THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Domestication syndrome via indirect selection in simulated cereal grains

OPEN ACCESS

Domestication syndrome via indirect selection in simulated cereal grains

Chris Marriott[1] and Jobran Chebib[2]

[1]University of Washington
[2]University of Edinburgh

Abstract

Domestication syndrome in cereal grains is commonly thought to be the product of domestication through a combination of direct artificial selection and indirect natural selection by humans. We propose an agent-based model of grain domestication. We simulate cereal grains with four genes that impact their reproductive cycle undergoing harvesting and selective planting by simulated humans. When direct artificial selection is applied to one gene domestication syndrome emerges in the other genes as a result of indirect natural selection. In the absence of direct artificial selection no domestication syndrome emerged, consistent with periods of predomestication cultivation in human history. Domesticated variants are strongest when humans select for traits inconsistent with the wild type traits, and weakest when humans select for traits consistent with the wild type.

Introduction

Domestication is a coevolutionary process acting on the genetics of the domesticated species and the culture of the human group harvesting and planting the crop (Larson et al., 2014; Purugganan, 2019; Allaby et al., 2021; Angourakis et al., 2022). This commonly results in a similar suite of traits in the domesticated species called *domestication syndrome* (Meyer et al., 2012). For instance, domesticated cereal grains, like wheat, barley and rye, have: non-shattering seed heads, larger seed size, reduced dormancy and lose natural seed dispersal mechanisms, etc. (Brown et al., 2009).

Modern theories of domestication suggest that these traits arise through a combination of direct intentional artificial selection and indirect unintentional natural selection (Heiser, 1988; Zohary, 2004; Purugganan, 2019). In the case of the indirect selection, these traits arise due to the change in the fitness landscape introduced by humans interfering in their reproductive cycle. For instance, non-shattering seed heads are a commonly found trait of domesticated grains, and this trait is commonly used to differentiate the domesticated grains from wild grains in the archaeological record (Brown et al., 2009). Wild grains will naturally release their seeds after they have matured; they will shatter. Seed heads that don't shatter will be easier for humans to harvest. So,

the evolution of non-shattering seed heads can be seen to arise from two different forces: through intentional artificial selection by farmers that preferred easy-to-harvest grains or by unintentional natural selection imposed by the new reproductive cycle of harvesting and planting. Genetic analysis of the archaeological record indicates that the selection pressure applied to these traits was within the range of natural selection, suggesting indirect selection (Purugganan, 2019), though this interpretation remains controversial (Peleg et al., 2022).

Computer simulation has become a powerful tool in the biological and social sciences to compliment traditional analysis (Gerbault et al., 2014; Zhang and DeAngelis, 2020). Numerical models rely on mathematical equations for the behavior of the system over time. Typically the differential equations are difficult to solve analytically and so simulation is a means to understand the behavior of the system under many parametric settings. These models are most useful when studying simple systems that can be described by a small number of variables.

Numerical models offer a top-down approach, while *agent-based models* (also known as *individual based models*) approach the problem from the bottom-up by simulating the interactions between agents in the system. Agent-based models compliment numerical models by studying the dynamics of underlying complex systems (Zhang and DeAngelis, 2020). Many of the parameters and equations of numerical models exist as emergent properties in agent-based models and in real world systems. For instance, Angourakis et al. (2022) has a numerical parameter for *plant growth rate* that controls the rate at which plants grow in a growth model. As you will see in our simulation, the growth rate of a plant varies from individual to individual and emerges from the interaction of their genes and environment and several numerical parameters governing this process. So, while the plant growth rate is a fixed input parameter determining plant behavior in the numerical model, in the agent-based model it is a descriptive output value computed through aggregating observations of plant behavior, as in real world systems.

The computational models that have been used to study

domestication are commonly numerical models (Currat et al., 2008; Angourakis et al., 2022). Allaby et al. (2008, 2010) use agent-based models to support analysis of archaeological data sets. In Allaby et al. (2008) they estimate the strength of artificial selection in cultivation and in Allaby et al. (2010) they compare the genetic impact of single-origin versus multiple-origin domestication. Beside these simulations our model one of the first agent-based models studying domestication. We believe that agent-based models are well suited to studying the role of indirect selection in domestication syndrome because they allow for the modelling of the relevant gene-human-environment interactions at play in the explanation of indirect selection.

We have explicitly modeled the interactions of harvesting and planting along with natural dispersal mechanisms for seeds in the absence of human interference. This creates the possibility for both direct and indirect selection upon our genes. Numerical models are often too abstract to model these types of interactions. We have designed an artificial grain with four genes that control four traits inspired by natural grains. In particular, we have a gene inspired by the shattering/non-shattering variants discussed above. We expose this artificial grain species to various natural and artificial selection pressures ranging from no human interference to humans harvesting and selectively planting seeds. Each gene is subjected to different natural selection pressures to see how they respond to the artificial and natural selection pressures introduced by harvesting and planting.

We hypothesize that domestication syndrome will arise in our grains as the result of artificial selection. To test this we vary the artificial selection pressure by applying it to each gene targeting the minimum or maximum for each gene. We do find that domestication syndrome occurs as the result of indirect selection so long as an artificial selection pressure is applied to a gene that distinguishes the variant from the wild type.

