0.0009
Density at 25°C (g/cm)	0.9187 ± 0.02
Viscosity at 20°C (mPa×s)	35.4 ± 0.4

Values are means ± standard deviations of triplicate determinations

The oil obtained from sacha inchi seeds was studied by means of FTIR and 1H NMR (Fourier Transform Infrared Spectroscopy, and Proton nuclear magnetic resonance). It was found that sacha inchi oil has a high degree of unsaturation (Table 3). The same fact is deduced

from the ratio between the absorbance of the bands due to the stretching vibrations of the cis olefinic CH double bonds at 3015.5 cm^{-1} and to the methylene symmetrical stretching vibrations at 2855.1 cm^{-1} (Guillén et al. 2003).

Table 3. Fatty Acid profile of sacha inchi in comparison with other oil crops cultivated in Southern America (Hamaker et al., 1992).

Fatty Acid	Sacha Inchi	Soybean	Peanut	Cottonseed	Sunflower
Total Oil	54.0	19.0	45.0	16.0	48.0
Saturated					
$C_{14:0}$, Myristic	0.0	0.0	0.0	0.0	0.0
$C_{16:10}$, Palmitic	4.5	10.5	12.0	18.7	7.5
$C_{18:0}$, Stearic	3.2	3.2	2.2	2.4	5.3
Unsaturated					
$C_{16:0}$, Palmitoleic	0.0	0.0	0.3	0.6	0.0
$C_{18:0}$, Oleic	9.6	22.3	41.3	18.7	29.3
$C_{18:2}$, Linoleic	36.8	54.5	36.8	57.5	57.9
$C_{18:3}$, Linolenic	45.2	8.3	0.0	0.5	0.0
$C_{20:1}$, Gadoleic	0.0	0.0	1.1	0.0	0.0

* All values are shown in percents. Values for soybean, peanut, cottonseed and sunflower are taken from Bodwell and Hopkins (1985).

2.4 Cultivation of *Plukenetia volubilis*

The germination takes 11 – 14 days. The first true leaf appears after 16 – 20 days from sowing. Flowering starts after 86 – 139 days. The fruit appear in 119 – 182 days and the first harvest can be done in 202 – 249 days after planting. In can be cultivated on soils varying between loams and sandy soils and it tolerates acid soils. Sacha inchi can be sown directly to the soil or can be grown in the nursery. The best time for transplanting is within the rainy season between January and March. The major way of propagation is through seeds (Manco, 2006) but vegetative propagation is also possible (Cachique, 2006). Pesticide treatment of the seeds is recommended.

For one hectare plantation establishment the amount of 1 – 1.5 kg of seeds is suggested amount, recommended spacing is 3 m between the plants and 2.5 – 3 m between the rows (1,111 – 1,333 plants per hectare). Only one seed is sown to one hole and recommended depth is 2 – 3 cm. When growing in the nurseries, transplanting should be done after two months from sowing, which should be done between during the rainy season. Since sacha inchi is a vine, the supporting constructions should be constructed prior to sowing within the land preparation in order to avoid molting of the fruits later (Manco, 2006).

Since the spacing is wide, erosion may occur, to avoid it, intercropping with cover crops, such as *Indigofera* spp., *Arachis* spp. or *Desmodium* spp. can be beneficial, sacha inchi is sometimes also intercropped with cotton, bean, maize, manioc or peanuts (Manco, 2006). Also trees can be implemented instead of buttresses. Traditional tree for this purpose is *Erythrina* spp. (Cachique, 2006).

Sacha inchi is susceptible to diseases like *Fusarium* spp. or pests like *Meloidogyne* spp. the risk can be reduced by pruning, which is also very beneficial in terms of aeration of the growth, increasing the amount of light reaching the leaves and giving the plants shape making harvesting easier. However also fungicides against *Fusarium* and nematicides against *Meloidogyne* are recommended (Manco, 2006).

The plantations are the most productive until their ages reach 10 years, however they may be maintained for 20 years. The harvest is carried out manually every 15 – 30 days once the plants started tu fruit. The yield can reach 0.7 – 2 t/ha after the first year. Harvested fruits should be dried and are usually stored in jute bags in dry environments (Manco, 2006).

2.5 Genetic diversity

In 1992 the first molecular marker method was used in the Euphorbiaceae family for estimation of genetic diversity by Lefévre and Charrier. According to the Table 5 the most investigated genera within Euphorbiaceae are *Jatropha* and *Manihot*. Cassava, as crop cultivated for human consumption, is understandable to have such broad number of investigation, however physic nut is not just edible, it is toxic for humans. Nevertheless, it may serve as source of biodiesel, which may explain the numerous papers. But *Jatropha* and *Ricinus* have similar pollination system as *Plukenetia*, can be both allogamous and autogamous, are monoecious but with unisexual flowers (Meinders and Jones, 1950; Brigham, 1967; Jubera et al., 2009). Even the rubber tree has less genetic studies than physic nut. The first known investigation of genetic diversity in *Plukenetia* comes from 2008. Based on current literature sources, the most used methods witin Euphorbiaceae family according to our findings, are SSR (48 studies), RAPD (37 studies), AFLP (30 studies), ISSR (21 studies), SNP (15 studies), and Isozymes (10 studies). The methods were combined in some studies (Table 5).

The greatest degree of (morphological) variation in *P. volubilis* was found in collections from the eastern slopes of the Andes bordering the Amazon basin in Peru (Gillespie, 1993). Until now, the genetic diversity was not broadly studied. Some natural populations were studied by Corazon-Guivin et al. (2008) using DALP (Direct Amplification of Length Polymorphism)

markers in San Martín Region in Peruvian Amazon. They found the populations were structuralized and they attribute this fact to natural barriers disabling pollination (either through wind or insects) in greater distances. Corazon-Guivin et al. (2009) also studies accessions in National Germplasm Bank in INIA (National Institute of the Agrarian Inovations) using the informative markers discovered in the previous study and they found the highest diversity among the accessions from region San Martín, which was also a reason why we have chosen this area for the collection of samples. In 2010, Rodríguez et al. applied ISSR markers on a few species of *Plukenetia* and suggested a new species originally considered as *P. volubilis*. Křivánková et al. (2012) in her study had compared several cultivated populations in region Ucayali, Peru using ISSR markers and suggested *P. volubilis* to be an allogamous species based on their findings compared to other authors. Rodrigues (2013) had used AFLP to estimate genetic diversity in 60 accessions in Brazil and also found that individuals formed groups with other individuals of the same origin.

Since *P. volubilis* is not a major crop, it was neglected by the science for a long time (Křivánková et al., 2007) and the breeding was not documented. However, some attempts have occurred. Local farmers always sow the biggest seeds coming from the biggest fruits and eliminate the seedling which appeared independently (Danter Cachique, personal communication), which can be considered as breeding (positive selection). Gillespie (1993) noticed if the *P. volubilis* fruit were more than tetra-lobular, they were always coming from populations recorded as under cultivation. Since the flowers of sacha inchi are mostly pollinated by wind (Cachique, 2006), all fruits on the maternal plant do not need to have a common paternal plant. That is why the selection, local farmers do, does not always matter.

Molecular markers provide valuable data on diversity through their ability to detect variation at the DNA level (Somasundaram and Kalaiselvam, 2011). They offer numerous advantages over conventional phenotype-based alternatives as they are stable and detectable in all tissues regardless of growth, differentiation, development, or defense status of the cell, they are not confounded by the environment, or by pleiotropic and epistatic effects (Agarwal et al., 2008). For estimation of species diversity, it is essential that individuals can be classified accurately (Somasundaram and Kalaiselvam, 2011), however different markers have different suitability in studying genetic diversity (Abdel-Mawgood, 2012). Thus, the choice of a marker system is a compromise between the properties of the marker system and its availability. Further, marker choice must be based on the hypothesis that is being tested, the properties of the marker system, and the resources that are available for the research program (Lowe et al.,

2004). The DNA based marker systems, which are commonly used in genetic diversity are summarized in Table 4. We decided to use a dominant marker method - ISSR because this method possess many benefits. It is suitable for use in genetic diversity determination and populations differentiation. It does not require prior knowledge of the genome, is relatively cheap, is not very complex and has a high reproducibility. Also Křivánková et al. (2012) considered ISSR markers as suitable and adequately sensitive for polymorphism detection for this species.

Table 4. Comparison of some molecular marker systems using for plant genome analysis (modified from Farooq and Azam, 2002; Harris, 2003; Semagn et al., 2006; Park et al., 2009; Abdel-Mawgood, 2012) in thesis by Dvořáková (2014).

	RFLP	RAPD	AFLP	ISSR	SSR
Abundance	Medium	Very high	Very high	Very high	High
DNA Quality	High	Medium	High	Medium	Medium
DNA Sequence Information	Not required	Not required	Not required	Not required	Required
Level of Polymorphism	Medium	High	High	High	High
Inheritance	Co-dominance	Dominance	Dominance	Dominance	Co-dominance
Reproducibility	High	Low	Medium	High	High
Technical Complexity	High	Low	Medium	Low	Low
Development Costs	High	Low (none)	Low	Low	High
Major Applications	Genetic diversity, polyploidy, hybridization, phylogeny, mating system	Fingerptinting, genetic diversity, polyploidy, hybridization, phylogeny	Fingerprinting, genetic diversity, population differentiation	Genetic diversity, individual genotyping, population differentiation	Individual genotyping, gene flow, Population differentiation

Table 5. Molecular markers used in Euphorbiaceae family

Genus	Molecular marker method	Reference	Year
Macaranga	AFLP	Bänfer et al.	2004
Fontainea	cpDNA, nrDNA, RAPD	Rossetto et al.	2000
Mercurialis	Isozymes	Holderegger and Stehlik	1999
	ITS	Krähenbühl, et al.	2002
Excoecaria	ISSR	Zhang et al.	2005
	ISSR, RAPD	Das et al.	2011
Phyllanthus	ISSR	Upadhyay et al.	2015
	RAPD	Chaurasia et al.	2009
Chamaesyce	ITS, RAPD	Morden and Gregoritza	2006
Triadica	SSR	de Walt et al.	2011
Vernicia	SSR	Zhang et al.	2015
Hevea	AFLP, RFLP, SSR, isozymes	Lespinasse et al.	2000
	cpDNA, mDNA	Luo et al.	1995
	ILP	Li et al.	2013
	RAPD	Lam et al.	2009
	RAPD	Venkatachalam et al.	2004
	RFLP	Besse et al.	1994
	SSR	Mantello et al.	2012
		Le Guen et al.	2009
		Souza et al.	2009
Jatropha	morphology	Bombonato de Oliveira et al.	2015
	454 pyrosequencing, SNP, EST-SSR	Laosatit et al.	2015
	AFLP	Sinha et al.	2016
		Pioto et al.	2015
		Shen et al.	2012
	AFLP, RAPD, SSR	Mastan et al.	2012
	AFLP, SSR	Sinha et al.	2015
	cp SSR, ISSR, RAPD	Basha and Sujatha	2009
	cpDNA, genome sequencing	Shrikant et al.	2010
	EST-SSR	Kumari et al.	2013
	EST-SSR	Hemant et al.	2011
	EST-SSR, genomic SSR	Mingfu et al.	2010
	EST-SSR, ILP	Saisug and Ukoskit	2013
	genome sequencing	Sato et al.	2011
	ISSR	Mavuso et al.	2015
		Maghuly et al.	2011
		Grativol et al.	2011
		Tanya et al.	2011
		Vijayanand et al.	2009
		Kumar et al.	2009

Table 5 continues

Genus	Molecular marker method	Reference	Year
	MS-AFLP	Kanchanaketu et al.	2012
	nrDNA ITS	Pamidimarri et al.	2009 a
	nrDNA ITS, cpDNA	Guo et al.	2016
	RAPD	Dhakshanamoorthy et al.	2013
	RAPD	Rafii et al.	2012
	RAPD	Danquah et al.	2012
	RAPD	Kumar et al.	2009
	RAPD	Ram et al.	2008
	RAPD, AFLP	Pamidimarri et al.	2010 b
	RAPD, AFLP	Pamidimarri et al.	2009 c
	RAPD, AFLP, nrDNA-ITS	Pamidimarri and Reddy	2014
	RAPD, AFLP, SSR	Pamidimarri et al.	2009 d
	RAPF, ISSR, SCAR	Basha and Sujatha	2007
	RBIP, FISH	Alipour et al.	2013
	SNP	Gupta et al.	2012
	SNP, SSR	Ricci et al.	2012
	SSR	Siju et al.	2015
		Maurya et al.	2015
		Raposo et al.	2014
		Ouattara et al.	2014
		Bressan et al.	2013
		Na-ek et al.	2011
		Phumichai et al.	2011
		Pamidimarri et al.	2010 e
		Sirithunya and Ukoskit	2010
		Sudheer et al.	2010
	SSR, TRAP, AFLP	Montes Osorio et al.	2014
	morphology	Tripathi et al.	2013
Manihot	454-sequencing, EST-SNP, SSR	Prochnik et al.	2012
	AFLP	Fregene et. al.	2000
	AFLP	Wong et al.	1999
	AFLP	Roa et al.	1997
	cDNA	Chacón et al.	2008
	cDNA, SNP	Lopez et al.	2005
	cpDNA, rDNA	Fregene et. al.	1994
	DArT	Xia et al.	2005
	ESTs	Fregene at. al.	2001
	genome sequencing	Daniell et al.	2008
	IRTAP	Oliveira-Silva et al.	2014
	Isozymes	Zaldivar et al.	2004

Table 5 continues

Genus	Molercular marker method	Reference	Year
		Sumarani et al.	2004
		Montarroyos et al.	2003
		Resende et al.	2000
		Brondani	1996
		Lefévre and Charrier	1992
	nDNA	Olsen	2002
	QTL	Cortés et al.	2002
	RAPD	Asante and Offei	2003
	RAPD	Marmey et al.	1993
	RAPD, SSR	Castelo Branco Carvalho and Schaal	2001
	RFLP, RAPD	Fregene et. al.	1997
	SNP	Soto et al.	2015
		de Oliveira et al.	2014
	SNP, SSR	Kawuki et al.	2009
	SNP, SSR	Olsen	2004
	SSR	Moura et al.	2016
		Carrasco et al.	2016
		Alves-Pereira et al.	2011
		de Bang et al.	2011
		Siqueira et al.	2010
		Kizito et al.	2007
		Peroni et al.	2007
		Lokko et al.	2006
		Okogbenin et al.	2006
		Elias et al.	2004
		Fregene at. al.	2003
		Mba et al.	2001
		Chavarriaga-Aguirre et al.	1998
	SSR, Isozymes, AFLP	Chavarriaga-Aguirre et al.	1999
	TRAP	Carmo et al.	2015
	morphology	Agre et al.	2015
		Manu-Aduening et al.	2013
		Kombo et al.	2012
Plukenetia	cDNA	Wang et al.	2012
	DALP	Corazon-Guivin et al.	2009
	DALP	Corazon-Guivin et al.	2008
	ISSR	Křivánková et al.	2012
		Rodríguez et al.	2010
	AFLP	Rodrigues et al.	2013
Ricinus	SSR	Meilian el. al.	2014
		Kyoung-In et al.	2011
		Bajay et al.	2009

Table 5 continues

Genus	Molercular marker method	Reference	Year
	cDNA, mtDNA	Rivarola et al.	2011
	SNP	Foster et al.	2010

3. Hypotheses and objectives

Taking into consideration that PCR-based markers have already revealed close relationships within many plant species with exact similarity specification, there is a hypothesis that ISSR markers could reveal the differences among individual locations of *P. volubilis* growths. It can be expected that the variability among individual locations may be high as Peruvian Amazon is believed to be the place of origin of *P. volubilis*. Taking into account that the locations are sort of isolated there is a hypothesis that samples from individual location would be similar and therefore would tend to group together. The hypotheses are:

A) There is a high genetic diversity in its region of origin, mainly within the populations;

B) The genetic distance will be related to the geographical one;

C) Populations differing genetically will differ also in the proteins contents.

The main aims of this thesis are presented in the following paragraphs:

I) Assessment of genetic diversity within and among populations of *Plukenetia volubilis* L. using ISSR markers
   a. Optimization of DNA extraction through CTAB method for sacha inchi
   b. Optimization of ISSR protocol for sacha inchi

II) Determination of total seed protein content and content of protein fractions in collected samples of *Plukenetia volubilis* L.
   a) Analysis of total protein and protein fractions content in seeds using Kjeldahl's and Osbornes's methods
   b) Evaluate the potential of FT NIRS for estimation of protein content in sacha inchi seeds
   c) Analysis of seed storage protein pattern in sacha inchi using SDS-PAGE method

4. Materials and methods

4.1 Study area

The plant material analysed in this study was collected from individual plants growing in different regions of Peruvian Amazon. In cooperation with Instituto de las Investigaciones de la Amazonía Peruana (Peruvian Amazon Research Institute - IIAP) we have selected 10 locations from the provinces of San Martín, Lamas and Dorado in region San Martín, which are located in northern part of Peruvian Amazon (Figure 6), which were the locations of interest for IIAP. Location Santa Lucia (SLU) is located close to Tingo María (Huánaco Region) and is not indicated in the map. These locations' management ranged from semi wild growth to intensive plantation.

Figure 6. The map of Peru pointing San Martín Region (Perry-Castañeda Library, 2016)

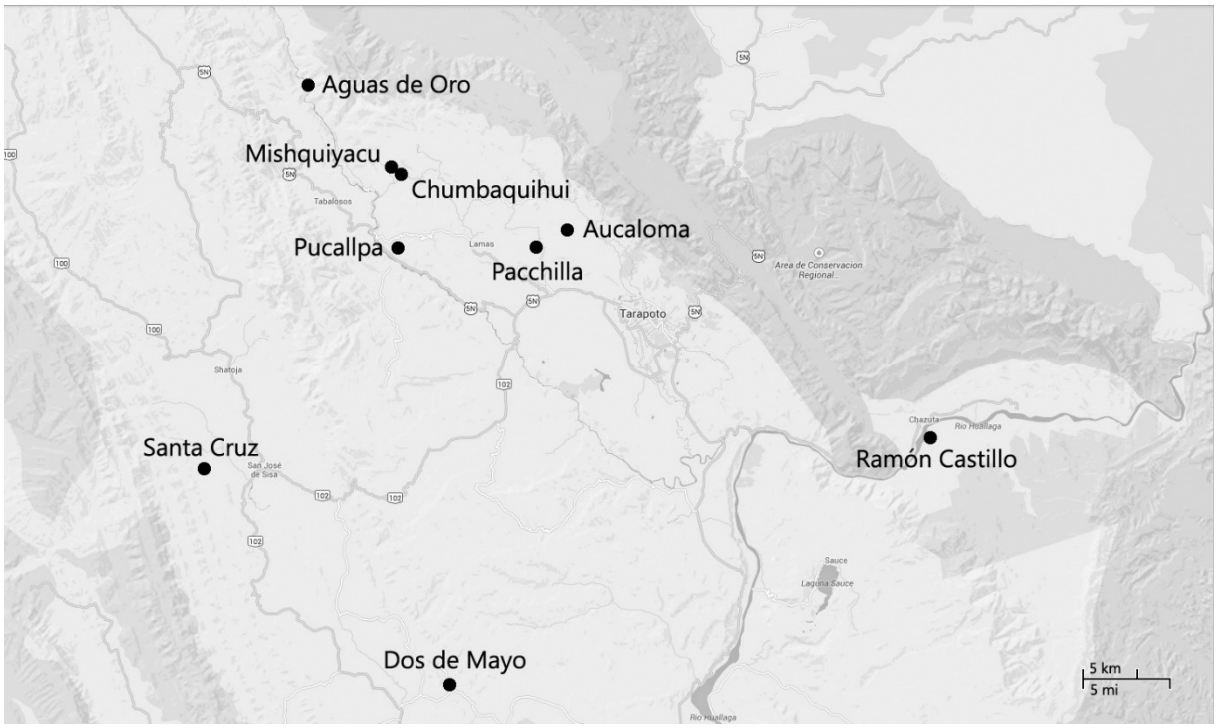

Figure 7. The samples collection sites in San Martín Region, Peru.(Google Earth, 2015)

4.2 Plant material collection

The plants from these provenances (Figure 7) were randomly selected for the sample collection of the total 173 individual plants (Table 6). Between 2 and 23 individuals (according to the availability) plants were selected, leaves and seed samples taken from each location. Collected leafy material of each individual plant in size of approximately 3 cm^2 was immediately stored in plastic tubes with 6 ml of silica gel (Carl Roth GmbH, Germany), in order to dry and preserve the samples. The meristematic vigorous tops with young leaves were selected and taken. Samples were then transported to Czech University of Life Sciences Prague (CULS) for further DNA analysis. Simultaneously with the leafy material, approximately 20 seeds from every plant were collected for protein and protein fractions analyses. The capsules were crushed in order to prevent the fungal attack during storage and the seeds were photographed with a linear scale to register the size. Unfortunatelly, not all of the selected plants in Ramón Castillo had mature fruits and seeds in the time of the collection, therefore samples RAC05 and RAC06 lacked the seeds. The leafy samples were done in two specimens, one was taken to CULS Prague, and the other remained in IIAP's office in Tarapoto, Peru.

