Original Article

Multivariate analysis of five chicken breed in Indonesia based on microsatellite allele frequency

Ferdy Saputra *, Tike Sartika, Anneke Anggraeni, Andi Baso Lompengeng Ishak, Komarudin, Nurul Pratiwi

Indonesian Research Institute for Animal Production
*Correspondence: ferdysaputra@pertanian.go.id; ferdy44saputra@gmail.com

Abstrak

Tujuan: Penelitian ini mencoba menguji beberapa metode multivariat dalam mengklasifikasikan keanekaragaman genetik menggunakan data frekuensi alel mikrosatelit.

Metode: Penelitian ini menggunakan data frekuensi alel mikrosatelit dari ayam White Leghorn (n = 48), Kampung (n = 48), Pelung (n = 24), Sentul (n = 24), dan Kedu Hitam (n = 25) yang berasal dari Balai Penelitian Ternak. Data frekuensi alel dianalisis dengan metode Neighbor-Joining (NJ) menggunakan program POPTREE2. Data tersebut juga dianalisis dengan metode *Principal Component Analysis* (PCA), *Correspondence Analysis* (CA), and *Hierarchical Clustering on Principal Components* (HCPC) menggunakan package factoextra dan FactoMineR dalam program R 4.0.0.

Hasil: *Correspondence Analysis* (CA) menemukan Sentul lebih dekat terhadap Kedu Hitam. Namun berdasarkan NJ, PCA, dan HCPC menunjukkan Sentul lebih dekat dengan Kampung. Berdasarkan nilai Dimensi 1, Analisis Korespondensi (80,7%) dapat menjelaskan variasi yang lebih besar daripada PCA (58,9%). Namun, metode CA memberikan hasil yang berbeda dibandingkan dengan NJ, PCA, dan HCPC. NJ, PCA, dan HCPC menemukan empat cluster ayam, yaitu cluster 1 (White Leghorn), cluster 2 (Pelung), cluster 3 (Kedu Hitam), dan cluster 4 (Kampung dan Sentul).

Kesimpulan: Disimpulkan HCPC merupakan metode multivariate yang lebih baik untuk menganalisis data frekuensi alel daripada PCA dan CA. HCPC dapat digunakan untuk menganalisis data frekuensi alel lebih baik daripada PCA, karena HCPC merupakan kombinasi metode dari clustering hierarkis dan komponen utama.

Kata Kunci: Analisis multivariat; Ayam Indonesia; Frekuensi alel; Mikrosatelit

Abstract

Objective: This study tries to examine several multivariate methods in classifying genetic diversity using microsatellite allele frequency data.

Methods: This study used microsatellite allele frequency data from White Leghorn (n = 48), Kampung (n = 48), Pelung (n = 24), Sentul (n = 24), and Black Kedu (n = 25) from Indonesian Research Institute for Animal Production. Allele frequency data were analyzed by the Neighbor-Joining (NJ) method using the POPTREE2 program. The data was also analyzed by the Principal Component Analysis (PCA), Correspondence Analysis (CA), and Hierarchical Clustering on Principal Components (HCPC) methods using the factoextra and FactoMineR packages in the R 4.0.0 program.

Results: Correspondence Analysis (CA) found Sentul is more closer to Black Kedu. However, based on NJ, PCA, and HCPC showed Sentul is closer to Kampung. Based on the value of Dimension 1,

Correspondence Analysis (80.7%) can explain greater variation than PCA (58.9%). However, CA method generated different results compared to NJ, PCA, and HCPC. NJ, PCA, and HCPC found four chicken clusters, namely cluster 1 (White Leghorn), cluster 2 (Pelung), cluster 3 (Black Kedu), and cluster 4 (Kampung and Sentul).

Conclusions: In conclusion, HCPC is a better multivariate method for analyzing allele frequency data than PCA and CA. HCPC can be used to analyze allele frequency data better than PCA, because HCPC is a combination of methods from hierarchical clustering and principal components.

Keywords: Multivariate analysis; Indonesian chickens; Frequency alleles; Microsatellites

INTRODUCTION

Chicken is the most livestock raised at home. Chicken meat and eggs are the most popular livestock products. The poultry business model is usually used as a poverty alleviation program as practised in Indonesia [1]. Furthermore, Indonesia has many local chicken genetic resources aimed at egg production, meat production, and ornamental animals. According to Nataamijaya [2], Indonesia has 32 native chicken breeds such as Kampung, Pelung, White kedu, Black kedu, Sentul, and Balenggek with various uses such as fighting, egg production, meat production, ornamental, traditional medicine, and there are also classified as endangered animals. In Indonesia, native chicken breeds are based on demographic differences so that genetic information needs to be studied further.

