

Elucidating Plant Genetic Diversity and Evolution through Bioinformatics: A Review of Selected Studies

Florence Lasalita- Zapico⁺ and Catherine Hazel Aguilar

Science Department, College of Natural Sciences and Mathematics, Mindanao State University-Fatima,
General Santos City, 9500, Philippines

Abstract. This paper discusses the use of bioinformatics in its simplest form to elucidate genetic diversity and evolution in selected crop plants. Data obtained from morphological, biochemical and molecular characterization were subjected to multivariate analyses using a wide array of statistical softwares. Results revealed very low genetic variability for upland rice, Philippine papayas and *Jatropha curcas* provenances studied. On the other hand, the genetic divergence of two *Maguindanao* varieties of *abaca* (Manila hemp) from T'boli, South Cotabato was underscored by principal component and cluster analyses. Morphological and isozyme data also revealed similar features for papaya inbred parents 5648 and Solo raising the possibility that they have phylogenetic affinities and/or common origins. For *Jatropha curcas*, no correlation was observed between spatial dispersal and phenotypic divergence implying that this semi-cultivated plant had previously evolved a wide range of adaptability to diverse conditions prior to its being introduced to other countries from its center of origin in Mexico. For Mt. Hamiguitan ferns, the divergence of *Sticherus hirtus* Copel, *Dipteris conjugata* and *Cyathea spp.* from the rest of the 55 fern species characterized was highlighted. These findings provide invaluable information that can be used in crafting conservation strategies and crop improvement programs that will prioritize the widening of the genetic bases of cultivated crops.

Keywords: bioinformatics, multivariate analysis, statistical software, factor analysis, cluster analysis, genetic diversity.

1. Introduction

While modern agriculture has made possible provision of food for the burgeoning world population today, it has also caused the narrowing of the genetic base of cultivated crops [1, 2]. This genetic uniformity, which is due to human selection, monoculture and evolutionary bottlenecks, results in loss of genes and increases vulnerability of crops to disease epidemics, pest infestation and abiotic stresses. One classic example is shown by the history of the European wine grape, *Vitis vinifera*. This grape variety is the finest in the world and the best vintage wines are produced from it. In the late 19th century, infestation by root-destroying insect Phylloxera, brought the European wine industry to the brink of collapse. One factor that led to this unparalleled disaster was the absence of resistance genes to this insect pest in European wine grapes due to centuries and even millennia of human selection for traits desirable to wine drinkers. It was the native American species of grapes that saved these unique and irreplaceable European varieties. The American species were resistant to Phylloxera and the European grapes were grafted onto rootstocks of their American relatives. Thus, the best wine grapes in the world today exist as genetic mosaics [3, 4]. This is one instance when a plant's diverse genepool has provided the necessary genes to save an agricultural industry.

Diversity studies using various marker systems have supported this general view about narrow genetic diversity in crop plants of today. Using a repertoire of methods such as conventional breeding, marker-assisted selection and genetic transformation, scientists attempt to infuse new genes to the depauperate gene

⁺ Florence L. Zapico. Tel: +63-083-3018349
Email address: florence.zapico@gmail.com

pools of cultivated crops. There are however, some sectors who claim that these new technologies will cause even more harm at the longer term. Horrific scenarios of ‘super bugs’ and bacteria resistant to a broad spectrum of antibiotics are all too possible in the near future due to pest/pathogen coevolution with cultivated crops. When subjected to heightened selective pressure, these pests/pathogens have to evolve in order to survive [5, 6].

2. Bioinformatics in Crop Diversity and Evolution

Bioinformatics is a multidisciplinary field which utilizes various forms of data and analyses to generate biological knowledge. In crop genetics and evolution, various forms of genetic markers viz. morphological, biochemical, cytogenetic and molecular markers are used to unveil genetic diversity in crop plants [7]. Diversity studies using genetic markers reveal polymorphisms and provide insights about crop genetic relationships and phylogenetic origins. Morphological markers rely on phenotypic variations which can be visually scored and measured. These markers are, however, very susceptible to environmental changes and are less reliable unless morphological structures with invariable phenotypic expressions are used. Isozyme polymorphisms result from variations in enzyme/protein structures while cytogenetic characterization detects variations in chromosome structure and numbers. As for molecular markers, variations in DNA sequences (i.e. point mutations) are detected. Furthermore, marker analyses such as these generate large volumes of data which are very cumbersome to analyze individually. Using multivariate analysis, focus shifts from individual factors taken singly to relationships among them. Since more information is simultaneously analyzed, multivariate analysis has greater resolving power, is more informative and makes possible better understanding of relationships between variables of the study.

