

Avidian Apocalypse: A Study of Invasive Genotypes in a Digital Environment

By: Sarah Dean, Lane Smith, and Chinedum Njoku (University of Texas at Austin)

I. Abstract

Existing invasive species are a major problem in every environment in the world, and cause inestimable ecological and economic damage on a global scale every year. As human networks grow ever more global and people continue to migrate to and from all corners of the earth---occasionally immigrating, intended or otherwise, with species native to their homeland but not present at their destination---the threat of introducing even more species to environments that are not their own grows ever stronger. It has been difficult, if not impossible, in the past to predict with any accuracy which foreign species have the capability to exploit an environment and overtake the native populations. By using Avida, software used in researching digital evolution, we investigate whether it is possible to predict the ability of an 'invasive' genotype, adapted to a different environment, to overtake the 'native' population in its own environment based on the condition of the invasive genotype having a distinct advantage over the native population when introduced. We conclude that there is a significant relation between the ability for the invasive genotype to become dominant and a distinct adaptational advantage over the native population at the time of introduction.

II. Introduction and Motivation

What happens when a species is introduced into an ecosystem where it doesn't occur naturally? Are ecosystems flexible and able to cope with change, or can a new arrival have extensive repercussions and do permanent damage? These are important questions which need to be addressed. The Center for Invasive Species and Ecosystems Health defines invasive species as any species (including its seeds, eggs, spores, or other biological material capable of propagating that species) that is not native to that ecosystem and whose introduction does or is likely to cause economic or environmental harm or harm to human health. History is rich with tales of the disastrous outcomes of some intentional introductions such as that of the Nile perch, which resulted in the extinction of more than 200 other fish species (Lowes et al. 2000). We can avoid repeating such mistakes by learning from history. Yet surprisingly, potentially damaging introductions continue to occur.

The United States suffers up to \$120 billion in economic loss per year (Pimentel et al. 2005) and this is being partly combated by having enormous fines and even jail time for any individuals caught smuggling invasive species into the country. As a result of 42% of the threatened or endangered species in the United States being at risk due to the introduction of invasive species (Pimentel et al. 2005), the National Invasive Species Council was established by the Presidential Executive Order EO 13112 to ensure that Federal programs and activities to prevent and control invasive species are coordinated, effective and efficient (The White House, 1999).

Introduction of non-native species to an ecosystem does not always have adverse effects and it does not always show any significant effect on the native population. If you study the dominant genotype in an ecosystem before and after a non-native species is introduced to an ecosystem, you can monitor the effects of introduction of such species to an ecosystem. And this is what we have done in our experiment. However, our experiment was done by using a digital environment and digital organisms in Avida, which make it a lot easier and faster to observe the any possible changes in the dominant genotypes.

In running our experiment, we made use of a null hypothesis and an alternate hypothesis. The null hypothesis states that the ability of the an invasive genotype's lineage to become the dominant genotype after 10,000 updates is independent of whether the invasive genotype could perform the task EQUALS prior to the invasion, as measured against our selected alpha value of 0.05. And the alternative hypothesis states that the ability of an invasive genotype's lineage to become the dominant genotype after 10,000 updates is dependent on whether the invasive genotype had the ability to perform the task EQUALS prior to the invasion, as measured against selected alpha value of 0.05. We have selected an alpha value of 0.05 because we have no reason to suspect that the data will be skewed in any manner or that there are any confounding variables that would cause us to interpret the p-values different.

III. Methods

Our project involved three main steps. First, we created the 'native' and 'invasive' populations, two different populations grown under different environments. Second, we introduced the dominant genotype from the invasive population into the native population in the native environment. The final step was to analyze all of the data produced in the previous step to obtain the results that would support test our hypotheses. The first two steps were performed independently by each of the group members, which allowed for a greater number of experimental trials in a shorter amount of time. In this section, each of these steps will be discussed in depth. In order to be efficient and reproducible in our methods, we wrote several Python scripts to automatically perform nearly all necessary steps to complete the methods in our experiments. More information about these scripts and where to obtain them can be found in *Supplemental Materials* (Section VII)