Genetic differences cause variations in phenotypes. Variations in phenotypes do not arise due to new mutations but can appear later from alleles that are segregating in the population [3]. Genetic diversity can be studied using molecular technology. Molecular markers commonly used as indicators of diversity are mitochondria DNA, Y chromosome, and microsatellite. Microsatellites were highly polymorphic due to their instability [4]. Because of its polymorphic nature, microsatellite is used as an excellent marker to determine genetic diversity within populations and between populations [5].

One indicator to study genetic diversity is to look at frequency alleles. Allele frequency is used to calculate the genetic distance of a population. It is from this genetic distance that a dendogram is built using methods known as Neighbor-Joining, UPGMA, Minimum Evolution, and Fitch–Margoliash [6]. Another

method that can explore genetic diversity is multivariate methods. Multivariate methods can summarize the genetic variability without making assumptions about an evolution model, the absence of linkage disequilibrium, and it does not rely on Hardy–Weinberg equilibrium [7]. Principal Component Analysis (PCA) is a multivariate method used in analyzing microsatellite data on Indonesian animal genetic resources and is able to distinguish between species of cattle[8]. The purpose of this study was to examine several multivariate methods in classifying genetic diversity, especially in Indonesian chickens and provides recommendations of multivariate methods that can be used to see genetic differences based on allele frequency data.

MATERIALS AND METHODS

Data analysis

Allele frequency data used are derived from Sartika [9]. The samples used were White leghorn (n = 48), Kampung (n = 48), Pelung (n = 24), Sentul (n = 24), and Black kedu (n = 25) from Indonesian Research Institute for Animal Production. The data were analyzed by the POPTREE2 [10] program to create Neighbor-Joining (NJ) trees with Nei's standard genetic distance (DST). Allele frequency data were also analyzed using three multivariate methods, namely PCA, Correspondence Analysis (CA), and Hierarchical Clustering on Principal Components (HCPC). Both analyzes were carried out using the factoextra [11] and FactoMineR [12] packages in the R 4.0.0 [13].

RESULT

Frequency allele data from microsatellite can be used by multivariate methods. The

results generated by Principle Component Analysis (PCA) and Hierarchical Clustering on Principal Components (HCPC) were the same as the results of Neighbor-Joining (NJ) (Figure 1-3). Kampung Chicken has a close relationship with Sentul chickens, while Black Kedu, Pelung, and White Leghorn chickens form their clusters. Using Correspondence Analysis (CA), Sentul is more relative to Black Kedu (Figure 4). Based on the value of dim1 (Dimension 1), Correspondence Analysis (80.7%) can explain greater variation than PCA (58.9%). However, CA result is different from that generated by NJ, PCA, and HCPC. NJ, PCA, and HCPC found four clusters, namely cluster 1 (White Leghorn), cluster 2 (Pelung), cluster 3 (Black Kedu), and cluster 4 (Kampung and Sentul).