Among the multivariate methods commonly used in genetic diversity studies are Principal Component Analysis (PCA) and Cluster analysis. PCA is a reducing algorithm which extracts few factors (known as principal components) from a large number of variables while retaining most of the original information. It allows the simultaneous evaluation of numerous variables, reduces variations displayed by organisms to a manageable level and displays the results in 2 or 3-dimensional graphs for easier interpretation and analysis [8]. Cluster analysis on the other hand, organizes sampling observations into discrete groups with high within-group similarity and low among-group similarity [7]. Data obtained from characterization are then run in various statistical platforms to construct similarity/dissimilarity matrices which would then form the bases for formation of scatterplots, phylogenetic trees and dendrograms that will subsequently be used in analysis. In this paper, statistical softwares used were the following: SAS-JMP ver. 8.0.1, SPSS ver 16.0, PASW Statistics ver 18, Kyplot 2b13, Minitab 13.2 and Phylip 3.695.

3. Crop Genetic Diversity and Evolutionary Patterns in Selected Crops

3.1. Abaca (*Musa textilis* L. Nee)

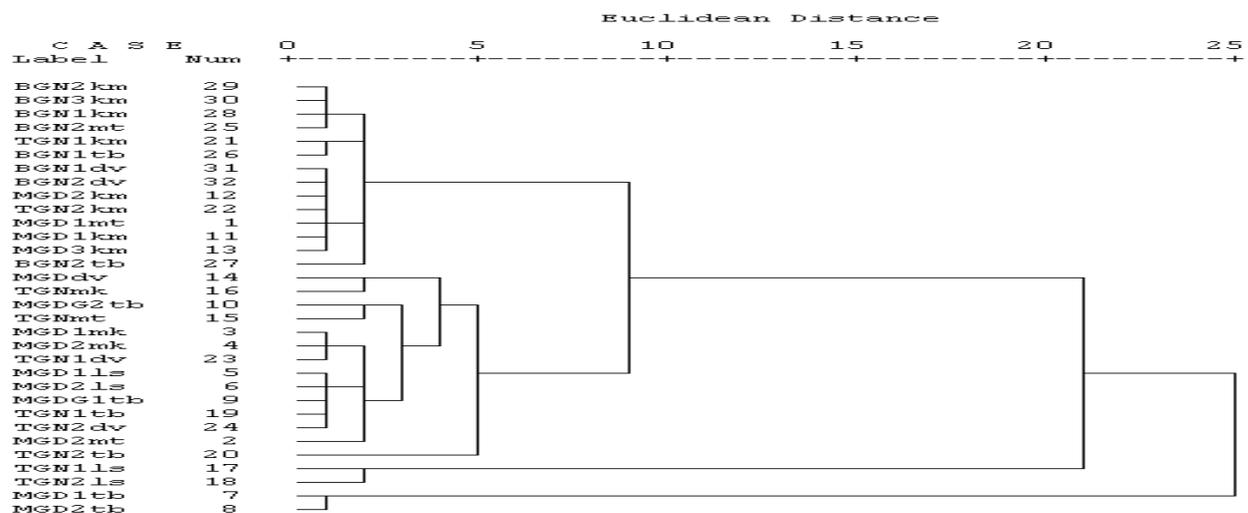


Fig. 1. Dendrogram of 32 Abaca (*Musa textilis* L. Nee) varieties using centroid method