Table 6. Sampled location in the Peruvian Amazon with basic specifications

No.of population	Name of population	No. of individuals	Coordinates and altitude	Description of the site	Weight of hundred seeds (WHS) in g
1	Mishquiyacu (MIS)	20	6°21,673' S 76°34,998' W 470 masl, southern orientation.	Approximately 400 m^2 plot close to a small village, about 200 plants in the plot, cultivated by local villagers	139.74
2	Chumbaquihui (CHU)	20	6°21,991' S 76°34,504' W 364 masl, eastern orientation.	Close to Mishquiyacu, approx. 200 plants on 400 square metres, cultivated together with banana plants by locals	177.78
3	Dos de Mayo (2DM)	22	6°47,573' S 76°32,108' W 335 masl, southern-western orientation	A population distant from the others, recognizable for its small seeds. Approximatelly 350 plants cultivated at a slope in the edge of the forest on an area of 1,000 m^2	91.29
4	Pucallpa (PUC)	21	6°25,676' S 76°34,689' W 455 masl, southern-eastern orientation	Cultivation on the top of the ridge, area of 1.5 ha with irregular spacing of plants (approximatelly 1,200) of a significantly shrubby growth	117.3
5	Aucaloma (AUC)	18	6°24,816' S 76°26,143' W 740 masl, southern-eastern orientation	A small plot (300 m^2) close to Pacchila plantation and intercropped together with banana and maize plants. A hundred plants in total.	120.14
6	Pacchilla (PAC)	20	6°25,694' S 76°27,729' W 703 masl, southern-western orientation.	Commercial plantation with an area of 15 ha, spacing 3 m x 3 m. Approximatelly 16,500 plants in total.	131.08
7	Ramón Castillo (RAC)	7	6°35,244' S 76°07,884' W 210 masl.	Several (7) plants scattered in the area of an abandoned sacha inchi field near the Huallaga River	102.11
8	Santa Cruz (SCR)	23	6°36,803' S 76°44,452' W 425 masl, southern orientation.	A garden on the edge of the village with robust plants cultivated in rows on wires. There were 150 plants on 0.1 ha.	131.8
9	Aguas de Oro (ADO)	20	6°17,570' S 76°39,200' W 385 masl, southern-western orientation.	Quite isolated cultivated population in the valley of Huallaga River, the seeds were transported from Churuzapa village. Approximatelly 300 plants on 0.4 ha.	126.56
10	Santa Lucía (SLU)	2	9°07,622' S 76°01,040' W 562 masl	Only two plants were encountered in this location several kilometres from Tingo María. Both were grown binding a tree and seemed as an old cultivation	131.91
Total number of individuals		173			

4.3 DNA analysis

4.3.1 DNA extraction

The DNA was extracted using CTAB method (Doyle and Doyle, 1987) with minor modifications according to Williams et al. (1992). From every sample, approximately 100 mg of dried leaves were put into individual 1.5 ml micro tubes (Eppendorf, Germany) and ground by plastic micro pestles until smooth powder was made. To every tube, 495 µl of 2% CTAB and 5 µl of 1% mercaptoethanol was added. The microtubes were shaked manually and placed into heating nest for 45 min, and were shaked three times (every 12 min) while heated to 65 °C. Samples were then centrifuged for 10 min at room temperature at 12,000 rpm. The supernatants were transferred to new 1.5 ml micro tubes, 500 µl of solution of chloroform and isoamyl alcohol (24:1) was added and shaked together for 10 min, then centrifuged for 5 min at room temperature at 12,000 rpm. There were three phases visible at this moment. The watery top one, where DNA was located, was transferred to the new 1.5 ml micro tubes and mixed together with 100 µl of 5% CTAB solution. Remaining two phases of undesired substances were discarded. After that, 500 µl of chloroform and isoamyl alcohol solution (24:1) was added and shaked together for 10 min, after shaking, the samples were centrifuged for 5 min at room temperature at 12,000 rpm. The top watery phase was transferred to new 1.5 ml mico tubes, 200 µl of ice cold isopropanol was added and the tubes were slowly overturned three times and put into the freezer (-20 °C) overnight. The other day, the centrifuge was cooled to 4 °C and the samples were centrifuged there for 5 min at 12,000 rpm. At this phase, the supernatants were removed and 300 µl of 1% TE buffer and 30 µl of sodium acetate were added to the pellets in order to dissolve them while shaking for 60 min at 37 °C. After shaking, 600 µl of ice cold 96% ethanol was added to the micro tubes which then were overturned three times and put to the freezer for 2 hours at -20 °C. Once again the centrifuge was cooled to 4 °C where the samples were then centrifuged for 10 min at 12,000 rpm. The supernatant was removed and 1000 µl of 70% ethanol was added instead, shaked gently and centrifuged at 4 °C for 2 min at 12,000 rpm. This step was repeated twice. Then the supernatant was removed and the pellets were left to dry in the heating nest at 45 °C for 20 min. Once dried, 100 µl of sterile and distilled H_2O was added and the pellet left to dissolve in the room temperature. The DNA quality was determined by 0.8% agarose gel electrophoresis and using a Nanodrop Spectrophotometer (Thermo Scientific, USA). The final concentration of half of the amount of isolated DNA of all samples was adjusted to 50 ng.μl^{-1} for PCR, and stored at -20°C. The other half of isolated DNA was stored in the freezer as a backup.

4.3.2 ISSR analysis

A set of 30 ISSR primers (University of British Columbia, Vancouver, Canada) was used for screening (Table 7). PCR amplification reactions were carried out in a total volume of 20 µl containing 0.5 µl of each primer, 10 µl PPP Master Mix (Top-Bio, Czech Republic), 0.2 µl of BSA - Bovine Serum Albumine (Thermo Scientific, Lithuania), 7.3 µl PCR Water (Top-Bio, Czech Republic) and 2.0 µl of individual samples. PCR amplification was performed in T100TM Thermal Cycler (Bio-Rad Laboratories, USA). The annealing temperatures in PCR were optimized for each primer (Table 7) using samples from ADO population as model. The cycling conditions were as follows: initial denaturation for 5 min at 94°C, followed by 40 cycles of denaturation for 1 min at 95°C, 1 min at specific annealing temperature in a range from 48°C to 55°C (according to primer), 2 min at 72°C (extension), these 40 cycles were afterwards followed by a final extension step for 10 min at 72°C. The PCR products were resolved in 2% agarose gels in 1xTBE buffer (Carl Roth GmbH, Germany) using the following programme: 180 minutes at 55 V and 120 mA. The gels were stained with ethidium bromide staining (Carl Roth GmbH, Germany) and the bands were visualized and acquired under UV light (Cleaver Scientific, UK). The size of the amplified products was estimated using 100 bp Plus DNA ladder (Thermo Scientific, Lithuania).

Prior to testing all samples, a population ADO (Aguas de Oro) was used as the model for annealing temperatures optimization, and for determination, which primers were polymorphic. The optimal annealing temperatures are listed in Table 7. Once the electrophoresis was completed, the gel was photographed using UV camera. The visible bands were recorded as presence (1) or absence (0) of the amplified fragments and entered in this way into excel (Microsoft Corporation, USA) version 2013. The matrix where every column was one sample and every line was one position on the gel, was created.

Table 7. ISSR primers selected for optimization of ISSR screening

	Primer	Sequence 5′- 3′	Annealing temperature (°C)
1	UCB 807	(AG) × 8 + T	48
2	**UCB 809**	**(AG) × 8 + G**	**50**
3	UCB 810	(GA) × 8 + T	49
4	UCB 812	(GA) × 8 + A	50
5	UCB 813	(CT) × 8 + T	50
6	UCB 814	(CT) × 8 + A	50
7	UCB 823	(TC) × 8 + C	54
8	**UCB 824**	**(AG) × 8 + YT**	**55**
9	**UCB 826**	**(AC) × 8 + C**	**48**
10	UCB 828	(TG) × 8 + A	48
11	UCB 829	(TG) × 8 + C	50
12	UCB 834	(AG) × 8 + YT	50
13	UCB 835	(AG) × 8 + YC	51
14	**UCB 836**	**(AG) × 8 + YA**	**48**
15	UCB 840	(GA) × 8 + YT	52
16	UCB 841	(GA) × 8 + YCY	48
17	UCB 843	(CT) × 8 + RA	48
18	**UCB 844**	**(CT) × 8 + RC**	**52**
19	**UCB 845**	**(CT) × 8 + RG**	**54**
20	UCB 846	(CA) × 8 + RT	48
21	**UCB 847**	**(CA) × 8 + RC**	**54**
22	UCB 848	(CA) × 8 + RG	48
23	UCB 851	(GT) × 8 + CT G	52
24	UCB 854	(TC) × 8 + RG	52
25	UCB 855	(AC) × 8 + YT	54
26	UCB 856	(AC) × 8 + YA	49
27	**UCB 859**	**(TG) × 8 + RC**	**50**
28	UCB 866	(CTC) × 6	50
29	UCB 873	(GACA) × 4	48
30	UCB 876	GATA GATA GACA GACA	50

Y=pyrimidines: cytosine or thymine; R=purines: adenine or guanine (Zietkiewicz et al., 1994), primers marked with bold letter were polymorphic

4.3.3 Statistical analysis of ISSR analysis

All fragments obtained from ISSR analysis were scored for presence (1) or absence (0) of homologous bands. A final binary matrix was created by assembling all resulting bands for each accession. Values obtained from scoring the ISSR data were used for the construction of a dissimilarity matrix by applying Dice's coefficient (Dice, 1945). Following formula was employed: $d_{ij} = (b+c)/2a+(b+c)$, where a represents number of variables where x_i is present and x_j is present; b stands for variables where x_i is present and x_j in absent; and finally c is the number of variables where x_i is absent and x_j is present. The d_{ij} means the dissimilarity between units i and j. Final dendrogram was constructed using a hierarchical clustering by Neighbour Joining (NJ). Data analysis was performed using DARwin5 software (Perrier and Jacquemoud-Collet, 2006). Shannon's index (*I,* LogBase=e) was estimated by FAMD 1.25 software for all accessions according to Hutchenson (1970) and normalised according to Ramezani (2012) (**Equation 1**). The percentage of polymorphic bands (PPB) and Nei's genetic distance (Nei, 1972 and 1979) matrix were calculated using FAMD software, version 1.25 (Schlüter and Harris, 2006). **Equation 1**: Shannon's diversity index. I is the Shannon's index, $_{pi}$ is the frequency of band presences in locus i, s is the number of loci and ln is the natural logarithms.

$$I \approx -\frac{\sum_{i=1}^{s} p_i \ln p_i}{\ln(s)}$$

The Principal Coordinates Analysis (PCoA) was performed by software DARwin 5.0.160 (Perrier and Jacquemoud-Collet, 2006) applying the Data obtained while calculating Dice's coefficient.

4.4 Protein and protein fractions analysis

4.4.1 Total protein analyses – Kjeldahl method and FT NIRS

Sample preparation

In average, eight seeds (10g) from every collected individual plant were crushed in a grinding mill (IKA A11 basic, IKA® Werke GMBH & Co.KG, Germany) to coarseness 0.8 mm. Similarly, ten seeds from each genotype (population) were selected randomly and then crushed and mixed together to form bulked samples (= laboratory test sample that is representative of all the specimens, each of at least 1 g, that are required).

Kjeldahl analysis

The dry matter content of seed samples (5 g of sacha inchi flour) was further dried in an electric hot-air drier at 133°C for 4h according to the standard method CSN EN ISO 662 (Czech State Norm, 2001). . the mineralized samples were ready for the proper estimation of total nitrogen on automatic analyser Kjeltec 2300 (Foss Analytical, Denmark) and calculated with conversion factor 6.25(Czech State Norm, 2012; FAO, 2002).

FT-NIR analysis

The results of Kjeldahl method were used for calibration of the device FT-NIR spectrophotometer (Antaris II, ThermoElectron Corporation, Madison, USA) in which approximately 1g of flour of each sample was analyzed according to methodology by Míka et al., 2008).

4.4.2 Protein fractions analyses – Osborne's method

Osborne fractionation

Seed flour from individual plant and /or bulked samples was defatted by washing 1 g of flour with 2.5 ml of hexane at 25°C overnight. The hexane fraction was discarded and the flour was lyophilized in aliquots and stored in a cold (4°C), dark, place. The defatted samples were subjected to Osborne fractionationed and/or SDS-PAGE.

The protein fractions were estimated using Osborne's method (1924) optimized for wheat (*Triticum aestivum* L.) by Dvořáček et al. (2001).

In this analysis, 1g of the flour of each sample was put into the 15 ml tube. Every sample was done in three specimens. The albumin-globulin fraction was being extracted for 15 min at 4°C in 5ml of 0.5M NaCl and then by centrifugation at 6,500 rpm for 15 min. Consequently, the specimens were "washed" with the same amount of the same extractant and centrifuged at 6,500 rpm for 5 min. The supernatants were merged together into the new 15 ml tube and stored in the refrigerator.

The gliadin fraction was obtained by 4 hours long extraction with 5ml of 60% ethanol at laboratory room temperature. This period was followed by 15 minutes of centrifugation at 6,500 rpm, two "washes" with 5 ml of 60% ethanol and final centrifugation at 6,500 rpm for 5 minutes.

Tubes containing protein fraction extracts and seeds pellets (glutelins) were freezed and lyophilized in a freeze dryer (Christ, Germany) for 24 h at 58°C and 0.018 mBar and stored in 4°C until analysis by SDS-PAGE was carried out.

The glutenin fraction was estimated by calculation. It was calculated as the difference between the total protein content and the amount of the two extracted fractions.

4.4.3 Statistical analysis of total proteins and protein fractions contents

From the laboratory measurements according to chapters 4.4.1 and 4.4.2 the primary data were obtained: measured values of nitrogen subsances in the dry matter (%) and contents of albumin-globulin fraction, gliadin fraction and glutenin fraction (by recalculation) – percentual share in the dry matter of the ground seeds of *P. volubilis* (all data available in Annex 1). One way ANOVA was used for total protein and glutenin contents, Kruskal-Walis test was used for albumins+globulins and gliadin contents. The set contains 169 variables from a total amount of 9 groups (populations) of different sizes. The number of individuals in the protein analyses is lower than in molecular analysis, because not all of the collected plants had mature fruits and seeds in the time of the collection (RAC05 and RAC06) Samples from Santa Lucia were not included in the statistics as they were only two from that location. There are two types of variables in this method – the dependent variable and independent variable. In this case, the independent variable is the location/population and the dependent variable (depending on the location) is the measured value (total protein, albumin-globulin fraction, gliadin fraction and glutenin fraction).

4.4.4 Protein solubilization

The lyophilized solid samples were mixed with 100 ml of extraction solution (0.0625M TrisHCl pH 8.8, 5% (w/v) 2-mercaptoethanol, 2% (w/v) SDS, 10% (w/v) glycerol, 0.01% (w/v) bromphenolblue) by vortexing several times in 1.5 ml tubes. The tubes were let stand at 4°C for 3 h. After this extraction time, the tubes were centrifuged at 12,000 g for 15 min and the supernatants were heated in boiling water for 2 min.

4.4.5 Protein separation by SDS-PAGE

The electrophoresis (SDS-PAGE) was carried out according to Laemmli (1970) with some minor changes and performed with the Hoefer SE 600 vertical unit (Hoefer, USA). The polyacrylamide gel preparation (180x160x0.75 mm, 10% (w/v) resolving gel, 4% (w/v) stacking gel; electrode Tris-glycine-SDS buffer of pH 8.3). The run was carried out at 45 mA

per gel until the bromphenolblue moved to the bottom of the gel (about 3 h). The gels were stained with a solution of 0.1% (w/v) Coomasie Brilliant Blue (CBB) R250, 50% (w/v) methanol, 10% (w/v) acetic acid, 0.02% (w/v) bromphenol blue salt. Destaining was performed with a solution of 25% (w/v) denatured alcohol and 3.5% (w/v) acetic acid. The gels were soaked in a solution of: 45% (w/v) denatured alcohol and 3% (w/v) glycerol then dried and stored between cellophane sheets.

5. Results

5.1 Molecular analysis

5.1.1 ISSR profile and analysis

In this study, a total amount of 30 ISSR primers were used for the screening of Sacha inchi populations. Eight of these ISSR markers showed clear and reproducible polymorphic bands. These 8 primers were then used to analyse the genetic diversity of all 173 accessions. A total amount of 97 fragments, ranging from 250 to 2500 bp, were amplified with a mean of 12.1 bands per primer, of which 90 (91.4%) were polymorphic (Table 8). Absolute polymorphism was observed at primers UBC836, UBC844, UBC845, and UBC847, while the lowest polymorphism was observed at primer UBC809 (69.2%). The highest number of polymorphic bands (18) was amplified by the primer UBC824, while the lowest number of polymorphic bands (5) was detected at the primer UBC826. The highest Nei's gene diversity (0.34) and Shannon information index (0.513) was exhibited by primer 844. In contrast, primer 809 showed the lowest Nei's diversity and Shannon information index with 0.081 and 0.147 values, respectively. The mean Nei's gene diversity and Shannon information index for all primers were 0.260 and 0.405, respectively. Gene diversity ranged from 0.103 for Santa Lucia to 0.238 for Santa Cruz with a mean of 0.18, and the same pattern was observed for the Shannon information index which ranged from 0.15 for Santa Lucia (SLU) to 0.357 for Santa Cruz (SCR) with mean of 0.280. The populations of Aucaloma (AUC), Pucallpa (PUC), Pacchilla (PAC), and Aguas de Oro (ADO) had gene diversity close to the mean of 0.18.

