



CHICAGO JOURNALS



Genetics and Domestication: Important Questions for New Answers

Author(s): Greger Larson

Reviewed work(s):

Source: *Current Anthropology*, Vol. 52, No. S4, The Origins of Agriculture: New Data, New Ideas (October 2011), pp. S485-S495

Published by: [The University of Chicago Press](#) on behalf of [Wenner-Gren Foundation for Anthropological Research](#)

Stable URL: <http://www.jstor.org/stable/10.1086/658401>

Accessed: 03/01/2012 05:09

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



The University of Chicago Press and Wenner-Gren Foundation for Anthropological Research are collaborating with JSTOR to digitize, preserve and extend access to *Current Anthropology*.

<http://www.jstor.org>

Genetics and Domestication

Important Questions for New Answers

by Greger Larson

The recent ability to extract genetic data from archaeological remains of wild and domestic animals has opened up a new window onto the history and process of domestication. This article summarizes the impact of that new perspective derived from both modern and ancient DNA and presents a discussion of the validity of both the methods and conclusions. In general I address the use of post hoc conclusions and the lack of starting hypotheses to inform what we know about domestication from a genetics perspective. I use three case examples (dogs, goats, and pigs) to exemplify fundamental aspects of the genetic data we still do not understand before specifically commenting on the use of molecular clocks to date domestication and the necessity of thinking about domestication as a process. I conclude on a positive note with a brief discussion about the future relationship between genetics and domestication.

Introduction

The hullabaloo really began in 1997. That year, an article appeared in *Science* with the title: “Multiple and Ancient Origins of the Domestic Dog” (Vila et al. 1997). The use of population-level DNA sequence data to reveal insights into animal domestication was not entirely novel. The year before, Bradley et al. (1996) explored the dynamics of African and European cows, but the high-profile nature of the dog article definitively consummated the marriage between genetics and domestication. The article created a stir for two reasons. First, it demonstrated the power of population genetic analysis to reveal details that were previously beyond the scope of an analysis based on either morphology or DNA restriction patterns. Sequences of As, Ts, Cs, and Gs possessed a degree of resolution that could not be matched by bones or differing molecular fragment lengths.

Second, the primary conclusion of the article, that dogs were domesticated 135,000 years ago, simultaneously sparked the imaginations of science journalists and in equal measure infuriated zooarchaeologists who knew that the oldest bones that could be safely ascribed to fully domestic dogs were no more than 10,000–12,000 years old (Clutton-Brock 1995). The estimated age of 135,000 years was nonsense. Archaeologists knew the dates were wrong, but a lack of familiarity

with genetic methods meant they could not say why. The sexy conclusions and the high impact that have often been generated from these kinds of studies (Vila et al. 1997 has been cited more than 300 times in 13 years) combined with the power to begin sentences with the words “in direct contrast to long-held beliefs” have led to a flood of domestication genetics papers.

In this essay I will review not only the conclusions of a number of publications in this vein but also the more general paradigms that geneticists have operated under in order to guide their research. I will then apply a thought experiment to demonstrate how little is known regarding the fundamentals of genetic data before addressing specific questions related to molecular clocks, the process of domestication, and the ramifications that domestication studies can have on other fields. First, however, in order to critique the validity of genetics-based assertions, it is worth discussing briefly the methods such studies employ and the strengths and weaknesses therein.

A Genetics Primer

The basic modus operandi of these studies is as follows. Hundreds if not thousands of (typically modern) samples of a given species are collected from as many different geographic locations, breeds, or populations as possible. These samples are usually derived from tissue, hair, or feathers, and each sample is bathed in a series of chemicals into order to isolate the DNA.

The extracted DNA possesses millions of copies of the entire genome of the organism as well as many more copies of the mitochondrial genome housed within the mitochondria

Greger Larson is Fellow and Principal Investigator in the Department of Archaeology, Durham University (South Road, Durham DH1 3LE, United Kingdom [greger.larson@durham.ac.uk]). This paper was submitted 13 XI 09, accepted 2 XII 10, and electronically published 13 VI 11.

organelle found in virtually every cell. Domestication geneticists have historically ignored the multibillion base pair nuclear genomes of their samples and instead focused on the 16,000 base pair mitochondrial genome. As sequencing technology becomes cheaper and faster, however, the nuclear genome is becoming increasingly accessible. Still, the mitochondrial genome remains attractive for several reasons.

First, it does not recombine. That is, the only changes that occur in the sequence of a mitochondrial genome are the result of mutations. Thus, modeling its linear evolution is far simpler than having to think about and account for diploid genomes that hybridize and swap genes every generation. Second, although most of the 16,000 base pairs that make up the mitochondrial genome code for genes crucial to an organism's basic survival, there is a small region, typically less than 2,000 base pairs, between these protein-coding genes where replication begins when a new mitochondrial genome is being built. This region, known as either the control region or the d-loop, is not part of the translation or transcription process and is thus not under the same restrictive selective regime to which the rest of the genome is subjected.

Mutations, or errors that occur during replication, occur all the time. It is the fate of those mutations that differ. If a mutation occurs in a protein-coding gene that negatively influences the functional properties of that gene, then the mitochondria carrying that mutation will not be replicated as often (if at all), and the mutation will disappear. This means that although mistakes during replication happen all the time and in an unbiased fashion across the genome, rates of substitution, or rather the potential of that mutation to survive, are distinctly nonrandom. Because the control region does not code for a protein and thus there are no ramifications (good or bad) for a new mutation, far more of those mutations become incorporated into the sequence. Evolution is simply change through time, and because mutations in this region generally do not affect the organism, the control region evolves at a faster rate than the rest of the genome, where most mutations are deleterious and are selected out before they can become incorporated and sequenced by researchers. This relatively speedier evolution allows for differences between populations and species to build up over short periods of time, thus enabling geneticists to differentiate between conspecific individuals and to draw inferences from the sequences regarding the demographic history of the species. Though mitochondrial DNA (mtDNA) data almost always distinguish between species and can often distinguish populations below the species level, some populations, such as dog breeds, do not possess diagnostic mtDNA signatures.