Model

Our human and plant agents are situated in a discrete 50×50 cell environment with an 8-cell-wide river running from the top to the bottom (see Fig. 1). The depth of the water table is 4 in the center two columns of the river and drops off by 1 per column as you move away from the center until reaching the minimum of -11. All cells in the left and right quarters of the map have this minimum water table depth. The water table depth of a cell is a factor in plant growth (details below). The left and right edge of the environment serve as shelters for the humans, and no plants can grown in these cells. Shelters are a shared resource store where humans store seeds for eating and planting. Humans can eat seeds only in shelters. Each cell in the environment can contain up to four (immature or mature) plants and any number of humans.
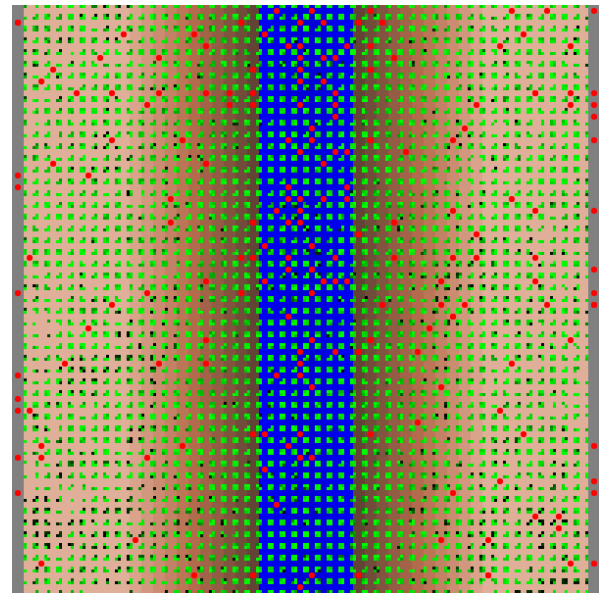

Figure 1: A snapshot of the simulation environment. The river occupies the center (in blue cells running from the top to the bottom of the environment) and water level deepens the further away from the river you get. The green squares (shallow roots) and black squares (deep roots) are the seeds and the red circles are the humans. The grey cells are the shelters along the left and right edge of the environment.

Seeds

In this paper when a *seed* takes root we will call it a *plant*. Up to four seeds can take root in a single cell to become plants. A plant will grow until it is mature. A mature plant will produce new seeds on its head. These seeds will remain on the head until harvested or until the plant is fully grown. At this point the plant will release the seeds and they may take root to form a new plant.

Seeds have four genes that determine four traits. Each gene is a real number in the range $[0, 1]$. When a new seed is asexually produced by a mature plant it mutates (100% mutation rate). A mutation adds a random value drawn from a uniform distribution of values between -0.04 and 0.04. This is applied independently to all genes in an individual.

Genes The four genes perform different functions in the seed life cycle. These functions are inspired by traits of cereal grains, though are abstractions and may not correspond to any particular gene or trait in cereal grains. Nonetheless we have named the genes for the trait that inspired them. The genes are *root depth* (impacts growth rate to maturity), *fecundity* (impacts the number of seeds produced), *abscission* (impacts when seeds drop), and *seed dispersal* (impacts where seeds fall). The impact of the genes on the reproductive success of the seed varies and thus so does the natural and artificial selection pressures on each gene during our ex-

periment.

Life Cycle Abstractly *root depth* is a gene that determines plant growth rate as a function of the environment. A plant will grow each discrete time step by an amount determined by the *root depth* gene and the water table depth of the cell. The *root depth* gene r_s blends two growth profiles. With shallow roots ($r_s = 0$) the amount of growth depends on the environment. Specifically, it is equal to the water table depth plus 11 ($v \in [0, 15]$). With deep roots ($r_s = 1$) the amount of growth does not depend on the environment as it is always 8. For values between the extremes the plants grow by a weighted blend of these two values determined by the gene's value. This means that plants in the middle half of the environment grow faster with shallow roots and plants in the outer two quarters grow faster with deep roots.

After growing during a time step, if the plant's total growth is greater than the maturity threshold then the plant matures and it produces new seeds. The threshold is equal to a base value (50) plus a growth penalty. The penalty is the sum of the gene values (max 4) multiplied by 50 plus a random variable between 0 and 50. This penalty is meant to capture a principle that more complex structures require more energy to grow (Lane and Martin, 2010) This principle introduces a selection pressure (other things being equal) on each of the four genes toward 0.

A mature plant produces new seeds on its head. The number of seeds produced at maturity depends upon the sum of three terms: a random integer R selected uniformly from the range $0 - 15$, the cell water depth $w_c - 4$ (producing a value ranging from -15 to 0), and the seed's *fecundity* gene f_s multiplied by 16. The sum r is calculated:

$$r = R + (w_c - 4) + f_s * 16$$

and the number of seeds produced is $\left\lceil \frac{r}{4} \right\rceil$. This will produce 0 to 8 seeds depending on the fecundity gene, the environment, and randomness.