Table 8. Results for ISSR markers used for screening of *Plukenetia volubilis*

			Scored Bands		Diversity			
	Primer	**Sequence 5´- 3´**	**Band Size (bp)**	**Total Bands**	**NPB**	**PPB (%)**	**h + SD**	**I + SD**
1	UBC809	(AG)8 G	220-2,000	13	9	69.2	0.081 ± 0.106	0.147 ± 0.168
2	UBC824	(AG)8 YT	200-2,000	18	18	100	0.329 ± 0.127	0.499 ± 0.159
3	UBC826	(AC)8 C	700-2,000	7	5	71.4	0.102 ± 0.120	0.181 ± 0.183
4	UBC836	(AG)8 YA	300-1,500	9	9	100	0.250 ± 0.148	0.400 ± 0.190
5	UBC844	(CT)8 RC	300-2,500	11	11	100	0.340 ± 0.129	0.513 ± 0.153
6	UBC845	(CT)8 RG	300-1,500	13	13	100	0.234 ± 0.080	0.391 ± 0.107
7	UBC847	(CA)8 RC	400-2,000	15	15	100	0.335 ± 0.134	0.506 ± 0.161
8	UBC859	(TG)8 RC	500-2,500	11	10	90.9	0.312 ± 0.155	0.471 ± 0.204
	Total		200-2,500	97	90			
	Mean			12.125	11.3	91.4	0.260±0.158	0.405±0.313

h Nei's gene diversity, I Shannon's information index, NPB number of polymorphic bands, PPB percentage of polymorphic bands

Data for Nei´s genetic diversity (H) values ranged from 0.103 to 0.231 with a mean 0.180 and the Shannon´s information index (I) ranged from 0.150 to 0.357 with mean value 0.280. Analysis of molecular variance (AMOVA) indicated a total of 64% within population and 36% among population variation. The Nei´s gene differentiation, G_{st} was calculated to be 0.290 (Table 8) with gene flow Nm estimated at 1.227 (Table 9).

Table 9. Measures of genetic diversity in the 10 populations of *P. volubilis.*

Population	NPB	PPB	h + SD	I + SD	Gst*	Nm*
Dos de Mayo (2DM)	70	77.78	0.173 ± 0.157	0.280 ± 0.228	-	-
Aucaloma (AUC)	65	72.22	0.174 ± 0.162	0.278 ± 0.237	-	-
Santa Lucia (SLU)	24	26.67	0.103 ± 0.180	0.150 ± 0.262	-	-
Pucallpa (PUC)	59	65.56	0.181 ± 0.190	0.277 ± 0.271	-	-
Ramón Castillo (RAC)	45	50.00	0.131 ± 0.167	0.207 ± 0.247	-	-
Mishquiyacu (MIS)	65	72.22	0.213 ± 0.196	0.322 ± 0.276	-	-
Santa Cruz (SCR)	69	76.67	0.238 ± 0.201	0.357 ± 0.279	-	-
Pacchilla (PAC)	69	76.67	0.184 ± 0.174	0.292 ± 0.246	-	-
Chumbaquihui (CHU)	69	76.67	0.212 ± 0.178	0.329 ± 0.253	-	-
Aguas de Oro (ADO)	67	74.44	0.188 ± 0.170	0.297 ± 0.245	-	-
Mean	60.2	66.89	0.180 ± 0.178	0.280 ± 0.254	0.290	1.227

* Nm estimate of gene flow from G_{st}. $Nm = 0.5(1-G_{st})/G_{st}$, h Nei's gene diversity, I Shannon's information index, SD Standard deviation, NPB number of polymorphic bands, PPB percentage of polymorphic bands

A Jaccard distance coefficient that ranged from 0.183 to 0.524 was obtained. The pairwise comparison of Jaccard value showed that Mishquiyacu (MIS) and Santa Cruz (SCR) as well as Aucaloma (AUC) and Santa Lucia (SLU) were the closest populations with Jaccard´s distance coefficient 0.183 and 0.197, respectively. Pucallpa (PUC) and Ramón Castillo (RAC) were the most distantly related populations with 0.524 distance coefficient (Table 10).

Table 10. Genetic distances among investigated populations

	2DM	AUC	SLU	PUC	RAC	MIS	SCR	PAC	CHU	ADO
2DM	0									
AUC	0.242	0								
SLU	0.231	0.197	0							
PUC	0.405	0.268	0.366	0						
RAC	0.426	0.438	0.258	0.524	0					
MIS	0.287	0.224	0.245	0.390	0.352	0				
SCR	0.273	0.229	0.273	0.374	0.379	0.183	0			
PAC	0.465	0.387	0.434	0.394	0.474	0.417	0.335	0		
CHU	0.332	0.348	0.229	0.356	0.419	0.327	0.303	0.472	0	
ADO	0.422	0.334	0.264	0.442	0.362	0.349	0.342	0.397	0.375	0

Coefficient: Standard Jaccard. Distance Transformation: d=1-s

5.1.2 Cluster analyses based on the ISSR genotyping profile

The NJ dendrogram of genetic distance among 10 populations clearly showed eight main clusters segregated according the localities where the samples were collected (Figure 8). Cluster 4 branched into two visible sub-clusters 4-A and 4-B, while Cluster 2 and Cluster 8 contained only four and eight individuals, respectively. In the analysis each cluster was dominated by samples belonging to a specific population with some intermixing with samples from other populations. Cluster 1 contained nearly the entire Chumbaquihui (CHU) population, with only samples 15 and 19 located in different clusters. Along with the CHU population, Cluster 1 contained three samples (2DM04, 2DM06 and 2DM08) from the Dos de Mayo (2DM) population. Cluster 2 was formed by only four individuals (CHU19, AUC14, AUC15 and AUC18). Cluster 3 contained samples mainly from the Dos de Mayo (2DM) and Aucaloma (AUC) populations but also two samples from Santa Lucía (SLU) and MIS01, PUC09, PUC12, CHU15 and SCR22 samples. Sub cluster 4A contained the sample PAC19 and all Pucallpa samples except PUC09 and PUC12, which were included in cluster 3. Sub cluster 4B contained the whole Pacchilla population, except PAC19, along with sample AUC19 from the Aucaloma (AUC) population. Cluster 5 was formed of 21 out of 23 samples from the Santa Cruz (SCR) population along with samples from the MIS03, AUC07, AUC08 and 2DM17 populations. Cluster 6 was formed exclusively by Mishquiyacu (MIS) samples, with the missing samples scattered among other clusters. Cluster 7 consisted of all twenty Aquas de Oro (ADO) samples. Finally Cluster 8 was composed of all six Ramón Castillo (RAC) samples and three samples from different locations MIS02, 2DM14 and SCR23. The dendrogram showed a clear differentiation of the five populations (CHU, SRC, MIS, ADO and RAC). The population collected in Pacchilla (PAC), which was the only location included to the study, where sacha inchi was cultivated on a commercial plantation, was closely related to the population from the Puccalpa (PUC) location. On the other hand samples from the most distant populations of Dos de Mayo (2DM) and Santa Lucía (SLU) were merged together with samples from the Aucloma (AUC) location. Neither cluster analysis nor principal coordinate analysis revealed any relation between the level of genetic diversity and geographical distance in studied populations. The Ramón Castillo (RAC) population from the abandoned field created a separate cluster and exhibited great divergence from the populations of Pucallpa (PUC), Dos de Mayo (2DM), and Cumbaquihui (CHU).

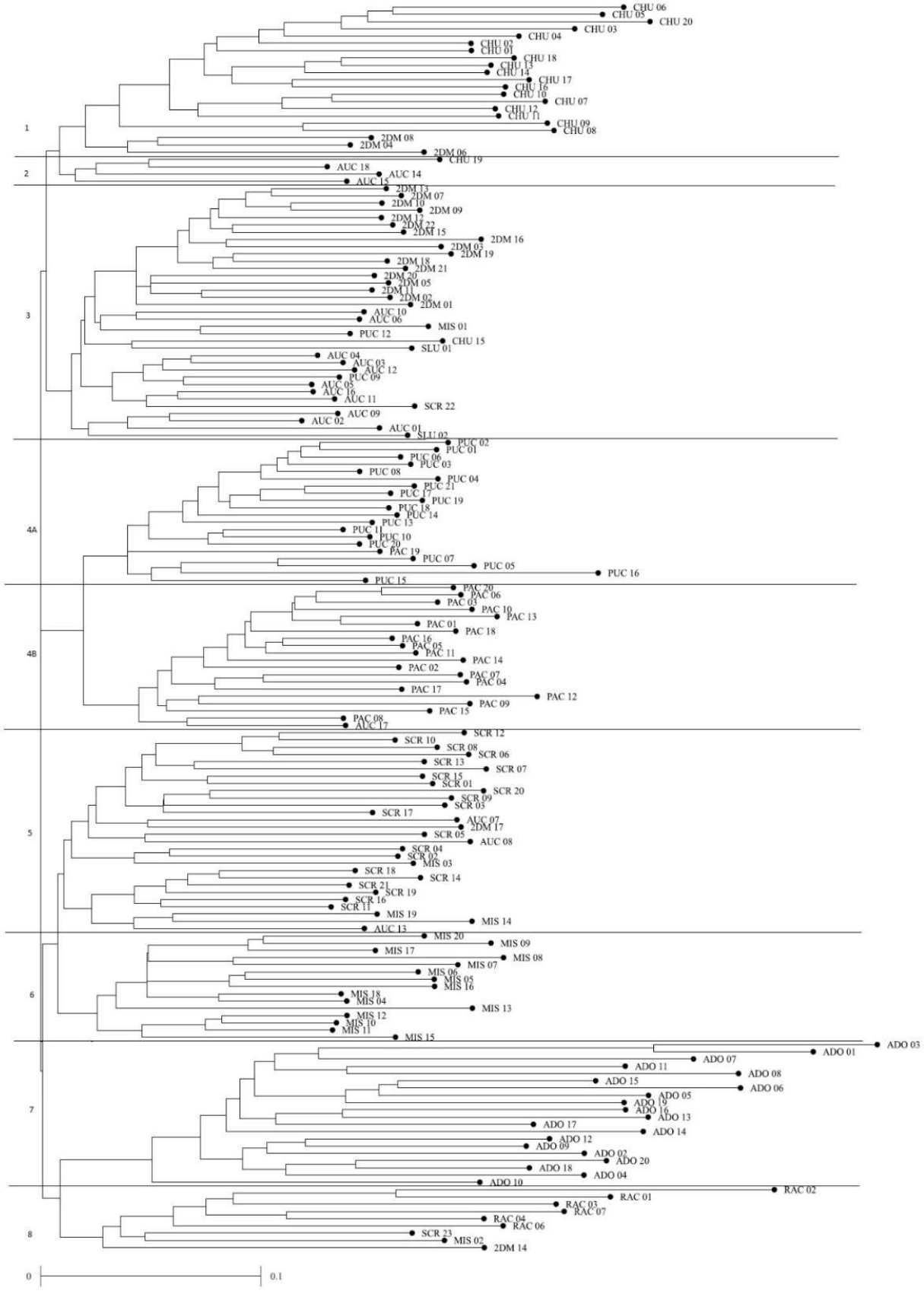

Figure 8. Neighbor-joining (NJ) based analysis of 173 individuals of *P. volubilis* using Jaccard's dissimilarity coefficient

5.1.3 Principal coordinate analysis

The results of three dimensional PCoA analysis indicated relatively clear differentiation of sacha inchi individuals from four localities Chumbaquihui (CHU), Pucallpa (PUC), Dos de Mayo (2DM) and Aguas de Oro (ADO) (Figure 9). On the other hand PCoA revealed the tendency of samples from other localities Mishquiacu (MIS), Pachilla (PAC), Ramón Castillo (RAC), Aucloma (AUC), Santa Lucía (SLU) and some admixed individuals from remaining localities to cluster together.

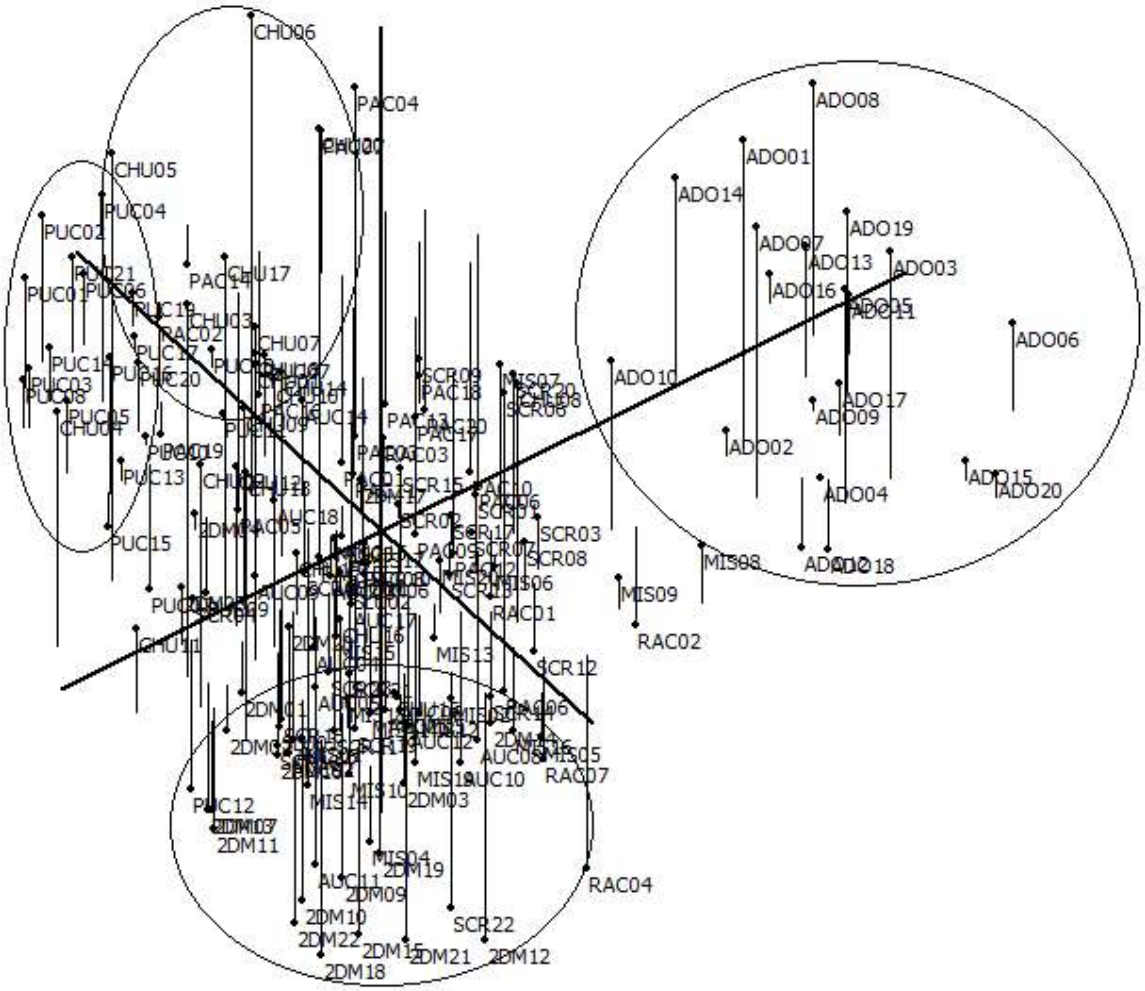

Figure 9. Three-dimensional principal coordinate analysis (PCoA) of the genetic data based on Nei's genetic coefficients for tested 173 individuals of *Plukenetia volubilis* marking four distinguished compact populations

5.2 Protein and protein fractions analysis

This chapter describes the results of total protein and protein fractions contents analyses carried out by Kjeldahl‘s and Osborne’s method, respectively. The homogenity of variances for all variants of total protein content albumins and globulins, gliadin, and glutenins was calculated using Cochran, Hartley, Bartlett test (Table 11). If the calculated level of significance is higher than the selected one $p > \alpha = 0.05$, the variances are conclusively different (homogennous). The result of analysis of variance is not burdened by the error, which would be caused by the nonhomogenity of the variances, and therefore ANOVA could be used, that applied for the total protein content and glutenins. In case the calculated level of significance was lower than the selected one $p < \alpha = 0.05$ the variances are therefore statistically different (nonhomogenous), the premise for using ANOVA is not fulfilled, and the analysis of variances therefore must be calculated using Kruskal-Wallis test, that applied for Albumins + Globulins and Gliadins.

Table 11. Cochran, Hartley, Bartlett test of the homogeneity of variances for total protein content, albumins and globulins, gliadin, and glutenins

	Hartley. F-max	Cochran C.	Bartlett χ^2	DF	p
Total protein content	3.654825	0.218733	9.400961	8	0.309608
Albumins + Globulins	7.503356	0.342191	15.87049	8	0.044272
Gliadins	7.380707	0.28065	32.84318	8	0.000066
Glutenins	4.083699	0.165622	10.20412	8	0.250991

DF = Degree of freedom

The total protein content and glutenin content could be evaluated by ANOVA, albumin + globulin content and gliadin content were evaluated by Kruskal-Wallis. For more detailed evaluation of the analysis of variances the Scheffé’s test was selected for albumin + globulin and gliadin fractions contents, it is more appropriate for diverse groups, meanwhile Tukey‘s test is better for groups of the same sizes and that was employed for total protein content and glutenin content, in order to find, which means (populations) significantly differ from each other.

Table 12. Overall results of the protein and protein fractions analysis (%).

Population	Total Protein Content	Albumins + Globulins	Gliadins	Glutenins
PUC	18.07 ± 1.10^{d}	13.29 ± 1.11^{d}	0.32 ± 0.05^{ab}	4.46 ± 1.28^{bc}
AUC	19.49 ± 0.96^{b}	14.07 ± 0.83^{ad}	0.27 ± 0.12^{ac}	5.16 ± 0.77^{abc}
2DM	20.28 ± 1.35^{ab}	14.53 ± 1.06^{a}	0.29 ± 0.05^{abc}	5.46 ± 1.21^{ab}
PAC	20.77 ± 0.88^{ab}	16.32 ± 0.91^{c}	0.26 ± 0.06^{ac}	4.20 ± 1.41^{c}
CHU	20.86 ± 1.30^{ab}	15.95 ± 1.32^{bc}	0.22 ± 0.10^{ac}	4.69 ± 1.53^{bc}
RAC	20.99 ± 1.60^{abc}	13.74 ± 2.27^{ad}	0.38 ± 0.04^{ab}	6.87 ± 1.18^{a}
SCR	21.26 ± 1.01^{ac}	14.93 ± 0.89^{ab}	0.33 ± 0.09^{ab}	5.99 ± 1.23^{a}
MIS	21.63 ± 1.03^{ac}	15.19 ± 1.36^{abc}	0.36 ± 0.07^{b}	6.08 ± 1.14^{a}
ADO	22.43 ± 0.84^{c}	15.91 ± 1.29^{bc}	0.31 ± 0.04^{ab}	6.20 ± 1.56^{a}
Mean	**20.64 ± 1.12**	**14.88 ± 1.27**	**0.25 ± 0.07**	**5.46 ± 1.26**

Values with different superscript within a row are significantly different

5.2.1 Total protein content analysis

The seeds of sacha inchi contained 20.64% of proteins in average. The means of the populations were ranging between 18.07% and 22.43% (Table 12). The statistical evaluation detected 4 groups of means for the total protein content. The significantly different least content of proteins was detected in population Pucallpa (18.07%). The highest content of total proteins was detected in population Aguas de Oro (22.43%) which was in group together with populations Mishquiyacu (21.63%), Santa Cruz (21.26%) and Ramón Castillo (20.99%). Population RAC was contained in 3 groups. The largest group contains populations Dos de Mayo (20.28%), Pacchilla (20.77%), Chumbaquihui (20.86%), Ramón Castillo (20.99%), Santa Cruz (21.26%) and Mishquiyacu (21.63%). The group "b" contains populations Ramón Castillo (20.99%), Chumbaquihui (20.86%), Pacchilla (20.77%), Dos de Mayo (20.28%) and Aucaloma (19.49%) which was only in this group.

