UC Davis
UC Davis Previously Published Works

Title
Storytelling and story testing in domestication

Permalink
https://escholarship.org/uc/item/8zk5m9j2

Journal
Proceedings of the National Academy of Sciences of the United States of America, 111(17)

ISSN
0027-8424

Authors
Gerbault, Pascale
Allaby, Robin G
Boivin, Nicole
et al.

Publication Date
2014-04-29

DOI
10.1073/pnas.1400425111

Peer reviewed

Storytelling and story testing in domestication

Pascale Gerbault[a], Robin G. Allaby[b], Nicole Boivin[c], Anna Rudzinski[a], Ilaria M. Grimaldi[c], J. Chris Pires[d], Cynthia Climer Vigueira[e], Keith Dobney[f], Kristen J. Gremillion[g], Loukas Barton[h], Manuel Arroyo-Kalin[i], Michael D. Purugganan[j,k], Rafael Rubio de Casas[l], Ruth Bollongino[m], Joachim Burger[m], Dorian Q. Fuller[i], Daniel G. Bradley[n], David J. Balding[o], Peter J. Richerson[p], M. Thomas P. Gilbert[q], Greger Larson[r], and Mark G. Thomas[a,1]

[a]Research Department of Genetics, Evolution, and Environment and [o]University College London Genetics Institute, University College London, London WC1E 6BT, United Kingdom; [b]School of Life Sciences, Gibbet Hill Campus, University of Warwick, Coventry CV4 7AL, United Kingdom; [c]Research Laboratory for Archaeology and the History of Art, School of Archaeology, Oxford OX1 3QY, United Kingdom; [d]Division of Biological Sciences, University of Missouri, Columbia, MO 65211; [e]Department of Biology, Washington University in St. Louis, St. Louis, MO 63130; [f]Department of Archaeology, University of Aberdeen, Aberdeen AB24 3UF, United Kingdom; [g]Department of Anthropology, Ohio State University, Columbus, OH 43210; [h]Department of Anthropology, Center for Comparative Archaeology, University of Pittsburgh, Pittsburgh, PA 15260; [i]Institute of Archaeology, University College London, London WC1H 0PY, United Kingdom; [j]Department of Biology, New York University, New York, NY 10003-6688; [k]Center for Genomics and Systems Biology, New York University Abu Dhabi Research Institute, Abu Dhabi, United Arab Emirates; [l]Departamento de Ecología, Universidad de Granada, E-18071 Granada, Spain; [m]Institute of Anthropology, Johannes Gutenberg University, D-55099 Mainz, Germany; [n]Smurfit Institute of Genetics, Trinity College, Dublin 2, Ireland; [p]Department of Environmental Science and Policy, University of California, Davis, CA 95616; [q]Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, 1350 Copenhagen, Denmark; and [r]Durham Evolution and Ancient DNA, Department of Archaeology, Durham University, Durham DH1 3LE, United Kingdom

The domestication of plants and animals marks one of the most significant transitions in human, and indeed global, history. Traditionally, study of the domestication process was the exclusive domain of archaeologists and agricultural scientists; today it is an increasingly multidisciplinary enterprise that has come to involve the skills of evolutionary biologists and geneticists. Although the application of new information sources and methodologies has dramatically transformed our ability to study and understand domestication, it has also generated increasingly large and complex datasets, the interpretation of which is not straightforward. In particular, challenges of equifinality, evolutionary variance, and emergence of unexpected or counter-intuitive patterns all face researchers attempting to infer past processes directly from patterns in data. We argue that explicit modeling approaches, drawing upon emerging methodologies in statistics and population genetics, provide a powerful means of addressing these limitations. Modeling also offers an approach to analyzing datasets that avoids conclusions steered by implicit biases, and makes possible the formal integration of different data types. Here we outline some of the modeling approaches most relevant to current problems in domestication research, and demonstrate the ways in which simulation modeling is beginning to reshape our understanding of the domestication process.

model | inference | evolution | agriculture | Neolithic

The emergence of agriculture beginning some 10,000 y ago marked more than a change in human patterns of subsistence. The beginnings of food production ushered in an era of radically new relationships between humans and other species, dramatic new evolutionary pressures, and fundamental transformations to the earth's biosphere. The evolutionary process of plant and animal domestication by humans led to morphological, physiological, behavioral, and genetic differentiation of a wide range of species from their wild progenitors (1, 2). The selection pressures that were placed on such species continue today, sometimes through direct genetic modification, and both the processes and their outcomes are accordingly of significant broader interest. Domestication is also part of a cultural evolutionary process (3, 4), and some human genes have evolved in response to cultural innovations (5–8), much as the genes of domesticated species have changed under the impact of human artificial selection. The study of domestication today is a multidisciplinary enterprise in which archaeologists and agricultural scientists have been joined by evolutionary biologists and population geneticists (2, 9).