Once mature, a plant continues to grow until a second threshold is reached (it is during this time the new seeds can be harvested by humans). Specifically the second threshold is $10 + (1 - a_s) \cdot 50$, where a_s is the value of the *abscission* gene. This means that a mature plant with $a_s = 1$ will cross this threshold almost immediately whereas a seed with $a_s = 0$ will have a second threshold of 60 plus the initial maturity threshold.

When the mature plant crosses the second threshold it dies and disperses its new seeds. Where the new seeds land depend on the value of the *seed dispersal* gene d_s. New seeds are deposited in the current cell with probability $1 - d_s$. Otherwise, the seeds are deposited in one of the 8 neighboring cells (selected uniformly). Seed dispersal for seeds on the same plant are carried out independently (even though they are related so their gene values should be correlated).

Humans

In this simulation the humans have no genetics and are behaviorally homogeneous, so can be considered environmental agents with respect to the plants. While we respect the coevolutionary nature of domestication, we have elected to limit the model of our humans in order to study the genetic dynamics of our simulated plant species. So, human behavior is a controlled variable of our model.

Depending on the experimental run, humans can differ in their harvesting and planting behaviors. Humans in our model must subsist on the seeds in addition to planting them so they have a simple metabolism.

During each time step outside of a shelter a human spends energy and attempts to gather (and possibly plant) seeds. If the energy spent increases beyond a threshold (90) the human returns to the shelter. Otherwise they attempt to gather seeds and on some settings plant seeds.

When gathering seeds the human first looks in the nine local cells for mature plants. They select a mature plant randomly and move into the cell with that plant and harvest the seeds. (If no mature plants are found in the local cells the human wanders randomly, more often towards the river.) Harvesting seeds is not perfect. For each new seed on the mature plant there is a 75% chance the seed is harvested. Otherwise the seed is dropped as though dispersed by a fully grown plant.

A human in the shelter will eat seeds to reduce its energy spent back to 0. Each time step spent in the shelter will reduce the spent energy by 15 assuming seeds are available in the shelter. One seed replenishes three spent energy and at most five seeds can be eaten each time step. Recall that the shelters share seeds gathered by all humans. This is important since it simulates granaries where a population shares its food resources and also uses this to selectively plant crops.

Under some experimental settings humans also plant seeds. To facilitate this they store seeds in the shelter in two separate stores: the food stores and the planting stores. If the seeds in the planting stores get below $\frac{1}{4}$ of total seeds stored then some seeds (about $\frac{1}{4}$) are transferred from the food stores to the planting stores. Which seeds are transferred is determined by the planting strategy. Then, when humans leave the shelter, they select some planting seeds (50) from the planting stores. If they encounter a free spot for a plant while foraging they plant one of the seeds from their planting pouch. The planting seeds and gathered seeds are kept separate from one another when gathering.

It is important to note a liberty of interpretation we are taking with our description of our simulated human planting activity. We are implying the process of selecting seeds as deliberate conscious process with intention upon the part of our humans. Of course, our simulated humans are incapable of such processes. So, if we prefer, we can also interpret the selection process as being largely unconscious and unintentional on the part of the humans.

Experimental Setup

Each of our experiments begins with *wild type* seeds before we add humans. There are three populations of seeds we deem *wild type* in this experiment. The first kind, *wild type I*, are the seeds that experience natural selection without harvesting by humans. The second kind, *wild type II*, are similar to the first kind, except they experience harvesting with no planting. The final kind, *wild type III*, is the population with harvesting and non-selective planting. More on these types below.

In order to establish *wild type I* we allow our seeds to evolve in the absence of a human population. Preliminary experiments indicated that *wild type I* had stabilized genetically before time step 10,000 from an initially random population. So we begin our experimental runs by adding in humans at time step 10,000 and terminate the experiment on time step 20,000.

When adding humans we add 2 humans to each shelter (for a total of 200 humans). New humans have spent no energy so they begin harvesting seeds immediately. Humans differ from one another only in different experiments, where different seed planting strategies are set as parameters. We have 11 runs (each with a different experimental setup) including the three *wild type* control runs.

The planting strategies are labelled *none*, *random*, *max gene* and *min gene* for each of the four genes. When no planting is selected, humans harvest seeds, but they do not plant them, resulting in *wild type II*. If humans plant randomly, without a planting strategy, then they do not differentiate when adding seeds to the planting stores producing *wild type III*. In experimental runs with planting strategies, humans add seeds to the planting stores based on their genetic values (maximizing or minimizing) for each of the four genes.

All experiments are run for an additional 10,000 time steps after humans are added at time step 10,000. During this time we gather data on the genetic distribution of all four genes as well as population data. Since each gene is real valued in the range $[0, 1]$ we plot the genetic distribution as a time series of histograms (see Figures 2 to 5). Each vertical slice represents the genetic distribution at that time step using a heat map to represent the concentration of population. All data represent the averages of 100 runs.

Wild Type I

Figure 2 (left column) shows the genetic distribution of *wild type I* seeds after time step 10,000. It is clear that these distributions have reached an equilibrium since they do not appear to change between time step 10,000 and 20,000. *Wild type I* represents the starting conditions of our experimental runs.