Once the DNA has been extracted, geneticists select and amplify a fragment rich in variability, typically a few hundred base pairs of the control region of the mitochondrial genome, from every individual. As stated above, this region evolves quickly, and as a result, geneticists expect to find differences (substitutions) at numerous positions along the sequenced fragment, the patterns of which differentiate the individual

samples. Not every individual, however, will possess a unique sequence. Numerous samples will often carry the identical sequence of base pairs along the entire fragment. The set of samples that share such identical sequences are said to possess the same haplotype. Although the word "haplotype" has different meanings depending on the type of genetic marker amplified by geneticists, in the studies discussed below, a "haplotype" simply refers to a unique combination of base pairs across the amplified fragment. It is with the haplotypes that the next stage of analysis begins.

A set of haplotypes can be used as raw material to build either a phylogenetic tree or a haplotype network. Both kinds of diagrams visually depict the relationships between the haplotypes. Of the two, networks are the easiest to understand. They typically consist of circles and lines in which each circle represents a haplotype, and the size of the circle corresponds to the number of individuals in the data set that possess that haplotype. Larger circles result when more individuals are all identical across the sequenced region, and the smallest circles represent those individuals that possess haplotypes not shared by any other individual. Two circles connected by a line differ by a single substitution no matter where in the sequence that substitution sits, and hash marks are often placed across the line to indicate additional substitutions. These kinds of figures are often simple enough to be drawn by hand, but software programs also exist to first identify the haplotypes in a data set and to draw the corresponding network depicting how closely or distantly related all the haplotypes are from one another (fig. 1).

Phylogenetic trees are generally more complex than networks because they employ models of evolution to infer the evolutionary relatedness of the haplotypes. The evolutionary models place differing weights on the kinds of substitutions encoded in the sequence, and these weights alter the mathematical distances between the haplotypes. Because trees are depictions based on those distances, different models can produce differently shaped trees or trees with the same shape but with differing levels of statistical support for the branches. Adding additional individuals from previously unsampled populations that possess novel haplotypes can also alter the shape of the networks and trees.

Once networks and trees have been generated, two additional terms are often used to discuss their shapes. A "haplogroup" on a network is a cluster of closely related haplotypes that together create an easily recognizable group that is more or less (it is hoped more) differentiated from other haplotypes and haplogroups. A "clade" on a tree is more or less the same thing, although its technical definition is a group of haplotypes that are more closely related to each other than any one is to any other haplotype. In human-relationship terms, this would be called a family. Because a single tree is just one of many that could possibly be drawn from the data, phylogeneticists prefer to generate numerous trees to see how frequently the same patterns of clades and haplogroups appear. The more often they appear, and the more robust they are

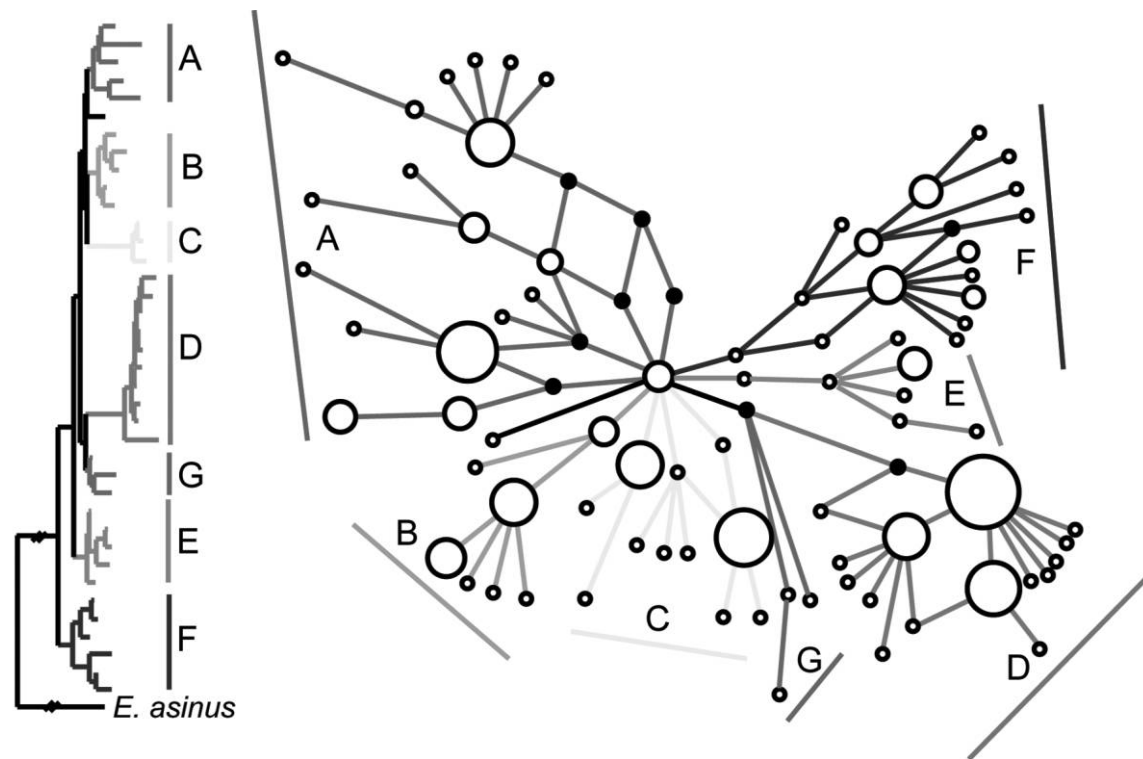


Figure 1. Example of a phylogenetic tree on the left (rooted with a donkey, *Equus asinus*) and an unrooted network on the right, both of which were generated using mitochondrial control region haplotypes of modern horses. The colors and letters associated with the clades on the tree correspond to the colored and lettered haplogroups on the network. (A color version of this figure is available in the online edition of *Current Anthropology*.) Note the correlations between the relative positions of the haplogroups on the network and on the tree. The tree and network are adapted respectively from Vila et al. (2001) and Jansen et al. (2002).

to different parameter values within separate models of evolution, the more confidence phylogeneticists have that the relationships are “real” and not just artifacts of the data (fig. 1).