Abaca, known internationally as Manila hemp, is a widely cultivated yet under-researched crop that is endemic to the Philippines. This close relative of banana is grown all over the country for the fiber that can

be extracted from its pseudostem. In this study, 110 abaca cultivars from Southern Mindanao were subjected to morphological characterization using 25 qualitative and 6 quantitative traits [9]. PCA identified pseudostem and leaf characteristics as highly discriminant traits while hierarchical cluster analysis sorted abaca varieties into two big groups with very low intracluster variability but very high intercluster variability. This latter analysis also resulted in the separation of two Maguindanao strains from the Municipality of T'boli implying that these two strains are genetically divergent to other abaca cultivars in Southern Mindanao. These results are borne out by a related study on the multi-locational evaluation of 3 abaca varieties (Maguindanao, Tanggonon and Bongolanon) present in all study sites. It can be seen in Figure 1 that two Maguindanao strains from T'boli (Mdg1tb and Mgd2tb) formed a separate cluster apart from all other abaca cultivars. The very high altitude of the T'boli upland farms as well as their vastly different climatic conditions when compared to those in the lowlands could have resulted to the variant phenotypes exhibited by T'boli abaca. The isolation of these farms due to their remoteness and inaccessibility could also have contributed to the evolution of Maguindanao abaca into morphologically distinct strains. Additional evidence to the proposed genetic divergence of the two Maguindanao varieties is provided by their purplish pseudostem sap color as opposed to the whitish/milky sap of other abaca varieties.

3.2. Upland rice (*Oryza sativa*)

Lake Sebu in South Cotabato is home to the tribal Tbolis who engage in rice farming in the marginal upland areas. Due to lowland influences, traditional rice landraces kept by the Tbolis for many generations are being discarded for hybrid corn and other cash crops. This study was conducted to ascertain the status of genetic diversity of rice cultivars extant in the Lake Sebu upland farms [10]. Sixteen upland rice cultivars and one lowland cultivar (RC 18) were subjected to molecular characterization using Randomly Amplified Polymorphic DNA (RAPD). Gel runs generated 33 distinct bands of which 29 were common to all cultivars implying low variability in the rice germplasm tested (Fig. 2a). Cluster analysis formed two groups and separated RC 18 from the upland cultivars (Fig. 2b). Low intra-cluster variability suggests, albeit tentatively that the rice genepool in the upland farms of Lake Sebu has a narrow genetic base. These findings about the shrinking gene pool of cultivated rice had been reported by various scientific literatures [11, 12]. To unequivocally determine the true state of rice genetic diversity in Lake Sebu, more robust DNA marker systems must be utilized.

Thirty two (32) upland rice landraces cultivated by the tribal *Blaans* in Kihan, Malapatan, Sarangani Province were subjected to morphological characterization using 10 quantitative traits [13]. PCA revealed significant contributions of plant height, panicle length, leaf and culm characteristics to total phenotypic variability. The dendrogram based on morphometric characters resolved the upland rice varieties into two divergent groups with high intracluster similarity (Fig. 3). This very high morphological homogeneity within the two clusters suggests that there were originally two ancestral progenitors which evolved into the present number of rice landraces because of isolation of the farms and the diverse and oftentimes harsh conditions in the uplands. Moreover, apparent distinction was noted in rice cultivars within clusters. For instance, Cluster 1 varieties (*Azucena*, *Mlikat samlaka* etc) were generally tall, high tillering and had high values for leaf length, leaf area and culm length. Cluster 2 varieties, on the other hand, registered moderate to low values for these morphological parameters.

3.3. Papaya (*Carica papaya*)

One serious threat to papaya production worldwide is papaya ringspot virus (PRSV) infestation which can decimate entire papaya plantations. The drastic decline in papaya production in Hawaii during the early 1900s spurred biotechnologists from the University of Hawaii to produce the first transgenic papaya with resistance to PRSV [14]. In the Philippines, plant breeders from the Institute of Plant Breeding (IPB) undertook the incorporation of tolerance to the virus using conventional breeding methods. In this study, the five papaya inbred parents (3 foreign, 2 local) used in the IPB breeding program as well as the products of their intercrosses (hybrids) were subjected to morphological and biochemical characterization [15]. Inbred parent 5648 (from Florida) and Solo (from Hawaii) were found to be genetically similar implying that they have common geographic and phylogenetic origins. Solo germplasm from Hawaii could have been brought to Florida during the expansion of the papaya breeding program to the US mainland during the first part of