At least five major sets of questions tend to reoccur in the domestication literature. The first three are demographic: (*i*) When, where, and in how many geographic locations was a given species domesticated? (*ii*) What were the dispersal routes from the original domestication centers? (*iii*) To what extent did hybridization between domesticates and local wild relatives occur? The remaining questions relate to adaptation: (*iv*) To what extent, and how rapidly, were domestic traits fixed? (*v*) How well did domesticates adapt to diverse anthropogenic environments?

Most of these questions can be at least partially addressed using population genetic data from both ancient and modern samples. This is because variation across the genome is shaped by—and thus reflects—past demography, whereas genetic variation in and around particular genes determining key phenotypic traits is shaped by adaptation history. These principles, in combination with the availability of increasing quantities of ancient and modern genetic data, have led to a profusion of studies on particular domestication scenarios (e.g., refs. 10 and 11). However, the relationship between genetic data and the demographic or adaptation history that shaped it is noisy and often difficult to predict. This difficulty is primarily because: (*i*) in any evolving system that includes stochastic processes, patterns in genetic (or archaeological) data could have been generated under a range of

Significance

Our knowledge of the domestication of animal and plant species comes from a diverse range of disciplines, and interpretation of patterns in data from these disciplines has been the dominant paradigm in domestication research. However, such interpretations are easily steered by subjective biases that typically fail to account for the inherent randomness of evolutionary processes, and which can be blind to emergent patterns in data. The testing of explicit models using computer simulations, and the availability of powerful statistical techniques to fit models to observed data, provide a scientifically robust means of addressing these problems. Here we outline the principles and argue for the merits of such approaches in the context of domestication-related questions.

different histories (equifinality); (*ii*) any particular history can potentially give rise to a wide range of different patterns in data [evolutionary variance (12)]; and (*iii*) certain demographic or adaptation histories can give rise to counter intuitive patterns in data (emergence) (e.g., refs. 13–16).

Evolutionary histories are rarely directly "revealed" by looking only at patterns in data, such as the distribution of particular markers (e.g., morphological traits, material culture, genetic variants). This is because such data may be only weakly constrained by those histories; many different histories may explain the same data equally well. Thus, instead of simply providing narratives based on interpretations, implicit assumptions, and preconceived ideas, domestication histories need to be tested to identify those scenarios that best explain observed data; to do this, domestication histories must be modeled explicitly. We outline a range of modeling techniques that can be used in domestication research and provide examples that illustrate their utility. Although most of the discussion and examples given in this paper are based on population genetic data, most of the principles and approaches can also be applied to other datasets used to explore domestication processes.

Types of Modeling Approaches

A model is an explicit and simplified representation of the underlying causative mechanisms in a system and is used to make predictions about the observed outcomes (data) of that system. We consider two classes of models: discriminative models, which fit directly observed data to predicted relationships (e.g., linear regression), and generative models, which are intended to capture the main real-world mechanisms that generate data, and are typically used to produce artificial datasets. Discriminative models make assumptions, sometimes but not always explicitly, on the ways aspects of the data are correlated without specifying the actual mechanisms that generate those correlations (e.g., refs. 17–21). Generative models aim to explicitly replicate key hypothesized (i.e., assumed) processes that generate the data. Because all evolutionary processes include stochastic elements, a range of different outcomes—or patterns in empirical data—can be generated from any particular scenario or model. For this reason, when using generative models, it is often necessary to produce many datasets by simulation.

In population genetics, a powerful means of simulating data is the "retrospective" approach of coalescent simulation (22), where the joining (or coalescence) of lineages is simulated backward in time under specific assumptions about such variables as population size, structure, migration, and admixture. This approach is highly efficient because it only simulates the lineage history of the sample, not of the whole population, so simulation can be very fast. However, coalescent approaches are limited in the demographic and selection scenarios that can be modeled, and some researchers instead favor a more flexible, but computationally demanding forward-in-time simulation (e.g., ref. 23), or simulations with a combination of forward- and backward-in-time elements (e.g., ref. 24).

Generative models can be agent-based, whereby agents with prescribed interaction behaviors are simulated as individual units. However, agent-based models tend to be computationally demanding, so their application is usually restricted to revealing some emergent, sometimes counter-intuitive, properties of a modeled system (e.g., ref. 25) rather than making inferences by fitting to existing data (e.g., ref. 26). An additional level of resolution in evolutionary modeling can be achieved through spatially explicit simulation, sometimes reducing continuous space to a number of cells (or "demes") with defined neighbor relationships. These spatial refinements can be computationally challenging, particularly when geographic features (e.g., elevation, climate) or population dynamics (e.g., varying carrying capacities) are introduced.

A shared characteristic of these modeling approaches is that they are made up of a number of components reflecting the real-world processes that are hypothesized to have shaped the data. These components can be explicitly modified and combined with one another.

Models and Data

To be useful in evolutionary inference, models should be fitted to observed data. Often the most important aspect of model-fitting is deciding how to deal with unknown parameters, such as migration rates or selection coefficients.

