VOL. 15, NO. 8, APRIL 2020 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

PREDİCTİON OF *Sitophilus oryzae* GENDER USİNG LOGİSTİC REGRESSİON MODEL

Tun Mohd Firdaus Azis[1,2] and Khairul Farihan Kasim[2]
[1]Faculty of Applied Sciences, UiTM Cawangan Perlis, Arau, Perlis, Malaysia
[2]School of Bioprocess Engineering, Universiti Malaysia Perlis, Kompleks Pusat Pengajian Jejawi, Arau, Perlis, Malaysia
E-Mail: tunmohdfirdausbinazis@gmail.com

ABSTRACT

Sitophilus oryzae gender been discriminated using rostrum size and confirmation require the dissection of internal reproductive organ. This dissection technique is destructive and difficult to perform. Thus, the objective of this study is to enhance gender identification process by developing the model using Logistic Regression. This model been develop based on the numerical information of *S. oryzae* morphological features *per se*. The results show that rostrum measurement (length and width) was adequate to be used in developing a model for identification and prediction of *S. oryzae* gender with 100% correct classification.

Keywords:*Sitophilus oryzae*, logistic regression, gender identification.

1. INTRODUCTION

Sitophilus oryzae is one of the most severe stored grain pests in the world. This pest had been recorded to attack wheat, corn, oats, rye, barley, sorghum, buckwheat, dried beans, cashew nuts, wild bird seed, cereal products, and macaroni [1]. Besides that, this pest also causes a massive problem to rice storage because of it offers sufficient food. Thus, the pest may develop and reproduce in the rice grain [2], [3]. The larvae of this pest develop internally in the grain and eat the inside of the grain up to 60% of the individual kernel weight [4]. As the results to their behaviour, life cycle and habitat in the rice storage, it caused to drop of commercial value, seed germination and fungal infestation to the rice [1], [5].

At present, the fumigation of stored grain insect pests is the standard method to control this insect pest that relies on methyl bromide and phosphine [6]. However, the wide usage of pesticide increases the pesticide resistance, harm to human and cause environmental pollution [7]. For example, the use of methyl bromide was banned by the European since January 2005 due to its depleting properties on the stratospheric ozone layer [6].

To overcome the use of pesticides, Insect Pest Management (IPM) suggests a holistic approach to pest control that seeks to optimise the use of a combination of methods to manage a whole spectrum of pest within a particular cropping system. The concept of IPM was based on restricting pesticide use through the use of economic thresholds and utilisation of alternative control options such as biological products or methods [7].

In pest population assessment, gender proportion in the population is crucial information to understand the population growth. Female proportion of *S. oryzae* in the population could badly affect the grain compared to male because of their ability to lay about 150 eggs in their lifetime inside the grain [1]. The eggs will develop into larvae inside the grain and further damage the rice grain [4], [8], [9].

Understanding the male density is also important especially during sampling and monitoring. According to [10], male density is one of important features that affecting the reproductive fitness of insect. The study conducted by [11] shows that female genital tract was more likely to be damaged when male density increased. This damage then will lead to disruption of the female reproduction system and affect the population of the next generation.

Sex ratio and the total fertility rate is essential to understand the population growth, whether the pest population is decreasing, increasing or in stable condition [12], [13]. If the number of female offspring is larger than a previous female generation, it shows that the population is in increasing condition and vice versa. Therefore, it is to identify the males and females of *S. oryzae* to understand and forecast its population growth.

However, at present, there is no specific model that can be used to differentiate the gender of *S. oryzae*. According to [1] males and females of *S. oryzae* are similar in appearance but the males can be identified by having shorter, wider and irregular indentations on their rostra. This information is not definite numerically and not proven statistically. Therefore, this study focused on developing a model to differentiate *S. oryzae* gender based on its morphological characteristics.

