The Role of Lactase Persistence in Precolonial Development*

C. Justin Cook†

September 4, 2014

Abstract

This paper argues that a genetic adaptation to the Neolithic Revolution led to differential levels of development in the precolonial era. The ability to digest milk, or to be lactase persistent, is conferred by a gene variant that is unequally distributed across the Old World. Milk provided qualitative and quantitative advantages to the diet that 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 lactase persistence is positively and significantly related to population density in 1500 CE; specifically, a one standard deviation increase in the frequency of lactase persistent individuals (roughly 24 percentage points) is associated with roughly a 40 percent increase in precolonial population density. This relationship is robust to a large number of sample specifications and potentially omitted variables.

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 Quamrul Ashraf, James Feyrer, Oded Galor, David Weil, participants at the 2013 Deep Determinants of International Comparative Development Conference at Brown University, participants at the 2011 Integrating Genetics and the Social Sciences Conference at the University of Colorado, participants at the LSU 3rd year paper presentation, and three anonymous referees for helpful comments and suggestions; and to Stelios Michalopolous for sharing data on agricultural suitability and Anastasia Litina for suggesting the use of pastoral suitability. All errors and omissions are my own.

†Assistant Professor, School of Social Sciences, Humanities, and Arts, University of California-Merced, United States; email: jcook9@ucmerced.edu. The majority of this work was completed as part of my dissertation while at Louisiana State University.

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 precolonial 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 economic development in the precolonial era.

The Neolithic Revolution changed the environment in which humans lived.[1] Furthermore, this change occurred at different times for different peoples; implying, certain groups have had a longer time to adapt to the new environment. In the words of Clark (2008, P. 6; Galor and Moav 2002), "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 agricultural lifestyle is the ability to consume milk, or to be lactase persistent.[2] Milk is an additional resource that some could consume, while others could not. In the Malthusian economy of the precolonial era, this variation in the consumption of milk is hypothesized to be associated with the productivity of workers. In accordance with the Malthusian theory of Ashraf and Galor (2011), the increased productivity of lactase persistent workers led to a temporary increase in income and a permanent increase in populations. I therefore seek to explain differences in population density in 1500 CE as a function of a country's frequency of lactase persistence.

Frequencies of lactase persistence are available by ethnicity for the second half of the twentieth century (Ingram et al. 2009a). A central assumption is that these frequencies have not changed much over the past 500 years.[3] However, it is unlikely that country-level ethnic compositions are unchanged since 1500 CE; therefore, a measure of ethnic compositions for 1500 CE is needed. Through the use of migration data for the past 500 years, I am able to back out ethnic compositions for the year 1500 CE. And assuming gene frequencies have remained relatively stable over this period, this allows for the creation a country-level measure for the frequency of lactase persistence in 1500 CE.[4]

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

[2]As explained in Sec. 1.3, lactase persistence is equivalent to lactose tolerance.

[3]Section 1.3 and the supplemental appendix give more discussion to the validity of this assumption.

[4]This is discussed with greater detail in Sec. 2.1.1. In order to confirm my results, I also use a second, cruder strategy of assigning majority ethnic groups to represent countries in the 1500 CE. This strategy is pursued in similar research–i.e.,

The constructed measure of lactase persistence has a positive and significant association with population density in 1500 CE. My baseline estimate states that a one standard deviation increase in the fraction of lactase persistent individuals within a country is associated with roughly a 40% increase in population density. This relationship is robust to the inclusion of a number of valid controls as well as differing sample specifications. In particular, the effect of lactase persistence is not driven by the overarching advantages of an earlier transition to agriculture, which have been documented extensively (see e.g., Hibbs and Olson 2004 and Putterman 2008).

1.1 Previous Literature

A number of papers have established an empirical link between the past and current economic events (for review, see Nunn 2009).[5] The current work seeks to build upon this research by contributing to the explanation of historical differences in development.