This distribution allows us to see the natural selection pressures applied to our four genes in absence of human interference. Each gene has a different distribution pattern in
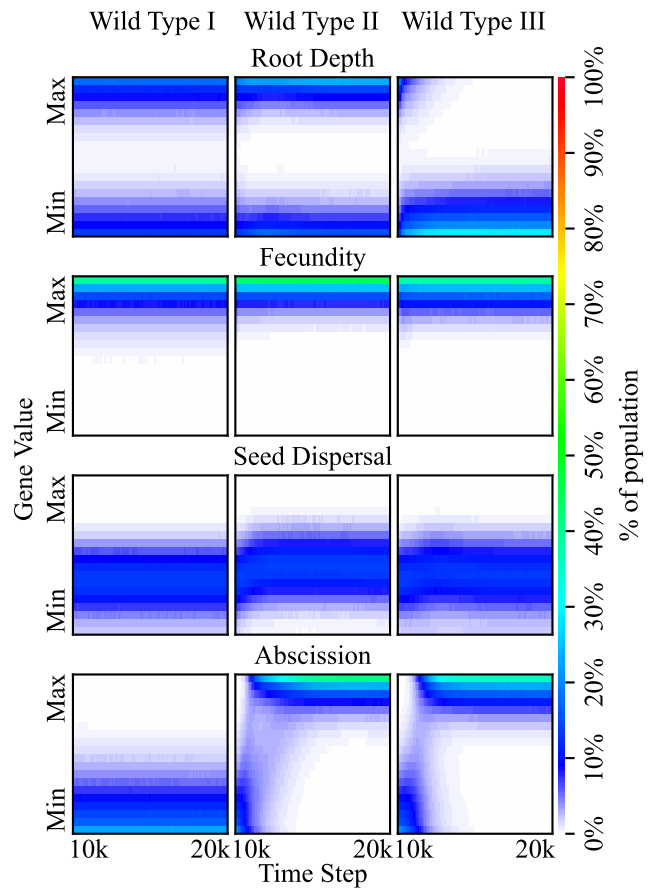

Figure 2: The genetic profiles of Wild Type I, II and III. Our experiments begin with Wild Type I and we use Wild Type III as the profile of wild variants in our experiments.

the population due to different selection pressures. The root depth gene forms two sub-types. The shallow roots wild type evolves to occupy the wet river cells and banks and the deep roots wild type evolves to occupy the dry regions on the edges of the environment.

The fecundity gene is subject to strong selection pressure to produce as many seeds as possible and so optimizes near 1. The seed dispersal gene evolves to a mean value around 0.35 which will allow some seeds (about 65%) to drop down into the current cell and others to spread to neighboring cells.

The abscission gene is selected towards 0 by the selection pressure applied due to the growth penalty. In our model there is no evolutionary advantage to early abscission in the absence of predators that will eat seeds from mature plants. As a result under natural selection with no predation the abscission gene is only subject to the growth penalty selection pressure.

Wild Type II

Our second control run establishes a population of seeds under random harvesting by humans (see Figure 2 middle column). This simulates humans in a pre-cultivation stage where humans, or other predators, harvested the seeds without deliberately planting them. In our simulation, with humans present, mature seeds may be plucked before they would normally disperse. This creates selection pressure on seeds to disperse as soon as possible after maturing. As a result we observe the *abscission* gene evolving towards the maximum value of 1 (rapid abscission). In our model we will treat this characteristic as the defining characteristic of wild type seeds following some archaeologists who use shattering as the identifying characteristic of wild type grains (Brown et al., 2009).

The equilibrium for the dispersal gene shifts towards spreading more seeds into adjacent cells. This is an adjustment due to more empty cells thanks to humans harvesting seeds. The remaining genes experience a slight strengthening of existing natural selection under the pressure added by human harvesting.

Wild Type III

Our third control run establishes a population of seeds under random harvesting and random (non-selective) planting by humans (see Figure 2 right column). This experimental run correlates to a period of predomestication cultivation, that in some locales lasted for centuries prior to the domestication of cereals (Weiss et al., 2006; Fuller et al., 2007, 2011; Willcox and Stordeur, 2012; Larson et al., 2014; Angourakis et al., 2022). The primary impact of random planting on the genetics of our simulated crop is the loss of the deep rooted sub-type. Random planting means the humans plant slow growing, shallow rooted seeds near the shelters, which take up space, have a low yield and clog out the deep rooted variant. Over time the deep rooted variant disappears. This is a non-intended, detrimental result of planting from both the perspective of the humans and the seeds (except the shallow rooted variant which is thriving).

We count this final type a *wild type* because, despite planting, there are no genetically or phenotypically distinct wild and domesticated variants. Seeds planted by the humans, and those distributed naturally, have the same genetic features. As a result humans are merely serving as a secondary dispersal mechanism for the seeds. This artificial dispersal mechanism has no further impact on the genetics, after considering the impact of harvesting, and the wider dispersal that humans grant the seeds. This can be considered a type of epizoochory or endozoochory dispersal mechanisms that parallels the distribution of cereal grains on the coats of animals that brush by the seeds or in the feces of herbivores that eat the seeds. This can result in new selection pressures on the plants and result in a mutualism between the humans and plants, but need not result in domestication as this control run demonstrates.