the 20th century. On the other hand, isozyme banding patterns revealed incomplete fixation in the inbred parents despite many generations of inbreeding (Fig. 4a). This persistence of heterozygosity in papaya inbred parents was also reported in Caribbean and Venezuelan papayas [16]. Isozyme analysis also yielded 13 loci of which only eight were polymorphic. Dendrograms from morphological and isozyme data (Fig. 4b) underscored phenotypic similarities of 5648 and Solo and the divergence of these two inbreds and their hybrid (5648xSolo) from the other papaya cultivars. This very narrow genetic base of Philippine papayas can be ascribed to the combined effects of migration, human selection for PRSV-tolerant varieties and genetic drift due to small numbers of surviving stands which interbred after widespread PRSV (papaya ringspot virus) infestation in the early 1980s. Similar findings relating to poor genetic diversity in common papaya had been reported by Coppens d'Eeckenbrugge et al [17] in Costa Rica despite the mountain ranges (cordilleras) separating Pacific and Caribbean coasts.

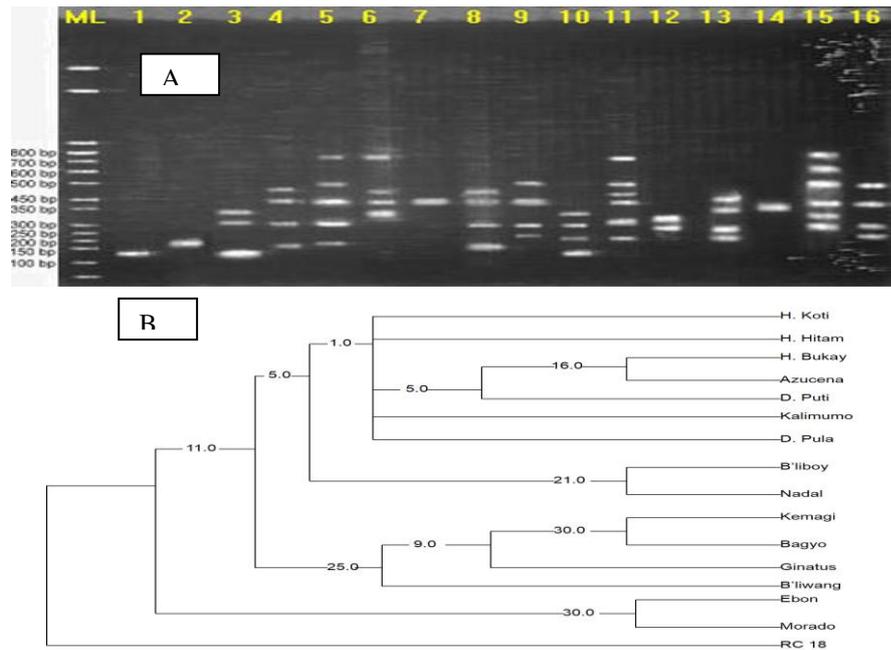


Fig. 2a-b. DNA amplification products of upland rice in Lake Sebu (a) and phylogenetic tree constructed with discrete character parsimony. (Rice Cultivars from Lake Sebu, South Cotabato Province)