At this stage—armed with a network, a tree, or both—geneticists are ready to begin the process of interpreting the images and gleaning the implications for our understanding of how, when, where, and how many times domestication of a particular species has occurred. Having laid out the basic methods of these studies, what follows is a short critique of the ways in which inferences and conclusions have been drawn.

The Appeal of the Post Hoc Narrative

The majority of studies of the ilk I describe above do not contrast their observed data with an expected result. For the most part, there are no expectations regarding the shapes of networks, the number of haplotypes or haplogroups, or the

structure of trees derived from the data. These studies do not follow the textbook scientific method that begins with a falsifiable assumption and dichotomous easily defined expectations and ends with a comparison between the generated data and the expected result. This is not necessarily a bad thing. Scientific enquiry, especially at the early stages of data gathering using a newly available technique, is often investigative and explorative. Newly derived categories of data cannot be expected to be acquired or interpreted within the confines of explicitly stated hypotheses. Given the relatively short period of time that population-level sequence data have been available, it is perhaps no surprise that within the field of domestication genetics, there has been little explicit hypothesis testing. My own work on pigs is no exception.

Though many domestication studies strive to interpret the data within the context of what is already known about their study animals, many of them are content to report only descriptive accounts of the generated data. Not every study has

done this, of course, but the general trend was recently exemplified by a 2007 article, the title of which was “Large-Scale Mitochondrial DNA Analysis of the Domestic Goat Reveals Six Haplogroups with High Diversity” (Naderi et al. 2007). This particular article does, in fact, present insights regarding goat domestication, but the plainly descriptive title hints at the nature of many of these studies that generate data in theoretical vacuums.

In order to elucidate this trend, I present three case studies that focus on dogs, goats, and pigs. All three employed the general methodology discussed above, and taken together they demonstrate the potential and limits of genetic domestication studies. I have chosen these three because populations of the respective wild ancestors—wolves, bezoars, and wild boar—remain extant, thus allowing for a comparative analysis of the genetic patterns found in both wild and domestic animals. For studies of animals whose wild ancestors are either extinct (e.g., cows and camels) or uncertain (e.g., sheep), the genetic differences between the wild and domestic forms can only be revealed by generating DNA sequences from archeological material. A large number of studies have attempted to do just that with respect to cows (Edwards et al. 2007), but the first ancient sheep and camel DNA article are still forthcoming.

Because wolves, bezoars, and wild boar are still around, geneticists are able to sample them and place both wild and domestic variants into the same network or tree. Though more recent studies have been published, a 2002 study of dogs (Savolainen et al. 2002) is instructive. This study typed more than 600 domestic dogs and nearly 40 wolves, numbers that in 2002 were relatively large. Two articles focused on goats were published in 2007 and 2008 (Naderi et al. 2007, 2008), but in the intervening 5 years, the acceptable standard for sample numbers had increased, and these studies analyzed 2,430 domestic goats and 473 wild bezoars.

The first statistic normally generated in these articles is the number of haplotypes found among all the samples. In these cases, the authors identified 110 unique haplotypes in dogs and 17 in wolves. The goat studies, based on fourfold more domestic samples and tenfold more wild samples, identified a total of 1,783 unique haplotypes in both populations. The issue of what those numbers mean and whether they are significant is difficult to answer for the simple reason that no one knows how many haplotypes to expect from a given number of populations or individuals.

Even without this understanding, a comparative approach can be used across species to ask new questions that will form the basis of future studies. The first question worth asking is how many haplotypes are found in both the wild and domestic samples. In canines, out of a total of 127 haplotypes, only one was identified in both dogs and wolves. In caprines, of 1,783 unique haplotypes, only three were shared by both wild and domestic goats, and those three were found only on the island of Sicily, where the status of the goats and the timing of their arrival is uncertain. According to the authors, the domestic goats found on this island could be ancestors of

wild animals only recently transported there (Naderi et al. 2007). If true, the number of shared haplotypes between truly wild and truly domestic would be 0.

This observation has not gone unnoticed, and the near universal lack of shared haplotypes between dogs and wolves has been exploited as a means to identify recent hybrids by observing stereotypically dog haplotypes in modern wolves (Randi and Lucchini 2002). Still, beyond the use of this observation as a conservation tool, no one has yet questioned why wild and domestic animals of these species share so few haplotypes.

A Thought Experiment Involving Haplotypes

We know that dogs and goats are derived exclusively from wolves and bezoars, respectively. Thus, the earliest domestic populations must have shared 100% of their mitochondrial haplotypes with their wild counterparts. Given this, the question must be why and how has the shared proportion dropped to virtually 0%. One explanation could be that the original wild populations that gave rise to domestic stocks are now extinct and the sampled extant wild populations in these studies were not involved in the domestication process. This could be especially true for wolves, which have suffered a long history of persecution. Under this scenario, however, the expected networks and trees would generate haplogroups that consist of either wild or domestic animals. In the dog study, the tree did in fact demonstrate that some clades consisted only of wolves or dogs consistent with the extirpation scenario, but the majority of clades contained haplotypes of both dogs and wolves even if that was because some of the haplotypes were shared (Savolainen et al. 2002). In goats, though several bezoar-only clades are evident, every single domestic goat sample is found within a cluster of bezoars, though again, none of the haplotypes are shared (Naderi et al. 2008). These patterns do not fit any simple scenario of domestication that focuses on demographic patterns of limited sampling from wild populations and periodic bottlenecks for both wild and domestic animals.

The explanation above rests on an assumption that though the control region of mtDNA does evolve quickly relative to both other genes in the mitochondrial genome and the nuclear genome of the organism, it is traditionally not thought to be fast enough for mutations to accumulate over the relatively short time frame of domestication (10,000 years). If true, this would mean that haplotypes found in modern wild and domestic animals have not changed since the beginning of the Holocene and that the observed substitutions not only occurred long before domestication but also reflect population structuring that resulted from a long-term lack of gene flow between geographically partitioned groups. This assumption may not always hold, however, and a series of articles has suggested that substitution rates are not fixed (Ho and Larson 2006; Ho et al. 2005).