Frequentist approaches to hypothesis testing or estimation treat the unknown parameters as fixed; the model specifies imaginary random repetitions of the data generation process (e.g., refs. 13, 25, 27, 28). There is therefore no probability distribution for the parameters, but instead statements are made about the frequency of future datasets satisfying certain conditions given assumed parameter values. Often the data are reduced to summary statistics (e.g., means, variances) intended to capture the most important information about the processes modeled. One of the simplest forms of inference is to consider the distribution of a summary statistic under a given model in comparison with the observed value of that statistic. This procedure can be used to reject models or parameter values as implausible, but is not useful for more quantitative comparisons.

Given some assumed—or known—model of the processes at play, the parameter values that maximize the probability of observing the data can be obtained (maximum likelihood). The main requirement of maximum-likelihood approaches is a likelihood function: a mathematical formula that specifies the probability of the data as a function of the parameter values. This function can be used in a frequentist setting, but is more commonly used directly to identify ranges of plausible parameter values (e.g., refs. 29, 30). Likelihood-based (and full-Bayesian, see below) approaches usually use the full information content of data and not just some aspect of it, such as summary statistics. However, (*i*) the likelihood function can be difficult to formulate for anything but the simplest models, (*ii*) if there are many parameters, maximizing the likelihood can be computationally demanding, and (*iii*) there can be multiple maxima of the likelihood function.

Full Bayesian methods also make use of the likelihood function but they allow the incorporation of "prior information" about the model parameters, which can help to focus on the most plausible regions of parameter space. Computational techniques, such as Markov chain Monte Carlo (MCMC), have made Bayesian methods more tractable and more popular. A wide variety of MCMC techniques exist. They are all "samplers" because they sample parameter values at random from their prior or posterior (i.e., target) distributions. More specifically, the aim of MCMC techniques in Bayesian inference is to use prior probabilities and the likelihood function to condition a random walk through parameter space. This process results in the distribution of parameter values "visited" numerically approximating the posterior distribution (i.e., the updated knowledge of the parameters given the data). However, for large parameter spaces MCMC can still be computationally very expensive (31, 32).

Because likelihood functions are only workable for relatively simple models, there can be a tension between fitting more elegant and powerful statistical methods assuming simplistic models, and assuming more general models that only permit crude modes of inference. For example, in population genetics a likelihood-based method may only be available for a single population model (e.g., ref. 33), and researchers may accept this limitation uncritically even when the data are clearly from multiple populations, or subsets of the data are selected to fit the model [e.g., selecting DNA sequences from only one branch of a phylogeny (34, 35)]. This uncritical acceptance will lead to

misleading inferences unless the data-selection step is incorporated in the model specification.

The difficulty of computing likelihoods for all but the simplest models has led to the development of a family of techniques known as approximate Bayesian computation (ABC) (36, 37). In its simplest form, ABC works by simulating data from a generative model with parameter values chosen at random from their prior distributions. A simulation is accepted if the simulated data resemble the observed data, and rejected otherwise, where the "resemblance" of datasets is measured using one or more summary statistics. The proposed parameter values that are accepted in this algorithm form a sample from an approximation to the posterior distribution. Thus, the approach is very similar to MCMC, with the exception that the latter samples are from the true posterior distribution. Several variants of this process have been developed to improve accuracy and computational efficiency (38–40). ABC provides a framework for estimating parameters of interest and comparing relative support for different models based on the same data (37, 41).

The big advantage of ABC is modeling flexibility because almost any generative model can be used, but this comes at the cost of only approximate answers, because ABC does not use all of the information in the data. In addition, the accuracy of the resulting approximation is hard to assess and choosing appropriate summary statistics can be difficult (42, 43). Although the development of full-likelihood Monte Carlo methods, particularly those addressing issues of statistical intractability, continues apace (e.g., ref. 44, 45), ABC provides a useful adjunct to these approaches (e.g., ref. 46), permitting currently intractable problems to be side-stepped, even if only temporarily. In addition, methodologies such as ABC allow for the integration of distinct sources of data (47). This integration will become increasingly important as new types of data (e.g., paleoclimatic, archaeological, genetic) accumulate and demand statistically informed comparison and integration. Given the many factors that are important in shaping data patterns in domesticates, ABC provides the most promising means of democratizing simulation modeling for the domestication research community.

Modeling Domestication History

Explicit modeling-based studies of domestication are relatively new and mostly confined to inference from population genetic data, but have nonetheless begun to transform our understanding of the five major domestication questions outlined above. Here we highlight some examples in which initial inference about domestication processes based on direct interpretation of patterns in data were later demonstrated problematic when tested using modeling approaches.

