

Exploration of Machine Learning for the Analysis of Generational Tick Colony Variability

Brian T. Rich

Abstract

Studies involving new tick management techniques necessitate access to large numbers of ticks from populations of interest to have clean and uniform populations for testing. Tick species may have peculiar life-history strategies or evolutionary pressures that are expected to leave distinct genetic signatures (Araya-Anchetta et al., 2015) leading to distinct populations over time. Rearing methods like those employed by Sonenshine et al. (1999) can be used to safely rear many generations of ticks in a laboratory setting satisfying the needs of researchers. In many settings, dozens of generations of a single strain may be reared without the need to add additional genetic diversity that might either require constant resupply of wild-caught ticks or contaminate the colony with incidental pathogens that have been intentionally screened out.

It has been theorized that long term inbreeding in tick colonies may eventually result in successive generations beginning to differ from the strains of interest in the wild. We used 63 generations of tick drop data from a single strain of Southern Cattle Tick, *Rhipicephalus (Boophilus) microplus*, to fit a supervised machine learning model and checked for generational variability generation over generation. Our model looked at generation, the time when females first began to drop from the host, 50% drop, ending drop, and seasonality. The results of the ML model showed that generation is a primary factor in variability within the colony.

Introduction

Cattle ticks are important ectoparasites that pose a significant threat to livestock sustainability through damage to animals and health due to the large number of pathogens they may vector to their hosts (Levin and Schumacher, 2016). The Southern cattle ticks have three major life stages after emerging from the egg (larvae, nymphs, and adults) and are obligate blood feeders that remain attached to their hosts for several days during each stage in order to obtain the nutrients they need to molt to the next stage of their development, mate, and produce eggs. The Southern cattle tick is a one-host tick and spends the entirety of the parasitic phase of their life cycle on the host. Average time to complete the life cycle is approximately 21 days under temperate conditions (Nuñez, 1982). Colder weather can lengthen the time spent on host and warmer temperatures shorten it (Hitchcock, 1955). Upon molting into adults the male ticks will seek out feeding females and mate. Unmated females are not capable of producing offspring and will not fully feed until replete (Stone, 1963). Mated female ticks will feed until completely replete (engorged) before dropping from their host to crawl away into the environment to lay their eggs in a suitable place (Needham and Teel, 1991).

Observing ticks in the wild is difficult at best due to the factors involved in access to the hosts they prefer for the long periods of time in which they remain attached. This is also a limiting factor in the collection of ticks for the purposes of research in that sufficient numbers of specific strains may be difficult to obtain from the wild. It has been noted that colony ticks may

speed up their life cycle, needing to spend less time on the host before completing their life cycle generation over generation in vitro. Those colony ticks may then no longer be representative of tick cohorts in the wild over time, degrading their usefulness and potentially introducing novel features. An analysis of tick drop data over the course of many generations through machine learning may permit the prediction of when those colonies vary too far from their base strain to be useful. Analysis and model creation for ticks to simulate spread and life cycle have been explored (Wang et al., 2017) but not for colony ticks due to a variety of factors including the difficulty in obtaining colony records from the labs that maintain them.

We were able to get the tick drop data for *R. (B.) microplus* raised in colony at the department of agriculture's Cattle Fever Tick Research Laboratory as well as permission to analyze them. This presents an opportunity to progress our understanding of variability in colony ticks by analyzing those drop data using machine learning to determine if there is any difference that can be observed generation over generation. This is useful in determining when to look at the genetics of those colonies in order to see how far genetic drift has actually taken them when compared the the initial colony characteristics from the wild.

Data

The data used in this analysis were obtained from a single strain of *R. (B.) microplus* raised in colony over the course of 63 generations. Drop data contained information about the date of the infestation, the animal identification number, total number of females collected, egg and female weights, number of females collected per day, strain, and generation. This initial analysis extrapolated the days until first drop occurred, the days until 50% of the total number of ticks that would eventually drop did drop, and the day when tick drop concluded. The dataset we analyzed also included the seasonality (temperature data was not recorded) based on the months each infestation occurred and tick generation.