One of the most comprehensive works in explaining precolonial populations, and therefore, precolonial 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 large populations, the development of cities and states, the specialization of labor, and, ultimately, a head start in the acquisition of material prosperity. Diamond's hypothesis is tested by Putterman (2008) 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 archaeological 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.

Instead of archaeological evidence or environmental estimates, I 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. In other words, the use of genetic variations provide a quantitative measure of historic differences that may be used to measure the usage or availability of

Spolaore and Wacziarg (2009). The correlation between the two measures of the frequency of lactase persistence is 0.98. Estimation with the alternative measure is found in the supplemental appendix.

[5] See, e.g., Acemoglu et al. 2001; Bockstette et al. 2002; Chanda and Putterman 2007; Comin et al. 2010; Engerman and Sokoloff 1997, 2002; La Porta et al. 1999; Nunn 2008.

a cultural or environmental advantage that is conferred to some societies and not others.

A number of recent papers explore the effect that genetics may have on aggregate economic outcomes (see, e.g., Ashraf and Galor 2013; Spolaore and Wacziarg 2009). 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 by the use of a particular gene variant, not differences in the general genetic make-up of a population. The current work uses variation in an expressed genetic trait which has been naturally selected for since the Neolithic Revolution. To my knowledge, this is the first paper to explore the effect of a single genetic adaptation on aggregate economic conditions.

A similar work by Nunn and Qian (2011) explores how the introduction of the potato to the Old World has affected populations in the 18th and 19th centuries. 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 Qian use soil conditions, whereas I use the observed differences of an underlying genetic variation.

The role of natural selection since the Neolithic Revolution has also been studied in recent research. Theoretically, Galor and Moav (2002) establish a unified model that captures the evolution of both man and economic outcomes, while Galor and Michalopolous (2012) show the selection for particular traits since the beginning of agriculture. Empirically, Galor and Moav (2007) 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 (2007) implies that differences have developed since the Neolithic Revolution and that these differences may be correlated with differing economic outcomes.

1.2 Population Advantages of Milk Consumption

The first direct evidence of milk consumption is from Northwest Anatolia in 7,000 BCE (Evershed et al. 2008). The practice of dairying diffused to Eastern Europe by 6,000 BCE and to Britain by 4,000 BCE (Craig et al. 2005; Copley et al. 2003). Outside of Eurasia, the first evidence of dairying occurred roughly 7,000 years ago in North African grasslands, which are presently found in the Saharan Desert (Dunne et al. 2012).

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).[6] 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. Along with quality advantages to diet, 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.

In terms of the Malthusian theory of Ashraf and Galor (2011), workers able to consume milk were able to produce greater levels of output (e.g., milk, cheese, etc.). The increased level of output associated with lactase persistent workers led to a temporary wage increase for these workers. This increase in wage for a segment of the population resulted in a greater demand for offspring amongst this segment, which ultimately increased the population while keeping output per worker, or income, constant. This had two effects: First, this provides a theoretical relationship between lactase persistence and population density in the precolonial era. Second, this is the basic premise for the natural selection of lactase persistence.

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 precolonial 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, Nunn and Qian state (2011, 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."[7] Considering two societies with equal

[6]For simplicity I reference milk to be from cattle.

[7]This idea from Nunn and Qian is supported by a companion paper that shows the introduction of the potato following the Columbian Exchange had larger effects in countries with large frequencies of lactase persistence (Cook *forthcoming*).

resources, the society that is able to digest milk gains a qualitative dietary advantage that improves health and therefore population.

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; WHO 1998). The use of animal milk as a substitute for mother's milk reduces weaning time, thereby shortening the postpartum infertility period.[8] 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, providing another mechanism for the hypothesized positive relationship between dairying and populations.

1.3 Selection for Lactase Persistence

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.[9] If this 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.[10] Certain populations of humans, however, have acquired an allele, or gene variant, that allows for the production of lactase throughout their adult lives; this is known as lactase persistence.[11] 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. 2009a).