When, Where, and in How Many Geographic Locations Was a Given Species Domesticated? A common but questionable interpretation of lineage divergence date estimates for Y chromosome or mitochondrial DNA (mtDNA) data are that they represent founding events in species or populations. However, the choice of which lineages to estimate divergence dates for can be arbitrary, and there is little reason to expect demographic processes, such as domestication, to correlate with lineage ages, unless those founder events involved very small numbers of individuals; population genetic models show that lineage coalescent dates can predate major demographic episodes. This finding is well-illustrated with domestic dogs; an early estimate of 135,000 y for the coalescence age of the major mtDNA lineage (clade I, see ref. 48) was interpreted as indicating a domestication founding event around that time. More recently, modeling approaches based on diffusion approximations (45) and the generalized phylogenetic coalescent sampler (49), both conditioned on whole-genome sequence data, estimated domestic dog-wolf divergence between 32,000 y ago (50) and 11,000–16,000 y ago (51). Although these model-based

date estimates differ (most probably because of the assumed evolutionary rate), they have concordance with those from fossil canids currently considered morphologically more similar to dogs than wolves (e.g., refs. 52 and 53).

Goat domestication has also been reevaluated. mtDNA sequences in domesticated goat have been assigned to five major haplogroups (54), the first three of which have expansion age estimates in the range of 10,000–841 y ago, based on DNA sequence mismatch distributions (55). The coalescent date estimates between these haplogroups are considerably older (103,000 and 597,800 y ago) (54). Initially, these haplogroups were interpreted as representing independent domestication events (34, 55), and the overall patterns of mtDNA divergence as only being consistent with an implausibly high number of initial domesticates [38,000–82,000 females (55)]. However, application of coalescent simulation and ABC fitting to published ancient and modern mtDNA data indicated that these data could be equally well explained by a single domestication episode of smaller size, or successive founding events as domestic goat populations expanded into Europe (56).

The extraordinary phenotypic range of the common bean (*Phaseolus vulgaris*) has made it a particularly interesting target for domestication research (57). A range of genetic studies (e.g., ref. 58) indicates two highly diverged gene pools, one hypothesized to originate in Mesoamerica and the other in the Andes. Within the Mesoamerican gene pool, random amplified polymorphic DNA (59) and chloroplast data (60) have been interpreted as indicating independent domestication events. However, using coalescent simulation and ABC, Mamidi et al. (57) showed that single domestication episodes for both the Mesoamerican and Andean gene pools, with strong bidirectional gene flow between domesticated and wild species, provided the best fit to data on 13 loci.

Modeling approaches have also altered our views of the domestication of rice (30). Previous genetic studies had inferred that rice was domesticated twice, in China and in India, giving rise to the *japonica* and *indica* cultivars, respectively (61). When demographic modeling using a diffusion approximation-based approach (45) was applied to SNP data from three rice chromosomes, however, only one domestication was indicated. In conjunction with archaeological data, a more nuanced view of rice domestication has emerged, suggesting that *japonica* was domesticated in China, and that *indica* arose possibly as a result of subsequent introgression of *japonica* into wild rice or proto-*indica* populations in India (62).

What Were the Dispersal Routes from the Original Domestication Centers? Traditionally, investigating the geographical origin and subsequent dispersal pathways of a domesticate rely heavily on identifying the genetically closest wild progenitor populations. However, shifts through time in the location of such ancestral populations, extinctions, and undersampling can all weaken such an approach. Although application of modeling techniques to the delineation of pathways of dispersal is in its infancy, preliminary geospatial modeling has been conducted to infer the dispersal of maize in the Americas (20). This approach accounted for landscape, radiocarbon dates of crop remains, and genetic diversity. The model-fitting was performed by multiple-criteria regression analysis over archaeobotanical and genetic data. This allowed tensions between archaeological and genetic data to be explicitly modeled and explored, and points the way forward for the more systematic and statistically informed examination of dispersal pathways, particularly in domesticates like rice, for which the geography of both domestication and subsequent dispersal remains a source of significant debate (63, 64).

To What Extent Did Hybridization Between Domesticates and Local Wild Relatives Occur? Gene flow from wild populations can have important effects on patterns of genetic variation. Low Y chromosomal diversity in modern horses has been interpreted as the result of a single geographically restricted area of domestication (65), whereas the high diversity and low phylogeographic structure in mtDNA has been interpreted as support for multiple origins of domesticated horses (66, 67). Using a spatially explicit forward simulation model, conditioned on autosomal genotype data using ABC (68), the Western Eurasian steppe has been identified as the most likely origin for modern horses, with a model of repeated introgression from local wild to domesticated horses offering the best fit to observed patterns of diversity.

Genetic interactions can be particularly complex when introgression between domesticates and several related wild taxa is possible. The cultivated apple (*Malus domestica*), for example, has been proposed to derive from multiple wild relatives: most notably *Malus sieversii* and *Malus sylvestris*, with potential contribution of other taxa, such as *Malus orientalis*. A recent study used demographic modeling with ABC to compare the introgression scenarios between the different taxa (69). Results supported *M. sieversii* as the primary source of the domesticated apple, but also with frequent and widespread introgression from *M. sylvestris*, potentially contributing characters relevant for the adaptation to novel environments and human use. In this case modeling was essential to disentangle the biological complexity of the domestication process, which involved at least two different species and repeated hybridization events over a long period (69).

