E. J. OURSO COLLEGE OF BUSINESS
Department of Economics

DEPARTMENT OF ECONOMICS WORKING PAPER SERIES

The Role of Lactose Tolerance in Pre-Colonial Development

Charles J. Cook
Louisiana State University

Working Paper 2011-12
http://bus.lsu.edu/McMillin/Working_Papers/pap11_12.pdf

Department of Economics
Louisiana State University
Baton Rouge, LA 70803-6306
http://www.bus.lsu.edu/economics/

The Role of Lactose Tolerance in Pre-Colonial Development*

Charles J. Cook†

September 2011

Job Market Paper

Abstract

This paper establishes a link between natural selection since the Neolithic Revolution and economic conditions in the pre-colonial era. The ability to digest milk, or to be lactose tolerant, is conferred by a gene variant, which is unequally distributed across the Old World. Digesting milk conferred qualitative and quantitative advantages to early farmers's diets, which ultimately, led to differences in the carrying capacities of respective countries. It is shown through a number of specifications that country level variation in the frequency of the ability to consume milk is positively and significantly related to population densities in 1500 CE; specifically, a one standard deviation increase in the frequency of lactose tolerant individuals (24% points) is associated with roughly a 60% increase in pre-colonial population densities. This relationship remains while controlling for agricultural transition dates, other measures of genetic distance, and a wide array of environmental controls. Additionally, the basis for the relationship between dairying and population density is confirmed with the use of instrumental variables estimation.

JEL Classification: O13, N5, Z13.

Keywords: Historical Development, Genetic Diversity, Neolithic Revolution, Population Density.

---

*I owe thanks to Areendam Chanda for thoughtful discussion and direction; to James Feyrer, David Weil, Quamrul Ashraf, participants at the 2011 Integrating Genetics and the Social Sciences Conference at the University of Colorado, and participants at the LSU 3rd year paper presentation for comments and suggestions; and to Stelios Michalopolous for sharing data on agricultural suitability. All errors and omissions are my own.

†Ph.D. Candidate, Department of Economics, Louisiana State University; Baton Rouge, LA, 70803; e-mail: ccook2@lsu.edu.

1 Introduction

The great disparities in productivity that are seen throughout the world today are not new. As of 500 years ago great variations in technology, state development, and industry were obvious across states and continents; most notable is the distinction between Europe and Sub-Saharan Africa. Europe was in the middle of the Renaissance, had complex systems of state organization, numerous divisions of labor, and was making great strides in seafaring, while Africa was vastly under populated and relatively under developed. What are the causes of variations in historic development? It is known that Eurasia contained advantages in initiating and spreading agriculture, but are there other factors which led to larger pre-colonial populations? Why did Europe in particular have an advantage over other Eurasian states? This paper argues the variation in an important food source, milk, is significantly related to differences in pre-colonial development, or pre-colonial populations.

The Neolithic Revolution radically changed the environment for humans.[1] Furthermore, this change occurred at different times for different peoples; implying, certain groups have had a longer time to evolve, or adapt, to the new environment. In the words of Clark (2008, P. 6), "The Darwinian struggle that shaped human nature did not end with the Neolithic Revolution but continued right up until the Industrial Revolution." A major adaptation to the sedentary agricultural lifestyle is the ability to consume milk, or to be lactose tolerant. Milk was an additional resource that some could consume, while others could not. In the Malthusian economy of the pre-colonial era, this variation in the consumption of milk should be associated with variations in the productive capacity of land. Specifically, we seek to explain the differences in population density for 1500 CE using the fraction of lactose tolerant individuals within a country.

Lactose tolerance data is available by ethnicity for the second half of the twentieth century. A central assumption in our paper is that this has not changed much over the past 500 years (Section 2 includes a detailed discussion on the validity of this). Since our hypothesis concerns pre-colonial development, we also need a measure of ethnic composition for 1500 CE, which is not directly available. We follow two strategies. The first, and primary

[1] The Neolithic Revolution is the name given to the transition from hunting and gathering to agriculture

strategy, involves post-multiplying the matrix of current ethnic compositions countries with the inverse of a matrix that captures human migration from 1500 to 2000 CE (Putterman and Weil 2010). This, in theory, gives county level ethnic compositions for the year 1500 CE. In order to confirm our results, we also use a cruder strategy of assigning majority ethnic groups to represent countries in the 1500 CE.[2]

We show that our constructed measure of lactose tolerance has a positive and significant effect on population. Specifically, our baseline estimate states that a standard deviation increase in the fraction of lactose tolerant individuals within a country is associated with a 60% increase in pre-colonial population density. The results are robust to a large number of geographical and environmental variables. In particular, we show that the effect of lactose tolerance does not pick up the overarching advantages of earlier transitions into agricultural societies that have been documented extensively. The results are also robust to other measures of genetic distances that have been used to explain technological diffusion across countries, as well as variables that capture other environmental or cultural determinants of pre-colonial development. In addition to least squares estimation, we also consider an instrumental variables approach. Lower levels of sunlight result in a defficiency of vitamin D. A diet that is rich in milk can offset the harmful benefits of vitamin D deficiency through the addition of absorbable calcium (Flatz and Rotthauwe 1973). Therefore, we use a measure of solar radiation to instrument country level differences in the frequency of lactose tolerance. Due to concerns, however, our use of IV estimations are not meant to replace estimates through OLS; the use of instrumental variables is intended to supplement and confirm the relationship between dairying and pre-colonial population density.

An interest in the role of history in explaining economic disparities has recently been renewed. The idea that current development levels are path dependent has established the search for a more ultimate understanding of the long run causes of growth; knowing the causes of small differences in past growth rates gives valuable insights into the cross-country disparities in current economic conditions. According to Nunn (2009, P. 88): "The main fact . . . is that history matters." Specifically, a number of papers have established an

[2]This strategy is pursued in similar research, i.e., Spolaore and Wacziarg (2009).

empirical link between past and current economic events, where it is shown that variations in the past have economic repercussions that are felt today (see, e.g., Acemoglu et al. 2001; Bockstette et al. 2002; Chanda and Putterman 2004; Comin et al. 2007; Engerman and Sokoloff 1997, 2002; La Porta et al. 1997, 1998; Nunn 2008). The current work seeks to build upon this research.

One of the most comprehensive works in explaining pre-colonial populations and, therefore, pre-colonial development is Jared Diamond's *Guns, Germs, and Steel* (1997). Diamond's main argument is that societies on the Eurasian continent contained a geographical advantage in both initiating and spreading agriculture. In particular, the geographical advantages of Eurasia are the number of domesticable species (plants and animals) and the East-West orientation of the continent, where the former is associated with an ease of initiating agriculture and the latter an ease of agricultural diffusion. These advantages allowed for an earlier transition to, and a more widespread use of, agricultural practices; which in turn, allowed for mass populations, the development of cities and states, the specialization of labor, and, ultimately, a head start in the acquisition of prosperity. Diamond's hypothesis is tested by Putterman (2007) and Hibbs and Olsson (2004), who find a positive correlation between agricultural transition dates and wealth levels in 1500 CE. The most tangible difference between the two papers is in the way agricultural transition dates are calculated: Putterman uses archeological facts in calculating the dates for particular countries, while Hibbs and Olsson use biogeographic and geographic conditions in order to estimate the transition dates for regions. Diamond's argument, however, does not give reason as to why variations within Eurasia may develop. This paper seeks to supplement Diamond's by providing a possible explanation to within levels of development; particularly, we use the varied use of milk as an explanation of varied levels of development throughout the Old World.

Instead of archeological evidence or environmental estimates, we use an observed genetic difference between societies as a predictor of past economic development. This genetic difference is primarily driven by differences in culture; and through the process of natural selection, this information has been passed through generations of humans until today. Diamond states: "History followed different courses for different peoples because of differences among peoples' environments, not because of biological differences among peoples themselves." A difference in environments, however, is the main cause in divergent evolutionary paths, according to Darwin: "In the struggle for survival, the fittest win out at the expense of their rivals because they succeed in adapting themselves best to their *environment*."[3] Therefore, a difference in environments, including both cultural and geographical differences, allows for differences in genetic adaptations. Conversely, the use of genetic variation may be used as an indicator of the usage or availability of a cultural or environmental advantage conferred to some societies and not others.

The effects environmental changes have on evolution are numerous and well documented. The most common example involves the peppered moths of England before and after the industrial revolution (Kettlewell 1956). Before the revolution light colored moths were the vast majority due to camoflauge provided by light colored trees; however, the industrial revolution caused dark soot to form on the trees causing lighter colored moths to stand out. The darkening of the trees allowed for the darker variety of the peppered moth to have a greater relative probability of survival, thereby increasing the frequency of dark moths compared to light. Just as the dark colored moths had an advantage after the environmental shift, those peoples who were able to capitalize the additional resource of milk were able to increase their numbers relative to those who were unable to digest lactose.

A number of recent papers explore the effect that genetics may have on aggregate economic outcomes (see, e.g., Ashraf and Galor 2008; Spolaore and Wacziarg 2009; Michalopolous 2008). In general, these papers use broad genetic variation measures between, and within, particular countries to explore differing economic outcomes, historic and current. This paper differs in the use of a particular gene variant, not differences in the general genetic make-up of a population. In particular, the current work uses variation in an expressed genetic trait which has been naturally selected for since the Neolithic Revolution. To our knowledge, this is the first paper to explore the effect of a particular gene variant expression has on aggregate economic conditions.

A similar work by Nunn and Qian (forthcoming) explores how the introduction of the

[3] Emphasis my own.

potato to the Old World has affected populations in the 18th and 19th centuries. Specifically, they show that exogenously determined soil conditions, which are favorable for potato production, account for 25%-26% of the population increase from 1700 to 1900 and 27%-34% of the increased urbanization rate in the same time period. Both the current work and that of Nunn and Qian explore how the addition, or varied use, of a particular food source affects historic populations. A slight difference, however, is found in quantifying the spread of the respective food sources; Nunn and Quian use soil conditions, whereas we use the observed differences of an underlying genetic variation.

Galor and Moav (2008) show adaptation since the initiation of agriculture has a statistically significant relationship with contemporary variations in aggregate health measures. The work of Galor and Moav (2008) implies that differences have developed since the Neolithic Revolution and that these differences may be correlated with differing economic outcomes. It is this attitude that we seek to capture. Particularly, the variation in the timing of the Neolithic led to a variation in the genetic ability to consume milk.

1.1 Population Advantages of Milk Consumption

The consumption of milk today ranges from cows in Europe, America, Australia, and Africa to camels and goats in the Middle East, reindeer in the Arctic, mares and asses in the Eurasian steppe, and water buffalo in Southeast Asia (WHO 2009). [4] There is considerable evidence that milk stimulates growth, increases bone density, and provides essential vitamins and minerals (Hoppe et al. 2006). Milk is an incredibly complex liquid that contains fats, proteins, vitamins, and minerals; as the popular slogan states (McCann-Erikson 1990): "Milk: It Does a Body Good!" Along with these qualitative advantages, milking also allowed early farmers and pastoralists to obtain a greater number of calories from a fixed number of cattle. Through the qualitative and quantitative attributes of milk, greater populations could be supported for a fixed quantity of land.

A sugar found in milk, lactose, is responsible for the exclusivity in consumption. The enzyme required to break down lactose, lactase, is found within the small intestine.[5] If this

[4] For simplicity we reference milk to be from cattle.

[5] Lactose is found in all milk

enzyme is not present, the lactose will pass to the colon causing diarrhea or cramping to occur (Simoons 1969). Like all mammals, humans produce lactase from birth until the end of weaning in order to digest the numerous nutrients that are passed from mother to offspring.[6] Certain populations of humans, however, have developed an allele, or gene variant, that allows for the production of lactase throughout their adult lives; this is known as lactase persistence.[7] Considering that the vast majority of humans, and all other mammals, are unable to produce lactase beyond the weaning period, it must be the case that the inability to drink milk into adulthood is the original state (Simoons 1969). Accordingly, the ability to digest milk, or to be lactase persistent, is one of the most famous cases for continued evolution in humans (Ingram et al. 2009).

The quantitative advantages in the ability to digest lactose are apparent. Consider two farmers (or pastoralists) with identical numbers of cattle (or some other milk producer). One of the farmers is able to digest milk, while the other is not. The farmer who is able to digest milk immediately gains an additional resource from his set herd of cattle. Moreover, the farmer who is able to digest milk can now support a larger family, which in turn has the effect of increasing the population and increasing the percent of lactase persistence within the population.

It isn't necessarily the case that strict specialization in milk production is required to increase population densities. This paper argues that the supplementation of the additional resource is enough to improve pre-colonial population levels. Horticulture can supply vastly more calories per acre than any husbandry technique (Cooper and Spillman 1917). A homogenous diet of a few grains, however, led to adverse health effects in early farmers (Cohen and Armelagos 1984). The addition of fats, proteins, vitamins, and minerals found in milk provided a healthy balance to the early farmer's diet, which, in turn, allowed for longer lives and greater populations. According to the World Health Organization (2009, p. 3): "The profile of amino acids in milk complement those in grains and cereals, which is of considerable benefit in communities where grains and cereals predominate." Additionally,

[6] Weaning is the process of an infant taking nourishment other than by suckling.

[7] As is consistent with the literature, we will use lactase persistence instead of lactose tolerance. Although, the two terms have equivalent definitions.

Nunn and Quian state (forthcoming, p. 7): ". . . a single acre of land cultivated with potatoes and one milk cow was nutritionally sufficient for feeding a large family of six to eight."[8] Considering two societies with equal resources, the society that is able to digest milk gains a qualitative dietary advantage that increases health and, therefore, population.

Milk provided both quantitative and qualitative advantages to the early farmer's diet, which, respectively, can be seen as a substitute or a complement to a farmer's diet. Again consider two identical farmers: one can digest milk while the other cannot. The farmer who is able to digest milk is able to sustain solely on the caloric output that milk provides—i.e., milk is a substitute for other food sources. The farmer is also able to supplement needed vitamins, minerals, and other essential nutrients, which a staple crop provides an insufficient amount—i.e., complementing the farmer's current diet. Both effects would increase pre-colonial populations.

In addition to the direct effects of consumption, the availability of milk may have increased the fecundity of early sedentary women. Postpartum amenorrhea, or infertility, is positively related to the length of time an infant weans (Jain et al. 1970). The use of milk as a substitute for mother's milk would have reduced weaning time and, therefore, the postpartum infertility period.[9] Implying, a mother who had access to milk would have been able to give birth to a larger number of children over her life span, which corresponds to the positive relationship between dairying and populations.

1.2 Selection for Lactase Persistence

The Neolithic Revolution radically changed the environment for humans, and this change has occurred at different times for different peoples. This implies that certain groups have had a longer time to evolve, or adapt, to the new environment, and one adaptation is the continued production of lactase. Burger et al. (2007) have shown that the allele, or gene variant, that allows for lactase persistence in Europeans is absent, or rare, in early Neolithic Europeans. Considering that Europeans have the highest levels of lactase

[8] Potatoes are nutritionally advantageous to other Old World staple crops, which implies the inclusion of milk is complimetary no matter the nutritional value of the staple crop.