5.2.2 Albumins-globulins fraction analysis

The albumin and globulin fraction contents were ranging between 13.29% and 16.32%. The lowest content of albumin and globulin fraction was detected in population Pucallpa (13.29%), followed by populations Ramón Castillo (13.74%) and Aucaloma (14.07%), forming one group of means. The highest content of albumin and globulin fraction was detected in populations

Mishquiyacu (15.19%), Chumbaquihui (15.95%), Aguas de Oro (15.91%), and Pacchilla (16.32%).

5.2.3 Gliadin fraction analysis

The content ranged between 0.22% and 0.38%. Three groups on means were distinguished by the statistics. The group presenting the lowest contents of gliadin fraction contains populations Chumbaquihui (0.22%), Pacchilla (0.26%), Aucaloma (0.27%), and Dos de Mayo (0.29%). The highest contents are assembled in group consisted of Dos de Mayo (0.29%), Aguas de Oro (0.31%), Pucallpa (0.32%), Santa Cruz (0.33%), Mishquiyacu (0.36%), and Ramón Castillo (0.38%).

5.2.4 Glutenin fraction analysis

The content of glutenins estimated by ANOVA ranged between 4.20% and 6.87%. The statistics distinguished 3 groups. The lowest contents assembled in group consisted of populations Pacchilla (4.20%), Pucallpa (4.46%), Chumbaquihui (4.69%), and Aucaloma (5.16%). The group of the highest contents of glutenins contains populations Aucaloma (5.16%), Dos de Mayo (5.46%), Santa Cruz (5.99%), Mishquiyacu (6.08%), Aguas de Oro (6.20%), and Ramón Castillo (6.87).

5.3 Near Infrared Spectroscopy

The potential of near-infrared reflectance spectroscopy (NIRS) to detect within-plant differences for crude protein content in *Plukenetia* seeds was approved. The qualitative evaluation of new PLS calibration model for prediction protein content in *Plukenetia* seeds is shown in Table 13. The error of the group of prediction (RMSEP = 0.46) was very close to the error of calibration group (RMSEC = 0.47) and both can be considered low. This calibration model presented good correlation between reference values and the NIR predicted ones. The coefficient of determination (R^2) value 0.88 is considered as applicable in common agricultural practice. Standard error predictionStandard error of cross-validation value 0.53 is relatively high.

Table 13. Quality evaluation of calibration model

Parameter of calibration model	**Symbols**	**values**
correlation coefficient of calibration	R	0.95
standard error correlation of calibration	RMSEC	0.47
Parameter of validation		
correlation coefficient of cross validation	R_{cv}	0.94
standard error of cross-validation	RMSECV	0.53
standard error prediction	RMSEP	0.46
residual predictive deviation	RPD	0.66
coefficient of determination	$R^2 = (R_{cv})^2$	0.88

Figure 10 indicates that the calibrating model was not well balanced for estimated lower values of crude protein (for the range between 16-18%) because of lower incidence of standard samples in this region Prediction of protein content in samples with lower value will be influenced by major error. Figure 11 shows predicted value by NIRS against the values measured by standard Kjeldahl mineralization method for crude protein content.

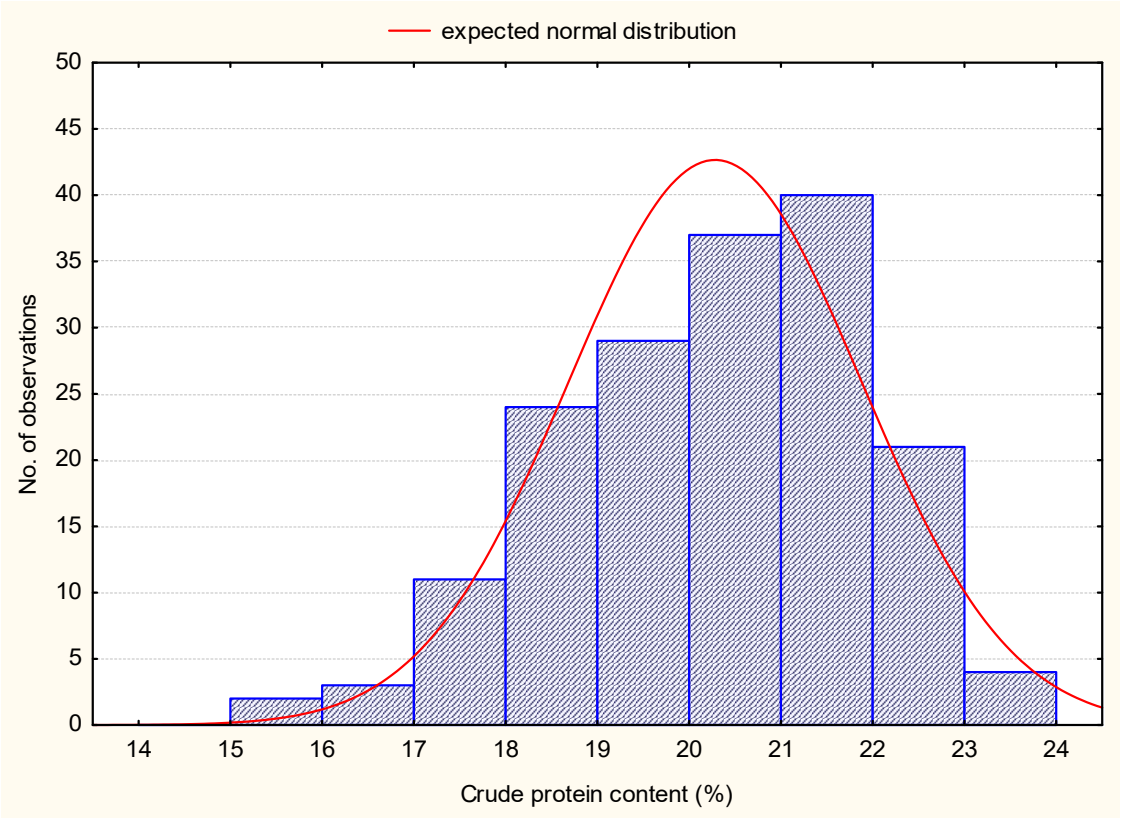

Figure 10. Distribution of crude protein content in sample set.

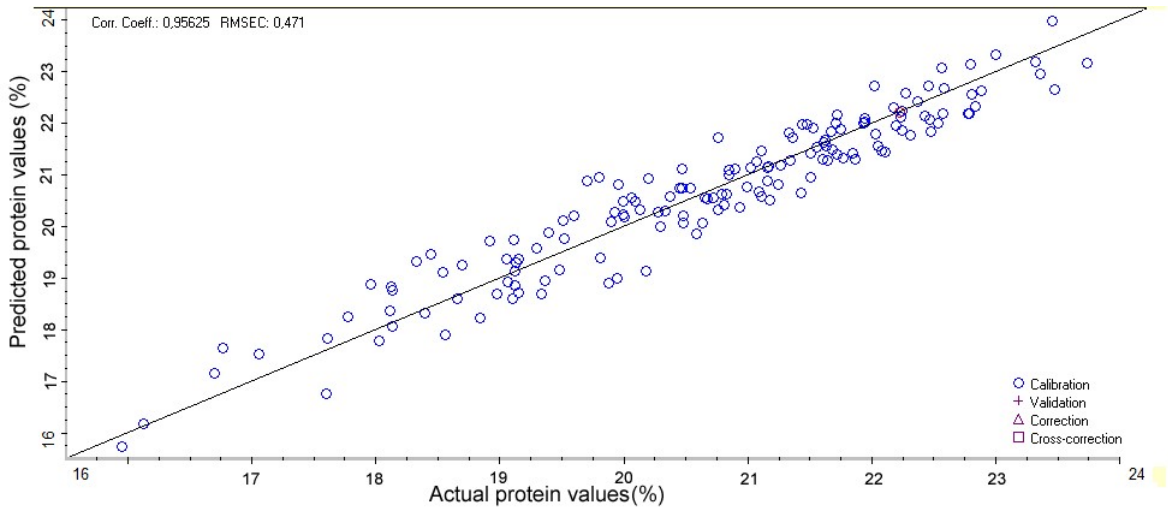

Figure 11. PLS calibration model for prediction protein content in collection of sacha inchi seeds

5.4 SDS PAGE analysis

In the study storage seed proteins and two protein fractions – albumins+globulins and prolamins in single seeds and bulked seed samples were characterized. Protein patterns were evaluated by the classical SDS-PAGE method. Protein bands of total seed protein were detected in the wide range of molecular weight 10-75 kDa (Figure 12) with detected seven bands in followed positions 8.29, 11.14, 16.45, 18.03, 29.70, 40.23 and 59.57 kDa. In the obtained spectra differences in intensity of protein bands (marked with red arrow) were detected, but the polymorphism in band position has generally been found low among all tested samples. In order to fit into the picture, the abreviations of individual populations were modified as follows: Dos de Mayo (2DM => D), Aucaloma (AUC => AU), Santa Lucia (SLU => L), Pucallpa (PUC => P), Ramón Castillo (RAC => R), Mishquiyacu (MIS => M), Santa Cruz (SCR => S), Pacchilla (PAC => PA), Chumbaquihui (CHU => C), Aguas de Oro (ADO => A)

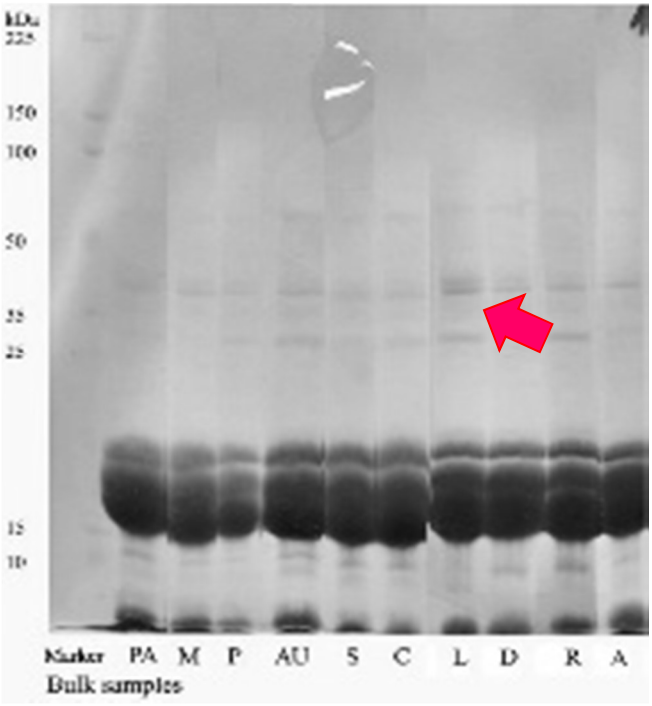

Figure 12. Electroforeograms of classical sodium dodecyl sulphate – polyacrylamide gel electrophoresis patterns of total seed protein (bulked samples) extracted from sacha inchi seeds

A large portion of sacha inchi protein was formed by albumins and globulins (Figure 13), abundant patterns were detected in position from 46.87to 59.14 kDa and 17.12, 18.88 and 18.17 kDa.

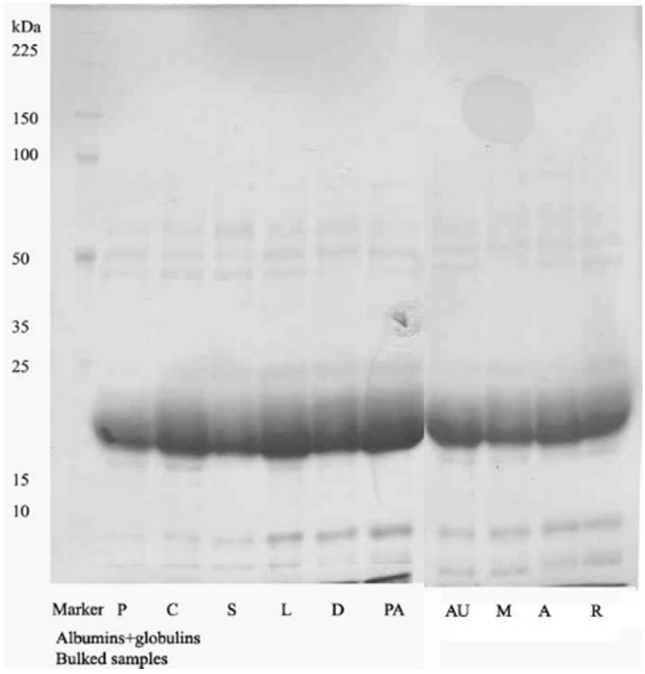

Figure 13. Electroforeograms of classical sodium dodecyl sulphate – polyacrylamide gel electrophoresis patterns of albumins+globulins (bulked samples) extracted from sacha inchi seeds

P. volubilis samples were very low in prolamin fraction (Figure 14). Seeds can be considered as a gluten-free due to low content of prolamin fraction. On the gel three prolamin subunits were found out with molecular weight from 19.43, 21.06 and 24.58 kDa.

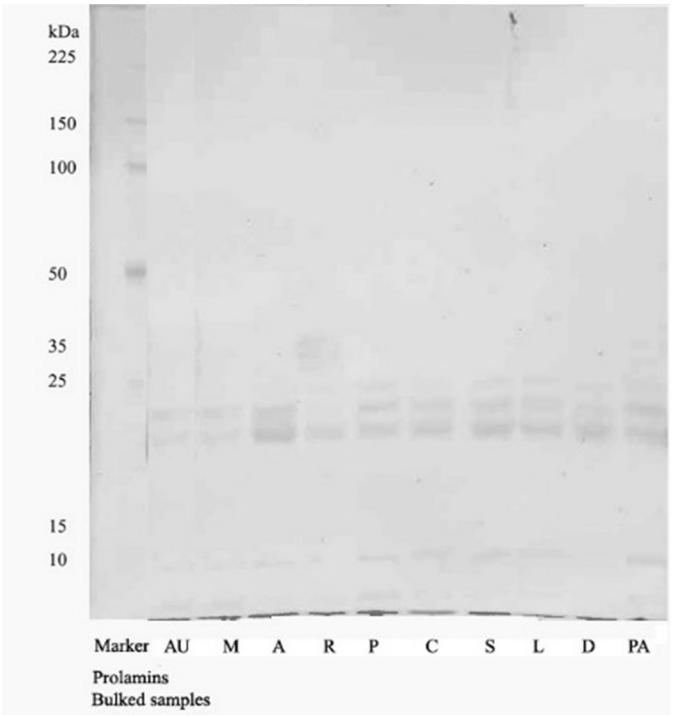

Figure 14. Electroforeograms of classical sodium dodecyl sulphate – polyacrylamide gel electrophoresis patterns of prolamins (bulked samples) extracted from sacha inchi seeds

Of the total soluble proteins, true albumin+globulin and glutelin were the major fractions in the seed flour soluble proteins with prolamin contributing in a small proportion. Typically, majority of the seed flour proteins could be solubilized by the aqueous solvents used for protein fractionation. The seed flour proteins were mainly composed of two types of polypeptides with estimated molecular weights in the range 17 – 19 kDa.

6. Discussion

6.1 Genetic analysis

This study uncovered a level of genetic diversity within 10 populations of *P. volubilis* collected in different locations of the region of San Martín, Peruvian Amazon. The estimation of genetic diversity is essential for choosing an adequate breeding programme and strategy for diversity conservation. As the estimation of the genetic diversity based on morphological and biochemical parameters, due to the influence of different environmental conditions, has limitations, the application of molecular biology method is desirable. According to the obtained results the application of ISSR DNA finger printing method was efficient and successful for disclosing the diversity among 173 samples of sacha inchi, scoring a range from 5 to 18 polymorphic bands. In total eight ISSR primers generated 90 polymorphic bands.

Our result differ noticeably from results by Křivánková et al. (2012) who used the same set of ISSR primers on samples from Ucayali region in the Peruvian Amazon, but found different primers to be polymorphic. In her study, primers 807, 810, 814, 824, 826, 834, 835, 845, 846, 847, 848 and 851 were found as polymorphic, meanwhile or study found polymorphic primers 809, 824, 826, 836, 844, 845, 847, and 859, therefore only three polymorphic primers were common for both studies. However, primer 824 which was polymorphic in both studies, had simultaneously the highest number of polymorphic bands, 16 in the study of Křivánková et al. (2012) and 18 in our study and can therefore be considered as informative for this species. The average percentage of polymorphic bands in her study was 70% and 67% in our study. Křivánková et al. (2012) were investigating sacha inchi in various agroforestry systems in region Ucayali, which is neighbouring with San Martín, but both areas are a few hundreds kilometres distant, besides, the origin of Ucayali's *P. volubilis* germplasm is unknown, most likely not original in this area, therefore the differences may be really caused by the unrelated genetic materials.

The obtained level of genetic variability, 36% among tested localities, can be considered as lower diversity among the samples. In addition, *P. volubilis* estimated level of genetic diversity among populations was observed lower than genetic diversity within populations. This result coresponds with fact that long-living, out-crossing species, such as *P. volubilis*, retain most of their diversity within populations. The lower genetic variability among populations than in populations were dicovered also in other study on *P. volubilis* (Corazon-Guivin et al.,

2008) and in other allogamous plants like *Thymus spp.* (Hadian et al., 2014), *Jatropha curcas* (Basha and Sujatha, 2007), where the diversity within population was also 64%.

Corazón-Guivin et al. (2008) studied the intravarietal and intervarietal level of genetic diversity in four populations of *P. volubilis* from the region of San Martín in the Peruvian Amazon by using DALP primers and noticed strong differentiation among four natural populations attributed by gene flow restriction due to several factors such as a presence of natural barriers, geographical distance and mixed system of pollination. Also, the effect of deforestation can cause isolation and fragmentation of tested populations with consequence of their strong genetic differentiation. In the case of using dominant ISSR markers in this study G_{st} was employed for measuring genetic differentiation among populations (Nybon, 2004). G_{st} was estimated as 0.29 with gene flow value N_m=1.227. Hartl and Clarc (2007) considered the values of N_m smaller than 2.0 presenting considerable opportunity for genetic divergence among populations. Obtained statistical data showed genetic diversity with adequate gene flow supported the consideration than the populations are genetically different but not as strong as described by Corazón-Guivin et al. (2008). According to obtained G_{st} (0.29) among sacha inchi populations from San Martin region there exists exchange of alleles (migration), which corresponded with G_{st} values published by Fisher et al. (2000). G_{st} for perennial plants was calculated to be 22% and G_{st} in outcrossing species was reported 19% with an employment of RAPD markers. The visible result of this value is the dendrogram, which has shown a nice differenciation of the locations indicating relations between some of them. The above discussed results of study by Corazón-Guivin et al. (2008) were probably influenced by the type of used methods. The use of DALP technique with only three polymorphic primers may be responsible for the inaccuracy of obtained results. Other reason result discrepancy could be caused by type of sampling. In our study samples of sacha inchi were collected on substantially larger area region of San Martin and we consider that obtained results reflects more realistic situation in genetic diversity of this species. Also Rodrigues et al. (2013) studied the genetic diversity in 37 *Plukenetia volubilis* samples from four localities, provided by the gene bank of Embrapa Amazônia Ocidental, using the AFLP technique. Their range of Jaccard´s dissimilarity coefficients among the tested samples ranged between 0.338 and 0.900. The results of our study revealed somehow lower level of genetic variability among all of the samples tested with the Jaccard´s dissimilarity coefficients, ranging between 0.183 and 0.524. The studied samples collected from the forests, home gardens and one plantation seem to be genetically more related together compared to those studied by Rodrigues et al. (2013) stored in gene bank and collected

at two different localities of Brazil. This may be influenced by relatively short distances among our sample locations, unlike the study from Brazil, where samples provided by the gene bank might have origins more distant to each other. On the other hand, Cai et al (2010) considered genetic variation in their set of samples of *Jatropha curcas* as relatively high when they calculated Shannon information index to be 0.292 and Nei's gene diversity 0.19, which is comparable with this study's finding: Shannon information index was 0.28 and Nei's gene diversity 0.18.