To What Extent, and How Rapidly, Were Domestic Traits Fixed? Monophyletic patterns in phylogenetic analyses of genome-wide markers, such as amplified fragment-length polymorphisms, have been interpreted as indicating a rapid fixation of domestication traits and spread of domesticates from one center (11, 70). Rapid fixation once seemed to find confirmation in experimental field studies (71). Increasingly, however, the archaeological record reveals that some traits require centuries or millennia to reach fixation (19, 72), as would be expected if those traits were determined by dominant advantageous alleles. These monophyletic patterns were investigated using an individual-based modeling approach conditioned on amplified fragment-length polymorphism data from crops (13). Researchers found that both multiple and single origins of domestic crops could explain the data, and that the multiple-origin model produced the monophyletic signal more rapidly. Simulation also demonstrated that a monophyletic signal alone need not indicate rapid fixation.

How Well Did Domesticates Adapt to Diverse Anthropogenic Environments? An additional area of interest in domestication studies has been how outbreeding plant systems have adapted to the human environment in situations when there is continuous gene-flow between wild and domesticated populations (73). Le Thierry d'Ennequin et al. (74) used an individual-based model to demonstrate the increased number of genes likely to be under selection with linkage because of genome architecture and mating strategy. Thus, some genome architectures may be more adaptable to the domesticated environment than others. Artificial and natural selection can pull traits in opposite directions [e.g., seed size (75)], resulting in weak selection in net effect. Such adaptation complexity is consistent with the large number of genes (i.e., 27–70 genes) thought to underlie domestication traits in wheat, maize, and sunflower (76–79). Furthermore, because the number of genes that can be under selection simultaneously is constrained (80), it is necessary to consider gene interactions in models of adaptation (75, 81). A case in point is the adaptation of crops to

higher latitudes as they were moved out of their location(s) of origin (82–84). A reduction in crop adaptability may have led to population collapse in mid-Holocene Europe (85), resulting into regional agricultural abandonment (86).

In addition to looking at the adaptation of domesticates to anthropogenic environments, researchers interested in the cultural impact of the domestication process can also use models to evaluate human response to culturally determined selection pressures (87). The dietary changes, population growth, increased sedentism, and new diseases that accompanied the domestication of plants and animals appear to have triggered a wave of genetically based adaptations in our own immune and digestive systems (5, 88). The rapid increase in the availability of human gene-sequence data are making model-driven data analysis increasingly feasible and attractive in studies of human cultural and genetic response to agriculture-associated innovations (e.g., ref. 47).

Concluding Remarks

Model-based statistical approaches are an essential tool in domestication research. When inferring past processes, explicit models are particularly important as, typically, data are the result of a single experiment (the past), and it is necessary to explore a landscape of hypotheses to test which could have given rise to those observed data. Because the hypothesis landscape is effectively infinite, it will always be the case that some unjustifiably complex model can be found to explain the data well (overfitting), so it is necessary to make explicit assumptions, to consider simple models first, and to penalize complexity (89, 90).

These requirements are often seen by nonstatisticians as key drawbacks of modeling approaches. However, assumptions are always present when inferring past processes, and making them explicit enables their recognition and evaluation. In addition, the advantages of using simplified explicit models—particularly statistical tractability and the avoidance of overfitting—outweigh their drawbacks. Furthermore, simple does not mean easy. To quote George Box (91): "Just as the ability to devise simple but evocative models is the signature of the great scientist so overelaboration and overparameterization is often the mark of mediocrity."

Because of the central role clear hypothesis formulation and testing play in scientific research, we suggest that the arguments presented here apply not only to the field of domestication research or population genetics, but to any discipline involving historical inference (85, 92). Modeling is not the only way to proceed and does not guarantee the right answers [indeed, "models are always wrong, and sometimes useful" (91)]. Likewise, interpretative approaches can be valuable in the scientific process and may lead to the correct, or nearly correct, explanation. However, interpretative inference is better thought of as a means of generating hypotheses (storytelling), whereas explicit models permit those hypotheses (or stories) to be tested. With advances in statistical modeling techniques and increases in computer power, the approaches discussed in this article are set to transform our understanding of domestication processes.