The Neolithic Revolution radically changed the environment for humans, and this change 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) show that the allele, or gene variant, that allows for lactase persistence in Europeans is absent, or rare, in early Neolithic Europeans. Additionally, Plantinga et al. (2012) explore the frequency of the gene variant associated with lactase persistence amongst Late Neolithic (5,000 to 4,500 YBP) populations of southwest Europe, finding a frequency well below modern levels. Considering that Europeans have the highest levels of lactase persistence in the world, the findings of

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

[9] Lactose is found in all milk.

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

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

Burger et al. and Plantinga et al. imply 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. This advantage is estimated in Bersaglieri et al. (2004) who 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 produce lactase.[12] Furthermore, the gene variant that confers lactase persistence is the "textbook" example of a selective sweep (Nielsen et al. 2005; Ingram et al. 2009a).[13]

If no cattle were available, and therefore no milk, then no advantage is had by producing lactase past the weaning period. Therefore, 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 (1970). 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).

One concern in establishing the relationship between the frequency of lactase persistence and pre-colonial development is that 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 Northern Europeans, yet these areas have significantly lower levels of lactase

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

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

persistence (Simoons 1978). Differences in dairying also have a cultural significance. The use of lactase persistence frequencies intends to measure the advantages of dairying, not animal husbandry.

A further, and potentially more problematic, concern is reverse causality. As will be explained in Section 2, the data for lactase persistence are for contemporary populations. The use of historic population density and contemporary frequencies of lactase persistence provides for the possibility that higher population density *led* to an increase in the frequency of lactase persistence. One mechanism for this potential reverse-causal relationship is tied to the role of infectious disease, which is more prevalent in dense populations. Holding constant relevant factors, population size remains constant during the Malthusian era. Infectious disease, which has a positive association with population density, leads to an increase in the mortality rate, and given the consistency of the population, this increase in mortality will also be associated with an increase in the birth rate. This process effectively leads to a greater turnover of the population within a country, thereby accelerating the selection of favorable traits such as lactase persistence. Therefore, population density in 1500 CE could potentially lead to differential selection of lactase persistence and in the process create a positive statistical relationship between the two variables that is independent of the hypothesized increase in carrying capacity from dairying. The potential for this source of bias, however, is unlikely given the relatively short period of 500 years, or 20 generations. From simple simulations it can be shown that the majority of the selection for the lactase persistence trait occurs shortly after the adoption of dairying, and that any changes in the last 500 years are small. Furthermore, other traits–lessened skin pigmentation and blue eyes– that have no hypothesized relationship to a country's carrying capacity and have also have undergone recent selection, show no positive relationship with population density, providing little support the acceleration of selection from population turnover due to increased infectious disease loads. However, given the concern of reverse causality and the possibility of omitted variables tied to the cultural adaptation of dairying, a strict interpretation of lactase persistence *causing* differences in population density should be avoided.[14]

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. 2009b; Tishkoff et al. 2006). Furthermore, the frequency of a particular gene variant is ethnic specific. In other words, the gene variant that allows for milk consumption in Northern Europeans is not identical to the allele

[14] The supplemental appendix provides a fuller discussion of the potential for reverse causality.

that allows for milk consumption in Western Africans.[15] It is for this reason that the observed, or phenotypic, ability to consume milk is the primary determinant of the measure of lactase persistence.[16]

The data for the frequencies of lactase persistence come from Ingram et al. (2009a), from which the authors aggregate data from past studies of lactase persistence frequencies for differing ethnicities. The lactase persistence frequencies are obtained by conducting ethnic specific lactose tolerance tests. 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 are assumed to have 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. (2009a):

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

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, I need to aggregate ethnic groups into countries. And secondly, I need ethnic compositions representative of the precolonial era, not for contemporary periods.