Cluster analysis in this study distinguished 10 groups and PCoA distinguished four populations. The most distant population Santa Lucia found in deep forest showed the lowest genetic diversity probably due to the small size of population which is mentioned by Dostálek et al. (2014). On the other hand in the dendrogram individuals of Santa Lucía population (SLU) were clustered to the same cluster with individuals Dos de Mayo (2DM) and Aucaloma (AUC) population. All three populations were cultivated on a large distance and were cultivated under very extensive conditions (slopes and intercropped), moreover size of seeds collected in Dos de Mayo location was significantly smaller. On the other hand samples from Pucallpa (PUC) and Pacchilla (PAC) populations were placed in proximity to each other in cluster analysis, when population PAC characterized by farmers as a new plantations with the bigger seed size and used in plantations, probably after selection by farmers. Also samples from Aguas de Oro (ADO) were comprised together in one cluster with Ramón Castillo (RAC), both localities were small and isolated. This is evidence that genetic diversity is presented in this region of Peruvian Amazon especially among extensively and intensively cultivated plants. PCoA and NJ showed the strong clustering in the most of individuals however in some clusters the admixed individuals were revealed as a possible consequence of gene flow through pollen or seed, migration of people from place to place, short and long distance marketing of seeds or as a consequence of fragmentation of the habitat. Rodrigues et al. (2013) found out in tested sacha inchi samples some degree of genetic similarity which may be attributed to the long cultivation history of this crop in the region. Aliyu and Awopetu (2007) described the situation in cashew nut and explained that during the pre-research era, exchange of planting materials must have taken place among farmers and this could be the reason why samples from different geographical regions were closely related then the samples from the samples from neighbouring locations. Rao et al. (2012) attributed the loss of genetic diversity in *Solanum pimpinellifolium* to the migration of the species from Peru to Ecuador resulting in selection towards autogamy. Heywood and Iriondo (2003) considered anthropological influence as crucial for biodiversity

conservation. In our study, both cluster analysis and principal coordinate analysis did not reflect any relation between level of genetic diversity and geographical distances in studied populations. Also Beebee and Rowe (2008), Rodrigues et al. (2013), Shilpha et al. (2013) and Bekele et al. (2014) did not find any clear pattern of clustering according their geographical locations.

Result of both cluster analysis and PCoA successfully identified diversity among samples from different locations. The study revealed difference among samples from abandoned or older cultivations places and new plantations independently on the geographical distance. It is possible to presume that anthropological effect have the strong influence on genetic diversity and dissimilarity. The distribution of such diversity is connected with human activities such as migration of people from place to place (such as for Aguas de Oro in our study), short and long distance marketing of seeds (Bekele et al., 2014).

6.2 Protein and protein fraction analysis

Mean value of the crude protein content reached 20.6% with range between 16% and 23.8%. ADO population was observed as the population with the highest content of crude protein content (mean value 22.43%), on the other side PUC population demonstrated the lowest content of crude protein (mean value 18.07%). These findings are comparable with results of Gutiérrez et al. (2011), on the other hand it differs from results obtained by Sathe et al. (2013) and Hamaker et al. (1992) that reported the total protein content in sacha inchi to be as much as 24.7% and 27%, respectively. Hamaker et al (1992) collected seeds of plants from a collection of the National University of San Martín, probably more same as in our study and yet they obtained higher contents of crude protein. However, it is unknown if it was a collection of random plants or of the best representatives found in the growths cultivated by the farmers or plants found in he wild. The farmers performed some selection of plants they want to grow. They basically use the largest seeds from fruit with as many lobes as possible, while the protein content remains unknown. It is possible that those plants really had higher content of protein. It might also be caused by the lab methodology used, as they performed the analyses more than 20 years ago.

Sathe et al. (2012) in *P. volubilis* found that the albumin, globulin and glutenin were the major fractions in the seed flour with prolamin contributing in a small proportion. In his earlier study, Sathe (2002) detected the total protein content in the seeds to be 20.64% and the albumin-globulin fraction to be 14.95%. Sathe did not mention whether the seeds in their study were

shelled or not, however, we analysed unsheld seeds. But these findings correspond with ours. The content of crude protein was found to be lower compared with soybean protein content ranging between 30.3 and 41.2% (Lee et al., 2013). Similar protein content was observed in almond kernel - 20.6%, gingerbread plum (*Parinari curatellifolia*) kernel - 20.4% (Amza et al., 2014), cashew kernel - 21.2% (Fetuga et al., 1974) and peanuts – 24.5% (Khalil and Chughtai, 1983). These seeds also have similar oil content 47.5%, 47.3%, 48.1% and 49.5%, respectively, compared to 54% reported in sacha inchi (Hamaker et al. 1992). That makes sacha inchi a beneficial alternative to those nuts.

In comparison, Alireza (2014) reported albumins to be 38.33% and globulins 39.04% of the total seed protein in sunflower, while it represents 35% of the seed meals. Further fractions were glutenin (17.09%) and prolamins (5.54%). In the gingerbread plum kernel had the largest fraction was glutenin with 40.6% of total seed protein, followed by albumin (27.6%), globulin (25.8%) and prolamin (6.48%) using the same method of fractionation (Amza et al., 2013).

All authors above mentioned isolation also of prolamins, but they probably were not necessary to be extracted in our study as Sathe et al. (2012) pointed them to contribute in a very small amount, which was consequently confirmed within the SDS-PAGE analysis.

Pucallpa population differed from other populations by statistically significantly lower detected content of protein (18.07%). If we compare it with results by Hamaker et al (1992), the difference would be as much as 9% of total protein content. Most of the samples were collected in somehow protected areas, such as valleys. Populations Pucallpa, Pacchilla and Aucaloma were collected on the hill ridge. These populations performed lower values of total protein content, while Pucallpa was the lowest one. It seems that the protein content can also be affected by the environmental conditions. The soil in Pucallpa was of a lower quality too as it was stony, the plants' growths were shrubbier which also may indicate the plants are used to adverse conditions. It would also be interesting to compare the oil contents, since Shi et al (2010) detected negative correlation between oil and protein content in soybean.

If we confront the protein contents with genetic relations, we can see the Pucallpa population, which performed the lowest content of total proteins, was one of the best distinguished populations by PCoA analysis. On the other side, PCoA also clearly distinguished population Aguas de Oro, which performed the highest protein content. Based on our analysis, we can say, the protein content might be affected affected genetically. The statistical methods attributed the populations to several groups based on the total protein content therefore a diversity was detected from this point o view as well, besides went out supported by the genetics.

NIR spectroscopy has been used to predict the protein content of single seeds of wheat and soybean (Abe et al., 1995) and rapeseeds (Velasco and Möllers, 2002) and it is considered as a nondestructive fast technique. In case of sacha inchi seeds it was necessary to use milled seeds because of their big lenticular and compressed seed size (the diameter of seeds was approximately one centimeter) in comparison with other crops. For more precise results of developed calibration model the smaller particles of measured samples were preferred. Similarly in soybean there were used milled samples for FT-NIR (Ferreira et al., 2013) and Raman spectroscopy (Hoonsoo et al., 2013). The calibration model in this study presented high coefficient of determination R^2 = 0.88, which is considered as applicable in common agricultural practice, values above 0.90 are considered excellent (Dvořáček et al., 2014) and low prediction error RMSEP = 0.46%, as Ferreira et al. (2013) considered their RMSEP = 1.61% low after comparison with further studies. Value of RDP = 0.66 is appropriate, because according to Williams (2001) a value of RPD > 2.4 is desirable to an appropriate model, while Williams and Sobering (1995) indicate that the value of 3 or more is the recommended value. This result depends on the errors of the prediction (RMSEP) which must be lower than the standard deviation of the data group. That makes our model appropriate, since our RMSEP = 0.46% and the standard deviation of the group is 1.63.

SDS–PAGE method was intended for genetic diversity assessment, as it was already used for detection of genetic variability (Akbar et al., 2012; Khurshid and Rabbani, 2012), however, no visible differences were observed with this method. This may be caused by using the bulked samples instead of individual sampling and the differences among the populations on this level might not be detected. However, this method did not reveal high level of genetic diversity in other studies either (Akbar et al., 2012; Ghafoor et al., 2002), and it is generally not a sufficiently powerful technique to distinguish a specific cultivar (Liang et al., 2006) anyway, therefore we did not intend to run such an extensive analysis on all of the samples individually as the method is quite time and money consuming. The populational diversity was evaluated by ISSR analysis and protein contents differences only.

Obtained result of protein bands of total seed protein were detected in the wide range of molecular weight 10-75 kDa corresponds to findings of Sathe et al. (2012) who detected sacha inchi seed soluble protein in range of 10 – 70 kDa. Our results, however, differ in the major molecular species. Meanwhile Sathe et al. (2012) informed their soluble seed flour proteins are mainly composed of two molecular species (32–35 kDa and ~60–62 kDa), and Sathe et al. (2002) informed about albumins being composed mainly of glycosylated polypeptides with estimated molecular weight of 32.8 – 34.8 kDa, we in this study detected the major portion

between 15 – 20 kDa. In their study, Sathe et al. (2012) compared the profiles of albumins and globulins next to each other and the albumins were really consisting mainly of 32 kDa proteins meanwhile the globulins consisted rather of the 20 kDa proteins which were detected in our study. Considering fact that in our study we evaluated albumins and globulin together (they were not separated) and they only formed one major portion of polypeptides of molecular weight 15 – 20 kDa, it is after a comparison with the previous studies by Sathe et al. (2002 and 2012) possible to say, that the albumins are in lower proportions in our samples and the major part of proteins consists of globulins.

7. Conclusion

So far, a few genetic studies about *P. volubilis* using any of molecular markers have been published in scientific journals. Thus our recent research aimed to increase available information about *P. volubilis* focusing on the genetic analysis in order to describe current genetic structure of selected populations and obtain information useful for conservation strategies and future breeding programmes. The method of ISSR markers, used in this study, provided segregation of tested individuals to appropriate clusters with attribution to place of their distribution in the field and with independence to their geographical distance. It therefore proved to be an appropriate tool for genetic diversity assessment. The relatively high genetic diversity found in the populations of sacha inchi might be caused by isolation of populations and limited but still existing gene flow. Also the present level of some similarities among populations was detected, caused by possible historic long term use and seed transmission by people.

The protein contents were rather lower in comparison with other studies, which might have been affected by different methodology used for analysis, but also by genetic differences and selection. Nevertheless, estimated protein content usually higher than 20% still makes sacha inchi an interesting source of protein for human consumption.

The detection and knowledge of the current genetic diversity could help to improve breeding and domestication of this plant, that could bebeneficial for human consumption in other areas. Mapping of useful genetic resources could help preserving the variability essential for improvements. The cultivation has already been spread to China, Cambodia and other countries in Southeast Asia. The future work on this species should be directed to the comparison of samples originating in other countries of South America and the employment other methods of molecular biology with aim to discover genetic diversity. As this plant's cultivation is still increasing and there is still lack of studies on its genetic diversity, more studies should be carried out for further breeding programmes focused either on higher protein content or resistance to pests and diseases. The propagation system is sort of complicated and more genetic studies are desireable. This study might help in mapping of appropriate cultivars in Peruvian Amazon.

8. References

Abdel-Mawgood AL. 2012. DNA Based Techniques for Studying Genetic Diversity. In: Caliskan, M. (Ed.), Genetic Diversity in Microorganisms, Available at http://www.intechopen.com/books/genetic-diversity-in-microorganisms/dna-based-techniques-for-studying-genetic-diversity Accessed [2016-06-25].

Abe H, Kusama T, Kawano S, Iwamoto M. 1995. Non-destructive determination of protein content in a single kernel of wheat and soybean by near infrared spectroscopy. In: A. M: C. Davies and P. Williams (Eds.). Near Infrared Spectroscopy: The Future Waves: 457 - 461. NIR Publication, Chichester, UK.

Agarwal M, Shrivastava N, Harish P. 2008. Advances in molecular marker techniques and their applications in plant sciences. Plant Cell Reports 27: 617–631. doi: 10.1007/s00299-008-0507-z.

Agre AP, Bhattacharjee R, Dansi A, Becerra Lopez-Lavalle LA, Dansi M, Sanni A. 2015. Assessment of cassava (*Manihot esculenta* Crantz) diversity, loss of landraces and farmers preference criteria in southern Benin using farmers' participatory approach. Genetic Resources and Crop Evolution: 1 - 14. DOI 10.1007/s10722-015-0352-1

Agroindustrias Amazónicas 2006. "El Aceite de Inca Inchi." Avalaible on: http://www.incainchi.com.pe/inca.htm Accessed [2016-06-25]

Akbar F, Yousah N, Rabbani MA, Shinwari ZK, Masood MS. 2012. Study of Total Seed Proteins Pattern of Sesame (*Sesamum indicum* L.) Landraces via Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE). Pakistan Journal of Botany 44 (6): 2009 – 2014.

Alipour A, Tsuchimoto S, Sakai H, Ohmido N, Fukui K. 2013. Structural characterization of copia-type retrotransposons leads to insights into the marker development in a biofuel crop, *Jatropha curcas* L. Biotechnology for Biofuels 6:129. DOI 10.1186/1754-6834-6-129

Aliyu OM, Awopetu JA. 2007. Assessment of genetic diversity in three populations of cashew (*Anacardium occidentale* L.) using protein-isoenzyme-electrophoretic analysis. Genetic Resources and Crop Evolution 54: 1489-1497. doi: 10.1007/s10722-006-9138-9

Alves-Pereira A, Peroni N, Gonçalves Abreu A, Gribel R, Clement C. R. 2011. Genetic structure of traditional varieties of bitter manioc in three soils in Central Amazonia. Genetica 139 (10): 1259 - 1271. DOI 10.1007/s10709-011-9627-4

Arévalo G. 1995. El cultivo del sacha inchi (*Plukenetia volubilis* L.) en la Amazonía. Programa Nacional de Investigación en Recursos Genéticos y Biotecnología – PRONARGEB, Estación Experimental El Porvenir – Tarapoto, Perú. 21 p.

Asante IK, Offei SK. 2003. RAPD-based genetic diversity study of fifty cassava (*Manihot esculenta* Crantz) genotypes. Euphytica 131 (1): 113 - 119. DOI 10.1023/A:1023056313776

Bajay MM, Pinheiro JB, Araújo Batista CE, Medeiros Nobrega MB, Imaculada Zucchi M. 2009. Development and characterization of microsatellite markers for castor (*Ricinus communis* L.), an important oleaginous species for biodiesel production. Conservation Genetics Resources 1:237. DOI 10.1007/s12686-009-9058-z

de Bang TC, Raji AA, Ingelbrecht IL. 2011. A Multiplex Microsatellite Marker Kit for Diversity Assessment of Large Cassava (*Manihot esculenta* Crantz) Germplasm Collections. Plant Molecular Biology Reporter 29 (3): 655 - 662. DOI 10.1007/s11105-010-0273-2

Basha SD, Sujatha M. 2009. Genetic analysis of Jatropha species and interspecific hybrids of *Jatropha curcas* using nuclear and organelle specific markers. Euphytica 168 (2): 197 - 214. DOI 10.1007/s10681-009-9900-0

Basha SD, Sujatha M. 2007. Inter and intra-population variability of *Jatropha curcas* (L.) characterized by RAPD and ISSR markers and development of population-specific SCAR markers. Euphytica 156 (3): 375 - 386. DOI 10.1007/s10681-007-9387-5

Bänfer G, Fiala B, Weising K. 2004. AFLP analysis of phylogenetic relationships among myrmecophytic species of *Macaranga* (Euphorbiaceae) and their allies. Plant Systematics and Evolution 249 (3): 213 - 231. DOI 10.1007/s00606-004-0219-4

Beebee TJC, Rowe G. 2008. An Introduction to Molecular Ecology. Population Genetics. 2nd edition. Oxford University Press, New York.

Bekele A, Feyissa T, Tesfaye K. 2014. Genetic diversity of anchote (*Coccinia abyssinica* (Lam.) Cong.) from Ethiopia as revealed by ISSR markers. Genetic Resources and Crop Evolutions 61: 707-719.

Besse P, Seguin M, Lebrun P, Chevallier MH, Nicolas D, Lanaud C. 1994. Genetic diversity among wild and cultivated populations of *Hevea brasiliensis* assessed by nuclear RFLP analysis. Theoretical and Applied Genetics 88 (2): 199 - 207. DOI 10.1007/BF00225898

Bisby FA, Roskov YR, Orrell TM, Nicolson D, Paglinawan LE, Bailly N, Kirk PM, Bourgoin T, Baillargeon G, eds (2010). Species 2000 & ITIS Catalogue of Life: 2010 Annual Checklist. Available at http://www.catalogueoflife.org/annual-checklist/2010. Species 2000: Reading, UK. Accessed [2016-06-25]

Bodwell CE, Hopkins DT. 1985. Nutritional Characteristics of Oilseed Proteins: 221 – 257. New Protein Foods 5: Seed Stotage Proteins. Orlando, Florida.

Bressan EA, Sebbenn AM, Ferreira RR, Lee TSG, Figueira A. 2013. *Jatropha curcas* L. (Euphorbiaceae) exhibits a mixed mating system, high correlated mating and apomixis. Tree Genetics and Genomes 9 (4): 1089 - 1097. DOI 10.1007/s11295-013-0623-y

Brigham RD. 1967. Natural Outcrossing in Dwarf-Internode Castor, Ricinus communis L. Crop Sciences 7:353–355. DOI 10.2135/cropsci1967.0011183X000700040022x

Brondani C. 1996. Isozymic analysis of two natural putative hybrids in two wild species of *Manihot* genera (Euphorbiaceae). Pesquisa Agropecuaria Brasileira 31 (9): 639 - 643.

Bussman RW, Téllez C, Glenn A. 2009. *Plukenetia huayllabamb*ana sp. nov. (Euphorbiaceae) from the upper Amazon of Peru. Nordic Journal of Botany 27: 313-315.

Bussman RW, Zambrana NP, Téllez C. 2013 *Plukenetia carolis-vegae* (Euphorbiaceae) – A New Useful Species from Northern Peru. Economic Botany 67 (4): 387-392.

Cai Y, Sun D, Wu G, Peng J. 2010. ISSR-based genetic diversity of *Jatropha curcas* germplasm in China. Biomass and Bioenergy 34 (12): 1739 – 1750. DOI 10.1016/j.biombioe.2010.07.001

Cachique DH. 2006. Biología floral y reproductiva de *Plukenetia volúbilis* L. (Euphorbiaceae) - (Sacha inchi). [M.Sc. Thesis]. Tarapoto: Universidad Nacional de San Martín. 67 p.

Carmo CD, Santos CD, Alves LB, Oliveira, G. A. F, Oliveira, E. J. 2015. Development of TRAP (Target Region Amplification Polymorphism) as New Tool for Molecular Genetic Analysis in Cassava. Plant Molecular Biology Reporter 33 (6): 1953 - 1966. DOI 10.1007/s11105-015-0887-5

Carrasco NF, Oler JRL, Marchetti FF, Carniello MA, Amorozo MCM, Valle TL, Veasey EA. 2016. Growing Cassava (*Manihot esculenta*) in Mato Grosso, Brazil: Genetic Diversity Conservation in Small–Scale Agriculture. Economic Botany 70 (1): 15 - 28. DOI 10.1007/s12231-016-9331-5

Castelo Branco Carvalho LJ, Schaal BA. 2001. Assessing genetic diversity in the cassava (*Manihot esculenta* Crantz) germplasm collection in Brazil using PCR-based markers. Euphytica 120 (1): 133 - 142. DOI 10.1023/A:1017548930235

Corazon-Guivin M, Castro-Ruiz D, Chota-Macuyama W, Rodríguez Á, Cachique D, Manco E, Del-Castillo D, Renno J-F, García-Dávila C. 2009. Caracterización Genética de Accesiones Sanmartinenses del Banco Nacional de Germoplasma de Sacha Inchi *Plukenetia volubilis* L. (E.E. El Porvenir – INIA). Folia Amazónica 18 (1-2): 23 - 31.