DISCUSSION

Kampung and Sentul chickens are in one cluster. This happens because the Sentul chicken is a Kampung chicken from the Ciamis Regency, which is used as a producer of meat and eggs [14]. Black kedu formed its cluster. Black kedu originated from the district of Temanggung Central Java which is one of Indonesia's rare chickens and has a high potential for egg production among native chickens [15]. According to Dharmayanthi et al [16], Black Kedu chicken had a close relationship with Silkie chickens based on *EDN3* gene. Therefore, Black kedu formed its cluster. Pelung chicken is a chicken that was developed in Cianjur [17]. Asmara et al [18] expected Red jungle fowl (Gallus gallus) was the ancestor of Pelung but, there is no data regarding the pedigree of this chicken. From the results obtained in this study, it is suspected that Pelung chickens are the result of cross-breeding. However, to support this hypothesis, comprehensive research must be carried out to see the genetic structure of pelung chickens. On the other hand, White Leghorn origin from Tuscany, Italy which was developed in the United States and Europe. White Leghorn have eye size and focal length match the larger body size compared to Red Junglefowls [19]. However, White Leghorn showed less intense fear-induced behaviours than the ancestors (Red Junglefowls) [20]. So, Less fear is characteristics of more-domesticated chicken breeds.
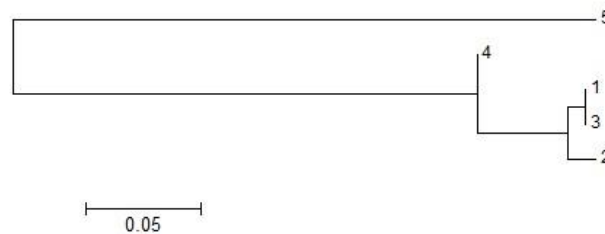
Figure 1. Neighbor-Joining based DST (corrected) (1: Kampung, 2: Pelung, 3: Sentul, 4: Black Kedu, 5: White Leghorn)
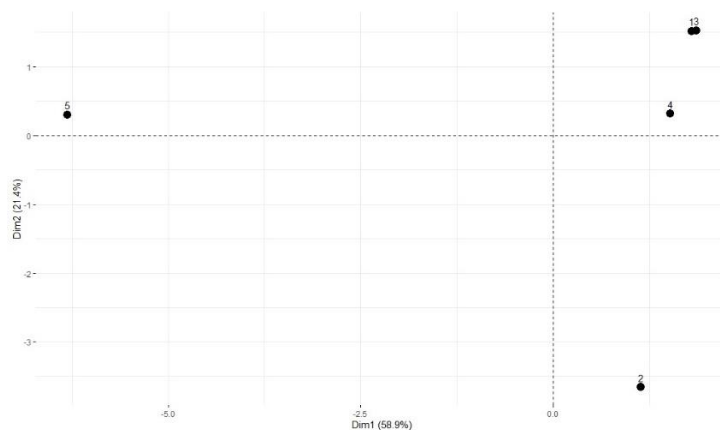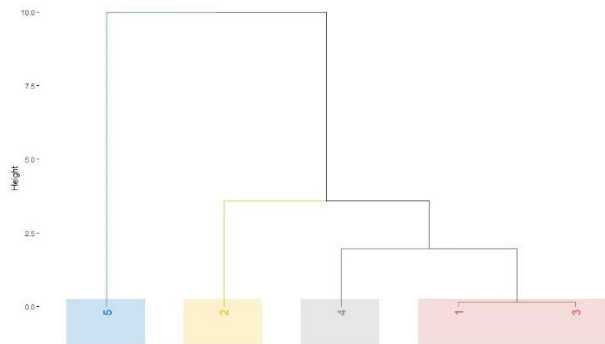

Figure 2. Principal Component Analysis (1: Kampung, 2: Pelung, 3: Sentul, 4: Black Kedu, 5: White Leghorn)

The Neighbor-Joining was a well known distance-based phylogenetic strategy that computes a tree metric from a dissimilarity from biological data [21]. The UPGMA method assumes that all taxa have constant evolutionary rates. to build a more accurate phylogenetic trees, the Neighbor Joining (NJ) method can be used [6]. When compared, UPGMA has a rooted tree type, while NJ has an unrooted tree type. In this study, using NJ with previous study has the same results. The Principal Component Analysis (PCA) and Hierarchical Clustering on Principal Components (HCPC) results showed similarities to NJ except Correspondence Analysis (CA). CA showed the closeness between Sentul and Black Kedu, but other results (NJ, PCA, and HCPC) showed Sentul closer to Kampung. Based on morphology, Sentul chickens have high similarities with native chickens [22]. This shows that sentul is a native chicken raised in Ciamis with certain criteria such as feather color.

The Principal Component Analysis (PCA) method is widely used to find out admixture in a population [23]. The advantages of PCA are identifying genetic structures in enormous datasets in computational time that can be ignored, and the absence of assumptions about the genetic model of the underlying population [24]. PCA itself has been widely used to analyze microsatellite data, and in this study, it can be used to analyze allele frequency data and have similar results with the NJ method. PCA has been carried out to analyze chicken genome data used to determine genetic structure in the population [25; 26; 27].

Correspondence Analysis (CA) is a method for visualizing the relationship between observations and variables by allowing their partitions to be sets that are interconnected, thus revealing which hypotheses can be proposed to help lead to the discovery [28]. Furthermore, CA is used to visualize relationships between genes for finding genetic links. In this study, CA results were different from NJ, PCA, and HCPC. Hence, CA analysis is not suitable for use in analyzing frequency allele data. Even though,