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 ethnic compositions for 190 countries from roughly 1990 to 1995. Using ethnolinguistic classifications, ethnic groups, which have lactase persistence frequencies from Ingram et al. (2009a; hereafter Ingram), have been matched to ethnic groups in Alesina et al. (Lewis 2009). For example, "Western Europeans" in Sweden from Alesina et al. are assigned the lactase persistence frequency of "Dane" from Ingram et

[15] 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 significant 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 gene's 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.

[16] A phenotype is the physical expression of a genotype (Hartl and Clark 2006).

al., "Filipinos" in Alesina et al. are assigned to the "Maori" ethnic group in Ingram et al., and the "Fon" people from Benin are assigned to "Yoruba" from Nigeria.[17] This matching yields data for 118 Old World countries (i.e., Europe, Asia, and Africa), of which 49 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 49 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.[18]

The aggregation from ethnic groups to countries gives a cross-country measure for the frequency of lactase persistence. This measure, however, is for the present period and may not be relevant in the prediction of variables in the precolonial period. A cross-country measure for lactase persistence 500 years in the past is needed. The 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).[19] 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 I 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:

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

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 CE China was entirely composed of the

[17] 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.

[18] See the supplemental appendix for all estimations involving the reduced, conservative sample.

[19] The supplemental appendix explores an alternative method of establishing ethnic compositions in 1500 CE.

Han ethnic group and Malaysia was entirely composed of the Maori ethnic group, Malaysia's current ethnic composition is 75% Maori and 25% Han. This is shown by:

$$A^{2000}_{m\times n} = E^{1500}_{m\times n} \times M^{1500-2000}_{n\times n} = \begin{array}{lcc} & \text{Malaysia} & \text{China} \\ \text{Han} & 0.25 & 1 \\ \text{Maori} & 0.75 & 0 \end{array}$$

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

$$\begin{aligned} E^{1500}_{m\times n} &= A^{2000}_{m\times n}(M^{1500-2000}_{n\times n})^{-1} \\ &= \left(\begin{array}{lcc} & \text{Malaysia} & \text{China} \\ \text{Han} & 0.25 & 1 \\ \text{Maori} & 0.75 & 0 \end{array}\right) \times \left(\begin{array}{lcc} & \text{Malaysia} & \text{China} \\ \text{Malaysia} & 1.33 & 0 \\ \text{China} & -0.33 & 1 \end{array}\right) \\ &= \left(\begin{array}{lcc} & \text{Malaysia} & \text{China} \\ \text{Han} & 0 & 1 \\ \text{Maori} & 1 & 0 \end{array}\right) \end{aligned}$$

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.

In summary, I use language groups to match similar ethnicities in Alesina et al. to those in Ingram et al. This allows us to create a country-level measure for the frequency of lactase persistence from the ethnic-level data in Ingram et al. The ethnic compositions, however, are based on contemporary populations. To correct for this, I use migration data from Putterman and Weil to create ethnic compositions representative for the precolonial period.[20]

[20] Another possible concern involves the monotonicity, or *relative* frequencies, of lactase persistence over the past 500 years. For this to bias estimation, one of two (or both) events must occur: Either highly populated countries obtained lactase persistence at a greater rate than less dense countries, or less populated countries have reduced their frequencies of lactase persistence relative to densely populated countries. Please see the supplemental appendix for a fuller discussion.

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 108 countries for the base estimation, of which 49 have exact ethnic matches. Table 1 presents the descriptive statistics for the frequency of lactase persistence as well as other variables used in the baseline regression model. The mean frequency of lactase persistence in the base sample is 41.52%, which is similar to the world mean of 35% given by Ingram et al. (2009a).[21] 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 from Simoons (1969). Comparing Fig. 1 and Fig. 2, there appears to be a relatively tight fit between historically non-milking areas and areas with 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. This theory is further established by the work of Ashraf and Galor (2011). Population density is a viable proxy for the level of economic development in the precolonial era; additionally, 1500 CE population densities are used regularly in similar research (see e.g., Acemoglu et al. 2002, Ashraf and Galor 2013, Chanda and Putterman 2007, Putterman 2008). The proposed hypothesis is that milking provided an extra resource to certain peoples that expanded the carrying capacity of their environment, thereby increasing population densities, or economic development in the Malthusian economy of the precolonial era. 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. Sources and explanations of all additional variables are given within the Variable Appendix.