Corazon-Guivin M, Rodríguez Á, Cachique D, Chota W, Vásquez G, Del-Castillo D, Renno J-F, García-Dávila C. 2008. Diversidad Genética en Poblaciones Naturales de Sacha Inchi *Plukenetia volubilis* L. (Euphorbiaceae) en el Departamento de San Martín (Perú). Folia Amazónica 17 (1-2): 83 – 90.

Correa JE, Bernal HY. 1992. Especies vegetales promisorias de los paises del Convenio Andres Bello. Especies Vegetales Promisorias 7: 577–596.

Cortés DF, Reilly K, Okogbenin E, Beeching JR, Iglesias C, Tohme J. 2002. Mapping wound-response genes involved in post-harvest physiological deterioration (PPD) of cassava (*Manihot esculenta* Crantz). Euphytica 128 (1): 47 - 53. DOI 10.1023/A:1020695719706

Czech State Norm (CSN) EN ISO 662, 2001. Živočišné a rostlinné tuky a oleje – Stanovení vlhkosti a těkavých látek. (in Czech).

Czech State Norm (CSN) 5983-1 (467035) 2012. Krmiva - Stanovení obsahu dusíku a výpočet obsahu hrubého proteinu - Část 1: Kjeldahlova metoda. (in Czech).

Daniell H, Wurdack K J, Kanagaraj A, Lee S-B, Saski C, Jansen RK. 2009. The complete nucleotide sequence of the cassava (*Manihot esculenta*) chloroplast genome and the evolution of atpF in Malpighiales: RNA editing and multiple losses of a group II intron. Theoretical and Applied Genetics 116: 723. DOI 10.1007/s00122-007-0706-y

Danquah EO, Akromah R, Quashie-Sam SJ, Oduro W, Falk D, Thevathasan NV, Gordon AM. 2012. The genetic diversity of *Jatropha Curcas* (L.) germplasm in Ghana as revealed by random amplified polymorphic DNA (RAPD) primers. Agroforestry Systems 83 (6): 443 - 450. DOI 10.1007/s10457-012-9488-6

Das AB, Jena S, Pradhan C, Chand PK. 2011. Genetic variability among male populations of a minor mangrove *Excoecaria agallocha* L. as evident by chromosome morphology and DNA markers. Nucleus (India) 54 (1): 39 - 47. DOI: 10.1007/s13237-011-0027-z

Dhakshanamoorthy D, Selvaraj R, Chidambaram A. 2013. Induced mutagenesis in *Jatropha curcas* L. using ethyl methanesulphonate (EMS) and assessment of DNA polymorphism through RAPD markers. Journal of Crop Science and Biotechnology 16 (3): 201 - 207. DOI 10.1007/s12892-012-0079-x

Dice LR. 1945. Measures of amount of ecological association between species. Ecology 26: 297–302.

Dostálek T, Münzbergová Z, Plačková I. 2014. High genetic diversity in isolated populations of *Thesium ebracteatum* at the edge of its distribution range. Conservation Genetics 15: 75-86. doi: 10.1007/s10592-013-0522-7

Doyle J, Doyle JL. 1987. Genomic plant DNA preparation from fresh tissue-CTAB method. Phytochemical Bulletin. 19:11.

Dvořáček, V., Moudrý, J., Čurn, V., 2001. Studies of protein fraction in grain of spelt wheat (*Triticum spelta* L.) and common wheat (*Triticum aestivum* L.). Scientia Agriculturae Bohemica 3 (4): 287–305.

Dvořáček V, Prohasková A, Štočková L. 2014. Efektivní využití blízké infračervené spektroskopie s Fourierovou transformací pro hodnocení technologických vlastností pšenice. Výzkumný ústav rostlinné výroby Praha – Ruzyně.

Dvořáková Z. 2014. Assesment of genetic variation in millet species and related species by application of molecular markers. Ph.D. thesis. Czech University of Life Sciences Prague. 67 p.

Elias M, Mühlen GS, McKey D, Roa AC, Tohme J. 2004. Genetic diversity of traditional South American Landraces of Cassava (*Manihot esculenta* Crantz): an analysis using microsatellites. Economic Botany 58 (2): 242 - 256. DOI 10.1663/0013-0001(2004)058[0242:GDOTSA]2.0.CO;2

FAO 2002. Food energy – methods of analysis and conversion factors. Fao, Food and Nutrition Paper 77. Report of a technical workshop, Rome, 3–6 December. Food and Agriculture Organization of the United Nations. ISSN 0254–4725.

Farooq S, Azam F. (2002). Molecular Markers in Plant Breeding-II. Some Pre-requisites for Use. Pakistan Journal of Biological Sciences 5 (10): 1141-1147.

Fisher M, Husi R, Prati D, Peintinger M, van Kleunen M, Schmid B. 2000. RAPD variation among and within small and large populations of the rare clonal plant *Ranunculus reptans* (Ranunculaceae). American Journal of Botany 87(8): 1128-1137. doi: http://dx.doi.org/10.2307/2656649

Foster JT, Allan GJ, Chan AP, Rabinowicz PD, Ravel J, Jackson PJ, Keim P. 2010. Single nucleotide polymorphisms for assessing genetic diversity in castor bean (*Ricinus communis*). BMC Plant Biology 10:13. DOI 10.1186/1471-2229-10-13

Fregene MAE, Bernal A, Duque M, Dixon A, Tohme J. 2000. AFLP analysis of African cassava (*Manihot esculenta* Crantz) germplasm resistant to the cassava mosaic disease (CMD). Theoretical and Applied Genetics 100 (5): 678 - 685. DOI 10.1007/s001220051339

Fregene MA, Suarez M, Mkumbira J, Kulembeka H, Ndedya E, Kulaya A, Mitchel S, Gullberg U, Rosling H, Dixon AGO, Dean R, Kresovich S. 2003. Simple sequence repeat marker diversity in cassava landraces: genetic diversity and differentiation in an asexually propagated crop. Theoretical and Applied Genetics 107 (6): 1083 - 1093. DOI 10.1007/s00122-003-1348-3

Fregene MAE, Vargas J, Ikea J, Angel F, Tohme J, Asiedu RA, Akoroda MO, Roca WM. 1994. Variability of chloroplast DNA and nuclear ribosomal DNA in cassava (*Manihot esculenta* Crantz) and its wild relatives. Theoretical and Applied Genetics 89 (6): 719 - 727. DOI 10.1007/BF00223711

Fregene MAE, Okogbenin, C. Mba, F. Angel, Maria Christina Suarez, Guitierez Janneth, P. Chavarriaga, W. Roca, M. Bonierbale, J. Tohme. 2001. Genome mapping in cassava improvement: Challenges, achievements and opportunities. Euphytica 120 (1): 159 - 165. DOI 10.1023/A:1017565317940

Fregene MAE, Angel F, Gomez R, Rodriguez F, Chavarriaga P, Roca W, Tohme J, Bonierbale M. 1997. A molecular genetic map of cassava (*Manihot esculenta* Crantz) 95 (3): 431 - 441. DOI 10.1007/s001220050580

Ghafoor A, Ahmad Z, Qureshi AS, Bashir M. 2002. Genetic relationship in Vigna mungo (L.) Hepper and V. radiata (L.) R. Wilczek based on morphological traits and SDS-PAGE. Euphytica 123 (3): 367 – 378. DOI: 10.1023/A:1015092502466

Gillespie LJ. 1993. A Synopsis of Neotropical *Plukenetia* (Euphorbiaceae) Including Two New Species. Systematic Botany 18 (4): 575-592.

Gillespie LJ. 1994. Pollen Morphology and Phylogeny of the Tribe *Plukenetieae* (Euphorbiaceae). Annals of Missouri Botanical Garden 81: 317-347.

Gillespie LJ. 2007. A Revision of Paleotropical *Plukenetia* (Euphorbiaceae) Including Two New Species from Madagascar. Systematic Botany, 32 (4), 780-802.

Google Earth. 2016. Google Inc. 1600 Amphitheatre Parkway Mountain View, CA 94043 USA. Available at https://www.google.com/intl/cs/earth/ Accessed [2016-05-17]

Le Guen V, Doaré F, Weber C, Seguin M. 2009. Genetic structure of Amazonian populations of *Hevea brasiliensis* is shaped by hydrographical network and isolation by distance. Tree Genetics & Genomes 5 (4): 673 - 683. DOI 10.1007/s11295-009-0218-9

Guillén MD, Ruiz A, Cabo N, Chirinos R, Pascual G. 2003. Characterization of Sacha Inchi (*Plukenetia volubilis* L.) Oil by FTIR Spectroscopy and ^{1}H NMR. Comparison with Lineseed Oil. Journal of the American Oil Chemists' Society 80: 755-762.

Guo G-Y, Wang S-H, Xu Y, Tang L, Yu M-Q, Chen F. 2016. Molecular phylogenetic analysis of key *Jatropha* species inferred from nrDNA ITS and chloroplast (trnL-F and rbcL) sequences. Genes & Genomics: 1 - 10. DOI 10.1007/s13258-016-0406-6

Gutiérrez, L. F, Rosada L. M, Jiménez, Á. 2011. Chemical composition of Sacha Inchi (*Plukenetia volubilis* L.) seeds and characteristics of their lipid fraction. Grasas y Aceites 62 (1): 76-83.

Gupta P, Idris A, Mantri S, Asif MH, Yadav HK, Roy JK, Tuli R, Mohanty CS, Sawant, SV. 2012. Discovery and use of single nucleotide polymorphic (SNP) markers in *Jatropha curcas* L. Molecular Breeding 30 (3): 1325 - 1335. DOI 10.1007/s11032-012-9719-6

Hadian J, Bigdeloo M, Nazeri V and Khadivi-Khub A (2014): Assessment of genetic and chemical variability in *Thymus caramanicus*. Molecular Biology Reports 41 (5): 3201-3210. doi: http://dx.doi.org/10.1007/s11033-014-3180-z

Hamaker BR, Valles C, Gilman R, Hardmeier RM, Clark D, García HH, Gonzales AE, Kohlstad I, Castro M (1992): Amino Acid and Fatty Acid Profiles of the Inca Peanut (*Plukenetia volubilis* L.). Cereal Chemistry 69: 461-463.

Harris SA. 2003. Molecular approaches to assessing plant diversity. In: Benson, E. E. (Ed.), Plant Conservation Biotechnology. Published by Taylor and Francis Ltd. UK: 11-19.

Hartl DL, Clark AC. 2007. Inbreeding subdivision, and migration. p. 257. In: Principles of population genetics. 4 th ed. Sinauer Associates, Inc. Publishers, Sunderland, Massachusetts, USA.

Heywood VH, Iriondo JM. 2003. Plant conservation: old problems, new perspectives. Biological Conservation 113: 321-335. DOI: http://dx.doi.org/10.1016/S0006-3207(03)00121-6

Hofmeijer I. 2010. Traditional Food Systems, Biodiversity, and Development: A Case Study on sacha inchi. Working paper, McGill University, Montreal, Canada: 11

Holderegger R, Stehlik I. 1999. Sibmating in a small, isolated population of the dioecious plant species *Mercurialis ovata*. Biochemical Systematics and Ecology 27 (7). DOI: 10.1016/S0305-1978(99)00010-1

Hoonsoo L, Byoung-Kwan C, Moon SK, Wang-Hee L, Jagdish T, Hanhong B, Soo-In S, Hee-Youn Ch. 2013. Prediction of crude protein and oil content of soybeans using Raman spectroscopy. Sensors and Actuators B: Chemical 185: 694-700 doi:10.1016/j.snb.2013.04.103

Hutchinson J. 1969. Evolution and Phylogeny of Flowering Plants. Academic Press, London, pp: 208–210.

Hutchenson K. 1970. A Test for Comparing Diversities Based on the Shannon Formula. Journal of Theoretical Biology 29: 151-154.

Chacón J, Madriñán S, Debouck D, Rodriguez F, Tohme J. 2008. Phylogenetic patterns in the genus *Manihot* (Euphorbiaceae) inferred from analyses of nuclear and chloroplast DNA regions. Molecular Phylogenetics and Evolution. 49 (1): 260 - 267. DOI: 10.1016/j.ympev.2008.07.015

Chaurasia AK, Subramaniam VR, Krishna B, Sane PV. 2009. RAPD based genetic variability among cultivated varieties of Aonla (Indian Gooseberry, *Phyllanthus emblica* L.). Physiology and Molecular Biology of Plants. 15 (2): 169 - 173. DOI 10.1007/s12298-009-0019-5

Chavarriaga-Aguirre P, Maya MM, Bonierbale MW, Kresovich S, Fregene MA, Tohme J, Kochert G. 1998. Microsatellites in cassava (*Manihot esculenta* Crantz): Discovery, inheritance and variability. Theoretical and Applied Genetics 97 (3): 493 - 501. DOI: 10.1007/s001220050922

Chavarriaga-Aguirre P, Maya MM, Tohme J, Duque MC, Iglesias C, Bonierbale MW, Kresovich S, Kochert G. 1999. Using microsatellites, isozymes and AFLPs to evaluate genetic diversity and redundancy in the cassava core collection and to assess the usefulness of DNA-based markers to maintain germplasm collections. Molecular Breeding 5 (3): 263 - 273. DOI 10.1023/A:1009627231450

Jiménez RJ, Martíenez MG, Cruz RD. 2000. El género *Plukenetia* (Euphorbiaceae) en México. Anales del Instituto de Biología de la Universidad Nacional Autónoma de México, Serie Botanica 71 (1): 11–18.

Joint FAO/WHO Expert Consultations. 1990. Protein Quality Evaluation. Food and Agriculture Organization: Rome, Italy.

Jubera MA, Janagoudar BS, Biradar DP, Ravikumar RL, Koti RV, Patil SJ. 2009. Genetic diversity analysis of elite Jatropha curcas L. genotypes using randomly amplified polymorphic DNA markers. Karnataka Journal of Agricultural Sciences 22 (2): 293 – 295.

Kanchanaketu T, Sangduen N, Toojinda T, Hongtrakul V. 2012. Genetic diversity analysis of *Jatropha curcas* L. (Euphorbiaceae) based on methylation-sensitive amplification polymorphism. Genetics and Molecular Research. 11 (2): 944 - 955. DOI: 10.4238/2012.April.13.2

Kawuki RS, Ferguson M, Labuschagne M, Herselman L, Kim D-J. 2009. Identification, characterisation and application of single nucleotide polymorphisms for diversity assessment in cassava (*Manihot esculenta* Crantz). Molecular Breeding. 23 (4): 669 - 684. DOI 10.1007/s11032-009-9264-0

Khurshid H, Rabbani MA. 2012. Comparison of Electrophoretic Protein Profiles frm Seed of Different Oilseed Brassica Cultivars. Journal of Public Health and Biological Sciences 1 (2): 36 – 42.

Kizito EB, Chiwona-Karltun L, Egwang T, Fregene MAE, Westerbergh A. 2007. Genetic diversity and variety composition of cassava on small-scale farms in Uganda: an interdisciplinary study using genetic markers and farmer interviews. Genetica 130 (3): 301 - 318. DOI 10.1007/s10709-006-9107-4

Kombo GR, Dansi A, Loko LY, Orkwor GC, Vodouhè R, Assogba P, Magema J. M. 2012. Diversity of cassava (*Manihot esculenta* Crantz) cultivars and its management in the department of Bouenza in the Republic of Congo. Genetic Resources and Crop Evolution. 59 (8): 1789 - 1803. DOI 10.1007/s10722-012-9803-0

Krähenbühl M, Yuan Y-M, Küpfer P. 2002. Chromosome and breeding system evolution of the genus *Mercurialis* (Euphorbiaceae): Implications of ITS molecular phylogeny. Plant Systematics and Evolution. 234 (1-4): 155 - 169. DOI: 10.1007/s00606-002-0208-y

Křivánková B, Polesný Z, Lojka B, Lojková J, Banout J, Preininger D. 2007. Sacha Inchi (*Plukenetia volubilis*, Euphorbiaceae): A Promising Oilseed Crop from Peruvian Amazon. Tropentag Conference, Witzenhausen, Germany.

Křivánková B, Hlásná Čepková P, Ocelák M, Juton G, Bechyně M, Lojka B. 2012. Preliminary Study of Diversity of *Plukenetia volubilis* Based on the Morphological and Genetic Characteristics. Agricultura Tropica et Subtropica 45 (3): 140 - 146. DOI 10.2478/v10295-012-0023-6

Křivánková B. 2012. Ethnobotany, diversity and cultivation potential of "Sacha inchi" (*Plukenetia* sp., Euphorbiaceae); the perspective oilseed crop from Peruvian Amazon. Ph.D. Thesis. Czech University of Life Sciences Prague. 39 p.

Kumar RV, Tripathi YK, Shukla P, Ahlawat SP, Gupta VK. 2009. Genetic diversity and relationships among germplasm of *Jatropha curcas* L. revealed by RAPDs. Trees 23 (5): 1075 - 1079. DOI 10.1007/s00468-009-0350-z

Kumari M, Grover A, Yadav PV, Arif M, Ahmed Z. 2013. Development of EST-SSR markers through data mining and their use for genetic diversity study in Indian accessions of *Jatropha curcas* L.: a potential energy crop. Genes & Genomics 35 (5): 661 - 670. 10.1007/s13258-013-0118-0

Laemmli UK. 1970. Cleavage of structural proteins during the assembly of the head of Bacteriophage T4. Nature 227 (5259), 608-685.

Lam LV, Thanh T, Quynh Chi VT, Tuy L. M. 2009. Genetic Diversity of *Hevea* IRRDB'81 Collection Assessed by RAPD Markers. Molecular Biotechnology 42 (3): 292 - 298. DOI 10.1007/s12033-009-9159-7

Laosatit K, Tanya P, Somta P, Ruang-Areerate P, Sonthirod C, Tangphatsornruang S, Juntawong P, Srinives P. 2015. De novo Transcriptome Analysis of Apical Meristem of *Jatropha* spp. Using 454 Pyrosequencing Platform, and Identification of SNP and EST-SSR Markers. Plant Molecular Biology Reporter: 1 - 8. DOI 10.1007/s11105-015-0961-z

Lee H, Cho B-K, Kim MS, Lee W-H, Tewari J, Bae H, Sohn S-I, Chi H-Y. 2013. Prediction of crude protein and oil content of soybeans using Raman spectroscopy. Sensors and Actuators B 185: 694-700

Lefèvre F, Charrier A. 1992. Isozyme diversity within African *Manihot* germplasm. Euphytica 66 (1): 73 - 80. DOI 10.1007/BF00023510

Lespinasse D, Rodier-Goud M, Grivet L, Leconte A, Legnate H, Seguin M. 2000. A saturated genetic linkage map of rubber tree (*Hevea* spp.) based on RFLP, AFLP, microsatellite, and isozyme markers. Theoretical and Applied Genetics 100 (1): 127 - 138. DOI 10.1007/s001220050018

Li D, Xia Z, Deng Z, Liu X, Feng F. 2013. Development, characterization, genetic diversity and cross-species/genera transferability of ILP markers in rubber tree (*Hevea brasiliensis*). Genes and Genomics 35 (6): 719 - 731. DOI: 10.1007/s13258-013-0122-4

Liang XQ, Luo M, Holbrook CC, Guo BZ. 2006. Storage protein profiles in Spanish and runner market type peanuts and potential markers. BMC Plant Biology 6:24. DOI:10.1186/1471-2229-6-24

Lokko Y, Dixon A, Offei S, Danquah E, Fregene M. 2006. Assessment of genetic diversity among African cassava *Manihot esculenta* Grantz accessions resistant to the cassava mosaic virus disease using SSR markers. Genetic Resources and Crop Evolution 53 (7): 1441 - 1453. DOI 10.1007/s10722-005-6841-x

Lopez C, Piégu B, Cooke R, Delseny M, Tohme J, Verdier V. 2005. Using cDNA and genomic sequences as tools to develop SNP strategies in cassava (*Manihot esculenta* Crantz). Theoretical and Applied Genetics 110 (3): 425 - 431. DOI 10.1007/s00122-004-1833-3

Lowe AJ, Harris SA, Ashton P. (2004). Ecological Genetics: Design, Analysis and Application Blackwells: Oxford:. 326.