1. Darwin C (1868) *The Variation of Animals and Plants Under Domestication* (John Murray, London), Vol I.
2. Larson G, Burger J (2013) A population genetics view of animal domestication. *Trends Genet* 29(4):197–205.
3. Cavalli-Sforza LL, Feldman MW (1981) *Cultural Transmission and Evolution: A Quantitative Approach* (Princeton Univ Press, Princeton, NJ).
4. Odling-Smee FJ, Laland KN, Feldman MW (2003) *Niche Construction: The Neglected Process in Evolution* (Princeton Univ Press, Princeton).
5. Gerbault P, et al. (2011) Evolution of lactase persistence: An example of human niche construction. *Philos Trans R Soc Lond B Biol Sci* 366(1566):863–877.
6. Laland KN, Odling-Smee J, Myles S (2010) How culture shaped the human genome: Bringing genetics and the human sciences together. *Nat Rev Genet* 11(2):137–148.
7. Leach HM (2003) Human domestication reconsidered. *Curr Anthropol* 44(3):349–368.
8. Richerson PJ, Boyd R, Henrich J (2010) Colloquium paper: Gene-culture coevolution in the age of genomics. *Proc Natl Acad Sci USA* 107(Suppl 2):8985–8992.
9. Zeder MA (2006) *Documenting Domestication: New Genetic and Archaeological Paradigms* (Univ of California Press, Berkeley, CA).
10. Edwards CJ, et al. (2007) Mitochondrial DNA analysis shows a Near Eastern Neolithic origin for domestic cattle and no indication of domestication of European aurochs. *Proc Biol Sci* 274(1616):1377–1385.
11. Heun M, et al. (1997) Site of Einkorn wheat domestication identified by DNA fingerprinting. *Science* 278(5341):1312–1314.
12. Wilson IJ, Balding DJ (1998) Genealogical inference from microsatellite data. *Genetics* 150(1):499–510.
13. Allaby RG, Fuller DQ, Brown TA (2008) The genetic expectations of a protracted model for the origins of domesticated crops. *Proc Natl Acad Sci USA* 105(37):13982–13986.
14. Edmonds CA, Lillie AS, Cavalli-Sforza LL (2004) Mutations arising in the wave front of an expanding population. *Proc Natl Acad Sci USA* 101(4):975–979.
15. Excoffier L, Foll M, Petit RJ (2009) Genetic consequences of range expansions. *Annu Rev Ecol Evol Syst* 40:481–501.
16. François O, et al. (2010) Principal component analysis under population genetic models of range expansion and admixture. *Mol Biol Evol* 27(6):1257–1268.
17. Fuller DQ, et al. (2009) The domestication process and domestication rate in rice: Spikelet bases from the Lower Yangtze. *Science* 323(5921):1607–1610.
18. Novembre J, Stephens M (2008) Interpreting principal component analyses of spatial population genetic variation. *Nat Genet* 40(5):646–649.
19. Purugganan MD, Fuller DQ (2011) Archaeological data reveal slow rates of evolution during plant domestication. *Evolution* 65(1):171–183.
20. van Etten J, Hijmans RJ (2010) A geospatial modelling approach integrating archaeobotany and genetics to trace the origin and dispersal of domesticated plants. *PLoS ONE* 5(8):e12060.
21. Vigne JD, Carrère I, Guilaine J (2003) Unstable status of early domestic ungulates in the near east: The example of Shillourokambos (Cyprus, IX–VIIIth millennia cal. B.C.). *Bulletin Correspondance Héllenique* Suppl 43:239–251.
22. Kingman JFC (1982) The coalescent. *Stochastic Process Appl* 13(3):235–248.
23. Hoggart CJ, et al. (2007) Sequence-level population simulations over large genomic regions. *Genetics* 177(3):1725–1731.
24. Currat M, Excoffier L (2004) Modern humans did not admix with Neanderthals during their range expansion into Europe. *PLoS Biol* 2(12):e421.
25. Kitchen JL, Allaby RG (2012) The limits of mean-field heterozygosity estimates under spatial extension in simulated plant populations. *PLoS ONE* 7(8):e43254.
26. Csilléry K, Blum MG, Gaggiotti OE, François O (2010) Approximate Bayesian Computation (ABC) in practice. *Trends Ecol Evol* 25(7):410–418.
27. Boyd R, Richerson PJ, Henrich J (2011) Rapid cultural adaptation can facilitate the evolution of large-scale cooperation. *Behav Ecol Sociobiol* 65(3):431–444.
28. Bramanti B, et al. (2009) Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. *Science* 326(5949):137–140.
29. Chikhi L, Nichols RA, Barbujani G, Beaumont MA (2002) Y genetic data support the Neolithic demic diffusion model. *Proc Natl Acad Sci USA* 99(17):11008–11013.
30. Molina J, et al. (2011) Molecular evidence for a single evolutionary origin of domesticated rice. *Proc Natl Acad Sci USA* 108(20):8351–8356.
31. Beaumont MA, Rannala B (2004) The Bayesian revolution in genetics. *Nat Rev Genet* 5(4):251–261.
32. Marjoram P, Tavaré S (2006) Modern computational approaches for analysing molecular genetic variation data. *Nat Rev Genet* 7(10):759–770.
33. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7:214.
34. Naderi S, et al. (2008) The goat domestication process inferred from large-scale mitochondrial DNA analysis of wild and domestic individuals. *Proc Natl Acad Sci USA* 105(46):17659–17664.
35. Brotherton P, et al.; Genographic Consortium (2013) Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. *Nat Commun* 4:1764.
36. Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics* 162(4):2025–2035.
37. Bertorelle G, Benazzo A, Mona S (2010) ABC as a flexible framework to estimate demography over space and time: Some cons, many pros. *Mol Ecol* 19(13):2609–2625.
38. Marjoram P, Molitor J, Plagnol V, Tavare S (2003) Markov chain Monte Carlo without likelihoods. *Proc Natl Acad Sci USA* 100(26):15324–15328.
39. Sousa VC, Fritz M, Beaumont MA, Chikhi L (2009) Approximate Bayesian computation without summary statistics: The case of admixture. *Genetics* 181(4):1507–1519.
40. Wegmann D, Leuenberger C, Excoffier L (2009) Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood. *Genetics* 182(4):1207–1218.