Domestication Syndrome

The goal of our experiment is to determine to what degree domestication syndrome can arise in our simulated plant species relative to different artificial selection pressures imposed by human planting strategies. In each of our eight experimental runs humans select one of the four genes and selectively plant the seeds with maximum (or minimum) value for that gene. In our experiment we consider this to be a direct or intentional selection pressure applied by the humans. However, as mentioned before, this is a liberty, and we could also interpret this to still be an indirect unintentional selection pressure applied as the result of some technique or tool used by the humans. Regardless of interpretation we expect, other things being equal, that this selection pressure introduced by the humans will be strong and will cause the selected gene to be selected towards the maximum (or minimum). We also expect it to behave differently with interference from natural selection pressures.

If indirect selection due to domestication occurs then we expect to observe a linkage disequilibrium between the genes, in particular between the selected gene and other genes. We will say domestication occurs in our grains if a population of non-shattering grains appears, that is, a population with late release of seeds. We will say domestication syndrome occurs in the seed population if a collection of common traits emerge as the result of artificial selection and linkage disequilibrium through indirect natural selection.

Observations and Discussion
Characterizing Domestication

In our analysis we define the wild type to be those with an *abscission* gene value above 0.6, as observed in *wild type II* and *wild type III*, and the domesticated type to be the seeds with *abscission* gene below 0.6. The *abscission* gene was selected to define domestication in our model for two reasons. For convenience some archaeologists rely on shattering/non-shattering to identify domesticated or wild variants in the archaeological record. Since our abscission gene was inspired by shattering and non-shattering variants this is a natural extension. Secondly, in our *wild type II* and *wild type III*, we see that human harvesting applies a selection pressure towards rapid abscission of seeds (to avoid this predation) but that human selective planting reverses this selection pressure in all but one case. This suggests our labels are not merely cosmetic, but are accurate labels of the two populations.

We have selected the arbitrary point 0.6 from observation of our model. Like the archaeological cases, our seeds often occupy a continuum from rapid abscission to delayed abscission. However, typically this gene's value is distinctly polarized and the area around 0.6 is sparsely occupied. Arbitrarily selecting our line means that we may accidentally
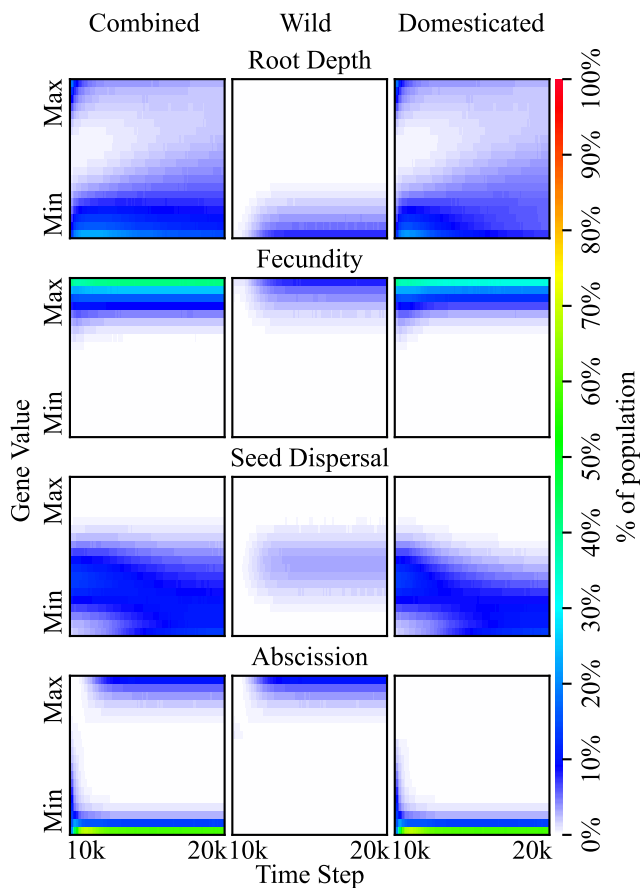
Figure 3: Genetic distributions for the wild and domesticated variants when humans attempt to minimize the *abscission* gene. The left column includes combined wild and domesticated variants. The other columns include the same variants in the combined column, but separated into wild (middle column) and domesticated (right column) variants.

draw it in the wrong place, but the small number of seeds that occupy this space means any error introduced will be small.

Coexistence of Wild and Domesticated Variants

The first experimental setup we will consider, is the case when humans select for seeds that have a delayed release (i.e. minimum abscission). Since we are using this gene to define domesticated and wild variants, this experiment represents direct selection for what we define as the domesticated variant. Figure 3 shows the genetic distribution of the seed population under this artificial selection strategy (left column), including the distributions of the separate wild and domesticated variants that emerge (middle and right column respectively) after applying our definition of wild and domestic.

We can see there are indeed two populations, one with quick release seeds (wild) and the other with delayed release seeds (domesticated). These populations coexist in the shared environment and compete for space. In this experimental setting the domesticated variant dominates, though under other settings the wild variant dominates.

Domestication Syndrome via Indirect Selection

In Figure 3 we can also see linkage disequilibrium between the *abscission* gene and the *root depth* and *seed dispersal* genes. That is, the domesticated variant experiences different selection pressures as a result of domestication, and that can be seen in differing genetic distributions between the wild and domesticated types in these genes. The selection pressure applied to the genes other than *abscission* are indirect selection pressures applied by the nature of human harvesting and planting, without intentional selection by the humans.