The native and invasive populations were all created in the same manner. Each population has their own environment configuration file that provides the crucial differences between the two populations. The native population is rewarded for all logical tasks *except* for AND, but performing AND has no effect on the fitness of the organism, positive or negative. Furthermore, the reward for the task of EQUALS requires that AND be performed prior to receiving a reward. All values of the reward for tasks performed were at the default value included with the distribution of Avida except for AND, which was set to '0'. In the invasive environment, all tasks are rewarded except for the reward for performing EQUALS, which carries no reward or punishment. All other values are also at their default value. To summarize and contrast the two environments, the native environment does not reward AND and AND is required for EQUALS, whereas the invasive environment rewards all tasks except for EQUALS. These conditions for

the native and invasive environments are stored in environment configuration files, *native_environment.cfg* and *invasive_environment.cfg*, respectfully. Because EQUALS is neither rewarded nor punished in the invasive environment, the dominant invasive genotype may or may not be able to perform EQUALS. We will later categorize the invasive populations according to this merit of being able to perform EQUALS in order to test our hypothesis

The populations are grown in Avida for 100,000 updates and the output from this process was a single population dump at the final update. All of these are specified in an events configuration file (*create_pop_events.cfg*) specific to creating the populations. This process is repeated once for each of the environments to produce the native and the invasive populations, which are saved in a directory of the experimenter's name in the *saved_data* directory. The second part of the experiment, the invasion of the dominant genotype of the invasive population into the native population in the native environment, is now ready to begin.

The first step in staging the invasion is to determine the dominant genotype of the invasive population. This process takes advantage of Avida's analyze mode, which allows us to determine the dominant genotype based on total fitness and the number of organisms with that genotype. This is performed by loading the population dump of the invasive population, using a custom analysis configuration file (*invasive_analyze.cfg*), and running in Avida in analysis mode. This analyze configuration file instructs Avida to produce a single file (*dom_genotype.dat*) with only the dominant genotype printed as output. This genotype is exactly the sequence of instructions performed by an Avida organism---analogous to codons of a biological organism. This sequence must be added to the events configuration file (*experiment_events.cfg*) that will be used to stage the invasion. The sequence is inserted using regular expressions at the line where it instructs Avida to inject a specific sequence into a population a specified number of times. For our purposes, we choose to inject the sequence into 5 cells, which avoids the possibility of the injected sequences being immediately overwritten by nearby replicating organisms. This injection, however, is performed after the event configuration file instructs Avida to load the entire native population dump file. This events configuration file further instructs Avida to produce summary data every 100 updates, to create a population dump at the start and the end of the run, and to run for exactly 10,000 updates.

The stage is now set for performing the invasion experiment. The native environment configuration file (*native_environment.cfg*) is used in the experiment in accordance with making this an 'invasion' of a foreign genotype into a native population in their native environment. Avida is then executed. It loads the native population, injects the invasive genotypes, runs for 10,000 updates, and produces all specified summary data files and population dumps. All output from Avida is moved to a new directory for storage. This process of executing Avida and storing output files is repeated for the specified number of trials; in the case of our experiment, 20 trials were performed for each pair of population files (3 group members, 3 pairs, 60 total trials).

The final step was to analyze the data to determine in which trials the invasive species became

dominant. Because our hypotheses specify categorical data, whether or not the invasive species could perform EQUALS and whether the invasive genotype's lineage became the dominant genotype are boolean categories. Our first step was determining whether the dominant genotype of the invasive species was able to perform equals. It should be noted, due to time constraints, no script was written to perform this task automatically; all data analysis was performed 'by hand'. This was done by running Avida in analysis mode with an analyze configuration file (*invasive_task_analyze.cfg*) that contains the dominant genotype sequence used in the invasion. This output a file which stated whether the genotype could perform AND and EQUALS. While we were interested specifically whether the genome could perform EQUALS, it was important to make sure the genome could also perform AND. If it could not, then it would be hardly different from the native population and thus an invalid invasive genotype. Fortunately, this did not occur in our invasive populations.