These authors demonstrated that the evolutionary rate derived from a data set is dependent on the time depth of the most recent common ancestor of the studied sample set. A data set consisting of a group of humans known to have had a common ancestor on the order of hundreds or thousands of years will possess a great deal more variation than what would be expected using standard evolutionary rates. When the data set is increased to include chimpanzees and other primates, the date of the most recent common ancestor is pushed back to a scale of millions of years, and the evolutionary rate tumbles. This so-called time dependency of evolutionary rates could result from the retention of slightly deleterious mutations over a sufficient time frame to be included in population-level data sets. Over longer time frames, those mutations are eliminated, which would then reduce the observed variability in the data set and give the appearance of a slower evolutionary rate. Additional studies of different species have thus far confirmed the phenomenon (Burridge et al. 2008) even if a fully satisfactory explanation remains elusive.

What this might mean is that we should not necessarily expect wild and domestic haplotypes to be identical. Instead, wild and domestic individuals that shared a common ancestor around the time of the origins of domestication would possess substitutions that have accumulated since they split. So long as this pattern was generalizable across different animal domesticates, this would explain why wild and domestic dog and goats fail to share any common haplotypes.

Pigs, however, are different. The pig data contradict the dog and goat data in at least two key ways. First, wild boar and domestic pigs share at least 17 haplotypes (Larson et al. 2005, 2007a, 2007b, 2010). This could easily result if the pig data were based on shorter sequences than dogs or goats, thus reducing the chances of finding substitutions that differentiate individuals, which would lead to a reduction in the number of overall haplotypes. The number of base pairs amplified for dogs, goats, and pigs, however, is 582, 469, and 662, respectively. All else being equal, pigs should therefore possess more total haplotypes and fewer shared haplotypes between wild and domestic animals. This is not the case.

Not only are the two most frequent domestic haplotypes found in Europe also found in European wild boar, more than 15 haplotypes are found in both East Asian wild boar and Chinese domestic breeds, an additional haplotype was shared by Indian wild boar and domestic pigs, and another was shared by wild boar from Vietnam and domestic and feral pigs found in Island Southeast Asia (Larson et al. 2007b, 2010). The most obvious explanation for this pattern is that it results not from distinct instances of domestication but that like the dog scenario discussed above, the shared haplotypes are the result of recent hybridizations between introduced domestic pigs and indigenous wild boar. Though this explanation cannot be ruled out, there are two significant factors that make it less likely. First, mtDNA is passed solely along the maternal line. Thus, in order for domestic pigs to share

the same haplotype as an indigenous wild boar that was never part of a domestication process, male domestic pigs would have to mate with female wild boar, and the piglets would have to be incorporated into the domestic stock. The opposite scenario is common practice in many cultures, especially in New Guinea, where females are often left tied to a stick at the edge of a village overnight and are subsequently impregnated by feral males from the forest. In this case the resulting piglets retain their mother's domestic mitochondrial signature.

Second, an argument that assumes a high degree of hybridization must explain why so many populations of indigenous wild boar—including those in India and on islands such as Japan, the Ryukyu chain, and Taiwan—retain their genetic distinctiveness (Larson et al. 2005) and why domestic pigs introduced to these areas have not acquired the local DNA haplotypes. Neither the model that associates shared haplotypes with independent domestication nor the model that assumes all instances of shared haplotypes are the result of recent hybridizations explains the data. The truth, of course, probably lies somewhere in the middle, although observing and describing DNA evidence is only the first step to uncovering it.

This issue touches on a second key difference between the wild boar and wolf and bezoar data sets, and for this discussion it is worth explaining another common term. "Phylogeography" is the study of the association of phylogenetic signals with the geographical provenance of the samples. A strong phylogeographic signal is the result of a high degree of reciprocal correlation between a geographic region and a specific haplogroup or clade. If an analysis of a wild population demonstrates that hypothetically, highly differentiated haplogroups are found in Spain, Italy, and Greece but that animals carrying all three types are present in northern Europe, geneticists would be tempted to suggest that the strong phylogeographic pattern in southern Europe suggests a genetic differentiation that took place in refugial regions during ice ages and a mixing of haplotypes when those populations migrated north after a climactic amelioration. By assigning colors to specific haplogroups or clades and by pinning the colors onto a map, geneticists are able to ascertain the relative strength of the phylogeographic signal.

A strong signal is desirable because it allows authors to pinpoint hypothetical centers of domestication. Unfortunately, most wild animals involved with domestication lack strong signals, at least when the data sets consist only of mtDNA. (As sequencing techniques become cheaper, data sets studies that interrogate and analyze nuclear genomes [e.g., vonHoldt et al. 2010] may reveal more geographically proscribed and genetically distinct populations of wild and domestic populations.) Three wolves, for example, sampled from China, Mongolia, and Saudi Arabia all possessed the same mitochondrial haplotype, as did individual wolves from Turkey, Sweden, and Portugal (Vila et al. 1997). Thus, assigning an origin to dogs who possessed haplotypes closely related to

these wolves is problematic or at least lacking in precision. Modern bezoars are significantly more geographically circumscribed than modern wolves, and though some haplotypes seem to be found only in small regional pockets, almost all of the full complement of haplotypes were identified in bezoars sampled exclusively between Turkey and Iran (Naderi et al. 2008).

In many ways this makes sense. Wolves migrate long distances during their lifetimes, and thus different haplotypes are expected to be present at many locations across the Old and New Worlds. Humans, too, have been responsible for the movement of both wild and domestic animals, thus smearing and blurring any phylogeographic pattern that may have existed in the Pleistocene. In addition, the history of the ice ages, as described in the hypothetical above, has also played a role by forcing populations apart where they begin to diversify before reuniting them. This has been shown to play a role in yaks, herds of which often contain individuals with highly variable and differentiated haplotypes (Ho et al. 2008).