Figure 3. Hierarchical Clustering on Principal Components (1: Kampung, 2: Pelung, 3: Sentul, 4: Black Kedu, 5: White Leghorn)
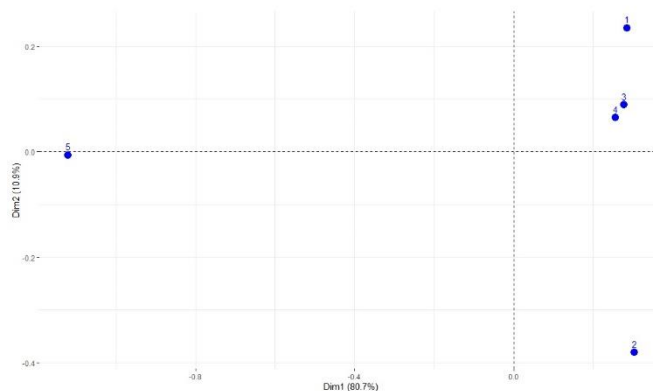
Figure 4. Correspondence Analysis (1: Kampung, 2: Pelung, 3: Sentul, 4: Black Kedu, 5: White Leghorn)

in dimension 1 the explainable variation of CA is higher than PCA, CA is more suitable for categorical data. CA is usually used to analyze two-way data tables, including several sizes of relationships between rows and columns [29]. Hierarchical Clustering on Principal Components (HCPC) is a combination of Hierarchical Clustering with Principal Components of the PCA model [30]. HCPC is objective grouping techniques on the results of principal component analysis, which leads to better cluster solutions [29]. Therefore, the HCPC results are similar to PCA and the addition of clusters in the PCA results.

CONCLUSION

The Hierarchical Clustering on Principal Components (HCPC) results found four clusters, namely, cluster 1 (White Leghorn), cluster 2 (Pelung), cluster 3 (Black Kedu), and cluster 4 (Kampung and Sentul). HCPC, PCA and NJ produced similar results compared to CA. HCPC can be used to analyze allele frequency data better than PCA, because HCPC is a combination of methods from hierarchical clustering and principal components.

CONFLICT OF INTEREST

The authors declare no conflict of interest with any financial organization regarding the material discussed in the manuscript.

REFERENCES

1. Rusdiana, S., and T. Sartika. 2020. Peningkatan Nilai Ekonomi Pada RTM (Rumah Tangga Miskin) Melalui Program Bekerja Kementerian Pertanian. SOCA J. Sos. Ekon. Pertan. 14:146–157. doi: 10.24843/SOCA.2020.v14.i01.p12.

2. Nataamijaya, A. G. 2000. The Native Chicken Of Indonesia. Bull. Plasma Nutfah. 6:1–6.

3. Orgogozo, V., B. Morizot, and A. Martin. 2015. The differential view of genotype-phenotype relationships. Front. Genet. 6:179. doi:10.3389/fgene.2015.00179.

4. Vieira, M. L., L. Santini, A. L. Diniz, and C. Munhoz. 2016. Microsatellite markers: what they mean and why they are so useful. Genet. Mol. Biol. 39:312–328. doi:10.1590/1678-4685-GMB-2016-0027.

5. Abdul-Muneer, P. M. 2014. Application of Microsatellite Markers in Conservation Genetics and Fisheries Management: Recent Advances in Population Structure Analysis and Conservation Strategies. Genetics Research International. 2014: 691759. doi:10.1155/2014/691759

6. Jin Xiong. 2006. Essential Bioinformatics. Cambridge University Press. New York.

7. Jombart, T. 2008. Adegenet: A R package for the multivariate analysis of genetic markers. Bioinformatics. 11:1403–1405. doi:10.1093/bioinformatics/btn129.

8. Agung, P. P., F. Saputra, M. S. A. Zein, A. S. Wulandari, W. P. B. Putra, S. Said, and J. Jakaria. 2019. Genetic diversity of Indonesian cattle breeds based on microsatellite markers. Asian-Australasian J. Anim. Sci. 32:467–476. doi:10.5713/ajas.18.0283.

9. Sartika, T., S. Iskandar, L. Prasetyo, H. Takahashi, and M. Mitsuru. 2004. Kekerabatan Genetik Ayam Kampung, Pelung, Sentul dan Kedu Hitam dengan Menggunakan Penanda DNA Mikrosatelit: I. Grup Pemetaan pada Makro Kromosom. JITV. 9:81–86.

10. Takezaki, N., M. Nei, and K. Tamura. 2010. POPTREE2: Software for Constructing Population Trees from Allele Frequency Data and Computing Other Population Statistics with Windows Interface. Mol. Biol. Evol. 27:747–752. doi:10.1093/molbev/msp312.

11. Kassambara, A., and F. Mundt. 2020. factoextra: Extract and Visualize the Results of Multivariate Data Analyses. Available from: https://cran.r-project.org/package=factoextra

12. Husson, F., J. Josse, S. Le, and J. Mazet. 2015. FactoMineR: Multivariate Exploratory Data Analysis and Data Mining. R package version 1.31.4. 2015. 24:1–18.