Next, we had to determine whether the dominant genotype 10,000 updates after the invasion had descended from the injected genotype; that is, whether the invasion had swept the native population. This could be roughly determined by looking at the *task.dat* file from each of the trials. If the invasive population overtook the native population, then there would be a large number of either AND or EQUALS being performed, as the native population most likely can do neither. However, there is the rare possibility that the native population evolved the ability to do both AND and EQUALS, so, in order to be more exact, we again turn to the Avida analyze mode to give us a definitive answer. New analyze configuration scripts were created (*lineage_analyze.cfg*) to allow us to determine the lineage of the dominant genotype 10,000 updates after the experiment. These configuration files loaded the population dump at the 10,000th update and output to a file all of the sequences of the lineage of the dominant genotype. If the first sequence of this file matched the genotype that was injected, then we know that the invasive genotype overtook the native population and became dominant. Otherwise, it did not. This was further verified by graphing the number of times the AND and EQUALS tasks were performed over time, which should match whether the invasive genotype became dominant. These results from each trial were tabulated and categorized to whether the invasive genotype became the dominant genotype.

Finally, with all necessary data gathered, we used Fisher's Exact Test with our categorized data to test against our null hypothesis (Yates, 1984). This test was done using the R open-source statistical software, and the `fisher.test()` command on a 2x2 matrix of our two-way contingency table between the number of trials with invasive genotypes that could perform equals and the trials in which the invasive genotype swept the native population.