[9] All infants produce lactase in order to digest mother's milk.

persistence in the world, this implies that the ability to digest lactose into adulthood is a new phenomenon that gives a significant advantage to its possessors. Toward this end, Bersaglieri et al. (2004) find that the differences in lactase persistence frequencies are due to a strong positive selection of an allele that allows for milk consumption occurring in the past 5,000-10,000 years, a time range that is consistent with the domestication of cattle and other milk producing domesticates. Furthermore, the gene variant that confers lactase persistence is the "textbook" example of a selective sweep (Nielsen et al. 2005; Ingram et al. 2009).[10]

Most gene mutations that occur do not confer any type of advantage. If, however, a gene mutation gives an advantage, then its possessor is more likely to survive and, in turn, produce more children. This process continues over time, with a larger and larger portion of the population containing this mutation, i.e. the allele is being naturally selected. Or in the words of Darwin:

> Owing to this struggle for life, variations, however slight and from whatever cause proceeding, if they be in any degree profitable to the individuals of a species, in their infinitely complex relations to other organic beings and to their physical conditions of life, will tend to the preservation of such individuals, and will generally be inherited by the offspring. The offspring, also, will thus have a better chance of surviving, for, of the many individuals of any species which are periodically born, but a small number can survive. I have called this principle, by which each slight variation, if useful is preserved, by the term Natural Selection.

Given the fast increase in the frequency of lactase persistence, then it must be the case that digesting lactose did provide an advantage for the owners of a lactase producing gene variant. Bersaglieri et al. (2004) find that the ability to continually produce lactase has a selective advantage between .014 and .15: this implies that a population of 1,000 individuals that are able to produce lactase throughout their lives will have between 14 and 150 more offspring per generation compared to individuals without the ability to digest lactose.[11]

---

[10] A selective sweep is defined as, "The process in which a favorable mutation becomes fixed in a population (Hartl and Clark, P. 184)."

[11] This is dependent on the availability of milk. If no milk is available; no advantage exists.

If no cattle were available, and therefore no milk, then there would be no advantage to producing lactase. This implies further that the availability of milk is a necessary condition for the rise in frequencies of lactase persistence. This co-evolution of dairying and lactase persistence is formally known as the "Cultural Historical Hypothesis" and is attributed to Simoons (1969). According to Simoons:

> Such an advantage most likely would occur in groups, not necessarily pastoral, that not only enjoyed a plentiful milk supply, but that had other foods inadequate in amount and quality, and that did not process milk into products low in lactose. Under these conditions, the lactase aberrant adults would better multiply, and would more successfully defend their families against others. And in their numerous descendants, high levels of adult lactase activity would come to prevail.

The "Cultural Historical Hypothesis" has received considerable attention lately with the discovery that the origination of lactase persistent alleles have coincided with the proposed dates of the domestication of cattle (Coelho et al. 2005, Mulcare 2006, Bersaglieri et al. 2004, and Tishkoff et al. 2007).

This indicates that the frequency of lactase persistence may just be a proxy for the origination of animal husbandry; whereby the frequency of lactase persistence is an increasing function of the years since the domestication of a particular mammal. While it is true that the availability of milk, or cattle, is a necessary condition for the evolution of lactase persistence, it is not, however, a sufficient condition. Southern Europe, Eastern Europe, the Near East, and the Middle East have had access to milk for as long, or longer, than Western Europeans, yet these areas have significantly lower levels of lactase persistence (Simoons 1978). This indicates that differences in dairying also have a cultural significance. For this reason, the use of lactase persistence frequencies does not measure the initial advantages of obtaining cattle; it measures the initial advantages of milking.

In summary, the gene variant that allowed its possesors to consume milk did provide an advantage. One question this work seeks to answer is whether or not this advantage led to differential economic outcomes. The next section provides a detailed explanation of the cross-country measure of lactase persistence.

2 Data

2.1 The Frequency of Lactase Persistence

Milk consumption has independent origins across the Old World, which has resulted in a number of gene variants, or alleles, responsible for the production of lactase (Ingram et al. 2009; Tishkoff et al. 2006). Further, the frequency of a particular variant is ethnic specific. In other words, the gene variant that allows for milk consumption in Northern Europeans is not identical to the allele that allows for milk consumption in Western Africans. It is for this reason that the observed, or phenotypic, ability to consume milk is the primary determinant of our measure of lactase persistence.[12][13]

The data for the frequencies of lactase persistence come from Ingram et al. (2009), in which the authors aggregate data from past studies of lactase persistence frequencies. The data are given at the ethnic level. The lactase persistence frequencies are obtained by conducting lactose tolerance tests on samples from an indigenous population. The data are collected from 1965 to 2007. While the tests do span a relatively large time scale, the testing methods used remain constant, and the gene frequencies themselves should have also remained constant over this relatively short period. There are two ways to test for lactase persistence: blood glucose and breath hydrogen. In both tests individuals are given lactose after an overnight fast in order to accurately conduct the tests. A description of the two tests from Ingram et al. (2009):

> A baseline measurement of blood glucose or breath hydrogen is taken before ingestion of the lactose, and then at various time intervals thereafter. An increase in blood glucose indicates lactose digestion (glucose produced from the lactose hydrolysis is absorbed into the bloodstream), and no increase, or a 'flat line' is indicative of a lactose maldigester…

[12] A phenotype is the physical expression of a genotype (Hartl and Clark 2007).

[13] A measure of the frequency of lactase persistence has been calculated by using the frequency of the gene that allows for the continued production of lactase in European populations. Substituting this measure into the estimating equation specified above leads to a positive and signficant coefficient, but the use of the European gene frequency is sensitive to the inclusion of a number of controls. This is to be expected, due to the genes positive relationship with milk consumption in Europeans and nonexistent relationship with milk consumption in all other ethnic populations, which results in a large measurement error on the explanatory variable of interest and an attenuation of the coefficient.

An increase in breath hydrogen indicates maldigestion and reflects colonic fermentation of the lactose...

The arbitrary cutoff levels in defining digesters and maldigesters, or, respectively, lactase persistence and non-lactase persistence, imply that measurement errors will be present.

2.1.1 Estimating the Ethnic Composition of Countries in 1500 CE

In creating a country wide measure for lactase persistence frequencies, two problems need to be overcome. First, we need to aggregate ethnic groups into countries. And secondly, I will need to scale this measure back 500 years as to measure the effect of lactase persistence on pre-colonial development.

In order to aggregate ethnic groups into country level measures, data on the ethnic make-up of countries is used from Alesina et al. (2003). The data from Alesina et al. (2003) give the ethnic composition of 190 countries from roughly 1990 to 1995. Using ethno-linguistic classifications, ethnic groups, which have lactase persistence frequencies from Ingram et al. (2009), have been matched to ethnic groups in Alesina et al. (2003). For example, "Western Europeans" in Sweden from Alesina et al. (2003; hereafter Alesina) are assigned the lactase persistence frequency of "Dane" from Ingram et al. (2009; hereafter Ingram), "Filipinos" in Alesina are assigned to the "Maori" ethnic group in Ingram, and the "Fon" people from Benin are assigned to "Yoruba" from Nigeria.[14] This matching yields data for 118 Old World countries (i.e., Europe, Asia, and Africa) , of which 51 countries have a direct match between the majority ethnic group given by Alesina and ethnic data from Ingram. An additional level of measurement error is to be expected from using ethnolinguistic classification in the matching of ethnic groups. As a result the 51 countries that have exact matches are considered to be more conservative estimates of the country level lactase persistence frequencies, and separate estimations are performed using the reduced sample.

The aggregation from ethnic groups to countries gives a cross-country measure of the

[14] Swedes and Danes belong to the East Scandinavian branch of the Indo-European language group, Filipinos and Maori belong to Malayo-Polynesian branch of the Austronesian language group, and the Fon and Yoruba belong to the Volta-Niger branch of the Niger-Congo language group.

lactase persistence frequency; however, this measure is for the present period and may not be relevant in the prediction of variables in the pre-colonial period. A cross-country measure for lactase persistence 500 years in the past is needed. One way around this problem is to ascribe the largest ethnic group within a country as the country's sole ethnic group in the year 1500 CE(Spaloare and Wacziarg 2009). A cross-country lactase persistence frequency is calculated in this manner with one exception: if an ethnic group does not constitute over 60% of a countries present day composition and another ethnic group constitutes over 30% of the countries composition, the country's ethnic composition in 1500 CE is ascribed as 50% to each group. For example, Belgium's present ethnic composition from matching ethnic groups in Ingram to Alesina is given to be 58% German and 30% French, so in calculating ethnic composition in the year 1500, 50% is ascribed to German and 50% is ascribed to French. Lactase persistence frequencies for 126 countries are found in this manner with 54 countries having exact ethnic matches.

Our primary way of calculating country level ethnic compositions in 1500 CE involves using data on migration frequencies over the period 1500 to 2000 (Putterman and Weil 2010). If it is known where a county's current population has migrated from over the past 500 years, it is possible to effectively remove this fraction of immigrants from the current population, leaving a rough representation of the population in the year 1500 CE. Consider an $m \times n$ matrix, $E^{1500}_{m\times n}$, which contains the ethnic composition of countries in the year 1500 with m ethnic groups and n countries. If we take the product of $E^{1500}_{m\times n}$ and the $n \times n$ Putterman and Weil matrix of migration (denoted as $M^{1500-2000}_{n\times n}$), this should give a rough estimate of the ethnic composition today. For example, consider China and Malaysia, which were respectively composed of the Han and Maori groups in 1500:

$$E^{1500}_{m\times n} = \begin{array}{lcc} & \text{Malaysia} & \text{China} \\ \text{Han} & 0 & 1 \\ \text{Maori} & 1 & 0 \end{array}$$

The matrix $E^{1500}_{m\times n}$ states that in 1500 CE the entire population of China is ascribed to the Han ethnic group and the entire 1500 population of Malaysia is ascribed to the Maori ethnic group. Migration over the last 500 years is given by:

$$M_{n\times n}^{1500-2000} = \begin{array}{lcc} & \text{Malaysia} & \text{China} \\ \text{Malaysia} & 0.75 & 0 \\ \text{China} & 0.25 & 1 \end{array}$$

which says that 75% of Malaysia's population is derived from Malaysia and 25% of Malaysia's population has immigrated from China. And given that in 1500 China was entirely composed of the Han ethnic group and Malaysia was entirely composed of the Maori ethnic group, this implies that Malaysia's current ethnic composition is 75% Maori and 25% Han. This is shown by:

$$A_{m\times n}^{2000} = E_{m\times n}^{1500} \times M_{n\times n}^{1500-2000} = \begin{array}{lcc} & \text{Malaysia} & \text{China} \\ \text{Han} & 0.25 & 1 \\ \text{Maori} & 0.75 & 0 \end{array}$$

However, we are interested in finding $E_{m\times n}^{1500}$ given $A_{m\times n}^{2000}$, which is found through methods described above using data from Alesina et al. (2003), and $M_{n\times n}^{1500-2000}$, which is given in Putterman and Weil (2010). In particular, post multiplying $A_{m\times n}^{2000}$ by the inverse of $M_{n\times n}^{1500-2000}$ gives $E_{m\times n}^{1500}$. In our example with Malaysia and China:

$$\begin{aligned} E_{m\times n}^{1500} &= A_{m\times n}^{2000}(M_{n\times n}^{1500-2000})^{-1} \\ &= \begin{pmatrix} & \text{Malaysia} & \text{China} \\ \text{Han} & 0.25 & 1 \\ \text{Maori} & 0.75 & 0 \end{pmatrix} \times \begin{pmatrix} & \text{Malaysia} & \text{China} \\ \text{Malaysia} & 1.33 & 0 \\ \text{China} & -0.33 & 1 \end{pmatrix} \\ &= \begin{pmatrix} & \text{Malaysia} & \text{China} \\ \text{Han} & 0 & 1 \\ \text{Maori} & 1 & 0 \end{pmatrix} \end{aligned}$$

In theory, post multiplying current country level ethnic compositions by the inverse of the Putterman and Weil migration matrix should remove all migration that has occurred over the last 500 years. This process, however, assumes an equality of migration across ethnic groups. It is improbable that migrations were ethnically equal. This problem is partly mitigated due to the high correlation between ethnicity and state in 1500 CE; e.g., France was entirely composed of French, Zimbabwe was entirely composed of Bantu, etc.

Comparing lactase persistence frequencies calculated through inverting the migration matrix to frequencies calculated through majority ethnic groups yields a correlation of roughly 98%. Assuming equality in migration appears to be a minor issue.

2.1.2 Monotonicity of Lactase Persistence

Although the ethnic composition of countries is somewhat mitigated due to the inversion of the migration matrix, it still remains that the frequencies of lactase persistence themselves are found roughly 500 years after the dependent variable to be explained. The main issue concerns the monotonicity, or relative relationships, of lactase persistence frequencies over the last 500 years. In order to create a false, positive relationship, either countries that were lightly populated in 1500 CE should have had a comparative decline in lactase persistence over the past 500 years, or relatively rich countries should have had a comparative increase in the frequency of lactase persistence.

To understand any potential biases that may occur, it is important to understand how gene frequencies come about. According to population geneticists three main variables affect how the frequency of a gene evolves: the selective, or survival, advantage conferred by the gene variant, the initial population containing the gene variant, and time (Hartl and Clark 2007). Considering all countries in the sample have the same time constraints, any differences in the frequency of lactase persistence must be attributed to either differences in the initial population containing the gene variant or the selective advantage conferred by the allele.

There is no valid reason to suspect variation in the lactase persistent allele prior to the domestication of milk producing animals. The availability of milk determines whether or not lactase persistence provides an advantage; if there is no milk, then there is no advantage, and according to the laws of natural selection: if there is no advantage, then a gene will not rise in frequency (Hartl and Clark 2007). This principle is shown in the absence of the lactase persistent allele in Europeans prior to the Neolithic Revolution (Burger et al. 2007). The possibility does remain, however, that migrations over the past 500 years have distorted the respective genotype of a country. This potential source of bias is partially corrected for by in the methods described above; although, introgression, or the exchange of genes from interactions in the migrant and native populations, may have altered the respective native genotype for a particular country. For this to create a bias in my estimation, the lactase persistent allele would have to be passed only to densely populated countries, which seems unlikely. Further dimming the possibility of bias estimation is given by the inverse relationship between the size of a population and the speed at which a gene frequency rises (Hartl and Clark 2007).

Everything else constant, differences in the selective advantage of lactase persistence will cause differences in the speed in which the frequency of the population obtains the gene (Hartl and Clark 2007). Consider again the peppered moths of England. The advantage of the darker moths was dependent on the level of soot within a particular area: The greater the soot, the greater the advantage of having a dark complexion. Dark moths had a greater reproductive advantage relative to light moths in the darker areas, which in turn caused their numbers, or frequency, to increase at a faster rate in these areas. This same idea can be applied to the advantage conferred by the ability to digest milk, where differing areas could confer differing advantages which could cause a non-monotonic relationship to develop between lactase persistence frequencies today and lactase persistence frequencies in the year 1500 CE.