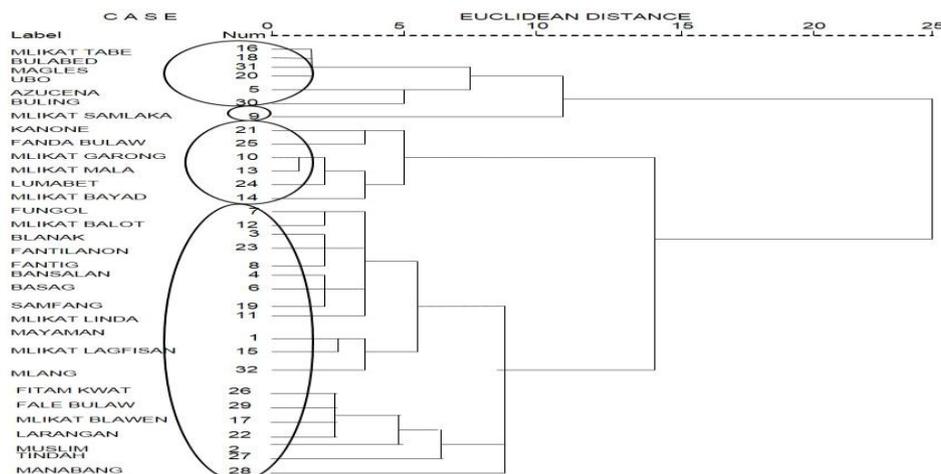


Fig. 3. Dendrogram constructed using UPGMC (Rice Cultivars from Kihan, Malapatan, Sarangani Province)

Source: (2) Zapico FL, Miranda JG and Pare MI 2006. Molecular characterization of selected rice varieties in Lake Sebu, Philippines using RAPD markers. *Proceedings of the International Conference on Biotechnology: Harnessing potentials of biological resources*, Baguio City, 9-11 August 2006. Pp11-12. (3) Zapico FL, Namocatcat JA and Carino-Turner JC (2010). Genetic Diversity Analysis of Traditional Upland Rice Cultivars in Kihan, Malapatan, Sarangani Province, *Philippine Journal of Science* 139(2): 177-18