Methods

Approximately 5000 larvae were allowed free release on a *Bos taurus* host and allowed to mature. Hosts were monitored daily to check for tick drop and female ticks were collected daily as they dropped from the host until drop was complete and for ten days afterwards. Female ticks were weighed and allowed a quiescent oviposition period before laying eggs. Eggs were then reserved in vials until hatch whereupon the larvae were apportioned out for experimental use and for

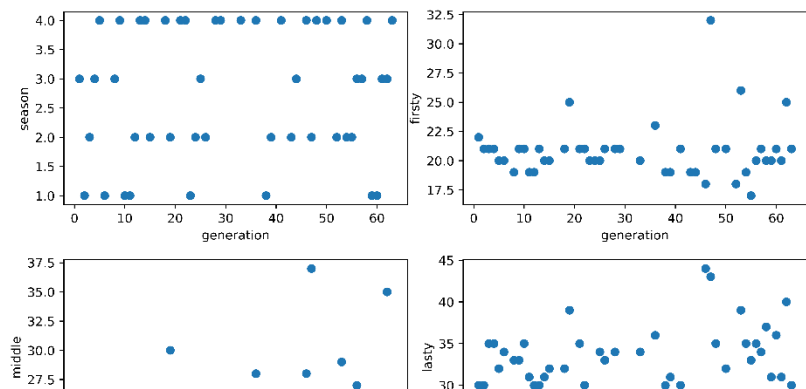


Figure 1. K-means output showing tick-drop variables in relation to tick colony generation.

continuance of the colony through the above procedure. Tick drop data was formatted to reduce the number of conflicting variables to only include the data above. Tick drop data was reduced to numerical data with seasonality represented by a number from 1 to 4 starting in Spring and ending in Winter.

K-means analysis was selected for the initial exploration. Model coding used the Python language and the “Jupyter Notebook” web based computing platform. Analysis used the k-means clustering code from the Sci-kit Learn package for Python. Data exploration (Fig. 1) did not reveal any immediate correlations between those data provided when compared to

generation. Fitting the data to k-means generated inertia (Fig. 2) suggesting that three centroids would be optimal for the model. K-means was then run using those three centroids generating the result shown in Figure 3. Cross tables were generated and were not indicative of meaningful interaction between the data (Fig. 4) until the last cross table generated for the generation where the F1-F19 generations formed a single cluster, the F21-F43 another cluster, and the third cluster comprised of the F44-F63 generations.

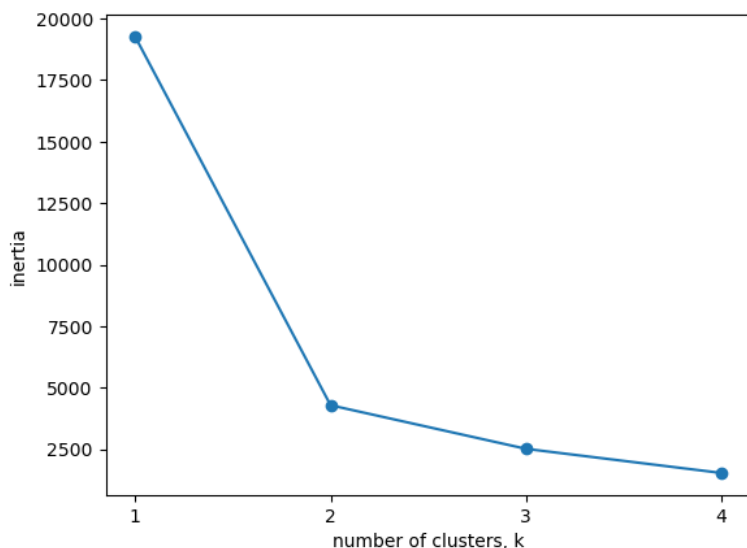


Figure 2. K-means output showing inertia in relation to predicted number of clusters, k .