The overall effect of these homogenizing forces should lead to a modern-day situation in which no wild population retains a strong phylogeographic signal. Yet unlike virtually every other wild animal involved in domestication and in defiance of both their natural migratory ability and a long history of human-assisted transport and reproductive meddling, wild boar do. This strong phylogeographic signal allows for a relatively straightforward identification of centers of origin. For instance, a handful of wild boar collected in India are all positioned in a haplogroup that is significantly different than all other groups. These haplotypes are only present in South Asia, and thus when an Indian pig identified as domestic also possessed the same signature, the most parsimonious explanation was that these wild boar were likely involved in domestication (Larson et al. 2005). Of course, it is possible that domestic pigs derived from a separate population in a different place were transported to India and then mated with an indigenous female wild boar, thus producing a litter that was retained by humans. The data cannot differentiate between these two scenarios, but the strength of the phylogeography at least allows the suggestion of an independent Indian domestication to be made, which can then be further investigated and corroborated by archaeological or historical sources.

When phylogeographic signals are weak, suggestions regarding the geography of domestication rest on more subtle arguments. The most popular one is based on a determination of the genetic variability present in domestic animals in different regions. The more variation a region possesses relative to the total diversity evident in an entire data set, the more likely that region was a center of origin because, as the argument goes, only a subset of the total diversity is generally transported by people away from the center. This is certainly true of human diversity, which is far higher on the African continent than it is in Australia, or at least it was until the fifteenth century, when large numbers of genetically diverse

people began migrating to Australia, the Americas, and other parts of the world. A modern analysis of this kind that did not take into account the historical migrations of people to the New World would conclude that the United States of America was the origin of all humans. That is to say, highly diverse regions can result not just from a legacy of origination but also from migration into the region by genetically diverse populations. Thus, demonstrating that a region is particularly diverse without also offering nongenetic evidence suggesting the region was in fact a center of domestication is problematic at best. All of the data related to cows is used in this argument (Troy et al. 2001) as well as in the argument that dogs were first domesticated in China (Savolainen et al. 2002). In the former case, archaeological evidence also supports a Near Eastern origin of cattle domestication and a subsequent Neolithic migration into Europe, suggesting that the genetic interpretation is correct. In the latter case, however, a supporting narrative based on archaeology remains elusive, and a recent publication using genetic data from a large number of geographically isolated wolves and domestic breeds concluded that because Near Eastern wolves also played a large role in the domestication of dogs (vonHoldt et al. 2010), China was likely not the sole center of dog domestication.

Confidence-Free Molecular Clocks

Numerous attempts have also been made to place an independent time frame on the history of domestication using genetic data sets. In many cases, the authors of these articles have concluded or at least implied that animal domestication began hundreds of thousands of years ago (Ho and Larson 2006). Perhaps the most famous of these attempts was the Vila et al. (1997) publication that pushed dog domestication back more than 100,000 years. The authors were able to conclude this by first determining that the average mitochondrial genetic difference between wolves and coyotes was 7%. By borrowing a date of one million years for the last shared common ancestor between the two species, they established a rate of 1% per 135,000 years. As discussed above, because most of the clades on their tree contained both wolves and dogs, the authors estimated the divergence between wolves and the one clade that only contained dogs. The figure, 1%, meant that dogs and wolves last shared a common ancestor 135,000 years ago (Vila et al. 1997).

There were a number of assumptions made during this exercise, and the decision to not bracket the estimate with error bars gave an unwarranted impression of precision. The conclusion proved intriguing, however, and for several years all genetic animal domestication articles included highly suspect molecular clock analyses (Ho and Larson 2006). Very few (if any) efforts have been made to combine all the possible sources of error associated with these kinds of analyses in an effort to confidently ascertain the precision of the estimates.

More recently, another source of error, the time dependency of molecular clocks discussed above, has only added to the

error, although the variable-rate issue does go some way toward explaining the discrepancy between the dates derived from molecules and from archaeology. In most cases, geneticists applied evolutionary rates derived from data sets whose most recent common ancestor existed millions of years ago to data sets of populations whose common ancestor was far more recent. By applying a slow clock to a data set that possessed substantial variation, they significantly overestimated the time it would take to produce that variation, thus pushing the timing of domestication deep into the past. Taking this effect into account only removes a single source of error, however, and the combination of all the others suggests that the error bars almost certainly encompass the present day. Thus, molecular clock efforts so far simply lack the precision to date Holocene phenomena. And this is true even after putting aside the issue of what a domestication date actually means.

The production of large mtDNA data sets to glean insights into domestication is no longer novel, and the number of species left to investigate in this manner is dwindling. The general approach has a great deal of merit, and this first stage of sequence generation is necessary to understanding how a genetics-based approach can help us to understand domestication. But it is just the first stage. If sequence data have thus far failed to revolutionize our understanding of the patterns and processes of domestication, I suspect this is because we may have hoped that the data themselves would be easily interpretable and provide robust conclusions. Without starting hypotheses about what the data sets would generate, however, easy interpretations were only possible if the trees and networks revealed something immediate and obvious. When they did not, we have been left to either simply describe what we see or tell post hoc stories sometimes using shaky assumptions.

I am confident, however, that the next stage will achieve a great deal more. Far from asking how many times was species X domesticated, we should be asking why are so few haplotypes shared between wild and domestic animals? What climatic conditions or landscape contexts are necessary to produce x number of clades or haplogroups? What exactly constitutes a high level of diversity? Is it appropriate to compare levels of diversity between species? By focusing not on how to interpret the data but instead on how many ways the data set can be generated and under what conditions and parameters, we can begin to replace post hoc explanations with a hypothesis-testing framework. Again, this is not to say that post hoc narratives are inferior; they are a vital prerequisite to further understanding, but they are limited in the degree to which they can ultimately inform the history of domestication.

Hypothesis-driven research in this vein is already yielding fascinating new conclusions. An article published by Allaby, Fuller, and Brown (2008) employed simulations to reveal that post hoc narratives used to support a rapid transition from wild to domestic crops were based on a false assumption

regarding how the data were derived. They demonstrated that counterintuitively, multiple origin scenarios of crop evolution are more likely to give the superficial impression of a single origin than a single origin scenario. This result demonstrates that our intuitions are not always valid and that we should therefore simulate data sets based on our assumptions of what is supposed to happen to see what other mistakes we might be making when divining the “obvious” story from the shapes of networks and trees.