2. METHODOLOGY

2.1 Culture of *S. oryzae*

The starter culture of *S. oryzae* was collected from infested stored rice. It was reared and maintained in a plastic container (13 x 10 x 20 cm) contained 2 kg of rice with 12.50% moisture content at a temperature between 15 to 34 °C and relative humidity between 58 to 89%. The container was properly closed and new rice was introduced periodically for proper development of the *S. oryzae*. This experiment was conducted at Biological Research Laboratory, Universiti Teknologi MARA Cawangan Perlis, Malaysia during 2017.

2.2 Morphological analysis

Total of 45 adults of *S. oryzae* was collected from the starter culture and was killed by using warm water.

The morphological features of *S. oryzae* i.e. rostrum length and width, prothorax length and width, club length and width, and abdomen width at third and last sternite were measured using Olympus SZX7 Zoom Stereomicroscope [14]. Confirmation of the gender was done by dissection of the abdomen part; the existence of aedeagus indicated the male, while the absence of aedeagus indicated female *S. oryzae*.

2.3 Prediction of *S. oryzae* gender using Logistic Regression Model

All the morphological features of *S. oryzae* were used in the prediction of *S. oryzae* gender. The gender data (male and female) were considered as binary response and the association between the binary response and set of explanatory variables i.e. *S. oryzae* morphological features were analysed using Logistic Regression (LR) Model in XLSTAT 2017: Data Analysis and Statistical Solution for Microsoft Excel Addinsoft, Paris, France (2017). A total of 45 *S. oryzae* were used in this study, where 35 had been used in developing the model and been through the classification estimation. While, another 10, does not include in the model but as classification validation sample. Five models have been tested. Model 1 consists of all combination of eight (8) morphological features in creating the model. Then, one by one dataset had been eliminated to observe the performance in differentiating the gender of the *S. oryzae* using the lesser morphological feature measurement. Model 2 consists of only rostrum width and length, Model 3 consists of only prothorax length and width, Model 4 consists of only third and last abdomen length and Model 5 consists of club width and length. Three approaches had been used in assessing the overall fit of the model i.e. i) based on the statistical measures of overall fit; ii) pseudo R^2 measured and, iii) classification accuracy as expressed in the hit ratio. The performance of each model in predicting the gender of *S. oryzae* was evaluated using the area under the receiver operating characteristic (ROC) curve or also called as the area under the curve (AUC) using the XLSTAT software. The AUC ranged from 0 to 1. A value of 1 indicated a perfect model agreement. A value of 0.5 indicated equal chance and value of 0 indicated complete disagreement of the model [15].

3. RESULTS AND DISCUSSIONS

3.1 Morphological analysis

In many insect species, the females tended to be larger and live longer compared to male. However, for some insect taxa, the body size dimorphism varies substantially. This can be seen in a most seed beetle (*Chrysomelidae*, subfamily Bruchinae) where females are larger than males, but in the genus *Stator,* male are generally larger than females [16]. For *S. oryzae* there are no studies conducted so far showing the differences in male or female body size. Figure-1 shows the morphological features, i.e. rostrum length and width, prothorax length and width, club length and width, and abdomen width at third and last sternite of male and female *S. oryzae*.The morphological characteristic of male and female are very identical, and there were no apparent differences in the body size of male and female of *S. oryzae*.

According to [14], the main differences between male and female of *S. oryzae* are the side view of the ventral surface; where male have curved posterior while the female has a straight line. Another difference is the rostrum of the male is smaller and thicker and less curved compared to female. However, this criterion is difficult to observe because it does not in numerical data rather been very subjective. This can lead to confusion and misidentification.

The only numerical data reported was the number of ommatidia. Male of *S. oryzae* have 167 ommatidia compared to the female with only 154 [14]. However, the process of counting the number of ommatidia is very tedious. Thus, dissection is required to confirm the gender of *S. oryzae* [1]. Figure 2 shows the dissection of the abdominal part of *S. oryzae*; the male was confirmed with the present of aedeagus and the absent of aedeagus confirmed the female. Dissection is not advisable because it is a destructive process and time-consuming.