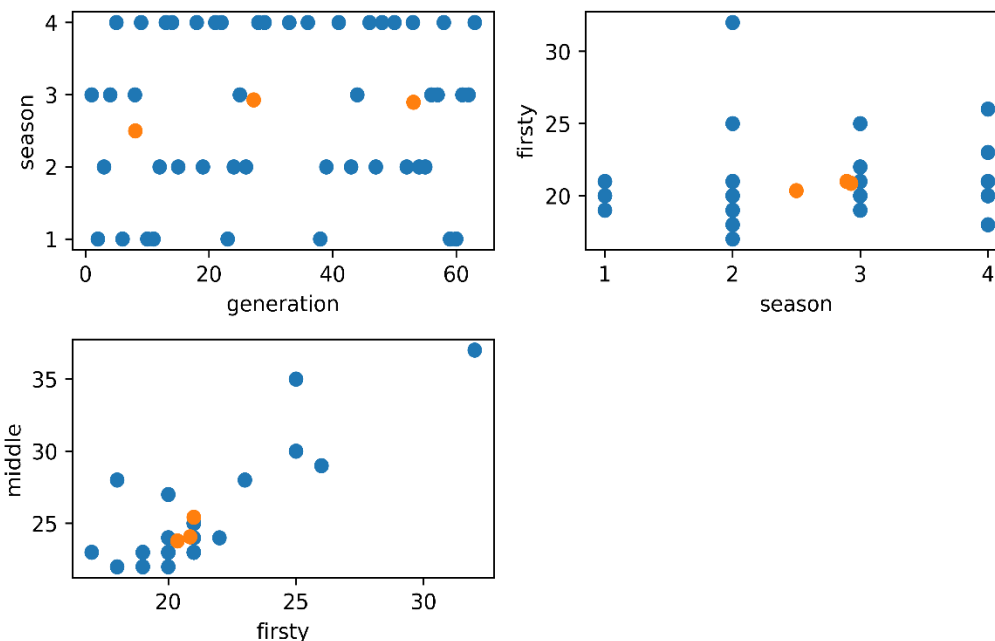


Figure 3. K-means output showing predicted centroids in tick-drop data.

"Season"	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	"First Drop"	<u>17</u>	<u>18</u>	<u>19</u>	<u>20</u>	<u>21</u>	<u>22</u>	<u>23</u>	<u>25</u>	<u>26</u>	<u>32</u>
Cluster					Cluster										
0	4	3	3	4	0	0	0	3	4	6	1	0	0	0	0
1	2	5	5	7	1	1	2	3	4	6	0	0	1	1	1
2	2	4	1	7	2	0	0	2	4	6	0	1	1	0	0

"50% Drop"	<u>22</u>	<u>23</u>	<u>24</u>	<u>25</u>	<u>27</u>	<u>28</u>	<u>29</u>	<u>30</u>	<u>35</u>	<u>37</u>
Cluster										
0	1	4	6	3	0	0	0	0	0	0
1	3	5	4	2	1	1	1	0	1	1
2	2	6	3	1	0	1	0	1	0	0

"Last Drop"	<u>23</u>	<u>25</u>	<u>27</u>	<u>28</u>	<u>29</u>	<u>30</u>	<u>31</u>	<u>32</u>	<u>33</u>	<u>34</u>	<u>35</u>	<u>36</u>	<u>37</u>	<u>39</u>	<u>40</u>	<u>43</u>	<u>44</u>
Cluster																	
0	0	0	0	0	0	4	2	2	2	1	3	0	0	0	0	0	0
1	0	1	0	1	1	2	2	1	1	1	3	1	1	1	1	1	1
2	1	0	2	0	0	2	1	1	1	3	1	1	0	1	0	0	0

Figure 4. Cross tables from a K-means analysis of tick drop data with key variables of "Season", "First Drop", "50% Drop", and "Last Drop". Tables indicate no strong interactions between clusters.