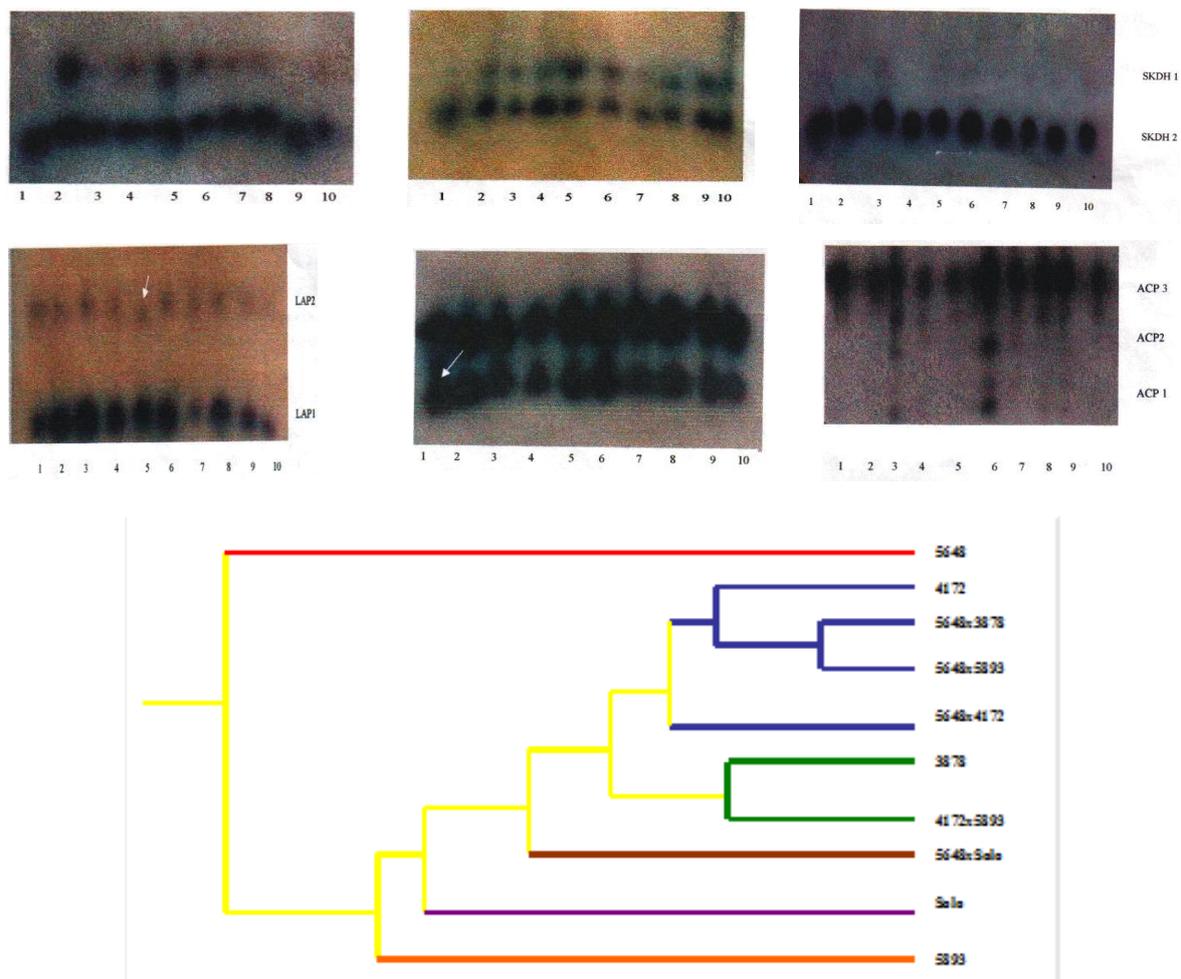


Fig. 4a-b Enzyme zymograms of MDH, SKDH, LAP, GPI AND ACP (a) and UPGMA dendrogram based on isozyme data (b)

Source: Zapico FL and Villegas VN (2010). Morphological and Isozyme Characteristics of Selected Papaya Inbreds and Hybrids. *Asia Life Sciences*. 19(2): 235-245

3.4. Jatropha curcas L.

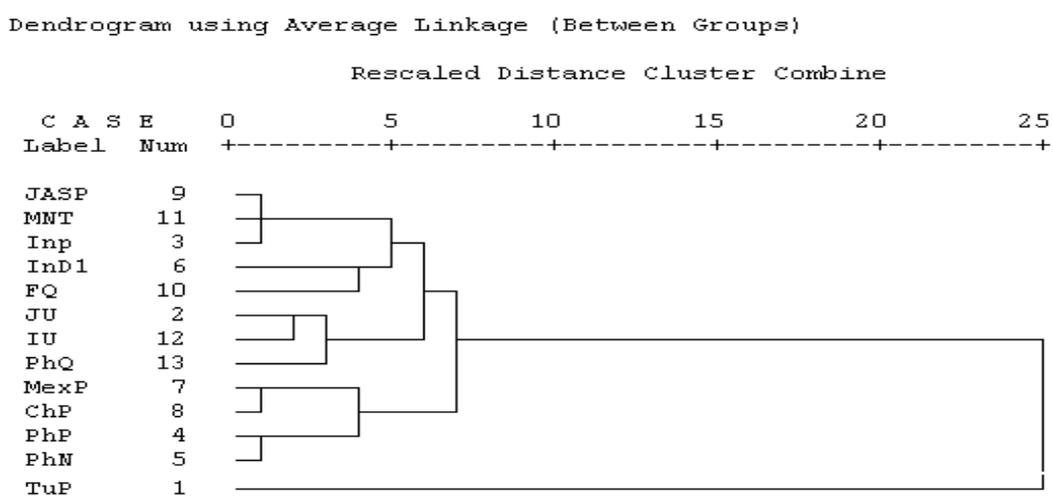


Fig. 5. J. curcas UPGMA dendrogram

Source : Zapico FL, Nival SK, Aguilar CHM and Eroy MN (2011). Phenotypic Diversity of Jatropha curcas L. from Diverse Origins. *Journal of Agricultural Science and Technology* 5(2): 215-219.

Jatropha curcas L. is a hardy plant that is best known for the oil that can be extracted from its seeds. One factor that has hindered optimum utilization of this plant is the toxicity of its seeds and seed oil. In this study, 13 *J. curcas* provenances (6 local; 7 foreign) were subjected to morphological characterization using 21 quantitative traits [18]. In the PCA scatterplot, groupings of provenances were mainly due to plant height and number of leaves. UPGMA (Fig. 5) and UPGMC dendrograms resolved *Jatropha curcas* provenances into two groups and highlighted separation of Mindanao variety *Tubao-Phil* from the rest of the provenances. Moreover, low variability within clusters implies a high level of genetic homogeneity for *J. curcas*. These results are in accordance with the findings of Danquah et al, Kanchanaketu et al and Pamidimarri et al [19]-[21] using molecular markers. Moreover, no association was detected between geographic dispersal and phenotypic divergence of the *Jatropha* provenances studied. Similar results were obtained by Pamidimarri et al and Gohil and Pandya [21, 22]. These findings signify that *J. curcas* had previously undergone evolutionary changes in the past and that presently it possesses a wide range of genes and gene complexes that enable it to survive even in the harshest of conditions.