One potential source of a differing selective advantage arises from the environment in which the gene evolved. Flatz and Rotthauwe (1973) theorize that differences in the frequency of lactase persistence are caused by differences in exposure to ultra violet light. Countries with low levels of sun exposure lack the necessary ultraviolet light to adequately synthesize Vitamin D. Deficiency in vitamin D is associated with rickets, or a weakening of bones. The inclusion of milk, which is high in calcium, offset the negative effects of vitamin D deficiency. This implies a greater advantage for milk in areas with lower sunlight; therefore, lactase persistence should rise to a greater frequency in these areas. A number of controls are used to account for this potential source of bias. These include a Western European dummy and the distance from the equator. In addition to the control variables, a sample truncation is conducted, in which all Western European states and all countries above and below the sample median distance from the equator are excluded. As an extension of this hypothesis, we consider sunlight to be an exogenous determinant of differences in the frequency of lactase persistence. With adequate controls to partial out the effect of sunlight on population density, we use a measure of solar radiation as an instrument for differences in the distribution of lactase persistence. This is further discussed in Section 3.2.2.

Conversely, it could be the case that moderately populated countries, which contained high frequencies of lactase persistence in 1500 CE, faced a situation in which the selective advantage to consuming milk became negative or nonexistent. There is currently no backing for any hypothesis suggesting a negative selective advantage associated with lactase persistence.[15] It is possible, however, that a particular country has lost its milk producing mammals in the past five hundred years, effectively giving no advantage to the ability of drinking milk. According to the Hardy-Weinberg principle, if a gene possesses no selective advantage its relative frequency should remain constant, not decline.[16] Indicating that if a country did lose its cattle stock in the last 500 years, the frequency of lactase persistent individuals within the country should have remained constant; further implying the improbability of a false relationship between lactase persistence frequencies and 1500 CE population density.

2.2 Data: Summary and Sources

Using the ethnic compositions given by the inversion of the migration matrix, I am able to create a lactase persistence measure for the year 1500 CE; this is the primary measure of lactase persistence to be used. This method yields 118 countries, of which 51 have exact ethnic matches. Table 1 presents the descriptive statistics for the frequency of lactase persistence as well as all control and dependent variables. The mean frequency of lactase

[15] There is a hypothesis that states riboflavin rich milk allows for an increased risk to the contraction of malaria (Anderson and Vullo 1994), but this hypothesis is unproven (Meloni et al. 1998).

[16] The Hardy-Weinberg equilibrium states that allele frequencies in a population remain constant, that is, they are in equilibrium from generation to generation unless specific disturbing influences are introduced. Those disturbing influences include non-random mating, mutations, selection, limited population size, ”overlapping generations”, random genetic drift and gene flow (Hartl and Clark 2007).

persistence in the base sample is 41.3%, which is similar to the world mean of 35% given by Ingram et al. (2009).[17] Figure 1 gives a shaded map of Old World lactase persistence frequencies. As expected lower frequencies of lactase persistence occur in Sub-Saharan Africa while higher frequencies are reported in Western Europe, Scandinavia in particular, with a max sample frequency of 96% in Sweden and a min of 2.33% in Zambia. Figure 2 gives historical areas of milking and non-milking. Comparing Fig. 1 and Fig. 2, there appears to be a relatively tight fit between historically non-milking areas and low levels of lactase persistence.

The main variable to be explained is population density in 1500 CE. This variable is from McEvedy and Jones (1978). Thomas Malthus's seminal work on the relationship between population and wealth has shown that any wealth increase prior to the Industrial Revolution was offset by an equivalent increase in population, thereby keeping income per capita constant. For this reason population densities are a viable proxy for wealth levels in 1500 CE; additionally, 1500 CE population densities are used regularly in similar research; e.g., Acemoglu et al. 2002, Ashraf and Galor 2008, Chanda and Putterman 2007, Putterman 2008. The hypothesis posed by this paper is that milking provided an extra resource to certain peoples that expanded the carrying capacity of their environment, thereby increasing population densities, or wealth. Figure 3 gives a simple plot with the natural log of population density on the y-axis and the country level frequency of lactase persistence on the x-axis.

As previously mentioned, the presence of mammals is a necessary, but not sufficient, condition for milking. This denotes that the frequency of lactase persistence may be picking up some of the effects of extended agricultural use. In order to show that milking itself increased population densities, agricultural transition dates need to be controlled for. As stated earlier, two different measures for agricultural transition dates have been used previously: the region specific measures from Hibbs and Olsson and the country specific measures from Putterman. Although Putterman's method is measured with greater certainty, the measure by Hibbs and Olsson may have the effect of capturing unseen technological

[17] The world lactase persistence frequency calculated by Ingram et al. (2009), however, is based on a flawed population weighted average.

similarities between countries assigned to the same region. To conserve space we use the millennia of agriculture measure given by Putterman (2007). The results have been checked using millennia of agriculture from Hibbs and Olsson (2004) with little difference in estimation. Ideally, the distribution of livestock within the Old World would also be controlled for; however, such a measure is unavailable.

In addition to the initiation of agriculture, the yield from agriculture is also extremely important to food production and, therefore, variations in pre-colonial populations. Controlling for land quality is necessary to the estimation of pre-colonial populations. The land quality measure used in this paper is the mean suitability of agriculture (Ramankutty et al. 2002, Michalopoulos 2008). The mean suitability of agriculture is constructed by the country average of 0.5 degree latitude by longitude grids that give a probability of cultivation. Additionally, the soil suitability of potatoes, Old World staple crops, and New World staple crops from Nunn and Qian (forthcoming) are used in the sensitivity analysis.

An additional genetic control comes from Spolaore and Wacziarg (2009) in which the authors measure the genetic distance, or variation, from the world's technological frontier. Using the genetic distance from the U.K. in the year 1500 CE gives a viable control for other alleles that may be highly correlated with lactase persistence. In other words, the frequency of lactase persistence may be accounting for a broad, underlying genetic capital possessed by Western Europeans; therefore, it is useful to see the effect of lactase persistence while controlling for other possible genetic variations.

When conducting sensitivity analyses for omitted variables, additional terrain, water access, environmental, cultural, and genetic controls are used.[18] The distance from the equator is intended to control for geographical variation that lactase persistence may be picking up; this variable is from Rodrik et al. (2002). Terrain and water access controls come from the Center of International Development. These include average elevation, average distance to the coast or navigable river, and the percent of land that is within 100 kilometers of the coast or navigable river. Terrain ruggedness and land within the tropics or deserts are from Nunn and Puga (forthcoming); to account for disease environments the

[18] Table 1 gives the source of all variables.

stability of malaria transmission is used from Kiszewski et al. (2004); and whether or not a particular country belonged to the Roman Empire is also used from Acemoglu, Johnson, and Robinson (2005).

3 Results

The main hypothesis presented in this paper, a higher frequency of lactase persistence is associated with greater population densities in the pre-colonial era, is tested with the following estimating equation:

$$ln(\text{Population Density})_i^{1500} = \alpha + \beta(\text{Frequency of Lactase Persistence})_i + \mathbf{\Phi'X}_i + \epsilon_i \quad (1)$$

where i is a country index, β is the coefficient of interest throughout the paper, and $\mathbf{X}_i$ is a vector of country specific relevant controls. Equation (1) is estimated by OLS with robust standard errors. Robustness exercises use varied samples and variations in $\mathbf{X}_i$.

3.1 Baseline Estimation

The baseline estimations of Equation (1) are given in Table 2. Table 2 establishes the empirical relationship between the frequency of lactase persistence calculated by inversing the Putterman and Weil migration matrix within a country and the log of the 1500 population density for that particular country while controlling for relevant variables.

Column (1) displays the simple bivariate regression of 1500 population density on the frequency of lactase persistence within a particular country. The explanatory variable has a positive coefficient that is significant at the 1% level and explains roughly 20% of the variance in the log of 1500 population density. To be more precise, column (1) reveals that a one standard deviation increase in the frequency of lactase persistence is associated with roughly a 63% increase in the number of people per kilometer. For the median country in the sample, the Sudan, this corresponds to a rough increase of two people per square kilometer. In column (2) a bivariate regression is run to show the impact of the millennia of agriculture within a country (Putterman 2008) on population densities; this is a direct test of the hypothesis proposed by Diamond. The coefficient is positive and significant at the 1% level with the explanatory variable accounting for roughly 14% of the variation in the dependent variable. Column (3) shows the effects of environmental variables, measured by the mean suitability of agriculture, distance from the equator, and dummies for Western Europe and Sub-Saharan Africa, on pre-colonial levels of development. The coefficient of our measure for the suitability of agriculture is positive and significant at the 1% level, which indicates improved land quality led to greater agricultural yields and larger populations. Column (3) also shows that a larger distance from the equator is associated with less dense populations in 1500 CE.[19] As expected, Western European countries had greater population densities, or wealth, relative to other countries, while Sub-Saharan African countries were relatively worse off.

Column (4) exhibits that when controlling for the millennia of agriculture, the coefficient of the frequency of lactase persistence remains significant at the 1% level, which further indicates that the frequency of lactase persistence is accounting for an additional advantage to a longer presence of agriculture. Column (5) shows the results of including the frequency of lactase persistence while controlling for environmental variables. The coefficient of the frequency of lactase persistence remains significant at the 1% level while also leading to a 10% increase in the explained variation of population densities in 1500. Col. (6) introduces millennia of agriculture while controlling for environmental variables; all signs are as expected, although the significance of the Sub-Saharan African dummy dissipates.

The baseline result is given by columns (7) and (8). As stated previously, the availability of cattle is a necessary condition for the development of a gene that allows for digesting lactose; this implies the frequency of lactase persistence may be only capturing the effects of the millennia of agriculture within a particular country. Column (7) shows that when controlling for the country specific measures of millennia of agriculture, as well as environmental controls, the coefficient on lactase persistence remains both positive and significant at the 1% level. Also, comparing columns (6) and (7), the addition of the lactase persistence frequency increases the explained variation of population density in 1500 by roughly 8%. The coefficient of interest in column (7) is consistent with the bivariate estimation of

[19] The coefficient of the distance to the equator is influenced by the use of only Old World countries.

column (1): an increase of one standard deviation in the ability of a population to digest lactose is associated with roughly a 60% increase in the population density in 1500. This suggests that the consumption of milk did indeed have a positive effect on the population density, or pre-colonial living standards, within a particular country.

Column (8) repeats the regression given by column (7); however, the sample is reduced to the countries in which the majority ethnic group is directly matched between Ingram et al. (2009) and Alesina et al. (2003). The coefficient of lactase persistence in Col. (8) is significantly larger than that in Col. (7); this is to be expected given the reduction in measurement error from using the more conservative sample. Also as expected the use of the smaller sample results in a larger standard error. In particular, a one standard deviation increase when using the coefficient in Col. (8) is associated with roughly a 87% increase in 1500 population density.

Table 3 performs the same estimations as Table 2, but instead uses the frequency of lactase persistence calculated by taking the majority ethnic groups within a country. The results, both magnitude and significance, are similar to those found in Table 2. For the baseline estimate of Col. (7), a one standard deviation increase in the frequency of lactase persistence corresponds to a 54% increase in 1500 population density; if we consider the conservative sample given in Col. (8), a one standard deviation increase in the frequency of lactase persistence corresponds to an increase in population density of 82%. Given the high correlation and the similarity of coefficients between the two lactase persistence measures, hereafter we will use the measure calculated with the inverse of the migration matrix.

Tables 2 and 3 corroborate our main hypothesis. Those societies who consumed milk had the advantage of an additional resource; this additional resource, in turn, allowed for the development of greater pre-colonial populations. This relationship remains stable and significant while controlling for agricultural transition dates, agricultural suitability, and other relevant geographic determinants of pre-colonial wealth.

Whether or not the relationship between the frequency of lactase persistence and pre-colonial population density is causitive, depends upon the source of the cross-country differences in lactase persistence. In some sense, lactase persistence is analogous to the land suitability of potatoes found in Nunn and Qian (forthcoming); in which, the frequency of lactase persistence can be seen as an exogenous suitability of consumption (rather than production) for a common good. Lactase persistence, however, has arisen in part due to cultural variation. The cultural cause of differences in lactase persistence creates an ambiguity in the exogeneity of our measure. In other words, did those cultures that adopted dairying have other unseen population advantages? The next section will attempt to alleviate the ambiguity in causation through sample adjustments, the inclusion of possible omitted variables, and instrumental variables estimation.

3.2 Sensitivity Analysis and Identification

The relationship between the frequency of lactase persistence and pre-colonial populations is established in Table 2; however, the nature of this relationship is unclear. The endogeneity of lactase persistence seems plausible: cultures which adopted dairying may have contained additional advantages that allowed for greater levels of pre-colonial development, geographic conditions that permitted dairying may have also permitted larger populations, etc. This suggests that OLS is unlikely to confirm a causative relationship between dairying and population densities. This section attempts to strengthen perceptions of the relationship between the frequency of lactase persistence and 1500 CE population densities. Firstly, we perform truncations and include possible omitted variables to control for a potential spurious relationship. Secondly, we use the average solar radiation a country receives as an exogenous determinant of cross-country differences in the frequency of lactase persistence in order to determine causation. In all specifications the coefficient on the frequency of lactase persistence remains positive, significant, and, for the most part, is consistent in magnitude to the baseline estimate.

3.2.1 Sensitivity Analysis

Table 4 restricts the baseline estimation to each of the three continents that makeup the Old World.[20] The purpose of this is to show that Europe is not responsible for the significance of the coefficient in the baseline estimate, and that the positive relationship between a greater

[20]The Western European and Sub-Saharan African dummies are excluded.

frequency of lactase persistence and pre-colonial population densities is seen within other continents. Column (1) performs the baseline estimation for countries contained only within Europe. The coefficient of lactase persistence in column (1) is positive, significant at the 1% level, and roughly double the magnitude of the baseline estimate given by column (7) of Table 2. This result implies the effects of milk consumption on population density are more pronounced within Europe; this is to be expected, since Europe has a greater history of milk consumption and, therefore, a greater exposure to the population advantages of milk (Simoons 1971). Column (2) constricts the sample to countries within Africa alone. The coefficient of interest is significant at the 10% level and the magnitude of the coefficient is lower than that given by the baseline estimate. The estimates of column (2), however, do show that milk consumption did have a positive effect on population density. The results are similar to those of column (3), which restricts the sample to only Asian countries. Within Asia, a greater frequency of lactase persistence is associated with a greater population density; this effect is significant at the 10% level and differs slightly in magnitude from the baseline estimate. Column's (2) and (3) provide support that it is lactase persistence itself that led to larger populations and not an externality associated with Europe. This result is further confirmed in column (4), in which only Asian and African countries are considered. In column (4), the coefficient of the frequency of lactase persistence is once again significant at the 1% level and the magnitude only differs slightly from that given in the baseline estimate. Table 4 provides substantial evidence that the effect of lactase persistence is not being driven by a European externality, narrowing the possibility of a spurious correlation and providing a better understanding of the role of lactase persistence in explaining variations in pre-colonial population density.

Table 5 conducts column specified sample truncations. Columns (1) and (2) of Table 5 give the results of the baseline regression (Col. (7) of Table 2) while omitting Western European countries from the sample.[21] The purpose of the omission of Western European countries is in the fact that Western European countries have both the highest population

[21] The baseline regression does include a Western European dummy, but the omission of Western European countries should further show that Western Europe is not the driving factor of the results given in the baseline case.

densities and the highest levels of lactase persistence. Additionally, Columns (3) and (4) drop Sub-Saharan African countries from the sample. The reasoning for the omission of Sub-Saharan African countries is due to the fact that these countries contain on average lower frequencies of lactase persistence and lower population densities; the opposite of Western Europe. Column (5) omits both Western Europe and Sub-Saharan Africa, in effect dropping the highest and lowest frequencies of lactase persistence and the highest and lowest regional averages of population density in 1500 CE. In all cases the significance of the coefficient on the frequency of lactase persistence remains at the 1% level, and all point estimates are similar to the baseline case.