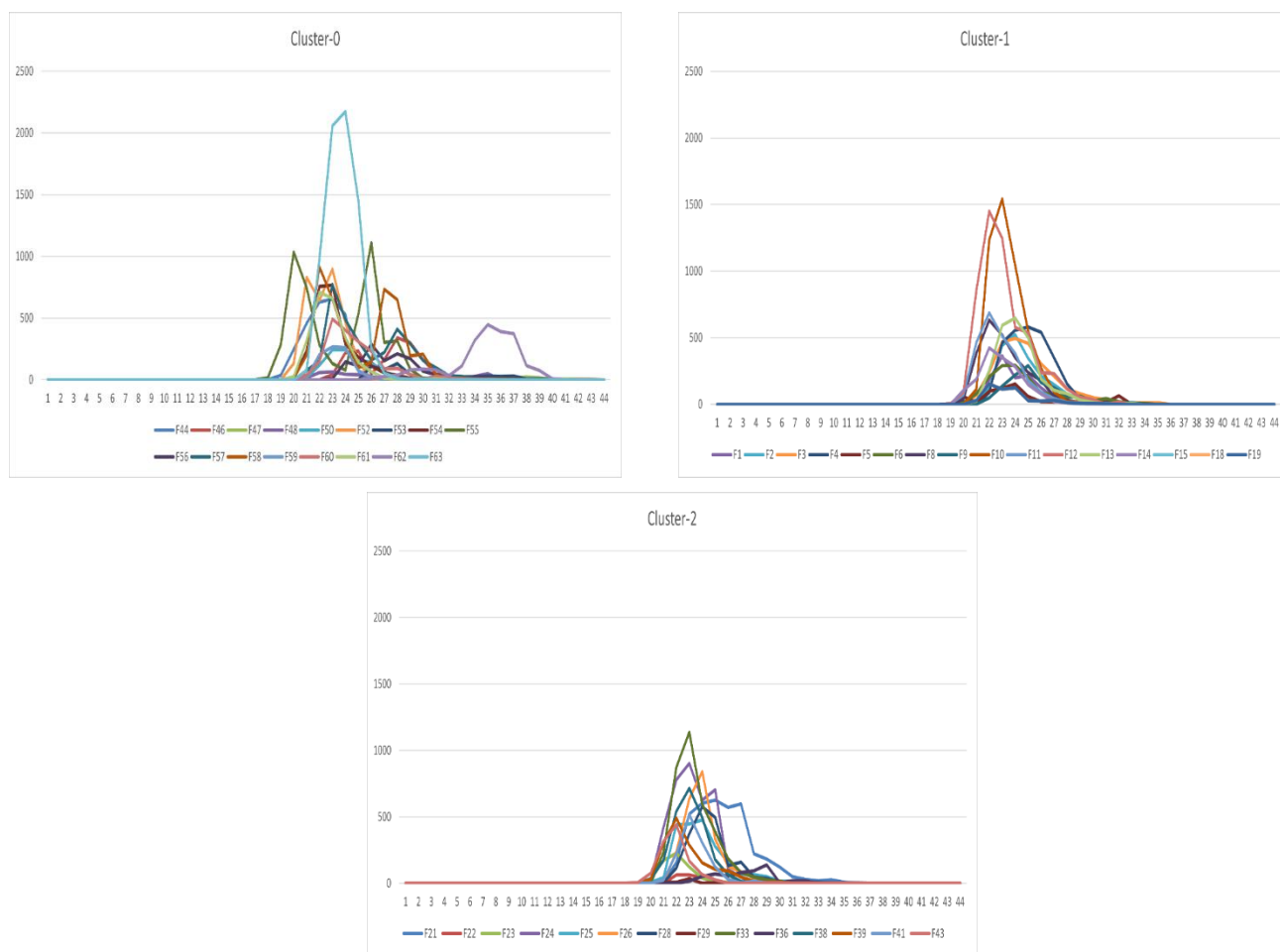


Figure 5. Tick drop generations grouped by predicted K-means clusters. Days of infestation shown along “x” axis with number of females collected along the “y” axis. Generations are colored and listed at bottom. Cluster shown at top.