13. R Core, T. 2020. R: A Language and Environment for Statistical Computing.

Available from: https://www.r-project.org/

14. Sudrajat, and A. Y. Isyanto. 2018. Keragaan peternakan ayam sentul di kabupaten Ciamis. J. Pemikir. Masy. Ilmuah Berwawasan Agribisnis. 4:237–253.

15. Nataamijaya, A. G. 2008. Karakteristik Dan Produktivitas Ayam Kedu Hitam. Buletin Plasma Nutfah. 14:85-89. doi:10.21082/ blpn.v14n2.2008.p85-89

16. Dharmayanthi, A. B., Y. Terai, S. Sulandari, M. S. A. Zein, T. Akiyama, and Y. Satta. 2017. The origin and evolution of fibromelanosis in domesticated chickens: Genomic comparison of Indonesian Cemani and Chinese Silkie breeds. PLoS One. 12:e0173147. doi:10.1371/journal.pone .0173147

17. Nataamijaya, A. G. 2010. Pengembangan Potensi Ayam Lokal Untuk Menunjang Peningkatan Kesejahteraan Petani. Jurnal Litbang Pertanian, 29: 131-138.

18. Asmara, I. Y., D. Garnida, W. Tanwiriah, R. Partasasmita. 2019. Qualitative morphological diversity of female Pelung Chickens in West Java, Indonesia. Biodiversitas 20: 126-133. doi:10.13057/ biodiv/d200115

19. Roth, L. S. V., and O. Lind. 2013. The Impact of Domestication on the Chicken Optical Apparatus. PLoS One. 8:e65509. doi:10.1371/journal.pone.0065509.

20. Abe, H., K. Nagao, A. Nakamura, M. Inoue-Murayama. 2013. Differences in responses to repeated fear-relevant stimuli between Nagoya and White Leghorn chicks. Behavioural Processes 99: 95– 99. doi:10.1016/j.beproc.2013.07.004

21. Davidson, R., and A. M. del Campo. 2020. Combinatorial and Computational Investigations of Neighbor-Joining Bias. Frontiers in Genetics. 11: 584785. doi: 10.3389/fgene.2020.584785

22. Mariandayani, H. N., D. D. Solihin, S. Sulandari, and C. Sumantri. 2013. Keragaman Fenotipik dan Pendugaan Jarak Genetik pada Ayam Lokal dan Ayam Broiler Menggunakan Analisis Morfologi. Jurnal Veteriner. 4: 475-484.

23. Ma, J., and C. I. Amos. 2012. Principal components analysis of population admixture. PLoS One. 7:e40115. doi:10.1371/journal.pone.0040115.

24. Jombart, T., S. Devillard, and F. Balloux. 2010. Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. BMC Genet. 11:94. doi:10.1186 /1471-2156-11-94.

25. Gärke, C., F. Ytournel, B. Bed'Hom, I. Gut, M. Lathrop, S. Weigend, and H. Simianer. 2012. Comparison of SNPs and microsatellites for assessing the genetic structure of chicken populations. Anim. Genet. 43:419–428. doi:10.1111/j.1365-2052. 2011.02284.x.

26. Kumar, V., S. Kumar Shukla, S. Kumar Singh, G. Goyal, J. Mathew, and D. Sharma. 2015. Genetic Diversity and Phylogenetic Relationship Analysis between Red Jungle Fowl and Domestic Chicken using AFLP Markers. J. Poult. Sci. 52:94–100. doi: 10.2141/ jpsa.0140145.

27. Nie, C., P. Almeida, Y. Jia, H. Bao, Z. Ning, and L. Qu. 2019. Genome-Wide Single-Nucleotide Polymorphism Data Unveil Admixture of Chinese Indigenous Chicken Breeds with Commercial Breeds. Genome Biol. Evol. 11:1847–1856. doi:10.1093/ gbe/evz128.

28. Tekaia, F. 2016. Genome Data Exploration Using Correspondence Analysis. Bioinform. Biol. Insights. 10:59–72. doi: 10.4137/BBI.S39614.

29. Argüellesa, M., C. Benavidesa, and I. Fernándeza. 2014. A new approach to the identification of regional clusters: hierarchical clustering on principal components. Appl. Econ. 46:2511–2519. doi:10.1080/00036846.2014.904491.

30. Chanana, S., C. S. Thomas, F. Zhang, S. R. Rajski, and T. S. Bugni. 2020. *hcapca*: Automated Hierarchical Clustering and Principal Component Analysis of Large Metabolomic Datasets in R. Metabolites. 10: 297. doi:10.3390/metabo10070297.