Columns (6) and (7) estimate the baseline regression while considering countries that are respectively above and below the median distance from the equator. The median absolute latitude of our sample is 33 degrees. This corresponds to an area just above the tropics or roughly equal to the Levant and slightly above North African states, India, and Southeast Asia. The truncation is done to control for any biases that may occur due to the relationships between milk consumption, vitamin D, and the availability of sun light.[22] After the respective truncations, the point estimates of the coefficient on lactase persistence remains significant at the 1% level and is similar in magnitude to the base line estimation. From the truncations, a selection bias seems improbable.

Our method for approximating ethnic compositions in 1500 CE is prone to measurement error. This is due to disparities in the current ethnic composition and country compositions in the migration matrix (Putterman and Weil 2010). Further, this error is larger in countries that have experienced large immigrations between 1500 and 2000 CE. To account for this potential error Table 6 truncates the base sample by the fraction of the current population that is derived from the 1500 CE population. Column (1), for example, excludes all countries which have less than 50% of the current population originating from the within country 1500 CE population. This results in the exclusion of only two countries from our baseline sample; as a result, the significance and magnitude of the coefficient of interest are analogous to those in column (7) of Table 2. Column (2) excludes countries in which less than 75% of

[22]This idea is further explored with the inclusion of a solar radiation variable into our baseline estimation.

the contemporary population is derived from the 1500 CE population. This results in the exclusion of 10 countries that are included in the baseline sample. The coefficient of the frequency of lactase persistence remains consistent in magnitude and significance. Column (3) performs the same truncation as columns (1) and (2) but sets the threshold of within country population to 85%; again, the estimates are similar to the baseline case. Column (4) excludes countries in which 95% of the current population is derived from 1500 CE populations. This results in excluding 50 countries from the baseline sample. The estimate of the coefficient of interest, however, remains roughly equivalent to the baseline estimate. As a further check, column (5) replaces the frequency of lactase persistence derived by post-multiplying by the inverse of the migration matrix with the measure calculated by assuming the majority ethnic group. Again, the coefficient of lactase persistence is positive, significant at the 1% level, and similar in magnitude to estimations with the full sample. The measurement error that results in our approximation of 1500 CE ethnic compositions does not appear to affect our results. This gives further credence to the relationship between milk consumption, measured by the frequency of lactase persistence, and population density posed in this paper.

Tables 7, 8, 9, 10, 11, and 12 explore whether additional controls can make the effects of lactase persistence frequencies disappear. Table 7 includes an additional genetic measure. Table 8 replaces the mean suitability of agriculture in the baseline estimation (Michalopolous 2008; Ramankutty 2002) with soil suitability measures from Nunn and Qian (2011); these include the suitability for potatoes, New World staples, and Old World staples. Table 9 includes additional environmental controls: elevation, ruggedness, whether a country is within the tropics or desert, a measure of malarial intensity, and whether or not a country belonged to the Roman Empire . Water access variables are included in Table 10. Table 12 includes biogeographic variables from Hibbs and Olsson (2004), while Table 13 includes all additional variables specified in the previous tables.

As noted earlier lactase persistence is a function of the genotype of a respective individual. It may be the case that a genotype that allows for lactase persistence may also allow for other growth promoting attributes, or, in other words, there may be some underlying genetic capital which is beneficial to development. Table 7 introduces the genetic distance from the technological frontier, Britain, in the year 1500 CE to the baseline model (Spolaore and Wacziarg 2009). Spolaore and Wacziarg argue that a smaller genetic distance (i.e. similar genotypes) allowed for an easier diffusion of technology. This is seen in the bivariate regression of Col. (2) in Table 7, where a greater genetic distance from Britain in 1500 CE is associated with lower population densities. The significance of genetic distance remains while controlling for the frequency of lactase persistence (Col. (3)); however, the inclusion of relevant agricultural and geographic controls makes the coefficient of Spolaore and Wacziarg's genetic distance statistically insignificant (Col. (4)). The additional genetic control does not alter the significance or magnitude of lactase persistence. Lactase persistence is of importance, not because it is part of some larger genetic package, but because lactase persistence allowed for the consumption of an additional resource. This singular genetic adaptation gave an advantage, which in turn, allowed for the development of larger historic populations.

Column (1) of Table 8 includes the average country-level soil suitability for potatoes while excluding the baseline soil suitability measure. The introduction of the potato in between the 18th and 19th centuries is associated with a large increase in population over this time period (Nunn and Qian forthcoming). The inclusion of this suitability measure is intended to capture any additional effects that this measure may be accounting for in regards to population variation. As seen in Col. (1) the potato suitability measure is positive and significant, indicating an additional relationship between the soil suitability and population density. The inclusion of this variable, however, does not affect the significance or magnitude of the coefficient of the frequency of lactase persistence. Columns (2), (3), and (4) respectively introduce the suitability for Old World staple crops, New World staple crops, and jointly controls for both measures of soil suitability. Again, the significance and magnitude of the coefficient of lactase persistence remain similar to the baseline estimation. Col. (5) replaces the measure for New World staple crops with that for potatoes; the role of lactase persistence is unaffected, while the suitability of both potatoes and Old World staples are positive and significantly related to pre-colonial population densities. Table 8 again confirms that dairying did have a strong association with historic population densities. This relationship is not the by product of soil suitability; rather, dairying was an important determinant to pre-colonial populations.

Columns (1), (2), and (3) of Table 9 introduce elevation (in km), ruggedness, and ruggedness squared into the estimation. Ruggedness is roughly the variation in elevation of particular cells within a country, which are then averaged to the country level (Nunn and Puga 2010). For our concerns, ruggedness and elevation may account for land variations that make farming difficult; and, therefore, may promote the use of animal husbandry, which increases the likelihood of milk consumption. The addition of these additional geographic controls should alleviate any potential biases that may occur due to land conditions that lead to an increased use of pastoralism. Col. (1) includes elevation into the estimation; results remain significant and similar to the baseline estimates. The inclusion of ruggedness and its square in column (2) produce trivial differences in the estimates of the coefficient of interest.

An argument has been put forward that extreme environments may contribute to variations in lactase persistence (Cook and al-Torki 1975). The idea being that extreme environments have fewer resources in which to support populations; therefore, the ability to drink milk becomes essential to surviving and will rise to a greater frequency within the population. Columns (3) and (4) of Table 9 control for environmental differences by including, respectively, the percent of land within the tropics and the percent of land which is desert. The percent of land within the tropics, for our purposes, represents an environment in which resources are rich; consequently, there should be little need for dairying. At the other extreme, deserts are poor in resources, implying a greater need for dairying. This is verified by the coefficients on the respective environments. Deserts have a negative and significant effect on pre-colonial population density, while the tropics have a positive but insignificant effect. Neither variable alters the effect of the frequency of lactase persistence. The coefficient of lactase persistence remains positive, significant, and similar in magnitude to the baseline estimate; this is true while including the environmental variables separately (Col.'s (3) and (4)) or jointly (Col. (5)).

An additional environmental effect that may act on the number of cattle (and, in turn, the number of milk drinkers) and population density is the disease environment. Cattle and other milk producers are extremely sensitive to the tsetse fly, while people are subject to malaria and other tropical disease from similar environments. Looking at Figures 1 and 2, areas with a low frequency of lactase persistence are similar to areas with historic levels of malaria. This indicates that the relationship between lactase persistence and historic populations may be driven by the disease environment. Column (6) controls for the disease environment by including the stability of malarial transmission within a particular country, which can also be seen as a proxy for the tsetse fly (Kiszewski et al. 2004). While this is a contemporary measure, we have little evidence to believe it is an ineffective control variable. The inclusion of the disease proxy does not affect the coefficient of lactase persistence. The estimated coefficient of the frequency of lactase persistence is unaffected by the inclusion of the malaria ecology index. Particularly, the coefficient remains significant at the 1% level and is of a consistent magnitude to the baseline estimate.

Column (7) of Table 9 includes a dummy for whether or not a country was part of the Roman Empire. Using historical evidence Acemoglu, Johnson, and Robinson (2005) argue that being included in the Roman Empire may have contributed to the advanced growth of Western Europe. This is a cultural variable that may be included with the diffusion of technology, development levels in the pre-colonial era, and, ultimately, the practice of dairying. Col. (7) shows that being a part of the Roman Empire did have a significant effect on population densities in 1500; however, this effect is not coming at the expense of lactase persistence. The inclusion of the Roman Empire dummy causes no meaningful difference in the magnitude or statistical significance in the coefficient of lactase persistence.

Column (8) in Table 9 introduces all environmental, disease, and cultural controls. Again, the significance and magnitude of the coefficient on lactase persistence are unaltered. The relationship between dairying and historic populations is not the result of a simultaneous correlation with an environmental or cultural variable.

Table 10 includes a number of water access controls. These include the distance from an ice free coast, the distance from a navigable river, the distance to either an ice free coast or a navigable river, the percent of land within 100 kilometers of an ice free coast, and the percent of land within 100 kilometers of an ice free coast or a navigable river. Neither individually nor jointly introducing water access controls affects the significance or magnitude of the coefficient on lactase persistence. Specifically, column (6) gives the baseline estimation while including both the distance from a coast or a river and the percent of land within 100 kilometers of a coast or river; the coefficient of lactase persistence is significant at the 1% level and resembles the baseline estimate.

Domesticable animals were a necessary condition for the development of lactase persistence. But domesticable animals also provide population benefits, e.g., meat, labor, etc. Table 11 uses the number of potential domesticate animals as a proxy for the additional benefits conferred by domesticate animals, as well as other biogeographic controls from Hibbs and Olsson (2004). Column (1) gives the baseline estimates with the sample reduction; results are similar to the larger sample in column (7) of Table 2. Column (2) includes the number of domesticable animals into the baseline estimation. The inclusion of this variable has a negligent effect on the coefficient of lactase persistence. This supports our main hypothesis that a greater level of milk consumption led to denser populations in the precolonial era. Columns (3) and (4) include a measure for the number of domesticable crops and a measure for the East-West orientation of a country respectively. The coefficient of interest remains roughly equivalent to the baseline estimate. Column (5) includes both the number of domesticable plants and animals, while column (6) includes all variables from Hibbs and Olsson into the baseline estimation. The inclusion of biogeographic controls does not influence the estimated relationship between the frequency of lactase persistence and population density in 1500 CE.

Table 12 simultaneously introduces the potential omitted variables discussed in Tables 7, 9, 10, and 11. [23] The inclusion of all additional variables does not affect the coefficient of lactase persistence; this is shown in column (5). Column (6) reproduces the estimate of

[23] The only soil suitability measure considered in Table 13 is the baseline measure from Michalopolous (2011). Inclusion of differing suitability measures has an insubstantial effect on the coefficient of the frequency of lactase persistence.

column (4), while only considering our conservative sample.[24] Again, neither the magnitude or significance of the coefficient are meaningfully affected. The effect of lactase persistence is robust to the inclusion of a large and theoretically important set of additional controls. Omitted variable bias seems to be insubstantial.

In summary, the coefficient on lactase persistence remains relatively constant throughout the numerous empirical specifications performed. Throughout the sensitivity analysis, the coefficient of the frequency of lactase persistence remains significant at the 1% level and is rarely different in magnitude from the bivariate or baseline estimations (Columns (1) and (7) of Table 2). This robustness is shown through differing samples and the inclusion of theoretically relevant variables, which should, in the least, mitigate a potential selection or simultaneity bias. A strong association exists between milk consumption and population densities in 1500 CE. This implies that the intensity of milk consumption did play some role in the development of larger pre-colonial societies. Those who were able, and did, consume milk gained both qualitative and quantitative advantages which led to larger populations; larger populations in turn led to greater armies, technological gains, and eventually a head start to prosperity differences seen today.

This works primary goal is to explore the role milk consumption, measured through the ability to digest lactose, had in the accumulation of pre-colonial populations. The coevolution of the ability to consume milk with the cultural adaptation of dairying, however, prevents the genetically given lactase persistence measure to be truly exogenous. The omitted reason as to why some cultures initiated dairying while others did not may also be correlated with the accumulation of pre-colonial populations, implying a potential simultaneity bias. Without the use of an exogenous instrument, causality cannot be established. The next section will attempt to alleviate the lack of causation with the use of an exogenous instrument.

[24]Biogeographic controls are omitted due to sample considerations.

3.2.2 Identification

In order to establish causation we consider the proposed relationship between lactase persistence and low sunlight areas (Flatz and Rotthauwe 1973). In adequate sunlight, the body is able to synthesize vitamin D; however, if sunlight is low, individuals may be deficient in vitamin D. A major disease associated with deficieny in vitamin D is rickets, which results in the softening of bones. A diet heavy in milk would increase calcium absorption, thereby partially offsetting the harmful effects of Vitamin D deficiency (Flatz and Rotthauwe 1973; Gueguen and Pointillart 2000).[25] Therefore, those societies in low sunlight countries, i.e. Western Europe, gained an additional benefit from the consumption of milk. With this understanding, we use a 22 year average of solar radiation as an exogenous determinant of the frequency of lactase persistence.

The measure of solar radiation comes from the Atmospheric Science Data Center of NASA (NASA Surface Meteorology and Solar Energy 2011). With the use of country latitude and longitude from the CIA World Factbook, we calculate the 22 year average of solar radiation of a horizontal surface, given in the kilowatts per hour of a squared meter, for all countries in our sample. Figure 5 plots the relationship between this measure of solar radiation and our measure of the frequency of lactase persistence. The relationship appears to be nonlinear. At low levels of sunlight, lactase persistence is widespread; however, as sunlight increases beyond an adequate amount, the frequency of lactase persistence becomes more varied. We therefore use solar radiation and its square in order to instrument the frequency of lactase persistence.

While the relationship between solar radiation and the frequency of lactase persistence is strong in our sample, the use of solar radiation as an instrument is problematic. First, solar radiation may correlate with factors that influence population density. This is partially alleviated by the inclusion of relevant controls, i.e. the mean suitability of agriculture, distance from the equator, and a Western European dummy, but the relationship between solar radiation and population density may not be fully accounted for.[26] Second, the proposed

[25] Milk also contains small amounts of vitamin D.

[26] After including relevant controls, neither solar radiation or its square are insignificant from zero at the 10% level. However, they are jointly significant.

relationship between sunlight and lactase persistence has come under recent criticism. Itan et al. (2009) simulate the evolution and spread of the gene associated with lactase persistence in Europeans. When controlling for relevant factors, they find that the low sunlight areas of Northern Europe do not correlate with a higher frequency of lactase persistence. Gerbault et al. (2009), however, find evidence supporting the relationship between solar radiation and lactase persistence. In short, the relationship between the frequency of lactase persistence and sunlight is still in question. Given the problems of our proposed instrument, we use IV estimation as a supplement to the estimates given by least squares. [27]

The baseline IV estimates are given in Table 13. Column (1) displays the bivariate regression of 1500 population densities on the frequency of lactase persistence. Solar radiation and its square have a strongly correlated with the frequency of lactase persistence; this is shown by the first stage F statistic of 113.39. The IV estimated coefficient of the frequency of lactase persistence is positive and significant at the 1% level. Additionally, the magnitude of the coefficient is similar to the bivariate, OLS estimate of Table 2.