Luo H, Coppenolle BV, Seguin M, Boutry M. 1995. Mitochondrial DNA polymorphism and phylogenetic relationships in *Hevea brasiliensis*. Molecular Breeding 1 (1): 51 - 63. DOI 10.1007/BF01682089

Macbride FJ. 1951. Flora of Peru. Parte 3A, 13 (1): 115-116.

Macbride FJ. 1990. Flora of Perú. Fieldiana Botany Vol XIII.: 115- 116.

Maghuly F, Jankowicz-Cieslak J, Calari A, Ramkat R, Till B, Laimer M. 2011. Investigation of genetic variation in *Jatropha curcas* by Ecotilling and ISSR. BMC Proceedings. 5:O50. DOI 10.1186/1753-6561-5-S7-O50

Manco CEI. 2006. Cultivo de Sacha Inchi. Tarapoto: Instituto Nacional de Investigación y Extensión Agraria, Subdirección de Recursos Genéticos y Biotecnología, Estación Experimental Agraria el Porvenir, Peru.

Mantello CC, Suzuki FI, Souza LM, Gonçalves PS, Souza AP. 2012. Microsatellite marker development for the rubber tree (*Hevea brasiliensis*): characterization and cross-amplification in wild *Hevea* species. BMC Research Notes 5: 329. DOI 10.1186/1756-0500-5-329

Manu-Aduening JA, Peprah BB, Agyeman A. 2013. Genetic variability of cassava progenies developed through introgression of cassava mosaic disease resistance into Ghanaian landraces. Journal of Crop Science and Biotechnology 16 (1): 23 - 28. DOI 10.1007/s12892-012-0053-7

Marmey P, Beeching JR, Hamon S, Charrier H. 1993. Evaluation of cassava (*Manihot esculenta* Crantz) germplasm collections using RAPD markers. Euphytica 74 (3): 203 - 209. DOI 10.1007/BF00040402

Mastan SG, Sudheer PDVN, Rahman H, Ghosh A, Rathore MS, Prakash CR, Chikara J. 2012. Molecular characterization of intra-population variability of *Jatropha curcas* L. using DNA based molecular markers. Molecular Biology Reports 39 (4):4383 - 4390. DOI: 10.1007/s11033-011-1226-z

Maurya R, Gupta A, Singh SK, Rai KM, Chandrawati, Katiyar R, Sawant SV, Yadav HK. 2015. Genomic-derived microsatellite markers for diversity analysis in *Jatropha curcas*. Trees 29 (3): 849 - 858. DOI 10.1007/s00468-015-1166-7

Mavuso C, Wu Y-P, Chen F-C, Huang B-H, Lin S-J. 2015. Genetic diversity analysis of *Jatropha curcas* L. accessions cultivated in Taiwan using inter simple sequence repeats (ISSR) markers. Agroforestry Systems: 1 - 15. DOI 10.1007/s10457-015-9864-0

Mba REC, Stephenson P, Edwards K, Melzer S, Nkumbira J, Gullberg U, Apel K, Gale M, Tohme J, Fregene MAE. 2001. Simple sequence repeat (SSR) markers survey of the cassava (*Manihot esculenta* Crantz) genome: towards an SSR-based molecular genetic map of cassava. Theoretical and Applied Genetics 102 (1): 21 - 31. DOI 10.1007/s001220051614

Meinders HC, Jones MD. 1950. Pollen shedding and dispersal in the Castor plant *Ricinus communis* L. Agronomy Journal 42:206–209

Míka V, Kohoutek A, Nerušil P. 2008. Spektroskopie v blízké infračervené oblasti (NIR). Výzkumný ústav rostlinné výroby Ruzyně. ISBN 978-80-87011-53-9

Montarroyos AVV, de Lima MAG, dos Santos EO, de França JGE. 2003. Isozyme analysis of an active cassava germplasm bank collection. Euphytica 130 (1): 101 - 106. DOI 10.1023/A:1022363122834

Montes Osorio LR, Torres Salvador AF, Etienne Jongschaap RE, Azurdia Perez CA, Berduo Sandoval JE, Miguel Trindade L, Franciscus Visser RG, van Loo EN. 2014. High level of molecular and phenotypic biodiversity in *Jatropha curcas* from Central America compared to Africa, Asia and South America. BMC Plant Biology 14:77. DOI 10.1186/1471-2229-14-77

Morden CW, Gregoritza M. 2006. Population variation and phylogeny in the endangered *Chamaesyce skottsbergii* (Euphorbiaceae) based on RAPD and ITS analyses. Conservation genetics 6 (6): 969 - 979. DOI 10.1007/s10592-005-9087-4

Moura EF, Sousa NRB, Moura MF, Dias MC, Souza ED, de Farias Neto JT, Sampaio JE. 2016. Molecular characterization of accessions of a rare genetic resource: sugary cassava (*Manihot esculenta* Crantz) from Brazilian Amazon. Genetic Resources and Crop Evolution: 1- 11. DOI: 10.1007/s10722-016-0378-z

Mouré L. 1967. Annals of the Missouri Botanical Garden. 54: 195. Accessed online at http://plantillustrations.org/illustration.php?id_illustration=117433&mobile=0 Accessed [26-05-2016]

Na-ek Y, Wongkaew A, Phumichai T, Kongsiri N, Kaveeta R, Reewongchai T, Phumichai C. 2011. Genetic diversity of physic nut (*Jatropha curcas* L.) revealed by SSR markers. Journal of Crop Science and Biotechnology 14 (2): 105 - 110. DOI 10.1007/s12892-011-0008-4

Nei M. 1972. Genetic Distance between Populations. The American Naturalist 106: 283-292.

Nei M, Li WH. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proceedings of the National Academy of Sciences, USA 76**:** 5269–5273.

Netmaps. 2016. Peru Political Map. Available at http://www.netmaps.net/netmaps/peru-political-map/ Accessed [2016-08-22].

Noriega H, Risco MY, Cachique D, Ruiz H, Solis R, Guerrero JC. 2010. Biología y autocompatibilidad de polen de Sacha inchi (*Plukenetia volubilis* L.). Proceeding de Primer Congreso Peruano de Mejoramiento Genético y Biotecnología Agrícola. Lima, Peru.

Nybon H. 2004. Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. Molecular Ecology 13: 1143-1155.

Okogbenin E, Marin J, Fregene MAE. 2006. An SSR-based molecular genetic map of cassava. Euphytica 147 (3): 433 - 440. DOI 10.1007/s10681-005-9042-y

Oliveira-Silva AM, Silva GF, Dias MC, Clement CR, Sousa NR. 2014. Inter-retrotransposon-amplified polymorphism markers for germplasm characterization in *Manihot esculenta* (Euphorbiaceae). Genetics and Molecular Research 13 (2): 3800 - 3804. DOI: 10.4238/2014.May.16.3

de Oliveira EJ, Fortes Ferreira C, da Silva Santos V, Nunes de Jesus O, Fachardo Oliveira GA, da Silva MS. 2014. Potential of SNP markers for the characterization of Brazilian cassava germplasm. Theoretical and Applied Genetics 127 (66): 1423 - 1440. DOI 10.1007/s00122-014-2309-8

Olsen KM. 2002. Population history of *Manihot esculenta*, (Euphorbiaceae) inferred from nuclear DNA sequences. Molecular Ecology 11 (5): 901 - 911. DOI: 10.1046/j.1365-294X.2002.01493.x

Olsen KM. 2004. SNPs, SSRs and inferences on cassava's origin. Plant Molecular Biology 56 (4): 517 - 526. DOI: 10.1007/s11103-004-5043-9

Osborne TB. 1924. The Vegetable Proteins, 2nd edn. Longmans, Green and Co., London, p 154

Ouattara B, Ndir KN, Gueye MC, Diédhiou I, Barnaud A, Fonceka D, Cissé N, Akpo EL, Diouf D. 2014. Genetic diversity of *Jatropha curcas* L. in Senegal compared with exotic accessions based on microsatellite markers. Genetic Resources and Crop Evolution 61 (6): 1039 - 1045. DOI 10.1007/s10722-014-0106-5

Pamidimarri DVNS, Chattopadhyay B, Reddy MP. 2009 a. Genetic divergence and phylogenetic analysis of genus *Jatropha* based on nuclear ribosomal DNA ITS sequence. Molecular Biology Reports 36 (7): 1929 - 1935. DOI: 10.1007/s11033-008-9401-6

Pamidimarri DVNS, Mastan SG, Rahman H, Reddy MP. 2010 b. Molecular characterization and genetic diversity analysis of *Jatropha curcas* L. in India using RAPD and AFLP analysis. Molecular Biology Reports 37 (5): 2249 - 2257. DOI: 10.1007/s11033-009-9712-2

Pamidimarri DVNS, Pandya N, Reddy MP Radhakrishnan T. 2009 c. Comparative study of interspecific genetic divergence and phylogenic analysis of genus *Jatropha* by RAPD and AFLP: Genetic divergence and phylogenic analysis of genus *Jatropha*. Molecular Biology Report 36 (5): 901 - 907. DOI: 10.1007/s11033-008-9261-0

Pamidimarri DVNS, Reddy MP. 2014. Phylogeography and molecular diversity analysis of *Jatropha curcas* L. and the dispersal route revealed by RAPD, AFLP and nrDNA-ITS analysis. Molecular Biology Reports 41 (5): 3225 - 3234. DOI: 10.1007/s11033-014-3185-7

Pamidimarri DVNS, Singh S, Mastan SG, Patel J, Reddy MP. 2009 d. Molecular characterization and identification of markers for toxic and non-toxic varieties of *Jatropha curcas* L. using RAPD, AFLP and SSR markers. Molecular Biology Reports 36 (6): 1357 - 1364. DOI 10.1007/s11033-008-9320-6

Pamidimarri DVNS, Rahman H, Mastan SG, Reddy MP. 2010 e. Isolation of novel microsatellites using FIASCO by dual probe enrichment from *Jatropha curcas* L. and study on

genetic equilibrium and diversity of Indian population revealed by isolated microsatellites. Moleular Biology Reports 37 (8): 3785 - 3793. DOI 10.1007/s11033-010-0033-2

Park YJ, Lee JK, Kim NS. (2009). Simple Sequence Repeat Polymorphisms (SSRPs) for Evaluation of Molecular Diversity and Germplasm Classification of Minor Crops. Molecules 14: 4546-4569.

Pereira de Souza AH, Gohara AK, Rodrigues AC, Evelázio de Souza N, Visentainer JV, Matsushita M. 2013. Sacha Inchi as Potential Source of Essential Fatty Acids and Tocopherols: Multivariate Study of Nut and Shell. Acta Scientiarum 35 (4): 757-763.

Peroni N, Kageyama PY, Begossi A. 2007. Molecular differentiation, diversity, and folk classification of "sweet" and "bitter" cassava (*Manihot esculenta*) in Caiçara and Caboclo management systems (Brazil). Genetic Resources and Crop Evolution 54 (6): 1333 - 1349. DOI 10.1007/s10722-006-9116-2

Perrier X, Jacquemoud-Collet JP. 2006. DARwin software http://darwin.cirad.fr/

Perry-Castañeda Library. 2016. Peru Shaded Relief Map. The University of Texas at Austin. United States of America. Available at http://www.lib.utexas.edu/maps/americas/peru.gif Accessed [2016-08-25]

Pioto F, Costa RS, França SC, Gavioli EA, Bertoni BW, Zingaretti SM. 2015. Document Genetic diversity by AFLP analysis within *Jatropha curcas* L.populations in the State of São Paulo, Brazil. Biomass and Bioenergy 80: 316-320. DOI: 10.1016/j.biombioe.2015.06.014

Prochnik S, Marri PR, Desany B, Rabinowicz PD, Kodira C, Mohiuddin M, Rodriguez F, Fauquet C, Tohme J, Harkins T, Rokhsar DS, Rounsley S. 2012. The Cassava Genome: Current Progress, Future Directions. Tropical Plant Biology 5 (1): 88 - 94. DOI 10.1007/s12042-011-9088-z

Phumichai C, Phumichai T, Kongsiri N, Wongkaew A, Sripichit P, Kaveeta R. 2011. Isolation of 55 microsatellite markers for *Jatropha curcas* and its closely related species. Biologia Plantarum 55:387. DOI 10.1007/s10535-011-0061-3

Rafii MY, Shabanimofrad M, Edaroyati MWP, Latif MA. 2012. Analysis of the genetic diversity of physic nut, *Jatropha curcas* L. accessions using RAPD markers. Molecular Biology Reports 39 (6): 6505 - 6511. DOI 10.1007/s11033-012-1478-2

Ram SG, Parthiban KT, Kumar RS, Thiruvengadam V, Paramathma M. 2008. Genetic diversity among *Jatropha* species as revealed by RAPD markers. Genetic Resources and Crop Evolution 55 (6): 803 - 809. DOI 10.1007/s10722-007-9285-7

Ramezani H. 2012. A Note on the Normalized Definition of Shannon's Diversity Index in Landscape Pattern Analysis. *Environment and Natural Resources Research* 2 (4): 54-60.

Rao ES, Kadirvel P, Symonds RC, Geethanjali S, Ebert AW. 2012. Using SSR markers to map genetic diversity and population structure of *Solanum pimpinellifolium* for development of a core collection. Plant Genetic Resources: Characterization and Utilization 10 (1): 38–48. doi: http://dx.doi.org/10.1017/S1479262111000955

Raposo RS, Souza IGB, Veloso MEC, Kobayashi AK, Laviola BG, Diniz FM. 2014. Development of novel simple sequence repeat markers from a genomic sequence survey database and their application for diversity assessment in *Jatropha curcas* germplasm from Guatemala. Genetics and Molecular Research 13 (3): 6099 - 6106. DOI: 10.4238/2014.August.7.25

Reed DH, Frankham R. 2003. Correlation between Fitness and Genetic Diversity. Conservation Biology 17 (3): 230 – 237. DOI: 10.1046/j.1523-1739.2003.01236.x

Reif JC, Zhang P, Dreisigacker S, Warburton ML, van Ginkel M, Hoisington D, Bohn M, Melchinger AE. 2005. Wheat genetic diversity trends during domestication and breeding. Theoretical and Applied Genetics 110 (5): 859 – 864. DOI 10.1007/s00122-004-1881-8

Resende AG, Vidigal Filho PS, Machado MFPS. 2000. Isozyme Diversity in Cassava Cultivars (*Manihot esculenta* Crantz). Biochemical Genetics 38 (7): 203 - 216. DOI 10.1023/A:1001963319965

Ricci A, Chekhovskiy K, Azhaguvel P, Albertini E, Falcinelli M, Saha M. 2012. Molecular Characterization of *Jatropha curcas* Resources and Identification of Population-Specific Markers. BioEnergy Research 5 (1): 215 - 224. 10.1007/s12155-011-9150-6

Rivarola M, Foster JT, Chan AP, Williams AL, Rice DW, Liu X, Melake-Berhan A, Creasy HH, Puiu D, Rosovitz MJ, Khouri HM, Beckstrom-Sternberg SM, Allan GJ, Keim P, Ravel J, Rabinowicz PD. 2011. Castor Bean Organelle genome sequencing and worldwide genetic diversity analysis. PLoS One 6 (7). e21743. DOI: 10.1371/journal.pone.0021743

Roa AC, Maya MM, Duque MC, Tohme J, Allem AC, Bonierbale MW. 1997. AFLP analysis of relationships among cassava and other *Manihot* species. Theoretical and Applied Genetics 95 (5): 741 - 750. DOI 10.1007/s001220050620

Rodrigues HS, Borem de Oliveira A, Gomes Lopes MT, Cruz CD, Chaves FCM, da Silva Bentes JL. 2013. Genetic diversity of sacha inchi accessions detected by AFLP molecular markers. Revista de Ciencias Agrarias - Amazon Journal of Agricultural and Environmental Sciences: 55-60 (in Portuguese, with abstract in English). doi: http://dx.doi.org/10.4322/rca.2013.081

Rodríguez Á, Corazon-Guivin M, Cachique D, Mejia K, Del Castillo D, Renno JF, García-Dávila C. 2010. Diferenciación morfológica y por ISSR (Inter simple sequence repeats) de especies del género *Plukenetia* (Euphorbiaceae) de la Amazonía peruana: propuesta de una nueva especie Revista Peruana de Biología 17 (39), 325-330.

Rossetto M, McNally J, Henry RJ, Hunter J, Matthes M. 2000. Conservation genetics of an endangered rainforest tree (*Fontainea oraria* - Euphorbiaceae) and implications for closely related species. Conservation Genetics 1 (3): 217 - 229.

Saisug W, Ukoskit K. 2013. Comparative analysis of EST-derived markers for allelic variation in *Jatropha curcas* L. and cross transferability among economically important species of Euphorbiaceae. Genes and Genomics 35 (1): 1 - 12. DOI: 10.1007/s13258-013-0064-x

Sathe SK, Hamaker BR, Sze-Tao KWC, Venkatachalam M. 2002. Isolation, Purification, and Biochemical Characterization of a Novel Water Soluble Protein from Inca Peanut (*Plukenetia volubilis* L.). Journal of Agricultural and Food Chemistry 50, 4906-4908.

Sathe SK, Kshirsagar HH, Sharma GM. 2012. Solubilization, Fractionation, and Electrophoretic Characterization of Inca Peanut (*Plukenetia volubilis* L.) Proteins. Plant Foods and Human Nutrition 67: 247 – 255. DOI 10.1007/s11130-012-0301-5

Sato S, Hirakawa H, Isobe S, Fukai E, Watanabe A, Kato M, Kawashima K, Minami C, Muraki A, Nakazaki N, Takahashi C, Nakayama S, Kishida Y, Kohara M, Yamada M, Tsuruoka H, Sasamoto S, Tabata S, Aizu T, Toyoda A, Shin-I T, Minakuchi Y, Kohara Y, Fujiyama A, Tsuchimoto S, Kajiyama S, Makigano E, Ohmido N, Shibagaki N, Cartagena JA, Wada N, Kohinata T, Atefeh A, Yuasa S, Matsunaga S, Fukui K. 2011. Sequence analysis of the genome of an oil-bearing tree, *Jatropha curcas* L.. DNA Research 18 (1): 65 - 76. DOI: 10.1093/dnares/dsq030

Semagn KB, Ndjiondjop MN. (2006). An overview of molecular marker methods for plants. African Journal Biotechnology 5: 2540–2568.