41. Wegmann D, Leuenberger C, Neuenschwander S, Excoffier L (2010) ABCtoolbox: A versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics* 11:116.
42. Joyce P, Marjoram P (2008) Approximately sufficient statistics and Bayesian computation. *Stat Appl Genet Mol Biol* 7(1):26.
43. Nunes MA, Balding DJ (2010) On optimal selection of summary statistics for approximate Bayesian computation. *Stat Appl Genet Mol Biol* 9:34.
44. Keith JM, Spring D (2013) Agent-based Bayesian approach to monitoring the progress of invasive species eradication programs. *Proc Natl Acad Sci USA* 110(33):13428–13433.
45. Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet* 5(10):e1000695.
46. Aandahl RZ, Stadler T, Sisson SA, Tanaka MM (2014) On the reliability of approximate Bayesian computation: Epidemiological parameter estimates from *Mycobacterium tuberculosis* genotypes. *Genetics*, 10.1534/genetics.113.158808.
47. Itan Y, Powell A, Beaumont MA, Burger J, Thomas MG (2009) The origins of lactase persistence in Europe. *PLOS Comput Biol* 5(8):e1000491.
48. Vilà C, et al. (1997) Multiple and ancient origins of the domestic dog. *Science* 276(5319):1687–1689.
49. Gronau I, Hubisz MJ, Gulko B, Danko CG, Siepel A (2011) Bayesian inference of ancient human demography from individual genome sequences. *Nat Genet* 43(10):1031–1034.
50. Wang GD, et al. (2013) The genomics of selection in dogs and the parallel evolution between dogs and humans. *Nat Commun* 4:1860.
51. Freedman AH, et al. (2014) Genome sequencing highlights the dynamic early history of dogs. *PLoS Genet* 10(1):e1004016.
52. Pionnier-Capitan M, et al. (2011) New evidence for Upper Palaeolithic small domestic dogs in South-Western Europe. *J Archaeol Sci* 38(9):2123–2140.
53. Larson G, et al. (2012) Rethinking dog domestication by integrating genetics, archeology, and biogeography. *Proc Natl Acad Sci USA* 109(23):8878–8883.
54. Naderi S, et al.; Econogene Consortium (2007) Large-scale mitochondrial DNA analysis of the domestic goat reveals six haplogroups with high diversity. *PLoS ONE* 2(10):e1012.
55. Luikart G, et al. (2001) Multiple maternal origins and weak phylogeographic structure in domestic goats. *Proc Natl Acad Sci USA* 98(10):5927–5932.
56. Gerbault P, Powell A, Thomas MG (2012) Evaluating demographic models for goat domestication using mitochondrial DNA sequences. *Anthropozoologica* 47(2):65–78.
57. Mamidi S, et al. (2011) Investigation of the domestication of common bean (*Phaseolus vulgaris*) using multilocus sequence data. *Funct Plant Biol* 38(12):953–967.
58. Kwak M, Gepts P (2009) Structure of genetic diversity in the two major gene pools of common bean (*Phaseolus vulgaris* L., Fabaceae). *Theor Appl Genet* 118(5):979–992.
59. Beebe S, et al. (2000) Structure of genetic diversity among common bean landraces of Middle American origin based on correspondence analysis of RAPD. *Crop Sci* 40(1):264–273.
60. Chacón S MI, Pickersgill B, Debouck DG (2005) Domestication patterns in common bean (*Phaseolus vulgaris* L.) and the origin of the Mesoamerican and Andean cultivated races. *Theor Appl Genet* 110(3):432–444.
61. Vitte C, Ishii T, Lamy F, Brar D, Panaud O (2004) Genomic paleontology provides evidence for two distinct origins of Asian rice (*Oryza sativa* L.). *Mol Genet Genomics* 272(5):504–511.
62. Fuller DQ (2011) Pathways to Asian civilizations: Tracing the origins and spread of rice and rice cultures. *Rice* 4(3–4):78–92.
63. Fuller DQ, et al. (2010) Consilience of genetics and archaeobotany in the entangled history of rice. *Archaeological and Anthropological Sciences* 2(2):115–131.
64. Vaughan DA, Lu B-R, Tomooka N (2008) The evolving story of rice evolution. *Plant Sci* 174(6):394–408.
65. Lindgren G, et al. (2004) Limited number of patrilines in horse domestication. *Nat Genet* 36(4):335–336.
66. Jansen T, et al. (2002) Mitochondrial DNA and the origins of the domestic horse. *Proc Natl Acad Sci USA* 99(16):10905–10910.
67. Vilà C, et al. (2001) Widespread origins of domestic horse lineages. *Science* 291(5503):474–477.
68. Warmuth V, et al. (2012) Reconstructing the origin and spread of horse domestication in the Eurasian steppe. *Proc Natl Acad Sci USA* 109(21):8202–8206.
69. Cornille A, et al. (2012) New insight into the history of domesticated apple: Secondary contribution of the European wild apple to the genome of cultivated varieties. *PLoS Genet* 8(5):e1002703.
70. Zohary D (1999) Monophyletic versus polyphyletic origin of crops found in the Near East. *Genet Resour Crop Evol* 46(2):133–142.
71. Hillman GC, Davies MS (1990) Domestication rates in wild-type wheats and barley under primitive cultivation. *Biol J Linn Soc Lond* 39(1):39–78.
72. Tanno K, Willcox G (2006) How fast was wild wheat domesticated? *Science* 311(5769):1886.
73. Pernés J (1986) L'allogamie et la domestication des cereales: L'exemple du ma (*Zea mays* L.) et du mil (*Pennisetum americanum* L.) K. *Schum Bull Soc Bot Fr* 133:27–34.
74. Le Thierry D'Ennequin M, Toupance B, Robert T, Godelle B, Gouyon PH (1999) Plant domestication: A model for studying the selection of linkage. *J Evol Biol* 12(6):1138–1147.
75. Allaby R (2010) Integrating the processes in the evolutionary system of domestication. *J Exp Bot* 61(4):935–944.
76. Chapman MA, et al. (2008) A genomic scan for selection reveals candidates for genes involved in the evolution of cultivated sunflower (*Helianthus annuus*). *Plant Cell* 20(11):2931–2945.
77. Peleg Z, Fahima T, Korol AB, Abbo S, Saranga Y (2011) Genetic analysis of wheat domestication and evolution under domestication. *J Exp Bot* 62(14):5051–5061.
78. Peng J, et al. (2003) Domestication quantitative trait loci in *Triticum dicoccoides*, the progenitor of wheat. *Proc Natl Acad Sci USA* 100(5):2489–2494.