Discussion

There is extreme difficulty in observing tick populations in the wild as they are hidden away when off their hosts and impossible to examine while on a host as most would be impossible to passively examine on a daily basis. A thorough examination of ticks in colony can add a wealth of information that may be extrapolated into wild populations and give us some insight into how gene flow functions outside of laboratory conditions. This isn't to say that colony ticks can be a suitable surrogate to wild populations but by further examination into the details of how those colonies change over time we can start to understand some foundations that may hold true across ticks as a whole.

Our lack of meaningful correlation between variable groups and generation was a bit surprising as when looking at Figure 5 there is an observable difference between the generations with the generations falling sequentially into each cluster. Something may be occurring in between the F20 generation that separates them from both the F44 generations and the generations that came before. The two items that stand out between the clusters appear to be the number of total number of ticks that drop as well as the lag time at the end of the drop where ticks take a few days longer to complete drop. More work should be done to take into consideration the full variability of both the tick drop on a generational basis as well as adding additional and more complete details about those drops.

Conclusions

The next step should be to run these data in a K-nearest neighbor analysis in order to pin down the specifics of where the differences lie and if they are statistically meaningful. Moreover, the full drop dataset should be analyzed and not just the first, middle, and last days of drop. Each generation should be its own array with the total ticks dropped per day and compared against all other generations. Another exploratory analysis needs to be performed on this full dataset to ensure that all variables are taken into consideration and bias removed. While it appears that the assumption that there is a difference between the earlier generations and the later generations, the question of why still needs to be answered.

References

- Araya-Anchetta, A., Busch, J. D., Scoles, G. A., & Wagner, D. M. (2015). Thirty Years of Tick Population Genetics: A comprehensive review. *Infection, Genetics and Evolution*, 29, 164–179.
- Hitchcock, L. F. (1955). Studies on the parasitic stages of the cattle tick. *Boophilus Microplus* (Canestrini) (Acarina: Ixodidae). *Australian Journal of Zoology*, 3(2), 145. <https://doi.org/10.1071/zo9550145>
- Levin, M. L., & Schumacher, L. B. (2016). Manual for maintenance of multi-host ixodid ticks in the laboratory. *Experimental and Applied Acarology*, 70(3), 343–367. <https://doi.org/10.1007/s10493-016-0084-8>
- Maramorosch, K., Mahmood, F. E., & Sonenshine, D. E. (1999). Maintenance of Ticks in the Laboratory. In *Maintenance of human, animal, and plant pathogen vectors*. essay, Science Publishers.
- Needham, G. R., & Teel, P. D. (1991). Off-host physiological ecology of ixodid ticks. *Annual Review of Entomology*, 36(1), 659–681. <https://doi.org/10.1146/annurev.en.36.010191.003303>
- Núñez Jorge Luis, Muños-Cobeñas Mario Enrique, Moltedo, H. L., & Bailie, H. (1982). *Boophilus Microplus: The Common Cattle Tick*. Springer.
- Stone, B. F. (1963). Parthenogenesis in the cattle tick, *Boophilus Microplus*. *Nature*, 200(4912), 1233–1233. <https://doi.org/10.1038/2001233a0>
- Wang, H.-H., Corson, M. S., Grant, W. E., & Teel, P. D. (2017). Quantitative models of *rhhipicephalus (boophilus)* ticks: Historical Review and synthesis. *Ecosphere*, 8(9). <https://doi.org/10.1002/ecs2.1942>