Shen J, Pinyopusarerk K, Bush D, Chen X. 2012. AFLP-based molecular characterization of 63 populations of *Jatropha curcas* L. grown in provenance trials in China and Vietnam. Biomass and Bioenergy 37: 265 - 274. DOI: 10.1016/j.biombioe.2011.12.003

Shi A, Chen P, Zhang B, Hou A. 2010. Genetic diversity and association analysis of protein and oil content in food-grade soybeans from Asia and the United States. Plant Breeding 129 (3): 250 – 256. DOI 10.1111/j.1439-0523.2010.01766.x

Shilpha J, Silambarasan T, Pandian SK, Ramesh M. 2013. Assessment of genetic diversity in *Solanum trilobatum* L., an important medicinal plant from South India using RADP and ISSR markers. Genetic Resources and Crop Evolution 60: 807-818.

Schlüter, PM, Harris, SA, 2006. Analysis of multilocus fingerprinting data sets containing missing data. Molecular Ecology Notes 6: 569–572.

Siju S, Ismanizan I, Wickneswari R. 2015. Genetic homogeneity in *Jatropha curcas* L. individuals as revealed by microsatellite markers: implication to breeding strategies. Brazilian Journal of Botany: 1 - 15. DOI 10.1007/s40415-014-0117-7

Sinha P, Islam MA, Negi MS, Tripathi SB. 2016. Analysis of genetic diversity and fatty acid composition in a prebreeding material of *Jatropha*. Journal of Plant Biochemistry and Biotechnology 25 (1): 111 - 116. DOI 10.1007/s13562-015-0301-2

Sinha P, Islam MA, Negi MS, Tripathi SB. 2015. Estimation of outcrossing rates in interspecific backcross plants of *Jatropha curcas* (L.) using AFLP and SSR markers. Physiology and Molecular Biology of Plants 21 (4): 605 - 609. DOI 10.1007/s12298-015-0318-y

Siqueira MVBM, Pinheiro TT, Borges A, Valle TL, Zatarim M, Veasey EA. 2010. Microsatellite Polymorphisms in Cassava Landraces from the Cerrado Biome, Mato Grosso do Sul, Brazil. Biochemical Genetics 48 (9): 879 - 895. DOI 10.1007/s10528-010-9369-5

Sirithunya P, Ukoskit K. 2010. Population genetic structure and genetic diversity of *Jatropha curcas* germplasm as investigated by 5′-anchored simple sequence repeat primers. Journal of Crop Science and Biotechnology 13 (3): 147 - 153. DOI 10.1007/s12892-010-0065-0

Somasundaram ST, Kalaiselvam M. 2011. Molecular tools for assessing genetic diversity. International Training Course on Mangroves and Biodiversity, Annamalai University, India, pp. 82–91.

Soto JC, Ortiz JF, Perlaza-Jiménez L, Ximena Vásquez A, Becerra Lopez-Lavalle LA, Mathew B, Léon J, Jimena Bernal A, Ballvora A, López CE. 2015. A genetic map of cassava (*Manihot esculenta* Crantz) with integrated physical mapping of immunity-related genes. BMC Genomics 16:190. DOI 10.1186/s12864-015-1397-4

Standley PC, Steyermark JA. 1949. *Polygalaceae*. In: Flora of Guatemala. Fieldiana, Botany 24 (6), 5-22.

Sumarani GO, Pillai SV, Harisankar P, Sundaresan S. 2004. Isozyme analysis of indigenous cassava germplasm for identification of duplicates. Genetic Resources and Crop Evolution 51 (2): 205 - 209. DOI 10.1023/B:GRES.0000020862.61748.26

Tanya P, Taeprayoon P, Hadkam Y, Srinives P. 2011. Genetic Diversity Among *Jatropha* and *Jatropha*-Related Species Based on ISSR Markers. Plant Molecular Biology Reporter 29 (1): 252 - 264. DOI: 10.1007/s11105-010-0220-2

Tripathi A, Mishra DK, Shukla JK. 2013. Genetic variability, heritability and genetic advance of growth and yield components of Jatropha (*Jatropha curcas* Linn.) genotypes. Trees 27 (4): 1049 - 1060. DOI 10.1007/s00468-013-0856-2

Upadhyay R, Kashyap SP, Tiwari KN, Singh K, Singh M. 2015. Micropropagation of *Phyllanthus fraternus* Webster (Euphorbiaceae) from field-derived shoot tip explant and assessment of its genetic fidelity. Brazilian Journal of Botany 38 (3): 517 - 525. DOI 10.1007/s40415-015-0171-9

Velasco L, Möllers Ch. 2002. Nondestructive assessment of protein content in single seeds of rapeseed (*Brassica napus* L.) by near-infrared reflectance spectroscopy. Euphytica 123: 89-93.

Venkatachalam P, Priya P, Saraswathy Amma CK, Thulaseedharan A. 2005. Identification, cloning and sequence analysis of a dwarf genome-specific RAPD marker in rubber tree [*Hevea brasiliensis* (Muell.) Arg.]. Genetics and Genomics 23 (5): 327 - 332. DOI 10.1007/s00299-004-0833-8

Vijayanand V, Senthil N, Vellaikumar S, Paramathma M. 2009. Genetic diversity of Indian *Jatropha* species as revealed by morphological and ISSR markers. Journal of Crop Science and Biotechnology 12:115. DOI 10.1007/s12892-009-0081-0

de Walt SJ, Siemann E, Rogers WE. 2011. Geographic distribution of genetic variation among native and introduced populations of Chinese tallow tree, *Triadica sebifera* (Euphorbiaceae). American Journal of Botany 98 (7): 1128 - 1138. DOI: 10.3732/ajb.1000297

Wang X, Xu R, Wang R, Liu, A. 2012. Transcriptome analysis of Sacha Inchi (*Plukenetia volubilis* L.) seeds at two developmental stages. BMC Genomics 13:716. DOI 10.1186/1471-2164-13-716

Williams JGK, Hanafey MK, Rafalski JA. Tingey SV. 1992. Genetics analysis using RAPD markers. Methods in Enzymology 260: 335-348.

Williams PC, Sobering D. 1995. How do we do it: A brief summary of the methods we use in developing near infrared calibrations. Near infrared spectroscopy: The future waves (pp. 185–188). Chichester, UK: NIR Publications.

Williams P. 2001. Near-infrared technology: In the agricultural and food industries. American Association of Cereal Chemists. pp. 296. St. Paul, Minn., USA: American Association of Cereal Chemists.

Wong H-L, Yeoh H-H, Lim S-H. 1999. Customisation of AFLP analysis for cassava varietal identification. Phytochemistry 50 (6): 919 - 924. DOI: 10.1016/S0031-9422(98)00628-1

Xia L, Peng K, Yang S, Wenzl P, Carmen de Vicente M, Fregene MAE, Kilian A. 2005. DArT for high-throughput genotyping of Cassava (*Manihot esculenta*) and its wild relatives. Theoretical and Applied Genetics 110 (6): 1092 - 1098. DOI 10.1007/s00122-005-1937-4

Zaldivar ME, Rocha OJ, Aguilar G, Castro L, Castro E, Barrantes R. 2004. Genetic variation of cassava (*Manihot esculenta* Crantz) cultivated by Chibchan Amerindians of Costa Rica. Economic Botany 58 (2): 204 - 213. DOI 10.1663/0013-0001(2004)058[0204:GVOCME]2.0.CO;2

Zhang L, Luo M, You FM, Nevo E, Lu S, Sun D, Peng J. 2015. Development of Microsatellite Markers in Tung Tree (*Vernicia fordii*) Using Cassava Genomic Sequences. Plant Molecular Biology Reporter 33 (4): 893 - 904. DOI 10.1007/s11105-014-0804-3

Zhang Z-H, Tang T, Zhou R-C, Wang Y-G, Jian S-G, Zhong C-R, Shi S-H. 2005. Effects of divergent habitat on genetic structure of population of *Excoecaria agallocha*, a mangrove associate. Acta Genetica Sinica 32 (12): 1286 - 1292. DOI: 0379-4172(2005)12-1286-07

Zietkiewicz E, Rafalski A, Labuda D. 1994. Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. Genomics 20: 176-183.

9. Annexes

Annex 1. The primary data from the laboratory measurements of the total protein and protein fractions contents in the dry matter of the seeds.

Sample	N (%) in DM Kjeldahl	The protein fractions in 1g of meal (% in DM)		
		Albumin - Globulin (extraction by 0.5M NaCl)	Gliadin (extraction by 60% EtOH)	Glutenin (recalculation)
ADO 1	22.223	16.577	0.253	5.393
ADO 2	22.456	15.977	0.261	6.218
ADO 3	23.831	16.153	0.246	7.432
ADO 4	21.524	17.561	0.276	3.687
ADO 5	23.316	16.168	0.268	6.880
ADO 6	23.470	17.411	0.255	5.804
ADO 7	22.663	18.320	0.271	4.072
ADO 8	23.451	15.239	0.316	7.896
ADO 9	22.537	15.231	0.293	7.013
ADO 10	20.268	16.140	0.376	3.752
ADO 11	22.475	14.457	0.334	7.684
ADO 12	22.559	17.998	0.379	4.182
ADO 13	21.241	14.946	0.364	5.931
ADO 14	21.942	15.191	0.349	6.402
ADO 15	22.470	14.916	0.334	7.220
ADO 16	22.777	14.050	0.365	8.362
ADO 17	22.107	14.967	0.308	6.832
ADO 18	22.581	14.037	0.343	8.201
ADO 19	22.999	15.474	0.328	7.197
ADO 20	21.645	17.451	0.350	3.844
SCR 1	22.177	14.322	0.434	7.421
SCR 2	21.338	14.673	0.224	6.441
SCR 3	21.720	13.843	0.336	7.541
SCR 4	22.274	15.062	0.426	6.786
SCR 5	22.028	13.712	0.404	7.912
SCR 6	19.058	14.784	0.268	4.006
SCR 7	19.878	13.734	0.313	5.831
SCR 8	22.571	16.967	0.223	5.381
SCR 9	20.852	14.174	0.238	6.440
SCR 10	20.093	15.082	0.216	4.795
SCR 11	20.272	16.198	0.262	3.812
SCR 12	21.869	15.503	0.284	6.082
SCR 13	22.195	15.181	0.210	6.804
SCR 14	20.058	14.216	0.427	5.415
SCR 15	20.851	14.523	0.307	6.021

Sample	N (%) in DM Kjeldahl	The protein fractions in 1g of meal (% in DM)		
		Albumin - Globulin (extraction by 0.5M NaCl)	Gliadin (extraction by 60% EtOH)	Glutenin (recalculation)
SCR 16	21.180	13.808	0.518	6.854
SCR 17	22.372	14.634	0.406	7.332
SCR 18	21.362	15.212	0.406	5.744
SCR 19	21.946	16.068	0.255	5.623
SCR 20	20.202	16.466	0.383	3.353
SCR 21	22.077	14.817	0.292	6.968
SCR 22	22.310	15.389	0.314	6.607
SCR 23	20.188	15.117	0.404	4.667
MIS 1	22.233	16.508	0.276	5.449
MIS 2	22.880	15.609	0.457	6.814
MIS 3	21.767	16.575	0.457	4.735
MIS 4	22.793	15.999	0.382	6.412
MIS 5	20.650	12.727	0.457	7.466
MIS 6	21.558	14.526	0.359	6.673
MIS 7	20.879	15.309	0.351	5.219
MIS 8	23.732	17.179	0.307	6.246
MIS 9	22.783	16.342	0.277	6.164
MIS 10	21.508	15.783	0.314	5.411
MIS 11	21.667	14.570	0.396	6.701
MIS 12	22.476	13.853	0.396	8.227
MIS 13	21.024	12.438	0.335	8.251
MIS 14	21.619	15.093	0.426	6.100
MIS 15	20.588	16.109	0.381	4.098
MIS 16	21.508	14.794	0.325	6.389
MIS 17	20.291	15.754	0.257	4.280
MIS 18	21.434	16.346	0.227	4.861
MIS 19	21.845	15.595	0.310	5.940
MIS 20	19.297	12.725	0.483	6.089
CHU 1	20.829	15.887	0.233	4.709
CHU 2	19.923	14.559	0.323	5.041
CHU 3	21.746	16.683	0.255	4.808
CHU 4	19.994	15.909	0.330	3.755
CHU 5	20.175	13.561	0.225	6.389
CHU 6	20.759	17.350	0.127	3.282
CHU 7	19.152	14.818	0.112	4.222
CHU 8	22.830	18.107	0.097	4.626
CHU 9	21.332	16.397	0.179	4.756
CHU 10	21.718	15.547	0.097	6.074
CHU 11	18.113	16.701	0.083	1.329
CHU 12	20.481	16.716	0.158	3.607

Sample	N (%) in DM Kjeldahl	The protein fractions in 1g of meal (% in DM)		
		Albumin - Globulin (extraction by 0.5M NaCl)	Gliadin (extraction by 60% EtOH)	Glutenin (recalculation)
CHU 13	22.429	16.400	0.098	5.931
CHU 14	20.058	15.551	0.376	4.131
CHU 15	21.935	14.333	0.443	7.159
CHU 16	21.109	17.351	0.201	3.557
CHU 17	19.130	14.597	0.290	4.243
CHU 18	21.088	14.338	0.290	6.460
CHU 19	23.354	15.908	0.193	7.253
CHU 20	21.112	18.384	0.230	2.498
AUC 1	18.391	15.062	0.192	3.137
AUC 2	18.120	13.855	0.163	4.102
AUC 3	19.953	14.226	0.481	5.246
AUC 4	20.535	14.573	0.407	5.555
AUC 5	19.893	12.818	0.37	6.705
AUC 6	19.485	13.856	0.393	5.236
AUC 7	19.210	13.848	0.415	4.947
AUC 8	19.118	13.200	0.311	5.607
AUC 9	21.069	14.885	0.304	5.880
AUC 10	19.154	13.418	0.333	5.403
AUC 11	19.594	14.716	0.289	4.589
AUC 12	18.132	12.642	0.311	5.179
AUC 13	20.377	14.316	0.185	5.876
AUC 14	20.334	15.315	0.118	4.901
AUC 15	19.802	14.669	0.17	4.963
AUC 16	20.760	15.179	0.184	5.397
AUC 17	17.772	13.163	0.066	4.543
AUC 18	19.120	13.444	0.125	5.551
PAC 1	19.992	14.486	0.135	5.371
PAC 2	20.813	15.820	0.210	4.783
PAC 3	21.163	15.436	0.394	5.333
PAC 4	21.714	15.128	0.253	6.333
PAC 5	21.685	16.818	0.312	4.555
PAC 6	20.482	16.415	0.326	3.741
PAC 7	21.436	15.042	0.267	6.127
PAC 8	22.463	16.363	0.193	5.907
PAC 9	21.157	16.886	0.193	4.078
PAC 10	19.339	16.237	0.230	2.872
PAC 11	21.262	16.728	0.200	4.334
PAC 12	19.809	15.663	0.304	3.842
PAC 13	21.604	16.032	0.275	5.297
PAC 14	19.700	17.132	0.252	2.316

Sample	N (%) in DM Kjeldahl	The protein fractions in 1g of meal (% in DM)		
		Albumin - Globulin (extraction by 0.5M NaCl)	Gliadin (extraction by 60% EtOH)	Glutenin (recalculation)
PAC 15	20.897	16.663	0.327	3.907
PAC 16	20.786	15.993	0.277	4.516
PAC 17	21.159	16.336	0.306	4.517
PAC 18	20.630	17.942	0.336	2.352
PAC 19	20.466	17.419	0.202	2.845
PAC 20	18.917	17.793	0.247	0.877
RAC 1	19.943	13.570	0.403	5.970
RAC 2	22.054	13.945	0.418	7.691
RAC 3	22.246	16.517	0.410	5.319
RAC 4	22.019	14.428	0.358	7.233
RAC 7	18.667	10.222	0.313	8.132
SLU 1	21.647	15.601	0.345	5.701
SLU 2	19.296	9.861	0.616	8.819
PUC 1	19.508	12.839	0.302	6.367
PUC 2	17.582	11.435	0.258	5.889
PUC 3	15.952	13.532	0.294	2.126
PUC 4	18.564	13.501	0.265	4.798
PUC 5	18.030	13.244	0.346	4.440
PUC 6	16.126	11.877	0.295	3.954
PUC 7	18.658	10.941	0.390	7.327
PUC 8	18.446	15.098	0.375	2.973
PUC 9	18.845	13.633	0.375	4.837
PUC 10	17.054	12.532	0.302	4.220
PUC 11	20.003	14.580	0.295	5.127
PUC 12	19.113	14.577	0.321	4.215
PUC 13	17.602	12.784	0.303	4.516
PUC 14	18.539	13.802	0.273	4.464
PUC 15	19.062	14.903	0.354	3.805
PUC 16	17.611	14.216	0.346	3.048
PUC 17	19.098	13.620	0.361	5.117
PUC 18	18.079	12.654	0.457	4.968
PUC 19	16.696	14.099	0.364	2.233
PUC 20	16.767	12.890	0.251	3.626
PUC 21	18.130	12.286	0.258	5.586
2DM 1	18.329	12.919	0.333	5.077
2DM 2	18.602	14.429	0.340	3.833
2DM 3	19.486	15.755	0.296	3.435
2DM 4	19.394	13.844	0.326	5.224
2DM 5	20.472	13.777	0.274	6.421
2DM 6	21.473	14.485	0.266	6.722

Sample	N (%) in DM Kjeldahl	The protein fractions in 1g of meal (% in DM)		
		Albumin - Globulin (extraction by 0.5M NaCl)	Gliadin (extraction by 60% EtOH)	Glutenin (recalculation)
2DM 7	17.956	12.685	0.251	5.020
2DM 8	18.975	14.971	0.369	3.635
2DM 9	19.515	13.895	0.310	5.310
2DM 10	20.128	14.967	0.221	4.940
2DM 11	20.672	15.313	0.231	5.128
2DM 12	20.931	15.328	0.239	5.364
2DM 13	22.801	15.740	0.254	6.807
2DM 14	20.995	15.723	0.239	5.033
2DM 15	21.628	14.992	0.254	6.382
2DM 16	20.717	13.588	0.221	6.908
2DM 17	21.820	17.177	0.258	4.385
2DM 18	20.447	14.170	0.442	5.835
2DM 19	22.247	13.538	0.295	8.414
2DM 20	18.697	13.721	0.317	4.659
2DM 21	21.612	14.649	0.258	6.705
2DM 22	19.361	14.094	0.281	4.986

10. List of author's publications

- **Ocelák, M**.; Hlásná Čepková, P.; Viehmannová, I.; Dvořáková, Z. Huansi, D. C.; Lojka, B. (2015). Genetic Diversity of *Plukenetia volubilis* L. assessed by ISSR markers. Scientia Agriculturae Bohemica 46 (4): 145 – 153 (SRJ) DOI doi: 10.1515/sab-2015-0029
- Křivánková, B.; Hlásná Čepková, P.; **Ocelák, M**.; Juton, G.; Bechyně, M.; Lojka, B. (2012). Preliminary Study of Diversity of *Plukenetia volubilis* Based on the Morphological and Genetic Characteristics. Agricultura tropica et subtropica 45 (3): 140-146. ISSN: 0231-5742.
- Hlásná Čepková, P.; Dvořáček, V.; Viehmannová, I.; **Ocelák M**.; Huansi, D. C.; Lojka, B. (2014). Use of lab-on-a-chip technology in characterisation of seed storage proteins and protein fractions in Inca peanut (*Plukenetia volubilis*) samples., Tropentag 17.-19.09.2014 Czech University of Life Sciences Prague, Czech Republic.
- **Ocelák, M**.; Prohasková, A.; Hlásná Čepková, P.; Viehmannová, I.; Cachique Huansi, D.; Dvořáček, V. (2013). Prediction of protein content in seeds of *Plukenetia volubilis* L. Populations using ft-nir spectroscopy. Mezinárodní Masarykova konference pro doktorandy a mladé vědecké pracovníky 2013 9. – 13. 12. 2013. Hradec Králové, Czech Republic.