In the bivariate case, however, the IV estimates may be bias. This is due to the agricultural benefits of sunlight. Therefore, columns (2) and (3) respectively add in millennia of agriculture and the suitability of agriculture, as well as other geographic variables. The inclusion of the additional controls does weaken the strength of our proposed instruments, but the instruments remain strong. The IV estimated coefficient of the frequency of lactase persistence is positive and significant while including the millennia of agriculture in column (2); however, the coefficient becomes insignificant in column (3), which includes the suitability of agriculture and other relevant geographic controls.

Column (4) gives the baseline IV estimate. Solar radiation and its square are highly related to the frequency of lactase persistence. The first stage F statistic of column (4) is 31.13, which satisfies the maximum Stock-Yogo criteria. The IV estimated coefficient of the frequency of lactase persistence is positive, significant at the 1% level, and roughy identical

[27] Given the shortcomings of solar radiation, we have also used the number of potential domesticate animals from Hibbs and Olsson (2004). While correlated with the frequency of lactase persistence, the number of potential domesticate animals is a weak instrument. This is especially true when including additional controls.

to the OLS estimate.[28] Column (5) reduces the sample to the conservative estimates, leading to a slight reduction in instrument strength and a larger estimated coefficient of interest. Again, the estimated coefficient is similar to that given by least squares estimation. As stated before, the use of IV estimates is meant to supplement the estimations by least squares. The consistency of the IV coefficient in magnitude and significance to the OLS estimates provided further evidence that the relationship between dairying and population density is substantial.

Table 14 performs IV estimations while including the additional controls of Tables 7, 9, 10, and 11. Column (1) includes genetic distance from the U.K. into the baseline IV estimation of column (4) of Table 13. The instruments remain strong, and the estimated coefficient is similar to the baseline IV estimate, as well as the baseline least squares estimate. All environmental variables of Table 9 are included in column (2). Again the coefficient remains similar to the baseline estimates. Column (3) includes water access controls given by column (6) of Table 10. This results in a reduction in magnitude in the coefficient of lactase persistence, which leads to the coefficient being insignificant at the 10% level. Column (4) includes biogeographic controls of Hibbs and Olsson (2004) into the baseline estimation. The coefficient of interest is significant at the 5% level and is similar to previous estimates in magnitude. Column (5) includes all additional controls. The coefficient remains similar in magnitude to previous estimates with statistical significance dropping to the 10% level. [29] Aside from the lack of significance in column (3), IV estimates of the coefficient of the frequency of lactase persistence remain similar in magnitude and statistical significance to the baseline IV estimates, as well as the baseline OLS estimates.

The use of solar radiation is potentially problematic. However, IV estimations provide further evidence for the relationship between dairying and population densities posed in this paper. Furthermore, given the uniformity in magnitude and significance of the IV and OLS estimates, we have no reason to suspect a potential spurious relationship. Milk consumption aided the diet of early farmers; these benefits appear to have resulted in denser populations.

[28]The OLS estimated coefficient is 2.34, while the IV estimated coefficient is 2.24.

[29]Due to the sample adjustment of the Hibbs and Olsson data, we exclude biogeographic controls from column (5); however, when these variables are included, no significant change is seen.

4 Conclusion

Diamond has stated that, "History followed different courses for different peoples because of differences among peoples' environments, not because of biological differences among peoples themselves." This paper does not intend to dispel this argument; rather this paper merely alters this view. Diamond is correct that the environment is the ultimate causal factor in the differing fates of humanity, but to assume the environment has not caused differences in people undermines one of the basic laws of evolution. The use of genetic frequencies above is merely an indicator for differing environments.

Toward this end, our work establishes an empirical relationship between milk consumption and pre-colonial development. Milk had the ability to improve both the quality and quantity of calories for Neolithic farmers and pastoralists. Both effects had the outcome of increasing populations. This relationship holds through a number of specifications and estimations, and gives important insights into the numerous advantages contained within Eurasian continent and Europe in particular.

References

[1] Nasa surface meteorology and solar energy, September 2011.

[2] Daron Acemoglu, Simon Johnson, and James Robinson. The colonial origins of comparative development: An empirical investigation. *The American Economic Review*, 91(5):1369–1401, 2001.

[3] Daron Acemoglu, Simon Johnson, and James Robinson. Reversal of fortune: Geography and institutions in the making of the modern world income distribution. *Quarterly Journal of Economics*, 117(4):1231–1294, 2002.

[4] Daron Acemoglu, Simon Johnson, and James Robinson. The rise of europe: Atlantic trade, institutional change, and economic growth. *The American Economic Review*, 95(3):546–579, 2005.

[5] Alberto Alesina, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg. Fractionalization. *Journal of Economic Growth*, 8(2):155–194, 2003.

[6] B. Anderson and C. Vullo. Did malaria select for primary adult lactase deficiency? *Gut*, 35(10):1487, 1994.

[7] Q. Ashraf and O. Galor. The out of africahypothesis, human genetic diversity, and comparative economic development. *Department of Economics Working Papers*, 2010.

[8] T Bersaglieri, P Sabeti, N Patterson, T Vanderploeg, S Schaffner, J Drake, M Rhodes, D Reich, and J Hirschhorn. Genetic signatures of strong recent positive selection at the lactase gene. *The American Journal of Human Genetics*, 74(6):1111–1120, 2004.

[9] G. Bloom and P.W. Sherman. Dairying barriers affect the distribution of lactose malabsorption. *Evolution and Human Behavior*, 26(4):301–312, 2005.

[10] V. Bockstette, A. Chanda, and L. Putterman. States and markets: The advantage of an early start. *Journal of Economic Growth*, 7(4):347–369, 2002.

[11] J. Burger, M. Kirchner, B. Bramanti, W. Haak, and MG Thomas. Absence of the lactase-persistence-associated allele in early neolithic europeans. *Proceedings of the National Academy of Sciences*, 104(10):3736, 2007.

[12] A. Chanda and L. Putterman. Early starts, reversals and catch-up in the process of economic development. *Scandinavian Journal of Economics*, 109(2):387, 2007.

[13] Gregory Clark. *A Farewell to Alms: A Brief Economic History of the World*. Princeton University Press, 2008.

[14] M. Coelho, D. Luiselli, G. Bertorelle, A.I. Lopes, S. Seixas, G. Destro-Bisol, and J. Rocha. Microsatellite variation and evolution of human lactase persistence. *Human genetics*, 117(4):329–339, 2005.

[15] M.N. Cohen, G.J. Armelagos, Wenner-Gren Foundation for Anthropological Research, and State University of New York College at Plattsburgh. *Paleopathology at the Origins of Agriculture*. Academic Press New York, NY, USA:, 1984.

[16] D.A. Comin, W. Easterly, and E. Gong. Was the wealth of nations determined in 1000 bc?, 2006.

[17] GC Cook and MT Al-Torki. High intestinal lactase concentrations in adult arbs in saudi arabia. *British Medical Journal*, 3(5976):135, 1975.

[18] M.O. Cooper and W.J. Spillman. *Human food from an acre of staple farm products*. US Dept. of Agriculture, 1917.

[19] J. Diamond. *Guns, Germs, and Steel: The Fates of Human Societies*. W.W. Norton, 1998.

[20] N.S. Enattah, T.G.K. Jensen, M. Nielsen, R. Lewinski, M. Kuokkanen, H. Rasinpera, H. El-Shanti, J.K. Seo, M. Alifrangis, I.F. Khalil, et al. Independent introduction of two lactase-persistence alleles into human populations reflects different history of adaptation to milk culture. *The American Journal of Human Genetics*, 82(1):57–72, 2008.

[21] S.L. Engerman and K.L. Sokoloff. Factor endowments: institutions, and differential paths of growth among new world economies: a view from economic historians of the united states, 1994.

[22] S.L. Engerman and K.L. Sokoloff. Factor endowments, inequality, and paths of development among new world economics, 2002.

[23] G. Flatz and H.W. Rotthauwe. Lactose nutrition and natural selection. *The Lancet*, 302(7820):76–77, 1973.

[24] J.L. Gallup, J.D. Sachs, and A.D. Mellinger. Geography and economic development, 1998.

[25] P. Gerbault, C. Moret, M. Currat, and A. Sanchez-Mazas. Impact of selection and demography on the diffusion of lactase persistence. *PLoS One*, 4(7):e6369, 2009.

[26] H.J. Greenfield, J. Chapman, AT Clason, A.S. Gilbert, B. Hesse, and S. Milisauskas. The origins of milk and wool production in the old world: A zooarchaeological perspective from the central balkans [and comments]. *Current Anthropology*, 29(4):573–593, 1988.

[27] L. Gueguen and A. Pointillart. The bioavailability of dietary calcium. *Journal of the American College of Nutrition*, 19(suppl 2):119S, 2000.

[28] D.L. Hartl and A.G. Clark. *Principles of population genetics*, volume 116. Sinauer associates Sunderland, MA, 4 edition, 2006.

[29] J. Hawks, E.T. Wang, G.M. Cochran, H.C. Harpending, and R.K. Moyzis. Recent acceleration of human adaptive evolution. *Proceedings of the National Academy of Sciences*, 104(52):20753, 2007.

[30] D.A. Hibbs and O. Olsson. Geography, biogeography, and why some countries are rich and others are poor. *Proceedings of the National Academy of Sciences of the United States of America*, 101(10):3715, 2004.

[31] E.J. Hollox, M. Poulter, M. Zvarik, V. Ferak, A. Krause, T. Jenkins, N. Saha, A.I. Kozlov, and D.M. Swallow. Lactase haplotype diversity in the old world. *The American Journal of Human Genetics*, 68(1):160–172, 2001.

[32] C. Hoppe, C. Mølgaard, and K.F. Michaelsen. Cow's milk and linear growth in industrialized and developing countries. *Annu. Rev. Nutr.*, 26:131–173, 2006.

[33] C.J.E. Ingram, C.A. Mulcare, Y. Itan, M.G. Thomas, and D.M. Swallow. Lactose digestion and the evolutionary genetics of lactase persistence. *Human genetics*, 124(6):579–591, 2009.

[34] C.J.E. Ingram, T.O. Raga, A. Tarekegn, S.L. Browning, M.F. Elamin, E. Bekele, M.G. Thomas, M.E. Weale, N. Bradman, and D.M. Swallow. Multiple rare variants as a cause of a common phenotype: several different lactase persistence associated alleles in a single ethnic group. *Journal of molecular evolution*, 69(6):579–588, 2009.

[35] Y. Itan, A. Powell, M.A. Beaumont, J. Burger, and M.G. Thomas. The origins of lactase persistence in europe. *PLoS computational biology*, 5(8):e1000491, 2009.

[36] A.K. Jain, TC Hsu, R. Freedman, and MC Chang. Demographic aspects of lactation and postpartum amenorrhea. *Demography*, 7(2):255–271, 1970.

[37] HB Kettlewell. D. 1956. further selection experiments on industrial melanism in the lepidoptera. *Heredity*, 10:287–301, 1956.

[38] A. Kiszewski, A. Mellinger, A. Spielman, P. Malaney, S.E. Sachs, and J. Sachs. A global index representing the stability of malaria transmission. *The American journal of tropical medicine and hygiene*, 70(5):486, 2004.

[39] R. La Porta. Lopez-de-silanes, f., shleifer, a., vishny, r., 1997. legal determinants of external finance. *Journal of Finance*, 52(3):1131–50, 1998.

[40] R. La Porta. Lopez-de-silanes, f., shleifer, a., vishny, r., 1998. law and finance. *Journal of Political Economy*, 106(6):1113–1155, 1998.

[41] T.R. Malthus. An essay on the principle of population. 1798. *Reprint. Amherst, NY: Prometheus Books*, 1998.

[42] R.D. McCracken. Lactase deficiency: an example of dietary evolution. *Current Anthropology*, 12(4/5):479–517, 1971.

[43] C. McEvedy and R. Jones. *Atlas of world population history.* Penguin Books Ltd, Harmondsworth, Middlesex, England., 1978.

[44] T. Meloni, C. Colombo, G. Ruggiu, M. Dessena, and GF Meloni. Primary lactase deficiency and past malarial endemicity in sardinia. *Italian journal of gastroenterology and hepatology*, 30(5):490, 1998.

[45] S. Michalopoulos. The origins of ethnolinguistic diversity: Theory and evidence. *Department of Economics, Tufts University*, 2008.

[46] CA Mulcare. The evolution of the lactase persistence phenotype. *London: University of London*, page 311, 2006.

[47] C.A. Mulcare, M.E. Weale, A.L. Jones, B. Connell, D. Zeitlyn, A. Tarekegn, D.M. Swallow, N. Bradman, and M.G. Thomas. The t allele of a single-nucleotide polymorphism 13.9 kb upstream of the lactase gene (lct)(c-13.9 kbt) does not predict or cause the lactase-persistence phenotype in africans. *The American Journal of Human Genetics*, 74(6):1102–1110, 2004.

[48] N. Nunn. The importance of history for economic development, 2009.

[49] N. Nunn and D. Puga. Ruggedness: The blessing of bad geography in africa, 2009.

[50] N. Nunn and N. Qian. The potato's contribution to population and urbanization: Evidence from an historical experiment, 2009.

[51] L. Putterman. Agriculture, diffusion and development: Ripple effects of the neolithic revolution. *Economica*, 75(300):729–748, 2008.

[52] L. Putterman and D.N. Weil. Post-1500 population flows and the long-run determinants of economic growth and inequality*. *Quarterly Journal of Economics*, 125(4):1627–1682, 2010.

[53] A. Sherratt. The secondary exploitation of animals in the old world. *World Archaeology*, 15(1):90–104, 1983.

[54] F.J. Simoons. Primary adult lactose intolerance and the milking habit: a problem in biological and cultural interrelations. *Digestive Diseases and Sciences*, 14(12):819–836, 1969.