3.2 Prediction of *S. oryzae* gender using Logistic Regression Model

An overall statistical measure of LR models in predicting *S. oryzae* gender is shown in Table 1. The first statistical measure analysed was the chi-square test in the -2 Log(Likelihood) value from the base model. Model 1 (all features) and Model 2 (rostrum) show the best goodness of fit test and statistically significant at .000 level. While Model 3, 4, 5 (prothorax, abdomen and club model -2Log(Likelihood) is high its negative and not statistically significant at .05. From the reduction of -2Log(Likelihood), the value R^2(Mcfadden), R^2(Cox and Snell), and R^2(Nagelkerke) can be obtained. For the model that consists rostrum and all morphological feature have a high value that near to 1. According to Hair *et al.*, (2010) higher values of R^2(Mcfadden), R^2(Cox and Snell), and R^2(Nagelkerke) indicating greater model fit.

The second statistical measures have been analysed was the Hosmer and Lemeshow measure of overall fit. This statistical test measures the correspondence of the actual and predicted values of the dependent variables. İn this study, the model that consists all morphological feature and rostrum only, was well fitted because the value of the Hosmer-Lemeshow test was 1. İt means there is no different in the observed and predicted classification. Another model also shows well-fitted condition where the value was higher than .05.

ARPN Journal of Engineering and Applied Sciences

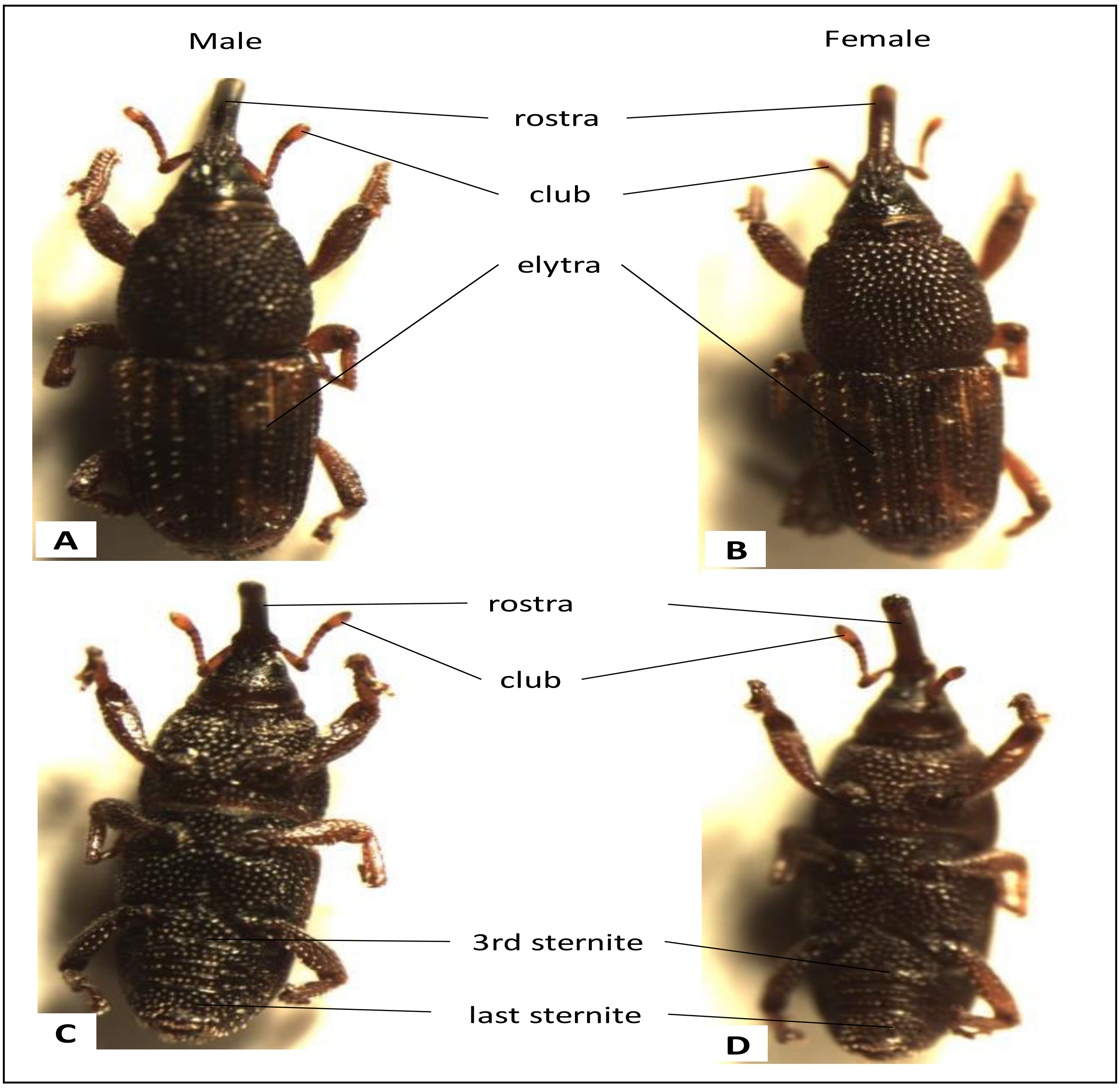

Figure-1. Morphological features of male and female *S. oryzae*. Dorsal image of (A) male *S. oryzae*, (B) female *S. oryzae*; Ventral image of (C) male *S. oryzae*, D) female *S. oryzae*.

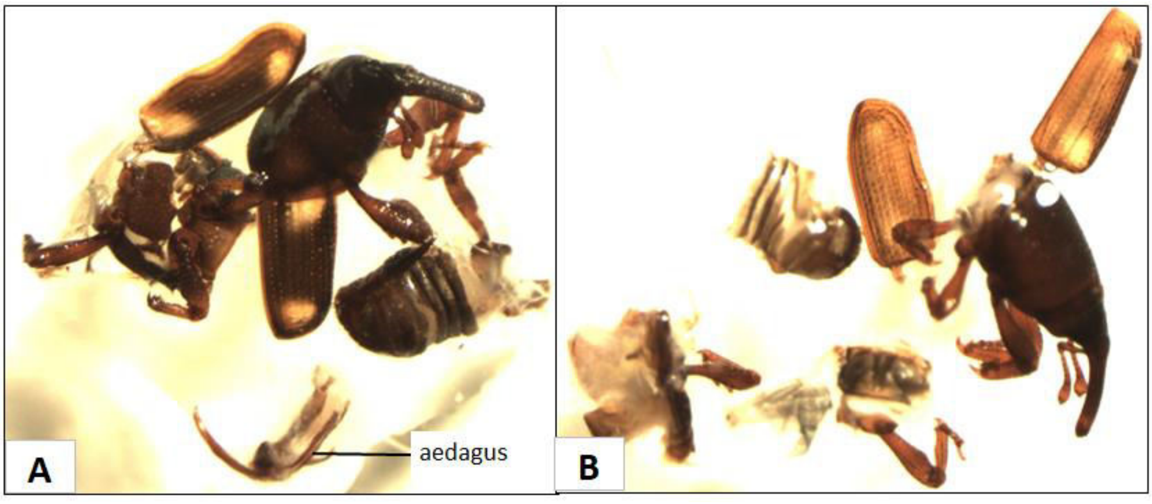

Figure-2. Dissection of S. oryzae; A) Male of S. oryzae with aedeagus. B) Female S. oryzae without aedeagus.