EVOLUTION

Gerbault et al.

79. Wright SI, et al. (2005) The effects of artificial selection on the maize genome. *Science* 308(5726):1310–1314.

80. Haldane JBS (1957) The cost of selection. *J Genet* 55:511–524.

81. Bekaert M, Edger PP, Hudson CM, Pires JC, Conant GC (2012) Metabolic and evolutionary costs of herbivory defense: Systems biology of glucosinolate synthesis. *New Phytol* 196(2):596–605.

82. Cockram J, et al. (2007) Control of flowering time in temperate cereals: Genes, domestication, and sustainable productivity. *J Exp Bot* 58(6):1231–1244.

83. Fu D, et al. (2005) Large deletions within the first intron in VRN-1 are associated with spring growth habit in barley and wheat. *Mol Genet Genomics* 273(1):54–65.

84. Wu W, et al. (2013) Association of functional nucleotide polymorphisms at DTH2 with the northward expansion of rice cultivation in Asia. *Proc Natl Acad Sci USA* 110(8):2775–2780.

85. Shennan S, et al. (2013) Regional population collapse followed initial agriculture booms in mid-Holocene Europe. *Nat Commun* 4:2486.

86. Stevens CJ, Fuller DQ (2012) Did Neolithic farming fail? The case for a Bronze Age agricultural revolution in the British Isles. *Antiquity* 86(333):707–722.

87. Richerson PJ, Boyd R, Bettinger RL (2001) Was agriculture impossible during the Pleistocene but mandatory during the Holocene? A climate change hypothesis. *Am Antiq* 66(3):387–411.

88. Hawks J, Wang ET, Cochran GM, Harpending HC, Moyzis RK (2007) Recent acceleration of human adaptive evolution. *Proc Natl Acad Sci USA* 104(52):20753–20758.

89. Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Automat Contr* 19(6):716–723.

90. Kass RE, Raftery AE (1995) Bayes factors. *J Am Stat Assoc* 90(430):773–795.

91. Box GEP (1976) Science and statistics. *J Am Stat Assoc* 71(356):791–799.

92. Allen TFH, Tainter JA, Pires JC, Hoekstra TW (2001) Dragnet ecology, "Just the facts ma'am": The privilege of science in a postmodern world. *Bioscience* 51(6):475–485.