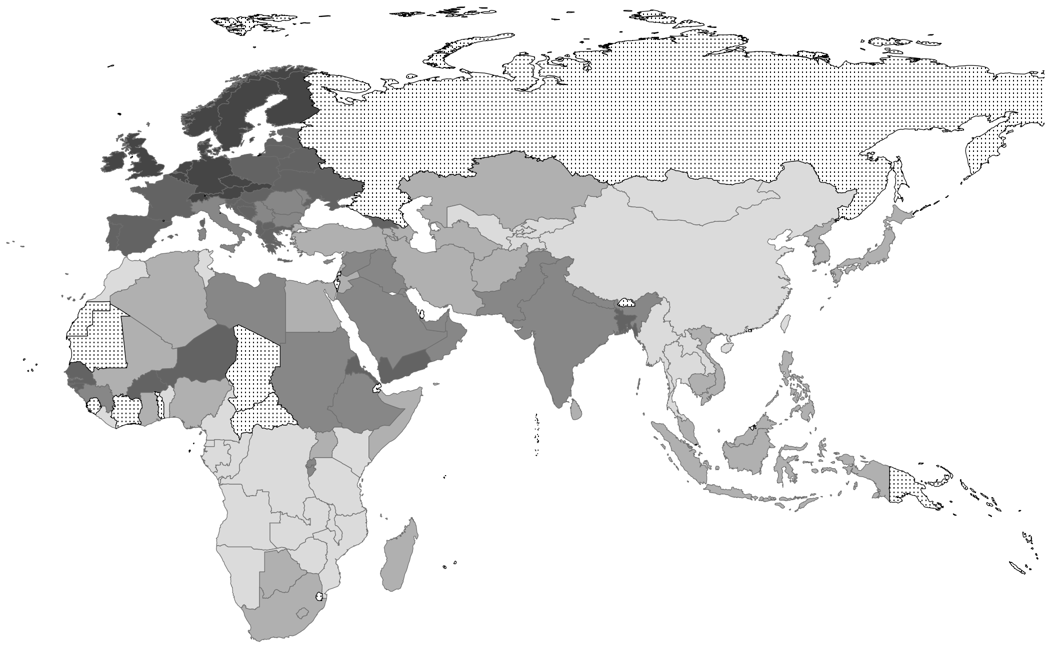
[55] F.J. Simoons. Primary adult lactose intolerance and the milking habit: a problem in biologic and cultural interrelations. *Digestive Diseases and Sciences*, 15(8):695–710, 1970.

[56] Frederick J. Simoons. The geographic hypothesis and lactose malabsorption. *Digestive Diseases and Sciences*, 23:963–980, 1978. 10.1007/BF01263095.

[57] E. Spolaore and R. Wacziarg. The diffusion of development*. *Quarterly Journal of Economics*, 124(2):469–529, 2009.

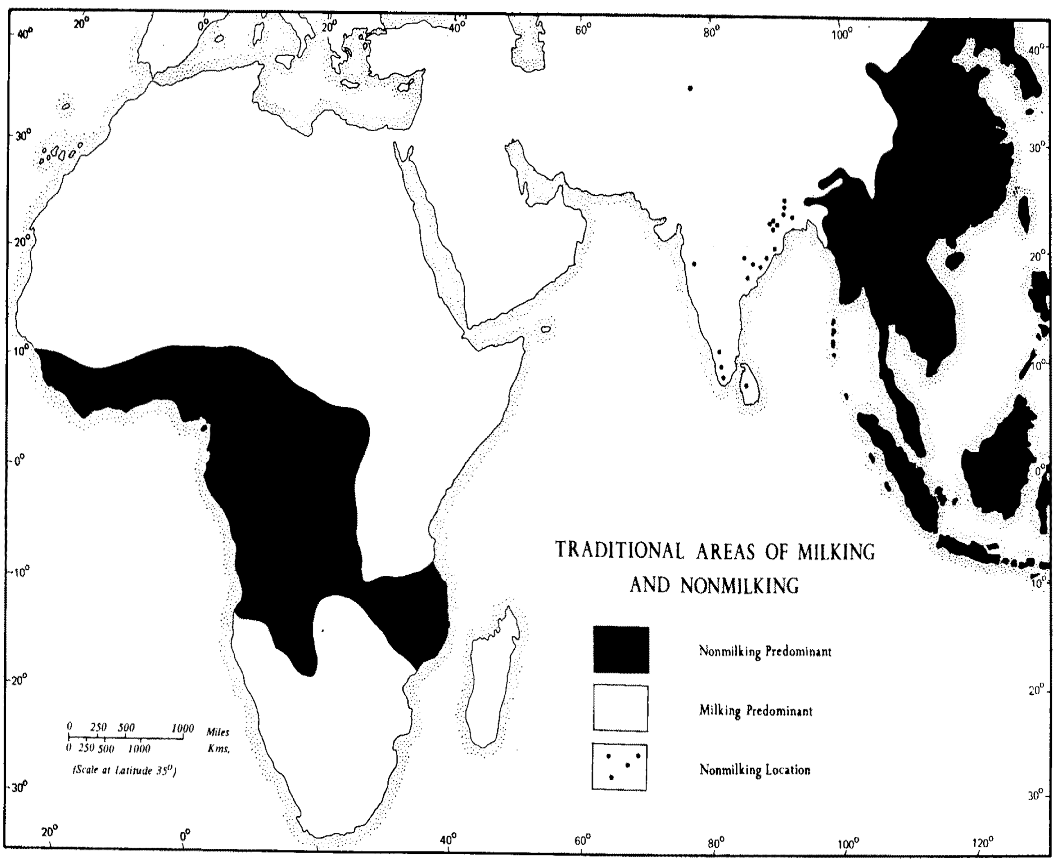
[58] S.A. Tishkoff, F.A. Reed, A. Ranciaro, B.F. Voight, C.C. Babbitt, J.S. Silverman, K. Powell, H.M. Mortensen, J.B. Hirbo, M. Osman, et al. Convergent adaptation of human lactase persistence in africa and europe. *Nature Genetics*, 39(1):31–40, 2006.

5 Figures



Note: Darker areas represent a greater frequency of lactase persistence. Dotted areas represent countries not in the data set. Western European countries are shown to have high levels of lactase persistence, while Sub-Saharan Africa and Southeast Asia have low levels of lactase persistence. This corresponds to the historical levels of milking from Simoons (shown in Fig. 2)

Figure 1
Distribution of Lactase Persistence



Note: Darker areas represent historically non-milking areas. There appears to be a high level of overlap of the historically non-milking areas and areas with low frequencies of lactase persistence shown in Fig. 1.

Figure 2
Historical Milk Consumption (Simoons 1969)

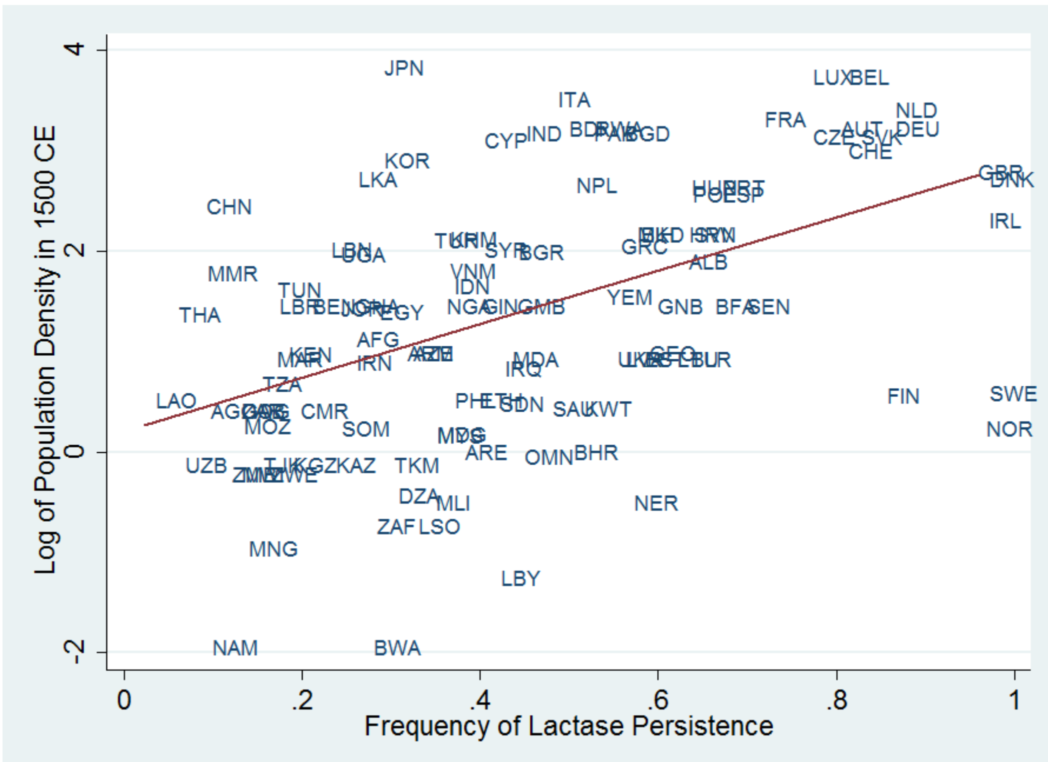


Figure 3
The Freq. of Lactase Persistence and the ln of Pop. Density in 1500 CE

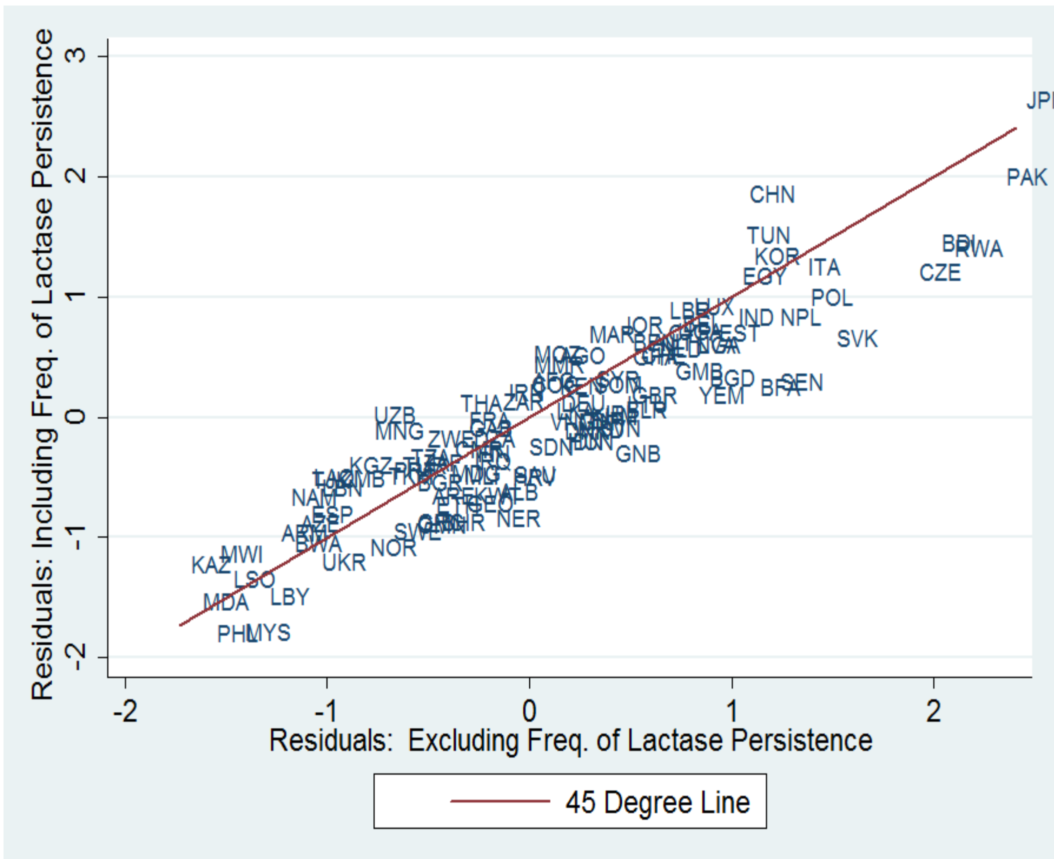


Figure 4
Residuals With and Without the Frequency of Lactase Persistence

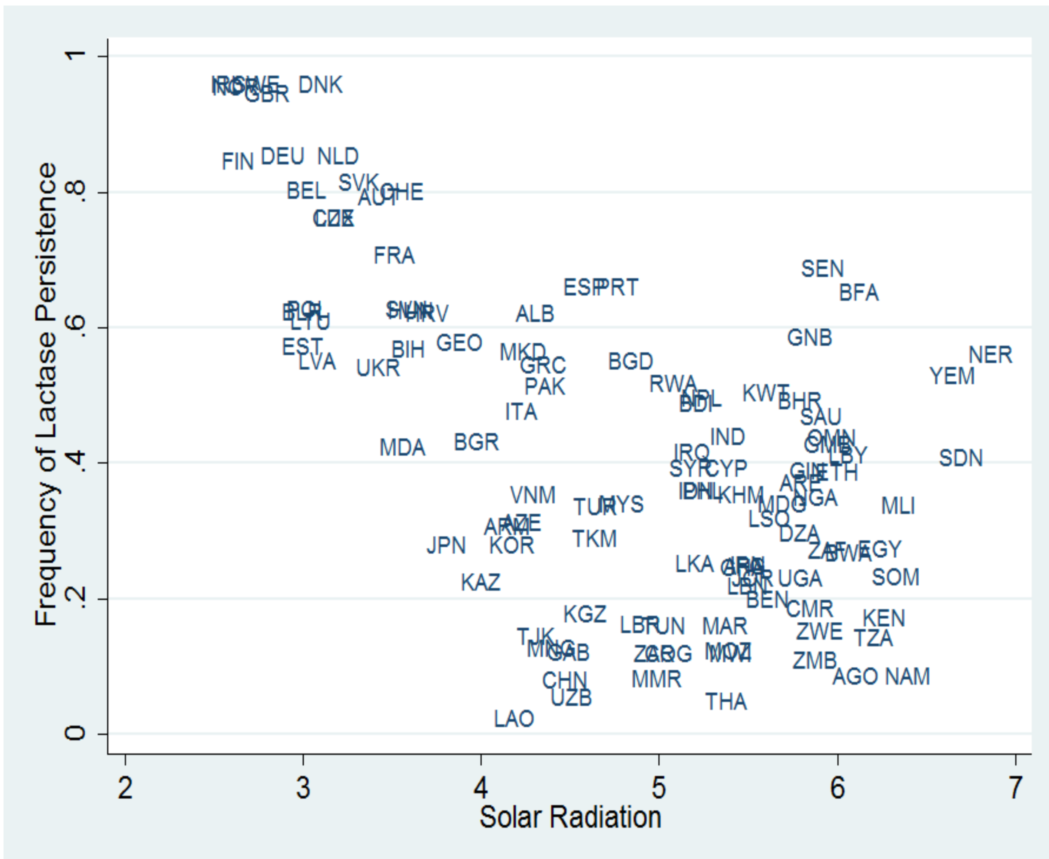


Figure 5
Solar Radiation and the Freq. of Lactase Persistence

6 Tables

Table 1
Summary Statistics

Variable:	N	Mean	Std. Dev.	Min	Max
Lactase Persistence Frequency (Inverse of Migration Matrix)	118	0.4133	0.2389	0.0233	0.96
Conservative Lactase Persistence Freq. (Inverse of Migration Matrix)	51	0.4324	0.2556	0.0486	0.96
Lactase Persistence Frequency (Majority Ethnic Group)	126	0.4069	0.256	0	0.96
Conservative Lactase Persistence Freq. (Majority Ethnic Group)	54	0.422	0.273	0	0.96
ln of Population Density in 1500 CE	126	1.3166	1.3173	-1.9459	4.1477
Millennia of Agriculture	119	5.4496	2.3616	1	10.5
Mean Suitability of Agriculture	122	0.4214	0.257	0.0029	0.9557
Distance from the Equator	124	28.9194	17.4079	0	64
Sub-Saharan Africa (Dummy)	126	0.3016	0.4608	0	1
Western Europe (Dummy)	126	0.1349	0.343	0	1
Genetic Distance from the U.K. in 1500 CE	124	0.869	0.7617	0	2.288
Mean Crop Suitability for Potatoes	113	1.7154	4.623	0.001	35.9686
Mean Crop Suitability for Old World Crops	113	6.4939	10.2131	.001	64.6213
Mean Crop Suitability for New World Crops	113	7.0571	15.3045	.001	116.2154
Solar Radiation ($kWh/m^2/day$)	126	4.5856	1.1135	2.0025	6.670
Elevation (Country Average in km)	118	0.6413	0.5844	0.0092	3.1859
Ruggedness (Country Average)	122	1.1413	1.3149	0.016	6.202
Mean Distance from Coast or River (in km)	118	0.3426	0.441	0.011	2.2917
% of Land within 100 KM of Coast or River	118	0.4536	0.3794	0	1
% of Land within the Tropics	121	24.974	39.0662	0	100
% of Land within a Desert	121	4.232	11.7341	0	77.28
Mean of Malarial Ecology Index	113	3.929	6.7954	0	31.639

Notes: Lactase Persistence Measures calculated from Ingram et al. (2009), Alesina et al. (2003), and Putterman and Weil (2010). Population Density data are given by persons per km^2 and are from McEvedy and Jones (1978). Mean suitability of agriculture is from Michalopolous (2010) and Ramankutty et al. (2002). Distance from the Equator comes from Rodrik et al. (2002). Genetic distance is from Spolaore and Wacziarg (2009). Mean crop suitability for potatoes, New World staples, and Old World staples come from Nunn and Qian (2011). Solar Radiation data come from the Atmospheric Science Data Center at NASA. Elevation, mean distance to a coast or river, % of land within 100 km of a coast or river, and other water access controls are from Gallup et al. (1999). Ruggedness is from Nunn and Puga (2011). The malarial ecology index is from Kiszewski et al. (2004). Genetic distance, crop suitability data from Nunn and Qian, distance to a coast or river, and elevation have all been scaled by 1/1000.