A Note on the Process

The Allaby, Fuller, and Brown (2008) article is also noteworthy because it does not ignore the long history of domestication. Domestication, like speciation, is not an event. Geneticists know this as well as archaeologists, but for a multitude of reasons including convenience, we often use the word “event” and describe wild and domestic as complementaries, that is, opposites that possess no intermediate form (Dobney and Larson 2006). This fallacy is maintained largely because processes are messier than events, and an event mind frame is a necessary fudge that must be assumed before analyses such as molecular clocks can be applied. None of the attempts to place a molecular time frame on the history of domestication differentiates between the beginning and the end of the process. Instead, a single date estimate is gleaned that is intended to be interpreted as the year in which wild became domestic.

If we are to embrace the process (see Denham 2011; Marshall and Weissbrod 2011; Piperno 2011; Vigne 2011; and Zeder 2011), we have to think differently about both the questions and the data sets. As the Allaby, Fuller, and Brown (2008) study demonstrated, an approach that replaces or at least supplements the mitochondrial genome with the full nuclear genome has enormous benefits. By looking at the genome of the organism, which contains the genes that code for the differences between wild and domestic individuals, it becomes possible not just to understand what genes are changing but precisely how those changes affect the total animal. Using these kinds of data sets, we can start asking deeper questions that focus not on the where and when but on the how. In other words, it may soon be possible to identify the genetic alterations that took place between the first steps of domestication (fig. 2) and today.

The Belyaev fox-farm experiments that began in Siberia in 1959 revealed that by selecting solely for tameness, it was possible to produce, in relatively few generations, a population of foxes that looked and acted like domestic dogs (Trut 1999). That much is well known. Two other aspects of these experiments have been less well publicized. First, the farm experimented not just with foxes but also with populations of rats, beavers, and other animals. Second, the goals were to produce both extremely tame animals and extremely aggressive ones as well. An anecdote from these later revelations stated that the Soviet army was ready to deploy large numbers of the



Figure 2. *Citizen Dog* cartoon by Mark O'Hara (reprinted by permission of Universal Press Syndicate, all rights reserved) that presents a more nuanced (and possibly more accurate) depiction of the process of domestication than is often built into genetic models of the same phenomenon.

most aggressive beavers on the Soviet borders in the event the U.S. military ever dared a land invasion.

The two colonies of tame and aggressive rats are now in residence in Leipzig, Germany. By first crossing individuals from both groups and then measuring 45 separate physiological and behavioral traits, a recent study (Albert et al. 2009) was able to identify two specific quantitative trait loci associated with tameness. This kind of study represents an important first step in revealing links between genetics and behavior and begins to test the hypothesis that a small number of genes are ultimately responsible for the large behavioral and phenotypic differences that divide wild and domestic animals (Dobney and Larson 2006; Stricklin 2001).

Despite a lack of access to parallel populations bred explicitly for this purpose, a number of geneticists have already developed a long history of insights into the genetic architecture underlying domestic phenotypic traits. These kinds of studies have generally been focused on single traits, many of which are commercially important. Geneticists first type a large number of known variable positions across the genome in two populations of animals, one that possesses one variant of a trait, such as a white coat, and one population that has a different coat color. A comparison of the regions of difference and similarity across the genome allows the geneticists to focus their search, and from there they use similar methods to isolate the fragment of DNA that possesses the causative mutation(s) underlying the trait. Actually identifying the mutation is often more difficult, although on occasion, such as in traits for muscle growth in pigs, a single mutation was pinpointed (Van Laere et al. 2003).

Occasionally these types of studies reveal insights into the history of domestication. After identifying the gene responsible for yellow legs in chickens, geneticists then sequenced

the region in a variety of wild jungle fowl. An alignment of four different wild species revealed that although the majority of the domestic genome was identical to the wild red jungle fowl, the gene responsible for producing yellow legs showed a far greater identity to the same region found in gray jungle fowl. This result resolved the paradox of how yellow legs, a trait never seen in red jungle fowl, could be so prevalent in domestic chickens, but in so doing it also revealed a somewhat unexpected conclusion that chickens are not derived from a single ancestor (Eriksson et al. 2008). This revelation opens up an entirely new set of questions related to the process of domestication, the frequency of hybridization and the creation of hybrid domestic animals, and the debate over the degree of human intentionality in selecting for specific traits at various stages.

Perhaps the best bet we have for using genetics to unravel the big questions surrounding domestication is to look for the newly identified changes that underlie key traits in the bones of domestic animals found in archaeological contexts. A recent study on coat colors in pigs demonstrated that the pattern of mutations that cause coat colors—including red, black, and white spotted—are the result of a strong selection pressure away from the camouflage coat colors selected for in the wild (Fang et al. 2009). The suggestion is that coat color variation has been a feature of domestication from the very beginning of the phenomenon. Armed with the causative mutations, this hypothesis can be tested by screening ancient bones for the genetic variants that underlie the specific coat colors. This method has already been used on both ancient mammoth (Rompler et al. 2006) and horse (Ludwig et al. 2009) remains, the latter of which revealed an explosion in the number of coat colors in horses around the fifth millennium BP.

Proxies for Domestication's Ramifications

Ancient DNA techniques will no doubt be employed in future studies to type phenotypic traits in subfossil material. As discussed above, uncovering a strong phylogeographic signal in domestic animals using an alignment of neutrally evolving DNA (the control region of the mitochondrial genome) has been rare, though wild boar possess an inexplicably strong relationship between the phylogenetic placement and their geographic provenance.

In his book *Guns, Germs, and Steel* (Diamond 1997), Diamond discusses the universal tendency for populations that have acquired agriculture and domestic animals to first develop a large population and then to move (see also Bellwood 2011). Diamond recounts migrations of people armed with a suite of domestic crops overtaking indigenous hunter-gatherers in, among other places, Europe, East Asia, sub-Saharan Africa, and New Zealand. The routes and timings of these migrations are often contentious, but given the fact that domestic animals were always a key part of the migratory package, the genetic signals derived from their remains can act as a proxy for human migration.