3.5. Ferns in Hamiguitan mountain range

Mt. Hamiguitan Range is a protected area in Davao Oriental which is known for its diverse array of flora and fauna. A study on fern diversity revealed 55 fern species which were characterized morphologically using 23 qualitative and 21 quantitative traits [8]. PCA underscored contributions of length of lateral pinnae, width of blade, receptacle position and true indusium to total variance. Dendrograms from cluster analysis revealed divergence of *Sticherus hirtus* Copel, *Dipteris conjugata* and *Cyathea spp.* from the rest of the fern varieties. *Sticherus hirtus* Copel occurs widely in the dipterocarp to montane forested zones of Mt. Hamiguitan. Its divergence into a solitary cluster can be ascribed to the highest values it obtained for number of pinnules in the lateral pinnae and in the terminal pinnae. Further discriminant features for this fern species include its large forked leaves and absence of true indusium. *Dipteris conjugata* and *Cyathea spp.*, on the other hand, inhabit open areas in montane forests. Both recorded highest values for petiole length, were the largest and tallest fern species and had extremely large fronds/blades. Finally, studies such as these utilizing morphological markers for diversity assessment in ferns had also been reported by Vasco et al and Creese et al [23, 24].

4. Conclusion

This study has demonstrated the effectiveness of bioinformatics in the elucidation of genetic diversity and phylogenetic origins of various plant species. Through the combined use of different marker systems and powerful statistical softwares, the underlying genetic properties abaca, upland rice, papaya, *Jatropha curcas* and Mt. Hamiguitan ferns were clarified and possible hypotheses as to their evolutionary patterns were tentatively proposed. Notwithstanding the simplicity of these presented studies, baseline information has been generated which can be utilized in the formulation of conservation strategies and in crop improvement programs aimed at broadening the narrow genetic bases of cultivated crops.

5. References

- [1] Y. B. Fu and D. J. Somers. Genome-wide reduction of genetic diversity in wheat breeding. *Crop Science*. 2009.
- [2] M. van de Wouw, C. Kik, T. van Hintum, R. van Teuren and B. Visser B. Genetic erosion in crops: concept, research results and challenges. *Plant Genetic Resources Characterization and Utilization*. 2009. **8**(1): 1-15.
- [3] A. Forneck, S. Kleimann, R. Blaich, and S. F. Anvari. Histochemistry and anatomy of phylloxera (*Daktulosphaira vitifoliae*) nodosities on young roots of grapevine (*Vitis* spp). *Vitis* 2002.**41**(2): 93-97.
- [4] Y. P. Du, H. Zhai, Q. H. Sun, and Z. S. Wang. Susceptibility of Chinese Grapes to grape phylloxera. *Vitis*. 2009. **48**(1): 57-58.
- [5] M. D. Rausher. Co-evolution and plant resistance to natural enemies. *Nature*. 2001. **411**: 857-864.
- [6] F. Gould. The evolutionary potential of crop pests. *American Scientist*. 1991. **79**: 496-507.
- [7] S. A. Mohammadi and B. M. Prasanna. Analysis of genetic diversity in crop plants- salient statistical tools and considerations. *Crop Science*. 2003. **43**: 1235-1248.
- [8] F. L. Zapico, C. H. Aguilar, B. Lon and N. Bagay. Phenotypic variations of extant ferns in Mt. Hamiguitan Range, Mindanao Island, Philippines. *World Applied Sciences Journal*. 2011. **12** (7): 979-987.
- [9] C. H. Aguilar, J. Aujero and F. L. Zapico. Morphological characterization of Abaca (*Musa textilis* L. Nee)

varieties and hybrids in selected areas in Southern Mindanao. Undergraduate thesis. 2010. Unpublished.