Table 2

Baseline Estimation

Dependent Variable: ln Population Density in 1500 CE								
	Extended Sample							Conservative Sample
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Freq. of Lactase Persistence	2.6548***			2.3295***	2.4492***		2.3432***	3.3804***
(Inverse of Migration Matrix)	(0.4672)			(0.4412)	(0.5115)		(0.5032)	(0.7773)
Millennia of Agriculture		0.2052***		0.1552***		0.1575***	0.1384**	0.0669
		(0.0458)		(0.0419)		(0.0529)	(0.0553)	(0.0843)
Avg. Suitability of Agriculture			2.7788***		2.6708***	2.7623***	2.6610***	2.5390***
			(0.3709)		(0.3533)	(0.3574)	(0.3453)	(0.5278)
Dist. from Equator			-0.0233**		-0.0364***	-0.0216**	-0.0343***	-0.0495***
			(0.0092)		(0.0098)	(0.0085)	(0.0090)	(0.0125)
Sub-Saharan Africa			-1.2218***		-1.2915***	-0.5445	-0.6935*	-1.8644***
			(0.3105)		(0.2927)	(0.3540)	(0.3818)	(0.5979)
Western Europe			1.5220***		0.6472**	1.6345***	0.7839***	0.0142
			(0.2517)		(0.2608)	(0.2637)	(0.2693)	(0.4093)
N	110	110	110	110	110	110	110	49
$R2$	0.2393	0.1395	0.5054	0.3154	0.5951	0.5402	0.6218	0.7054
$Adj.R^2$	0.2322	0.1315	0.4865	0.3026	0.5756	0.5181	0.5998	0.6633
F	32.2903	20.0730	37.1871	27.9543	41.6894	31.9421	41.6716	28.6852

Notes: OLS coefficients are reported in each column. *, **, and *** represent significance at the 10, 5, and 1% significance level, respectively. Robust standard errors are in parentheses. Lactase persistence frequency is the percent of a country's population that is able to digest milk. The extended sample includes the classification of ethnic groups based on language groups. The conservative sample represents direct matches in ethnic groups between Ingram et al. (2009) and Alesina et al. (2003). Country level lactase persistence frequencies are created by using the weighted average of a representative country's ethnic make-up in 1500 CE. Ethnic compositions in 1500 CE are found by the inverse of the Putterman and Weil migration matrix (2010). Millennia of agriculture come from Putterman (2007); separate estimations have been conducted using millennia of agriculture from Hibbs and Olsson (2004). Results are unchanged.

Table 3

Baseline Estimation

Freq. of Lactase Persistence Calculated through Majority Ethnic Group

Dependent Variable: ln Population Density in 1500 CE								
		Extended Sample						Conservative Sample
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Freq. of Lactase Persistence	2.5602***			2.2064***	2.2542***		2.1052***	3.0144***
(Majority Ethnic Group)	(0.4359)			(0.4114)	(0.4670)		(0.4539)	(0.6339)
Millennia of Agriculture		0.2097***		0.1493***		0.1848***	0.1612***	0.1085
		(0.0448)		(0.0408)		(0.0519)	(0.0524)	(0.0804)
Avg. Suitability of Agriculture			2.7249***		2.6318***	2.7387***	2.6500***	2.5457***
			(0.3614)		(0.3401)	(0.3446)	(0.3319)	(0.5077)
Dist. from Equator			-0.0173*		-0.0299***	-0.0175**	-0.0292***	-0.0382***
			(0.0089)		(0.0094)	(0.0080)	(0.0082)	(0.0129)
Sub-Saharan Africa			-1.0846***		-1.0911***	-0.3504	-0.4502	-1.3029**
			(0.3031)		(0.2777)	(0.3279)	(0.3449)	(0.5741)
Western Europe			1.4788***		0.6752***	1.6307***	0.8608***	0.1006
			(0.2506)		(0.2555)	(0.2638)	(0.2621)	(0.3788)
N	115	115	115	115	115	115	115	51
R^2	0.2469	0.1434	0.4834	0.3149	0.5715	0.5323	0.6082	0.6719
$Adj.R^2$	0.2402	0.1358	0.4647	0.3026	0.5518	0.5108	0.5865	0.6272
F	34.4992	21.8897	38.7139	28.3980	42.9698	33.6324	45.0906	31.0984

Notes: OLS coefficients are reported in each column. *, **, and *** represent significance at the 10, 5, and 1% significance level, respectively. Robust standard errors are in parentheses. Lactase persistence frequency is the percent of a country's population that is able to digest milk. Country-level lactase persistence frequencies are created by using the weighted average of a representative countries ethnic make-up in 1500 CE. Ethnic compositions in 1500 CE are simply ascribed to the largest current ethnic group. The extended sample includes the classification of ethnic groups based on language groups. The conservative sample represents direct matches in ethnic groups between Ingram et al. (2009) and Alesina et al. (2003). Millennia of agriculture come from Putterman (2007); separate estimations have been conducted using millennia of agriculture from Hibbs and Olsson (2004). Results are unchanged.

Table 4

Baseline Estimation: Within Continent Estimation

Dependent Variable: ln Population Density in 1500 CE				
	(1) Europe	(2) Africa	(3) Asia	(4) Asia + Africa
Freq. of Lactase Persistence	5.1419***	1.5544*	1.8872*	1.7967***
(Inverse of Migration Matrix)	(1.2427)	(0.8016)	(0.9582)	(0.5621)
Millennia of Agriculture	-0.0598	0.4831***	0.1275	0.2477***
	(0.2138)	(0.1131)	(0.1054)	(0.0500)
Avg. Suitability of Agriculture	2.7234*	3.2814***	2.6630***	2.8191***
	(1.3845)	(0.7878)	(0.4642)	(0.4487)
Dist. from Equator	-0.0805**	-0.0355**	-0.0205	-0.0319***
	(0.0361)	(0.0154)	(0.0164)	(0.0107)
N	32	38	40	78
R^2	0.5677	0.6101	0.4164	0.5184
$Adj.R^2$	0.5036	0.5628	0.3497	0.4920
F	13.7821	18.6864	11.7139	22.6639

Notes: OLS coefficients are reported in each column. *, **, and *** represent significance at the 10, 5, and 1% significance level, respectively. Robust standard errors are in parentheses. The baseline estimation is given by column (7) in Table 2. The sample in column (1) includes only European countries; column (2) includes only Asian countries; column (3) includes only African countries; and column (4) excludes all European countries.

Table 5

Baseline Estimation: Sample Truncations

	Dependent Variable: ln Population Density in 1500 CE						
Truncation:	Western Europe		Sub-Saharan Africa		WE and SSA	Dist. from Equator < Median	Dist. from Equator > Median
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Freq. of Lactase Persistence	2.3083***	2.4093***	2.6221***	1.7678**	1.9008***	2.1299***	2.6713**
(Inverse of Migration Matrix)	(0.5229)	(0.5178)	(0.5431)	(0.6759)	(0.7093)	(0.5895)	(1.1533)
Millennia of Agriculture	0.2061***	0.1346**	0.0976*	0.1120**	0.1027*	0.3792***	-0.0605
	(0.0399)	(0.0571)	(0.0566)	(0.0556)	(0.0572)	(0.0779)	(0.1293)
Avg. Suitability of Agriculture	2.4947***	2.4816***	2.4883***	2.5815***	2.3018***	3.5105***	2.5632***
	(0.3835)	(0.3729)	(0.3970)	(0.3912)	(0.4282)	(0.5048)	(0.6000)
Dist. from Equator	-0.0241***	-0.0331***	-0.0190**	-0.0209**	-0.0187**	-0.0497***	-0.0763**
	(0.0070)	(0.0090)	(0.0094)	(0.0094)	(0.0093)	(0.0146)	(0.0364)
Sub-Saharan Africa		-0.6843*				0.2147	
		(0.3878)				(0.4315)	
Western Europe				0.7959***			0.7327*
				(0.2960)			(0.4125)
N	95	95	77	77	62	54	51
R^2	0.5270	0.5466	0.5088	0.5390	0.4327	0.6920	0.6068
$Adj.R^2$	0.5059	0.5211	0.4815	0.5065	0.3929	0.6599	0.5631
F	29.8312	27.8345	23.0029	24.1589	14.3660	30.1561	18.0813

Notes: OLS coefficients are reported in each column. *, **, and *** represent significance at the 10, 5, and 1% significance level, respectively. Robust standard errors are in parentheses. Columns (1) and (2) drop Western European countries from the sample. Col.'s (3) and (4) drop Sub-Saharan African countries from the sample. Col. (5) drops both Sub-Saharan African and Western European countries from the sample. Col.'s (6) and (7) respectively drop data points below and above the median sample distance from the equator. The median distance from the equator corresponds to a latitude of 33 degrees, which corresponds to an area slightly above North African states, roughly equal to the Levant, and above India and Southeast Asia.

Table 6

Baseline Estimation: Truncations Due to Migration

	Dependent Variable: ln Population Density in 1500 CE				
% of Population Derived from 1500 CE Population:	(1) >50%	(2) >75%	(3) >85%	(4) >95%	(5)† >95%
Freq. of Lactase Persistence	2.3161***	2.5620***	2.3674***	2.2323***	2.0184***
	(0.5126)	(0.5254)	(0.5385)	(0.6830)	(0.6094)
Millennia of Agriculture	0.1422**	0.1279**	0.1284**	0.1022	0.1689
	(0.0560)	(0.0554)	(0.0617)	(0.1203)	(0.1046)
Avg. Suit of Agr.	2.5805***	2.4815***	2.7223***	2.7665***	2.7454***
	(0.3589)	(0.3862)	(0.4001)	(0.5076)	(0.4782)
Dist. from Equator	-0.0336***	-0.0404***	-0.0372***	-0.0401***	-0.0286**
	(0.0091)	(0.0085)	(0.0089)	(0.0147)	(0.0121)
Sub-Saharan Africa	-0.6580*	-0.8699**	-0.8086**	-0.9785	-0.3672
	(0.3825)	(0.3666)	(0.3760)	(0.7595)	(0.5751)
Western Europe	0.7754***	0.7016**	0.7155**	0.6666	0.7447*
	(0.2700)	(0.2775)	(0.2832)	(0.4212)	(0.4179)
N	108	100	93	58	63
R^2	0.6021	0.6114	0.6203	0.5929	0.5582
$Adj.R^2$	0.5784	0.5863	0.5938	0.5451	0.5108
F	37.6166	34.7570	32.6373	16.7502	18.9898

Notes: OLS coefficients are reported in each column. *, **, and *** represent significance at the 10, 5, and 1% significance level, respectively. Robust standard errors are in parentheses. The baseline estimation is given by column (7) in Table 2. The sample truncations are based on the percent of a countries 2000 CE population that is derived from the 1500 CE population (Putterman and Weil 2010).

† The frequency of lactase persistence in column (5) is calculated by the majority ethnic group.

Table 7

Additional Genetic Control

Dependent Variable: ln Population Density in 1500 CE					
	(1)	(2)	(3)	(4)	(5)
Freq. of Lactase Persistence	2.6445***		1.9352***		2.3251***
(Inverse of Migration Matrix)	(0.4684)		(0.5581)		(0.5482)
Genetic Dist. from U.K. in 1500 CE		-0.7567***	-0.4215**	-0.4048	-0.0243
(Spolaore and Wacziarg 2009)		(0.1564)	(0.1841)	(0.2644)	(0.2744)
Millennia of Agriculture				0.1112*	0.1358**
				(0.0569)	(0.0605)
Avg. Suitability of Agriculture				2.7934***	2.6651***
				(0.3607)	(0.3490)
Dist. from Equator				-0.0229***	-0.0343***
				(0.0081)	(0.0091)
Sub-Saharan Africa				-0.2433	-0.6741
				(0.4193)	(0.4730)
Western Europe				1.4404***	0.7777***
				(0.2898)	(0.2829)
N	109	109	109	109	109
R^2	0.2375	0.1889	0.2790	0.5506	0.6204
$Adj.R^2$	0.2303	0.1813	0.2654	0.5242	0.5941
F	31.8693	23.4171	20.9875	28.1937	35.1562

Notes: OLS coefficients are reported in each column. *, **, and *** represent significance at the 10, 5, and 1% significance level, respectively. Robust standard errors are in parentheses. Genetic distance from the United Kingdom is measured by differences in neutral genetic distances across peoples. The inclusion of this variable is intended to capture genetic capital differences that may affect population density.