The third assessment of overall model fit has been done by assessing the classification accuracy of the model. This classification accuracy had been converted into the form of the percentage (Table-1). The model that consisted of all morphological features and rostrum shows 100% correct classification for the classification estimation and classification validation. However, the model that consisted only prothorax, abdomen and club part, there are only 50% correct classification in the estimation and validation sample. Meaning that about 50% chances of wrong identification will be made by using this model.

Another analysis done to check the performance of each model is by observing the area under the curve (AUC). The model that consisted of the only rostrum and all morphological feature had the highest area under the curve. This showed a perfect model agreement. For the model that consists prothorax, abdomen and club the area under the curve less than 0.7.

Table-1. Overall statistical measure of LR models predicting *S. oryzae* gender.

Type of Statistical Measure	**Statistics**	**Model 1 (All)**	**Model 2 (Rostrum width & length)**	**Model 3 (Prothorax width & length)**	**Model 4 (3rd and last sternite of abdomen)**	**Model 5 (Club width and length)**
Goodness of fit statistics	-2 Log(Likelihood)	0	0	41.45	40.87	41.38
	R^2(McFadden)	1	1	0.11	0.12	0.11
	R^2(Cox and Snell)	0.75	0.75	0.14	0.16	0.14
	R^2(Nagelkerke)	1	1	0.19	0.21	0.19
Test of the null hypothesis	-2 Log(Likelihood)	< 0.0001	< 0.0001	0.07	0.06	0.07
Hosmer-Lemeshow test	Pr > Chi^2	1	1	0.99	0.92	0.43
Correct classification (%) estimation sample	Female	100%	100%	68%	68%	58%
	Male	100%	100%	60%	47%	67%
	Total	100%	100%	65%	59%	62%
Correct classification (%) validation sample	Female	100%	100%	20%	40%	0%
	Male	100%	100%	80%	60%	100%
	Total	100%	100%	50%	50%	50%
Performance of Model	ROC(AUC)	1	1	0.7	0.65	0.69

Based on the overall performance on each statistical measure, Model 1 and 2 that consist all morphological features and rostrum respectively perform well in goodness and fit test, percentage of correct classification and performance of model based on the area under ROC curve. Model 3, 4, and 5 that consist of the prothorax, abdomen and club its does not perform well. It also shows that Model 1 that consists of all morphological features, greatly influenced by the rostrum structure. By using elimination techniques, all other morphological features (prothorax, abdomen, and club) can be eliminated. Rostrum measurement is enough to be used to differentiate the gender of *S. oryzae*. Therefore, Model 2 which modelled using rostrum length and width was chosen as the best model due to its practical and easy to use. The model generated from the LR is:

Pred(gender) =1/(1+EXP(-(412.29181-7562*Rostrum length +33348*Rostrum width)))

For each observation, the LR techniques predict a probability value between 0 and 1. The predicted probability is based on the values of independent variables and the estimated coefficients [17]. For the model, if the predicted probability is greater than .5, then the prediction outcome is 1; therefore, it confirmed a male *S. oryzae*. On the other hand, when the predicted probability is less than .5, then the prediction outcome is 0; it proved a female *S. oryzae*.

The predicted model developed in this study is in line with results reported by [18] and [1] which found that the gender of *S. Oryzae* can be identified using rostrum. The male of *S. Oryzae* is wider and shorter, while the female has thinner and longer rostrum. However, the differences in the rostrum size are tiny and not distinct. Therefore, dissection of the abdomen of the insect is needed to confirm its gender. Dissection is a destructive method and not suggested [18] and practical in determining the population assessment of *S. Oryzae*. Thus, this developed model which based on the rostrum image of

S. oryzae are more robust and proven in predicting its gender with 100% correct classification.

4. CONCLUSIONS

Based on the overall performance on each statistical measure, LR Model 2, which modelled using rostrum length and width has been chosen as the best model in predicting the gender of *S. oryzae*. İt was able to differentiate the gender of *S. oryzae* with 100% correct classification. This model is accurate and liable because the identification is based on the numerical data compared to previous identification which needs to be confirmed with internal reproductive organ. The developed model not only help in predicting the gender of *S. oryzae* but will be beneficial in measuring the total fertility rate and sex ratio in order to understand the population pest assessment.