- [10] F. L. Zapico, J. G. Miranda and M. I. Pare. Molecular characterization of selected rice varieties in Lake Sebu, Philippines using RAPD markers. Proceedings of the International Conference on Biotechnology: Harnessing potentials of biological resources, Baguio City, 9-11 August 2006. pp. 11-12.
- [11] M. S. Lin. Changes in Genetic diversity in the Japonica rice region of Taiwan. *Botanical Bulletin of Academia Sinica*. 1993. **34**: 307-312.
- [12] S. A. Ogunbayo, D. K. Ojo, R.G. Guie, O. O. Oeylaki and K. A. Sanni. Phylogenetic diversity and relationships among 40 rice accessions using morphological and RAPD techniques. *African Journal of Biotechnology*. 2005. **4**: 1234-1244.
- [13] F. L. Zapico, J. A. Namocatcat and J. C. Carino-Turner. Genetic diversity analysis of traditional Upland Rice cultivars in Kihan, Malapatan, Sarangani Province, Philippines Using Morphometric Markers. *Philippine Journal of Science*. 2010. **139**(2): 177-180.
- [14] D. Gonsalves, C. Gonsalves, S. Ferreira, K. Pitz, R. Manshardt and J. Slighton. Transgenic virus resistant Papaya: From hope to reality for controlling Papaya Ringspot virus in Hawaii. APSnet Feature. July 2004
- [15] F. L. Zapico and V. N. Villegas. Morphological and Isozyme characteristics of Papaya inbreds and hybrids. *Asia Life Sciences*. 2010. **19**(2): 235-245.
- [16] J. Ocampo, G. C. d'Eeckenbrugge, S. Bruyere, L. D. de Bellaire and P. Ollitraul. Organization of morphological and genetic diversity of Caribbean and Venezuelan papaya germplasm. *Fruits*. 2006. **61**: 25-37.
- [17] G. C. d'Eeckenbrugge, M. T. Restrepo and D Jimenez. Morphological and Isozyme characterization of common Papaya in Costa Rica. *Acta Hort. (ISHS)* 740. 2007. pp.109-120.
- [18] F. L. Zapico, S. K. Nival, C. H. M. Aguilar and M. N. Eroy. Phenotypic diversity of *Jatropha Curcas* L. from diverse origins. *Journal of Agricultural Science and Technology*. 2011. **5**(2): 215-219.
- [19] E. O. Danquah, R. Akromah, S. J. Quashie-Sam, W. Oduro, D. Falk, S. V. Thevathasan and A. M. Godron. The genetic diversity of *Jatropha curcas* (L.) in Ghana as revealed by RAPD primers. *Agroforestry Systems*. 2012. **86**: 443-450.
- [20] T. Kanchanaketu, N. Sangduen, T. Toojinda and V. Hongtrakul. Genetic diversity analysis of *Jatropha curcas* L. (Euphorbiaceae) based on methylation-sensitive amplification polymorphism. *Genetic Molecular Research*. 2012. **11**(2): 944-955.
- [21] D.V. Pamidimarri, S. G. Mastan, H. Rahman and M. P. Reddy. Molecular Characterization and Genetic diversity analysis of *Jatropha curcas* L. in India using RAPD and AFLP analysis. *Molecular Biology Reporter*. 2010. **37** (5): 2249-2257.
- [22] R. H. Gohil and J. B. Pandya, Genetic diversity assessment in physic nut (*Jatropha curcas* L.), *International Journal of Plant Production*. 2008. pp. 2-4.
- [23] A. Vasco, R. C. Moran and B. A. Ambrose. The evolution, morphology and development of fern leaves. *Frontiers in Plant evolution and development*.4:345.doi:10.3389/fpls.2013.00345. 2013.
- [24] C. Creese, A. Lee and L. Sack. Drivers of morphological diversity and distribution in the Hawaiian fern flora: Trait associations with size, growth form and environment. *American Journal of Botany*. 2011. **98**(6): 956-966.