Table 8

Additional Soil Suitability Measures

	Dependent Variable: ln Population Density in 1500 CE				
	(1)	(2)	(3)	(4)	(5)
Freq. of Lactase Persistence	2.9583***	3.1373***	2.9832***	3.1337***	3.1172***
	(0.5510)	(0.5371)	(0.5518)	(0.5427)	(0.5362)
Millennia of Agriculture	0.1240*	0.1247**	0.1247*	0.1242**	0.1205*
	(0.0633)	(0.0618)	(0.0632)	(0.0622)	(0.0626)
Dist. from Equator	-0.0388***	-0.0333***	-0.0320***	-0.0332***	-0.0368***
	(0.0107)	(0.0099)	(0.0100)	(0.0099)	(0.0107)
Sub-Saharan Africa	-0.8981*	-0.9524**	-0.9195**	-0.9555**	-0.9675**
	(0.4622)	(0.4459)	(0.4517)	(0.4481)	(0.4545)
Western Europe	0.3360	0.2070	0.3124	0.2118	0.2419
	(0.3380)	(0.3400)	(0.3364)	(0.3496)	(0.3409)
Suitability of Potatoes	0.0560***				0.0311
	(0.0161)				(0.0214)
Suitability of Old World Crops		0.0266***		0.0249*	0.0180
		(0.0100)		(0.0148)	(0.0111)
Suitability of New World Crops			0.0145***	0.0014	
			(0.0044)	(0.0060)	
N	106	106	106	106	106
R^2	0.4224	0.4272	0.4140	0.4273	0.4342
$Adj.R^2$	0.3874	0.3925	0.3785	0.3864	0.3938
F	15.9889	15.9316	14.7533	13.7762	14.4565

Notes: OLS coefficients are reported in each column. *, **, and *** represent significance at the 10, 5, and 1% significance level, respectively. Robust standard errors are in parentheses. The baseline soil suitability measure (Michalopoulos 2009) is not included. Inclusion of the baseline soil suitability measure in the above table does not change the magnitude or significance of the coefficient on the frequency of lactase persistence.

Table 9
Additional Environment, Disease, and Cultural Controls

	Dependent Variable: ln Population Density in 1500 CE							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Freq. of Lactase Persistence	2.4417***	2.4726***	2.3919***	2.4186***	2.3954***	2.3811***	2.5181***	2.5007***
	(0.5825)	(0.5229)	(0.5007)	(0.5015)	(0.4952)	(0.5644)	(0.5167)	(0.6340)
Millennia of Agriculture	0.1390**	0.1273**	0.1583**	0.1510***	0.1604**	0.1383**	0.1303**	0.1409**
	(0.0554)	(0.0573)	(0.0645)	(0.0566)	(0.0643)	(0.0554)	(0.0554)	(0.0624)
Avg. Suitability of Agriculture	2.5782***	2.5219***	2.4382***	2.1557***	2.1063***	2.5962***	2.4377***	1.8259***
	(0.3598)	(0.3692)	(0.4079)	(0.4075)	(0.4409)	(0.3593)	(0.3667)	(0.4656)
Dist. from Equator	-0.0343***	-0.0321***	-0.0264**	-0.0348***	-0.0306***	-0.0332***	-0.0334***	-0.0219*
	(0.0094)	(0.0092)	(0.0112)	(0.0093)	(0.0113)	(0.0099)	(0.0091)	(0.0117)
Sub-Saharan Africa	-0.7211*	-0.6454*	-0.6156	-0.7657*	-0.7080*	-0.7967*	-0.7232*	-0.6759
	(0.3850)	(0.3866)	(0.4201)	(0.3942)	(0.4226)	(0.4077)	(0.3892)	(0.4326)
Western Europe	0.6424**	0.6216**	0.6009**	0.5502**	0.5333**	0.6572**	0.2735	0.0140
	(0.2706)	(0.2808)	(0.2600)	(0.2607)	(0.2576)	(0.2756)	(0.3488)	(0.3750)
Mean Elevation	0.0000							0.0000
	(0.0002)							(0.0003)
Mean Ruggedness		0.2633						0.2928
		(0.2241)						(0.2583)
Mean Ruggedness^2		-0.0557						-0.0617
		(0.0419)						(0.0407)
% in Tropics			0.0037		0.0019			0.0038
			(0.0037)		(0.0036)			(0.0046)
% in Deserts				-0.0160***	-0.0151**			-0.0143**
				(0.0059)	(0.0059)			(0.0066)
Malaria Ecology Index						0.0092		0.0103
						(0.0173)		(0.0157)
Member of the Roman Empire							0.6368**	0.8379**
							(0.3093)	(0.3244)
N	104	104	104	104	104	104	104	104
R^2	0.6169	0.6245	0.6204	0.6328	0.6337	0.6180	0.6263	0.6567
$Adj.R^2$	0.5890	0.5928	0.5927	0.6060	0.6029	0.5902	0.5991	0.6071
F	32.9801	27.2451	37.0206	35.0955	32.3632	33.6755	34.1439	19.9392

Standard errors in parentheses
* $p < .1$, ** $p < .05$, *** $p < .01$

Notes: OLS coefficients are reported in each column. *, **, and *** represent significance at the 10, 5, and 1% significance level, respectively. Robust standard errors are in parentheses. Tropics represent both a lush environment and a potential control for diseases, which may have affected cattle distributions. Deserts represent an extreme environment, in which reliance on milk may be necessary for survival. Roman occupation may have instilled both favorable institutions and an intogression of lactase persistent alleles.

Table 10

Inclusion of Water Access Controls

	Dependent Variable: ln Population Density in 1500 CE					
	(1)	(2)	(3)	(4)	(5)	(6)
Freq. of Lactase Persistence	2.3232***	2.2047***	2.1396***	2.3897***	2.1250***	2.0588***
	(0.5850)	(0.5412)	(0.6291)	(0.5435)	(0.6462)	(0.6842)
Millennia of Agriculture	0.1420**	0.1524**	0.1494***	0.1408***	0.1521***	0.1535***
	(0.0544)	(0.0592)	(0.0540)	(0.0529)	(0.0543)	(0.0539)
Mean Suitability of Agriculture	2.5576***	2.3422***	2.4659***	2.5943***	2.4014***	2.3918***
	(0.3734)	(0.4393)	(0.3769)	(0.3543)	(0.3888)	(0.3964)
Dist. from Equator	-0.0338***	-0.0374***	-0.0324***	-0.0349***	-0.0337***	-0.0324***
	(0.0102)	(0.0089)	(0.0101)	(0.0092)	(0.0094)	(0.0102)
Sub-Saharan Africa	-0.6821*	-0.7313*	-0.6090	-0.7079**	-0.5786	-0.5615
	(0.3804)	(0.3831)	(0.3818)	(0.3531)	(0.3703)	(0.3699)
Western Europe	0.7470***	0.7485***	0.7659***	0.7569***	0.7812***	0.7767***
	(0.2731)	(0.2648)	(0.2703)	(0.2730)	(0.2691)	(0.2712)
Mean Dist. to Coast	-0.0897					
	(0.2259)					
Mean Dist. to River		-0.2066				
		(0.1620)				
Mean Dist. to Coast or River			-0.2618			-0.1878
			(0.2143)			(0.2298)
% within 100 Km of Coast				0.0191		
				(0.3389)		
% within 100 Km of Coast or River					0.3373	0.1906
					(0.3416)	(0.3704)
N	109	109	109	109	109	109
R^2	0.6232	0.6311	0.6281	0.6226	0.6272	0.6292
$Adj.R^2$	0.5971	0.6056	0.6024	0.5964	0.6013	0.5995
F	36.0706	37.0203	38.4497	35.3034	36.5073	34.0285

Notes: OLS coefficients are reported in each column. *, **, and *** represent significance at the 10, 5, and 1% significance level, respectively. Robust standard errors are in parentheses. Additional water access controls are included into the baseline estimation. Easier access to water is intended to represent an ease to trade, which may affect 1500 CE population densities.

Table 11

Additional Biogeographic Controls

Dependent Variable: ln Population Density in 1500 CE						
	(1)	(2)	(3)	(4)	(5)	(6)
Freq. of Lactase Persistence	2.3929***	2.4414***	2.3673***	2.4257***	2.5728***	2.6226***
	(0.6166)	(0.6529)	(0.5985)	(0.6064)	(0.6284)	(0.6002)
Millennia of Agriculture	0.2327***	0.1792**	0.2254***	0.2557***	0.1896**	0.2178**
	(0.0747)	(0.0896)	(0.0764)	(0.0690)	(0.0858)	(0.0917)
Avg. Suitability of Agriculture	2.3611***	2.3424***	2.3578***	2.4031***	2.3492***	2.3923***
	(0.5056)	(0.4909)	(0.5075)	(0.5068)	(0.4997)	(0.5073)
Dist. from Equator	-0.0279**	-0.0405**	-0.0297*	-0.0266**	-0.0376*	-0.0350*
	(0.0128)	(0.0183)	(0.0161)	(0.0127)	(0.0190)	(0.0197)
Sub-Saharan Africa	-0.6883	0.4149	-0.6793	-0.8716	0.8204	0.6231
	(0.5162)	(0.8054)	(0.5279)	(0.6088)	(0.8371)	(1.1978)
Western Europe	0.3088	0.2363	0.2978	0.3003	0.2556	0.2563
	(0.3168)	(0.3312)	(0.3251)	(0.3257)	(0.3298)	(0.3335)
Number of Potential Domesticate Animals		0.2008			0.2818*	0.2793
		(0.1363)			(0.1677)	(0.2031)
Number of Potential Domesticate Plants			0.0044		-0.0190	-0.0222
			(0.0196)		(0.0238)	(0.0250)
East-West Orientation				-0.0003		-0.0003
				(0.0003)		(0.0003)
N	70	70	70	70	70	70
R^2	0.6718	0.6878	0.6722	0.6790	0.6926	0.6990
$Adj.R^2$	0.6406	0.6526	0.6352	0.6427	0.6523	0.6539
F	31.3909	28.7499	27.2766	27.6883	26.4689	25.0621

Notes: OLS coefficients are reported in each column. *, **, and *** represent significance at the 10, 5, and 1% significance level, respectively. Robust standard errors are in parentheses. Additional biogeographic controls—Number of Domesticable Animals, Number of Domesticable Plants, and East-West Orientation—are from Hibbs and Olsson (2004).

Table 12
All Controls

	Dependent Variable: ln Population Density in 1500 CE					
		Extended Sample				Conservative Sample
	(1)	(2)	(3)	(4)	(5)	(6)
Freq. of Lactase Persistence	2.8383***	2.2420***	2.6170***	2.7210***	3.0125***	3.9665***
	(0.6294)	(0.6895)	(0.7050)	(0.7036)	(0.7041)	(1.0285)
Controls:						
Baseline	Y	Y	Y	Y	Y	Y
Genetic	Y	Y	N	Y	Y	Y
Environment	Y	N	Y	Y	Y	Y
Water Access	N	Y	Y	Y	Y	Y
Biogeographic	N	N	N	N	Y	N
N	103	103	103	103	68	48
R^2	0.6765	0.6204	0.6789	0.6810	0.7387	0.7806
$Adj.R^2$	0.6250	0.5836	0.6236	0.6217	0.6353	0.6674
F	16.2606	27.9260	15.5175	14.9061	10.1204	11.9294

Notes: OLS coefficients are reported in each column. *, **, and *** represent significance at the 10, 5, and 1% significance level, respectively. Robust standard errors are in parentheses. Baseline controls include the millennia of agriculture, the average suitability of agriculture, the distance from the equator, and dummies for Sub-Saharan Africa and Western Europe. The additional genetic control is the genetic distance to the United Kingdom. Environmental controls include average elevation, average ruggedness and its square, the percent of a country within a desert or the tropics, a measure for the suitability of malaria, and whether or not a country belonged to the Roman Empire. Water access controls include the distance from the coast or a navigable river and the percent of a country within 100 kilometers from a coast or river. Biogeographic controls include the number of potential domesticate plants and animals and a measure for the country's East-West orientation. Inclusion of differing soil suitability measures does not change the magnitude or significance of the coefficient on the frequency of lactase persistence.

Table 13

Baseline Estimations: Instrumental Variables

Dependent Variable: ln Population Density in 1500 CE					
		Extended Sample			Conservative Sample
	(1)	(2)	(3)	(4)	(5)
			First Stage Estimates:		
Solar Radiation	-0.9289***	-1.0229***	-0.7828***	-0.8586***	-0.8636***
	(0.1042)	(0.0977)	(0.1177)	(0.1137)	(0.1569)
Solar Radiation^2	0.0902***	0.1022***	0.0832***	0.0894***	0.0818***
	(0.012)	(0.0113)	(0.0134)	(0.0129)	(0.0181)
First Stage F Statistic	113.39	129.19	23.92	31.13	21.67
			Second Stage Estimates:		
Freq. of Lactase Persistence	1.9345***	1.5988***	1.3068	2.2376***	4.2182***
	(0.6525)	(0.6041)	(0.9899)	(0.8502)	(0.9278)
Millennia of Agriculture		0.1709***		0.1392**	0.0535
		(0.0437)		(0.0556)	(0.0875)
Avg. Suitability of Agriculture			2.7212***	2.6656***	2.4884***
			(0.3497)	(0.3423)	(0.5442)
Dist. from Equator			-0.0303***	-0.0337***	-0.0529***
			(0.0107)	(0.0091)	(0.0126)
Sub-Saharan Africa			-1.2590***	-0.6868*	-2.0390***
			(0.3003)	(0.3839)	(0.6312)
Western Europe			1.0552**	0.8222**	-0.3643
			(0.4064)	(0.3888)	(0.4953)
N	110	110	110	110	49
R^2	0.2216	0.2981	0.5756	0.6216	0.6966
$Adj.R^2$	0.2144	0.2850	0.5552	0.5996	0.6532
F	8.7903	16.2761	35.7926	36.3610	29.4023

Notes: IV coefficients are reported in each column. *, **, and *** represent significance at the 10, 5, and 1% significance level, respectively. Robust standard errors are in parentheses. Table 13 is a re-estimates Table 2 with solar radiation and its square as an instrument for the frequency of lactase persistence.

Table 14

Additional Controls: IV Estimates

	Dependent Variable: ln Population Density in 1500 CE				
	(1)	(2)	(3)	(4)	(5)
			First Stage Estimates:		
Solar Radiation	-0.7717***	-0.7781***	-0.6521***	-0.7463***	-0.6736***
	(0.1206)	(0.1354)	(0.1258)	(0.1601)	(0.1492)
Solar Radiation^2	0.079***	0.0814***	0.0674***	0.0742***	0.0696***
	(0.0136)	(0.0151)	(0.0141)	(0.0194)	(0.0162)
First Stage F Statistic	22.83	16.62	14.08	18.62	10.25
			Second Stage Estimates:		
Freq. of Lactase Persistence	2.4261**	2.4606**	1.8378	2.3365**	2.4090*
	(0.9445)	(1.1148)	(1.1590)	(1.0749)	(1.3881)
Controls:					
Baseline	Y	Y	Y	Y	Y
Genetic	Y	N	N	N	Y
Environmental	N	Y	N	N	Y
Water Access	N	N	Y	N	Y
Biogeographic	N	N	N	Y	N
N	103	103	103	68	103
R^2	0.6154	0.6554	0.6185	0.7046	0.6600
$Adj.R^2$	0.5870	0.6051	0.5860	0.6588	0.5968
F	28.9754	17.6517	27.0862	19.1196	14.5441

Notes: IV coefficients are reported in each column. *, **, and *** represent significance at the 10, 5, and 1% significance level, respectively. Robust standard errors are in parentheses. Baseline controls include the millennia of agriculture, the average suitability of agriculture, the distance from the equator, and dummies for Sub-Saharan Africa and Western Europe. The additional genetic control is the genetic distance to the United Kingdom. Environmental controls include average elevation, average ruggedness and its square, the percent of a country within a desert or the tropics, a measure for the suitability of malaria, and whether or not a country belonged to the Roman Empire. Water access controls include the distance from the coast or a navigable river and the percent of a country within 100 kilometers from a coast or river. Biogeographic controls include the number of potential domesticate plants and animals and a measure for the country's East-West orientation.