Because wild boar indigenous to Europe possess such a divergent haplotype from those native to the Near East, a short fragment (less than 85 base pairs) of DNA was enough to ascertain the genetic legacy of an ancient pig bone. The wild or domestic status of pig bones was determined using a morphological analysis (though of course many remains could not be confidently assigned to either category) after which the diagnostic fragment was amplified. Not surprisingly, the bones identified as wild in European Mesolithic and Neolithic contexts were European in origin. The domestic bones from a number of sites stretching from Romania through Germany to France, however, displayed a Near Eastern signature. Although this pattern conformed to expectations based on the known history of the Neolithic migrations into Europe, what was a surprise was the speed with which the Near Eastern lineages were replaced by domestic pigs of European origin, first in Europe itself and then in the Near East (Larson et al. 2007a). European wild boar are now the primary (if not sole) progenitors of European domestic pigs, although whether this process was initiated independently of the Near Eastern pig domestication or whether it was kick-started by the introduction of Near Eastern pigs remains an open question.

Beyond demonstrating the use of genetics to reveal the patterns of movement among a key domestic animal and hence the movement of their human herders, the study by Larson et al. (2007a) also underlined the dangers of inferring historical patterns based on modern data alone. All modern continental pigs in Europe possess European-specific mitochondrial haplotypes. But they only do so today because the Near Eastern-specific pigs originally brought into Europe have been completely replaced, leaving no descendants in modern pig populations. Given the number of human mi-

grations and instances of animal transport that have taken place since the Neolithic, it is a certainty that domestic animal populations originally introduced into a new region have subsequently been replaced, perhaps several times over. A temporal perspective is thus a necessity for any study that pretends to a robust conclusion regarding the long-term history of population movements.

Conclusion

Given the relatively short period of time over which genetic methodologies have been applied to domestication questions, it is perhaps no surprise that the initial claims are now being tempered. This is the nature of youth. Practitioners of a new technique with the promise of novel data sets have the benefit of knowing that every result is potentially revolutionary. Journalists and academic journals alike are delighted to publish the rapidly generated conclusions of the new method, and the more often the new studies overturn conventional wisdom or directly contradict decades of findings based on more traditional methodologies, the better.

As the field eases beyond its teenage brashness, my position is that there is now time to take stock and to begin questioning the assumptions on which many of our early studies were based. The massive data sets that will be generated as part of the high throughput sequencing revolution will reveal fine-scale structure at the population level and new genes important in the domestication process. I suspect the new technologies will also generate insights not just into DNA sequences but RNA sequences as well. Insights at this level of organization will facilitate an understanding of not only what genes were key but in which tissues and when they are active. These kinds of studies will have significant ramifications not just for domestication but also for the nature of evolutionary change. In addition, a focus on simulation and modeling will reveal how demographic changes affect the patterns in population genetic data, which will better allow us to choose which of several competing scenarios best explains the early history and process of domestication. Finally, improved sequencing techniques will allow for an essential temporal component to be layered onto the data, and thus with any luck a complete understanding of the hows, whens, wheres, and maybe even the whys of domestication will be within our grasp.

The discussions that took place at the "The Beginnings of Agriculture: New Data, New Ideas" Wenner-Gren Foundation Symposium in March 2009 in Mexico went a long way toward solidifying my impression that the big questions are increasingly knowable. First, highly precise data regarding the specific temporal, geographic, and ecological circumstances in which domestic plants and animals became integrated into human settlements are accumulating at an unprecedented pace. This level of detail is allowing researchers to piece together the specific order of events (on a region-by-region basis) that first set the stage and then allowed for domestication to take place

(e.g., Zeder 2011). Second and equally impressive, the theoretical framework for understanding the process of domestication at the macrolevel is becoming ever more refined. These structures (see Denham 2011) will allow us to place the new data into a scaffold that will facilitate a genuine comprehension of the bigger themes of global domestication on top of their specific regional narratives. These are exciting times.