The *seed dispersal* gene is rendered largely useless for the domesticated variant that relies on humans for a dispersal mechanism. Only seeds that accidentally fall while gathered use this gene to determine where they fall. As a result the selection pressure due to the growth penalty causes the domesticated variant to evolve towards a less costly gene.

The *root depth* gene also displays linkage disequilibrium with the *abscission* and *seed dispersal* genes. While the wild variant collapses to the a shallow root variant, consistent with *wild type III*, the domesticated variant has roots of all depths and appears to be selected towards deep roots. This may be in response to the fact that domesticated seeds have no means of determining whether they will be planted in the wet or dry regions of the environment. While a mixed roots strategy is not good for wild seeds that "know" which region their offspring will grow in, it may be good when your environment is selected for you. In addition, deep rooted variants grow the same rate regardless of environment which is advantageous to domesticated seeds, but this also makes them worse off than mixed or wild strategies in wet environments.

The other remaining gene, *fecundity*, does not display clear linkage disequilibrium with the *abscission* gene in this case. If linkage occurs on these genes it is subtle and hard to notice in this experiment. However, when selecting for the *fecundity* gene (below) we do see some linkage disequilibrium with the other genes.

We consider the linkage disequilibrium between these genes as evidence of domestication syndrome occurring in our simulated grains as the result of indirect natural selection from the humans. Applying this methodology to our other experimental runs allows us to evaluate the strength of domestication syndrome under various selection strategies.

Strong Domesticated Variant

Domestication syndrome occurs in our experimental runs whenever humans selectively plant except for one case (dis-
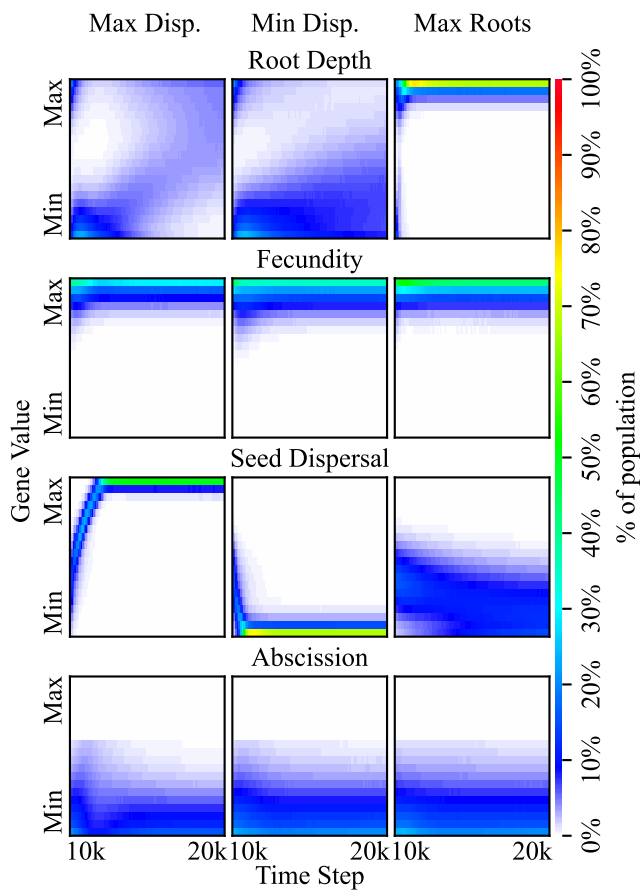
Figure 4: Genetic distributions of domesticated variant for when humans attempt to maximize the *seed dispersal* gene (left), minimize the *seed dispersal* gene (middle), and maximize the *root depth* gene (left).
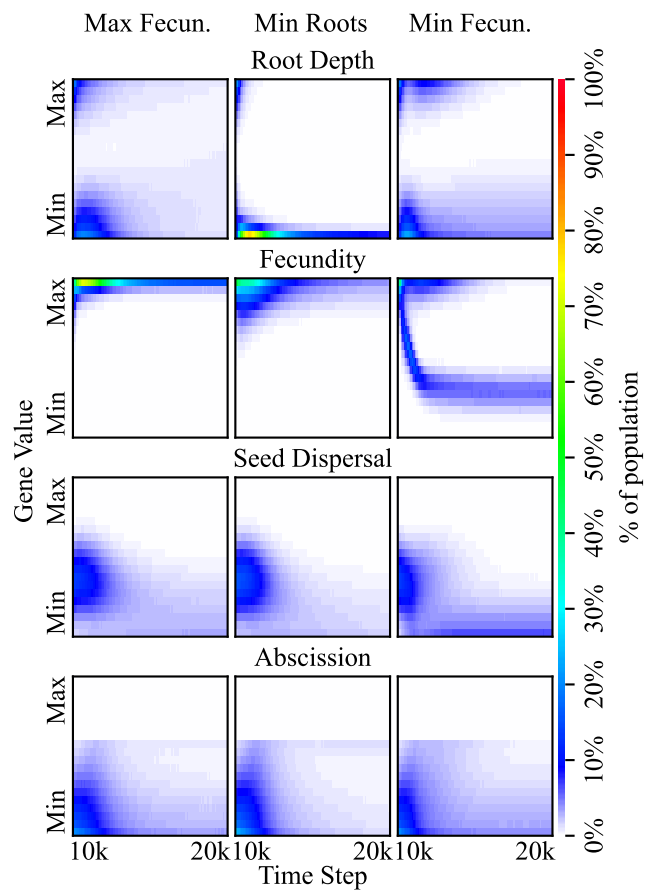