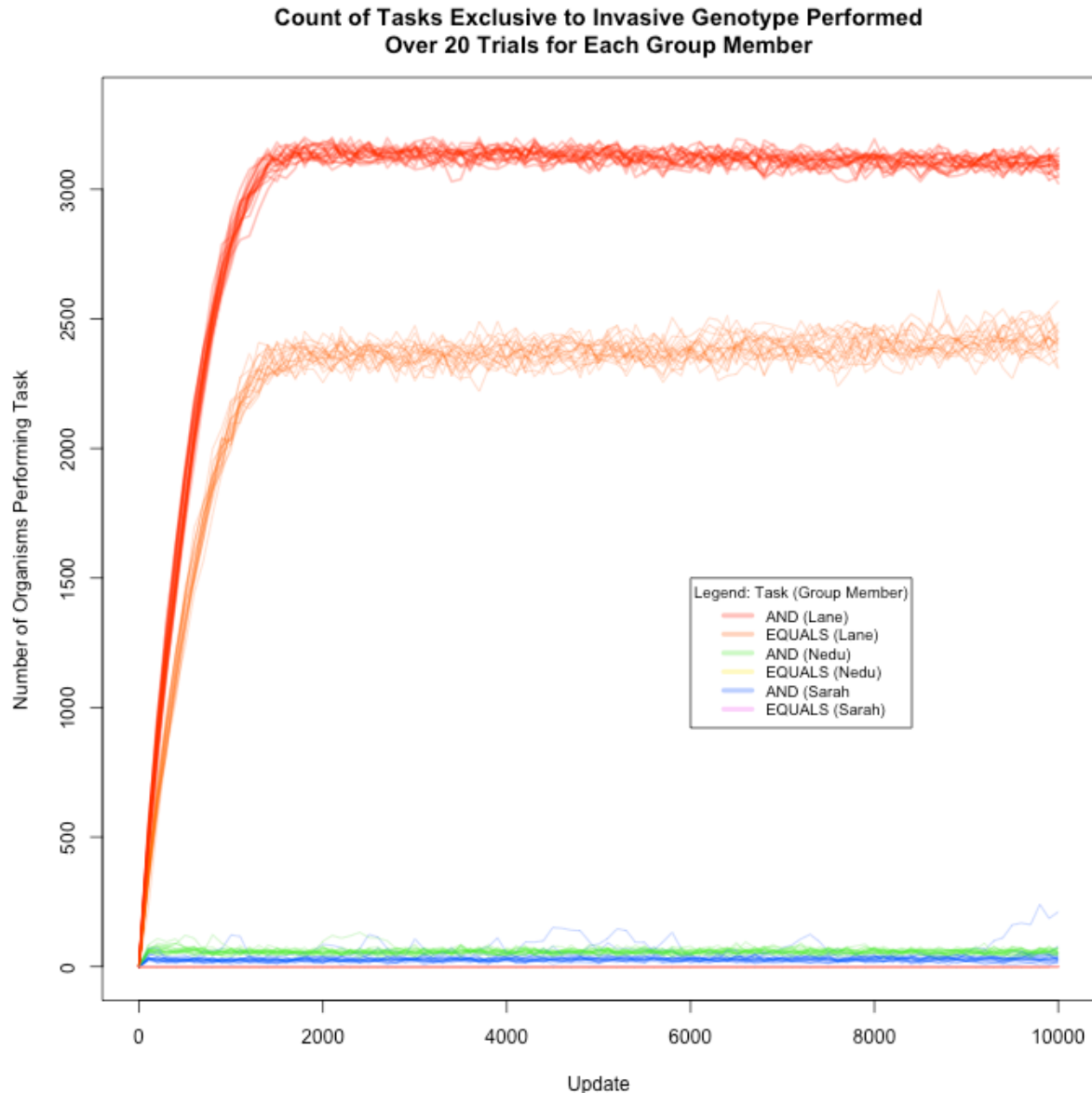
IV. Results

We first tabulated the results of our analysis, creating a 2x2 contingency table based on the criteria of whether the invasive genotype could perform EQUALS and whether the invasive genotype became the dominant genotype after 10,000 updates. Our results are available in the following table:

Table of the Relation of Performing Equals and Obtaining Dominance

	Performs Equals	Does Not Perform Equals	Total
Becomes Dominant Genotype	20	0	20
Does Not Become Dominant	0	40	40
Total	20	40	60

Immediately noticeable is the starkness of our results between the two variables. From our data, there was not a single trial in which the invasive genotype could perform EQUALS and did not become the dominant genotype; likewise, no genotype that could not perform EQUALS became the dominant genotype. We further corroborated these findings by graphing the counts of the tasks EQUALS and AND performed against updates since the introduction of the invasive genotype. Note that this method relies on the assumption that only the invasive genotype and its descendants can perform the tasks AND and EQUALS, and it is entirely possible that one organism from the native population could evolve AND and/or EQUALS. It is, however, unlikely, and there is no evidence of this occurring either in our table or in our graph. The following graph illustrates the relationship of performing EQUALS and becoming dominant.



As you can clearly see, the invasive organisms that could perform EQUALS quickly became the dominant genotype. Each continuous line is a single trial, and the lines were made to be semi-transparent to better illustrate the similarity of trend in each trial involving the same populations. Interesting to note is that the organisms who performed AND, but not EQUALS, integrated into the native population and, in most cases, did not die out.

Using Fisher's Exact Test in the R statistical software and the data listed in the contingency table, we obtained a calculated p-value equal to $2.386e-16$. We further validated this result using the Pearson Chi-Squared Test, which provided us with an Chi-Squared value of 73.3333, 3 degrees of freedom, and a p-value of $8.246e-16$. For the purposes of this project, we will use the value given by the Fisher's Exact Test.

Following the results of our data and its relation to our hypotheses, we can make the claim of rejecting the null hypothesis that the two variables of performing EQUALS and becoming the dominant genotype are independent. We have determined that the ability of the invasive genotype to become the dominant genotype after 10,000 updates after the introduction of the invasive genotype in the native population is dependent on the genotype's ability to perform the task EQUALS prior to being introduced to the native population. Comparing against our selected alpha of 0.05, our calculated p-value of 2.386×10^{-16} from Fisher's Exact Test shows our result to be strongly statistically significant, and thus, the null hypothesis is rejected.