References Cited

- Albert, F. W., O. Carlborg, I. Plyusnina, F. Besnier, D. Hedwig, S. Lautenschlager, D. Lorentz, et al. 2009. Genetic architecture of tameness in a rat model of animal domestication. *Genetics* 182: 541–554.
- Allaby, R. G., D. Q. Fuller, and T. A. Brown. 2008. The genetic expectations of a protracted model for the origins of domesticated crops. *Proceedings of the National Academy of Sciences of the USA* 105:13982–13986.
- Bellwood, Peter. 2011. Holocene population history in the Pacific region as a model for worldwide food producer dispersals. *Current Anthropology* 52(suppl. 4):S363–S378.
- Bradley, D. G., D. E. MacHugh, P. Cunningham, and R. T. Loftus. 1996. Mitochondrial diversity and the origins of African and European cattle. *Proceedings of the National Academy of Sciences of the USA* 93:5131–5135.
- Burridge, C. P., D. Craw, D. Fletcher, and J. M. Waters. 2008. Geological dates and molecular rates: fish DNA sheds light on time dependency. *Molecular Biology and Evolution* 25:624–633.
- Clutton-Brock, J. 1995. *A natural history of domesticated mammals*. Cambridge: Cambridge University Press.
- Denham, Tim. 2011. Early agriculture and plant domestication in New Guinea and Island Southeast Asia. *Current Anthropology* 52(suppl. 4):S379–S395.
- Diamond, J. 1997. *Guns, germs, and steel: the fates of human societies*. New York: Norton.
- Dobney, K., and G. Larson. 2006. Genetics and animal domestication: new windows on an elusive process. *Journal of Zoology* 269:261–271.
- Edwards, C. J., R. Bollongino, A. Scheu, A. Chamberlain, A. Tresset, J. D. Vigne, J. F. Baird, et al. 2007. Mitochondrial DNA analysis shows a Near Eastern Neolithic origin for domestic cattle and no indication of domestication of European aurochs. *Proceedings of the Royal Society B: Biological Sciences* 274:1377–1385.
- Eriksson, J., G. Larson, U. Gunnarsson, B. Bed'hom, M. Tixier-Boichard, L. Stromstedt, D. Wright, et al. 2008. Identification of the yellow skin gene reveals a hybrid origin of the domestic chicken. *PLoS Genetics* 4:e1000010, doi:10.1371/journal.pgen.1000010.
- Fang, M., G. Larson, H. S. Ribeiro, N. Li, and L. Andersson. 2009. Contrasting mode of evolution at a coat color locus in wild and domestic pigs. *PLoS Genetics* 5:e1000341.
- Ho, S. Y. W., and G. Larson. 2006. Molecular clocks: when times are a-changin'. *Trends in Genetics* 22:79–83.
- Ho, S. Y. W., G. Larson, C. J. Edwards, T. H. Heupink, K. E. Lakin, P. W. H. Holland, and B. Shapiro. 2008. Correlating Bayesian date estimates with climatic events and domestication using a bovine case study. *Biology Letters* 4:370–374.
- Ho, S. Y. W., M. J. Phillips, A. Cooper, and A. J. Drummond. 2005. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Molecular Biology and Evolution* 22:1561–1568.
- Jansen, T., P. Forster, M. A. Levine, H. Oelke, M. Hurles, C. Renfrew, J. Weber, and K. Olek. 2002. Mitochondrial DNA and the origins of the domestic horse. *Proceedings of the National Academy of Sciences of the USA* 99:10905–10910.
- Larson, G., U. Albarella, K. Dobney, P. Rowley-Conwy, J. Schibler, A. Tresset, J.-D. Vigne, et al. 2007a. Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. *Proceedings of the National Academy of Sciences of the USA* 104:15276–15281.
- Larson, G., T. Cucchi, M. Fujita, E. Matisoo-Smith, J. Robins, A. Anderson, B. Rolett, et al. 2007b. Phylogeny and ancient DNA of *Sus* provides insights into Neolithic expansion in Island Southeast Asia and Oceania. *Proceedings of the National Academy of Sciences of the USA* 104:4834–4839.
- Larson, G., K. Dobney, U. Albarella, M. Y. Fang, E. Matisoo-Smith, J. Robins, S. Lowden, et al. 2005. Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science* 307:1618–1621.
- Larson, G., R. R. Liu, X. B. Zhao, J. Yuan, D. Fuller, L. Barton, K. Dobney, et al. 2010. Patterns of East Asian pig domestication, migration, and turnover revealed by modern and ancient DNA. *Proceedings of the National Academy of Sciences of the USA* 107: 7686–7691.
- Ludwig, A., M. Pruvost, M. Reissmann, N. Benecke, G. A. Brockman, P. Castanos, M. Cieslak, et al. 2009. Coat color variation at the beginning of horse domestication. *Science* 324:485.
- Marshall, Fiona, and Lior Weissbrod. 2011. Domestication processes and morphological change: through the lens of the donkey and African pastoralism. *Current Anthropology* 52(suppl. 4):S397–S413.
- Naderi, S., H. R. Rezaei, F. Pompanon, M. G. Blum, R. Negrini, H.-R. Naghash, O. Balkiz, et al. 2008. The goat domestication process inferred from large-scale mitochondrial DNA analysis of wild and domestic individuals. *Proceedings of the National Academy of Sciences of the USA* 105:17659–17664.
- Naderi, S., H.-R. Rezaei, P. Taberlet, S. Zundel, S.-A. Rafat, H.-R. Naghash, M. A. El-Barody, O. Ertugrul, and F. Pompanon. 2007. Large-scale mitochondrial DNA analysis of the domestic goat reveals six haplogroups with high diversity. *PLoS One* 2:e1012, doi: 10.1371/journal.pone.0001012.
- Piperno, Dolores R. 2011. The origins of plant cultivation and domestication in the New World tropics: patterns, process, and new developments. *Current Anthropology* 52(suppl. 4):S453–S470.
- Randi, E., and V. Lucchini. 2002. Detecting rare introgression of domestic dog genes into wild wolf (*Canis lupus*) populations by Bayesian admixture analyses of microsatellite variation. *Conservation Genetics* 3:31–45.
- Rompler, H., N. Rohland, C. Lalueza-Fox, E. Willerslev, T. Kuznetsova, G. Robeder, J. Bertranpetit, T. Schoneberg, and M. Hofreiter. 2006. Nuclear gene indicates coat-color polymorphism in mammoths. *Science* 313:62.
- Savolainen, P., Y. P. Zhang, J. Luo, J. Lundeberg, and T. Leitner. 2002. Genetic evidence for an East Asian origin of domestic dogs. *Science* 298:1610–1613.
- Stricklin, W. R. 2001. The evolution and domestication of social behaviour. In *Social behaviour in farm animals*. L. J. Keeling and H. W. Gonyou, eds. Pp. 83–110. New York: CABI.
- Troy, C. S., D. E. MacHugh, J. F. Bailey, D. A. Magee, R. T. Loftus, P. Cunningham, A. T. Chamberlain, B. C. Sykes, and D. G. Bradley. 2001. Genetic evidence for Near-Eastern origins of European cattle. *Nature* 410:1088–1091.
- Trut, L. N. 1999. Early canid domestication: the farm-fox experiment. *American Scientist* 87:160–169.
- Van Laere, A. S., M. Nguyen, M. Braunschweig, C. Nezer, C. Collete, L. Moreau, A. Archibald, et al. 2003. A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. *Nature* 425:832–836.
- Vigne, Jean-Denis, Isabelle Carrère, François Briois, and Jean Gui-

- laine. 2011. The early process of mammal domestication in the Near East: new evidence from the pre-Neolithic and pre-Pottery Neolithic in Cyprus. *Current Anthropology* 52(suppl. 4):S255–S271.
- Vila, C., J. A. Leonard, A. Gotherstrom, S. Marklund, K. Sandberg, K. Liden, R. K. Wayne, and H. Ellegren. 2001. Widespread origins of domestic horse lineages. *Science* 291:474–477.
- Vila, C., P. Savolainen, J. E. Maldonado, I. R. Amorim, J. E. Rice, R. L. Honeycutt, K. A. Crandall, J. Lundeberg, and R. K. Wayne. 1997. Multiple and ancient origins of the domestic dog. *Science* 276:1687–1689.
- vonHoldt, B. M., J. P. Pollinger, K. E. Lohmueller, E. J. Han, H. G. Parker, P. Quignon, J. D. Degenhardt, et al. 2010. Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature* 464:898–902.
- Zeder, Melinda A. 2011. The origins of agriculture in the Near East. *Current Anthropology* 52(suppl. 4):S221–S235.