Figure 5: Genetic distributions of domesticated variant for when humans attempt to maximize the *fecundity* gene (left), minimize the *root depth* gene (middle), and minimize the *fecundity* gene (left).

cussed below). Domestication syndrome is strongest in cases where the humans select for traits inconsistent with the wild type. When this artificial selection pressure is applied we get two clear wild and domesticated variants displaying linkage disequilibrium among our genes. This is important, because in these runs we are not directly selecting for the *abscission* gene, but we are still relying on it to identify the wild and domesticated variants. So, for us to observe a domesticated variant in any run means that linkage disequilibrium developed due to indirect selection between the selected gene and the *abscission* gene in that run.

When humans attempt to minimize or maximize the *seed dispersal* gene or when they attempt to maximize the *root depth* gene we have a strong domesticated variant in competition with the wild variant (see Figure 4). Since the *seed dispersal* gene has an optimum in the middle of the gene space, selecting for either extreme allows for differentiation of the two types. Maximizing root depth leads to differen-

tiation for a similar reason, though the wild type optimizes to the minimum values. The wild and domesticated variants identified in these runs are similar to the ones discussed above except that artificial selection occurs on another gene.

In these cases the trait selected for is maximized (or minimized) as desired. This causes a domesticated variant to emerge and differential selection on this variant leads to linkage disequilibrium for the other genes. It is important to note that each of these artificial selection strategies aims to optimize one of the genetic traits away from values preferred by the wild variant. This allows artificial selection to produce a variant that is distinguishable from the wild type by the humans (or their techniques and tools) and this creates a positive feedback loop. The result is a strong domesticated population that also displays strong selection pressure to the domestication syndrome. In these runs the wild type is the sub-population that struggles to maintain itself with the dominant domesticated variant.

Weak Domesticated Variant

When humans attempt to select a gene towards values preferred by the wild type it makes it more difficult for the domesticated type to take hold in the population. This is most clear in the cases with non-selective planting (i.e. *wild type III*) or planting seeds with a high *abscission* gene (i.e. those that release seeds early, i.e. *wild type II* and *wild type III*). When seeds with a high *abscission* gene value are planted we get seeds that resemble *wild type III* for similar reasons. When humans select for traits favored by the wild variant both types have seeds that are equally good in the estimation of the humans. As a result humans plant both domesticated and wild variants. Thus human effort is expelled on planting the wild strain. This allows the wild strain to be parasitic on the human effort and to dominate the population.

In the most extreme cases no domesticated variant emerges at all, but in other cases a tiny, weak domesticated variant does emerge but struggles. As an example we look at the run where *fecundity* is maximized (see Figure 5 left column). Since *fecundity* is subject to a strong natural selection pressure to produce the most seeds, human efforts to artificially select seeds in this direction are redundant. We can see the domesticated variant still displays domestication syndrome but represents a much smaller percentage of the overall population. This pattern is repeated when we minimize the *root depth* gene (see Figure 5 middle column) with an even smaller domesticated population.

Interestingly this pattern is repeated when minimizing the *fecundity* gene for a different reason (see Figure 5 right column). The natural selection pressure on the *fecundity* gene is so powerful that the human added artificial selection pressure in the opposite direction has a hard time competing. Producing less seeds is not an advantage to any variants. In particular, if the domesticated variant produces fewer seeds, then wild seeds are likely gathered to make up the difference. The domesticated seeds will be selected for planting and it is possible all domesticated seeds are needed to replace the domesticated crop, or that wild seeds get planted along side domesticated seeds. Either situation makes it harder for the domesticated variant to flourish. As a result the humans are unable to minimize the genetic value as they can in the other cases. Instead the domesticate population settles into a normal distribution with a mean of around 0.3 (at this level some plants will be seedless). In this case the domesticated variant is weaker than other cases but due to the competition between the natural and artificial selection pressures.

Conclusions

We have presented an agent-based model of cereal crop domestication to provide a bottom-up analysis of the role of indirect selection on domestication syndrome. The model was well suited to identifying a clear role for indirect natural selection upon genes that were not exposed to direct artificial selection. As is becoming more accepted by archaeological theories (Purugganan, 2019), the replacement of natural reproductive cycles with artificial ones adjusts the natural selection pressures on key traits related to maintaining the natural cycle. In natural grains, the loss of shattering is an example of this phenomena. Our experiment helps confirm the growing consensus that indirect natural selection plays an important role in this transformation (and others). This analysis also applies to other common traits associated with domestication like seed size, loss of natural seed dispersal mechanisms and reduced dormancy, though our model did not test all of these functions. We plan to extend our model to incorporate more genetic traits inspired by domestication syndrome in grains like seed size and dormancy.