V. Conclusions and Discussion

Our experiment reveals several interesting aspects about the relationship of invasive organisms in native environments. First, as shown by our rejection of the null hypothesis, in most cases where invasive organisms quickly overtake an environment, the invasive species has a distinct advantage over the native populations or exploits a niche in the environment. This was shown in our experiment in the case of invasive genotypes that could perform EQUALS, which provided the genotype with a significant advantage over the native populations who could not. Furthermore, the speed at which real-life foreign species can take over an environment is also reflected in our experiment. Using the graph of the counts of tasks performed over updates, we can see that the invasive species that could perform EQUALS increases rapidly and becomes the dominant genotype around the 1,000th update, which is only about 10 generations. Finally, our results show that a large majority of invasive introductions have little effect on the native population, as shown with the introductions of the invasive genotypes that could not perform EQUALS. This is the case for many foreign species in real biological environments, which integrate into the native environment without taking over. However, there always remains the possibility that some occurrence could shift the balance and cause the invasive species to quickly become the dominant species.

All of the trials in which the invasive genotype could not already perform the task of EQUALS all had the possibility to evolve and perform EQUALS and, as a direct result, become the dominant genotype in the environment. However, as our results showed, this occurred in none of the trials. This could be due to a number of factors. The invasive genome could possibly be arranged in such a way that it is impossible to perform EQUALS without significant rearrangement of the operations the organism performs. This could be a result of having no selective pressure in the invasive environment for evolving EQUALS; there is no reason to arrange the genome to perform a task that is not rewarded. Furthermore, there's the consideration that there were only 5 invading organisms, all cloned from the same sequence. This number may be too small and the source too homogeneous to promote genetic variability, which would increase the likelihood that EQUALS would evolve.

It must be mentioned that the original goal of this project and the hypothesis that we first started with were much different from this final result. Our original goal was to whether or not it was possible to

simulate a situation of an invasive species within a native environment and population. However, we had a difficult time formulating a testable hypothesis with this train of thought, and it wasn't until after we had run our experiment that we saw the possibility for a slightly different experiment with a truly testable hypothesis. It was interesting that, of all three invasive populations created, one had the ability to perform EQUALS, despite the fact that it was not rewarded for performing EQUALS in its environment. Furthermore, it was interesting that only the invasion involving this population had become the dominant genotype. Modifying our experiment to fit our new hypothesis was relatively simple, and in order to increase the power of our statistical testing, we choose to redo the invasion part of the experiment, opting this time for 20 trials for each group member instead of our original 5.

Given the chance to further investigate the relationship between performing EQUALS and becoming the dominant genotype, we would propose several changes. Specifically, we would investigate the invasive populations that did not perform EQUALS to determine whether there is relationship between the number of invasive genotypes injected into the population or the number of updates the invasive genotype is allowed to 'grow' and the ability for the invasive genotype to become dominant. We also believe it would be of interest to study whether introducing a random selection from the invasive population would result in a successful invasion, as this would increase the genetic variability of the invaders in the population and simulate a more 'realistic' instance of introduction of invasive organisms.

VI. Literature Cited

Lowe S., Browne M., Boudjelas S., De Poorter M. , 2000. "Biological Invasion". 100 of the World's Worst Invasive Alien Species: A selection from the Global Invasive Species Database 2: 3

Pimentel, David, Rodolfo Zuniga, and Doug Morrison. 2005. "Update on the Environmental and Economic Costs Associated with Alien-invasive Species in the United States." *Ecological Economics* 52.3: 273-88. ScienceDirect. <<http://www.sciencedirect.com/science/article/B6VDY-4F4H9SX-3/2/149cde4f02744cae33d76c870a098fea>>.

The White House, 1999. "Invasive Species". Federal Register: Presidential Documents 64: 6183 - 6186 <http://invasivespecies.gov/home_documents/EO%2013112.pdf>

Yates, F. "Tests of Significance for 2×2 Contingency Tables." *Journal of the Royal Statistical Society* 147.3 (1984): 426-63. *Journal of the Royal Statistical Society*. Blackwell Publishing for the Royal Statistical Society. Web. <<http://www.jstor.org/stable/2981577>>.

VII. Supplemental Materials

All scripts, configuration files, and executables used in this project are available in a git version

control repository available at https://github.com/knyon/BIO321G_Project. The project is entirely self-contained, requires no dependencies or other programs, and allows you to perform all of our experimental procedures verbatim. Please read the README.md file included in the repository for detailed instructions on how to perform our experiments.