REFERENCES

[1] D. Rees. 2004. Insects of Stored Products. London: CSIRO.

[2] J. F. Campbell. 2002. Influence of seed size on exploitation by the rice weevil, Sitophilus oryzae. J. Insect Behav.15(3): 429-445.

[3] J. S. Kamara, S. M. Kanteh, S. M. Bockari-Gevao and S. Jalloh. 2014. Infestation, population density and sterilization effects on rice weevils (Sitophilus oryzae L.) in stored milled rice grains in Sierra Leone.Int. J. Agric. For.4(1): 19-23.

[4] J. E. Baker. 1991. Properties of glycosidases from the maize weevil, Sitophilus zeamais. Insect Biochem.21(6): 615-621.

[5] R. Yan, Z. Huang, H. Zhu, J. A. Johnson and S. Wang. 2014. Thermal death kinetics of adult Sitophilus oryzae and effects of heating rate on thermotolerance. J. Stored Prod. Res.59: 231-236.

[6] G. S. Germinara, G. Rotundo and A. De Cristofaro. 2007. Repellence and fumigant toxicity of propionic acid against adults of Sitophilus granarius (L.) and S. oryzae (L.). J. Stored Prod. Res.43(3): 229-233.

[7] D. Dent. 2000. Insect pest management, 2nd Edition. New York: CABI Publishing.

[8] B. C. Longstaff. 1981. Biology of the grain pest species of the genus. Sitophilus (Coleoptera: Curculionidae): a critical review. Prot. Ecol.2: 83-130.

[9] K. C. N. Swamy, G. P. Mutthuraju, E. Jagadeesh and G. T. Thirumalaraju. 2014. Biology of Sitophilus oryzae (L.) (Coleoptera: Curculionidae) on stored maize grains. Curr. Biot.8(1): 76-81.

[10] G. Arnqvist and T. Nilsson. 2000. The evolution of polyandry: Multiple mating and female fitness in insects. Anim. Behav.60(2): 145-164.

[11] C. D. Flay, X. Z. He and Q. Wang. 2010. Influence of multiple mating on female reproductive fitness in the rice weevil, Sitophilus oryzae. New Zeal. Plant Prot.63: 201-207.

[12] S. D. Johnson and W. College. 2016. Nordic Society Oikos Sex Ratio and Population Stability Authors (s): Stewart D . Johnson Published by : Wiley on behalf of Nordic Society Oikos Stable. 69(1): 172-176.

[13] D. J. Rankin and H. Kokko. 2007. Do males matter? The role of males in population dynamics. Oikos. 116(2): 335-348.

[14] Omar. 2012. Morphological studies on some external and internal structures of rice weevil, Sitophilus oryzae (L.) (Coleoptera: Curculionidae), A Major Pest Of The Stored Cereals In Egypt. 3(8): 843-863.

[15] C. Khatchikian, F. Sangermano, D. Kendell and T. Livdahl. 2011. Evaluation of species distribution model algorithms for fine-scale container-breeding mosquito risk prediction. Medical and Veterinary Entomology. 25(3): 268-275.

[16] C. W. Fox, L. Dublin and S. J. Pollitt. 2003. Gender Differences in Lifespan and Mortality Rates in Two Seed Beetle Species. Funct. Ecol.17(5): 619-626.

[17] J. F. Hair, W. C. Black, B. J. Babin and R. E. Anderson. 2014. Multivariate Data Analysis. Essex: Pearson.

[18] Hagstrum and Subramanyam. 2006. Fundamentals of Stored-Product Entomology David. MInnesota: AACC Internationla.