In addition, in our model this indirect natural selection only occurred when humans selectively planted seeds. Or stated negatively, when humans employed non-selective harvesting and planting no indirect natural selection occurred. This suggests that, in our model, some artificial selection of grains is necessary to trigger the indirect natural selection towards domestication syndrome. While our model suggests this selection must be present, our model cannot say that it must be conscious and intentional selection. This result helps to inform archaeological theories on the observed periods of predomestication cultivation in the archaeological record (Weiss et al., 2006; Fuller et al., 2007, 2011). Our model suggests that during these periods humans could have engaged in non-selective harvesting and planting before beginning to engage in selective planting and this transition could trigger selection pressures upon the domesticated variant. The transition could have been from deliberate selection, or from the emergence of a new tool or technique that unintentionally applied selection. Another hypothesis generated from anthropological theories is that selective harvesting with non-selective planting might also be sufficient to begin the domestication process. This is a hypothesis our model is well suited to test and we plan to engage in this experiment next.

Our model was also capable of showing the dynamics of the two crops, wild and domestic, coexisting together in the same space with relative strength depending upon the nature of the artificial and natural selection pressures. When the artificial selection pressure and the natural selection pressure attracted a trait to the same optima, it was difficult (sometimes impossible) for a differentiated domesticated crop to emerge. Whereas, when the artificial selection pressure attracted to a distinct optima a stronger domesticated variant took hold. This provides a secondary potential trigger for domestication in natural populations; selection for a trait not represented among the dominant wild type.

Acknowledgments

References

Allaby, R. G., Brown, T. A., and Fuller, D. Q. (2010). A simulation of the effect of inbreeding on crop domestication genetics with comments on the integration of archaeobotany and genetics: a reply to honne and heun. *Vegetation History and Archaeobotany*, 19:151–158.

Allaby, R. G., Fuller, D. Q., and Brown, T. A. (2008). The genetic expectations of a protracted model for the origins of domesticated crops. *Proceedings of the National Academy of Sciences*, 105(37):13982–13986.

Allaby, R. G., Stevens, C. J., Kistler, L., and Fuller, D. Q. (2021). Genetic revelations of a new paradigm of plant domestication as a landscape level process. *Plant breeding reviews*, 45:321–343.

Angourakis, A., Alcaina-Mateos, J., Madella, M., and Zurro, D. (2022). Human-plant coevolution: A modelling framework for theory-building on the origins of agriculture. *Plos one*, 17(9):e0260904.

Brown, T. A., Jones, M. K., Powell, W., and Allaby, R. G. (2009). The complex origins of domesticated crops in the fertile crescent. *Trends in ecology & evolution*, 24(2):103–109.

Currat, M., Ruedi, M., Petit, R. J., and Excoffier, L. (2008). The hidden side of invasions: massive introgression by local genes. *Evolution*, 62(8):1908–1920.

Fuller, D. Q., Harvey, E., and Qin, L. (2007). Presumed domestication? evidence for wild rice cultivation and domestication in the fifth millennium bc of the lower yangtze region. *Antiquity*, 81(312):316–331.

Fuller, D. Q., Willcox, G., and Allaby, R. G. (2011). Cultivation and domestication had multiple origins: arguments against the core area hypothesis for the origins of agriculture in the near east. *World Archaeology*, 43(4):628–652.

Gerbault, P., Allaby, R. G., Boivin, N., Rudzinski, A., Grimaldi, I. M., Pires, J. C., Climer Vigueira, C., Dobney, K., Gremillion, K. J., Barton, L., et al. (2014). Storytelling and story testing in domestication. *Proceedings of the National Academy of Sciences*, 111(17):6159–6164.

Heiser, C. B. (1988). Aspects of unconscious selection and the evolution of domesticated plants. *Euphytica*, 37:77–81.

Lane, N. and Martin, W. (2010). The energetics of genome complexity. *Nature*, 467(7318):929–934.

Larson, G., Piperno, D. R., Allaby, R. G., Purugganan, M. D., Andersson, L., Arroyo-Kalin, M., Barton, L., Climer Vigueira, C., Denham, T., Dobney, K., et al. (2014). Current perspectives and the future of domestication studies. *Proceedings of the National Academy of Sciences*, 111(17):6139–6146.

Meyer, R. S., DuVal, A. E., and Jensen, H. R. (2012). Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *New Phytologist*, 196(1):29–48.

Peleg, Z., Abbo, S., and Gopher, A. (2022). When half is more than the whole: Wheat domestication syndrome reconsidered. *Evolutionary Applications*, 15(12):2002–2009.

Purugganan, M. D. (2019). Evolutionary insights into the nature of plant domestication. *Current Biology*, 29(14):R705–R714.

Weiss, E., Kislev, M. E., and Hartmann, A. (2006). Autonomous cultivation before domestication. *Science*, 312(5780):1608–1610.

Willcox, G. and Stordeur, D. (2012). Large-scale cereal processing before domestication during the tenth millennium cal bc in northern syria. *Antiquity*, 86(331):99–114.

Zhang, B. and DeAngelis, D. L. (2020). An overview of agent-based models in plant biology and ecology. *Annals of Botany*, 126(4):539–557.

Zohary, D. (2004). Unconscious selection and the evolution of domesticated plants. *Economic botany*, 58(1):5–10.