3 Results

This section establishes the relationship between country-level lactase persistence and development in the precolonial era, measured by population density in 1500 CE. Firstly, I show that a positive and statistically significant relationship exists between these two variables, which provides support for

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

the main hypothesis of this paper. Going further, I show that this relationship is robust to sample adjustments and the inclusion of conceivable omitted variables. Finally, I explore the effect of lactase persistence on differing dependent variables.

3.1 Baseline Estimation

The baseline estimating equation follows a form similar to that given in Ashraf and Galor (2011). The frequency of lactase persistence allowed workers to produce more from their environments. This in turn increased production and population density in the Malthusian era. This hypothesis is tested with the following estimating equation:

$$ln(\mathrm{P})_i^{1500} = \beta_0 + \beta_1(\text{Freq. of Lactase Persistence})_i + \boldsymbol{\beta_2'}\mathbf{X}_i + \boldsymbol{\beta_3'}\mathbf{D}_i + \boldsymbol{\beta_4'}\mathbf{T}_i + \boldsymbol{\beta_5'}\mathbf{G}_i + \boldsymbol{\beta_6'}\mathbf{C}_i + \epsilon_i \quad (1)$$

where i is a country index, β_1 is the coefficient of interest throughout the paper, $\mathbf{X}_i$ is a vector of agricultural controls, $\mathbf{D}_i$ is a vector of controls for domesticate animals, $\mathbf{T}_i$ is a vector technological controls, $\mathbf{G}_i$ is a vector of controls for geography, $\mathbf{C}_i$ is a vector containing continent and region indicator variables, and ϵ_i is the cross-country error term. The coefficients of equation (1) are estimated with ordinary least squares and robust standard errors are reported.

Table 2 provides estimated coefficients of equation (1); the coefficient of the frequency of lactase persistence, the coefficient of interest, is included in all estimations, while the specified control vectors are piecemeal included. Column (1) displays the simple bivariate regression of 1500 population density on the country-level frequency of lactase persistence. 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 58% increase in the number of people per squared kilometer. For Egypt, the country with the median population density in the base sample, this corresponds to an increase of two people per square kilometer.

In column (2), continent and region dummies are included into the regression of column (1). Given the sample restraint of only Old World countries, the vector of continent dummies includes indicator variables for European and African countries. Considering the plot in Figure 3, Scandinavian countries are clear outliers in the relationship between the frequency of lactase persistence and precolonial population densities.[22] Additionally, the high frequency of lactase persistence associated with Scandinavian

[22] Finland, Norway, and Sweden are also statistical outliers in the relationship between the frequency of lactase persistence and population density as signified by a robust standardized residual larger than 2.25 (Verardi and Croux 2009). The coefficient of lactase persistence is unchanged in magnitude or significance when excluding the Scandinavian dummy. The Scandinavian dummy consists of markers for Denmark, Finland, Norway, and Sweden. While Finland is not culturally Scandinavian, it does exhibit characteristics for which the indicator is intended to control for, mainly high levels of lactase persistence and relatively low population density. The exclusion of Finland from the Scandinavia indicator, or the

countries may be the product of recent migration (Vuorisalo et al. 2012). I therefore include a regional dummy for Scandinavia. The inclusion of continent and Scandinavia effects does not significantly alter the effect of country-level lactase persistence. The coefficient remains positive and is statistically significant at the 1% level.

Controlling for the effects of agriculture are necessary to measure the added benefit of milk consumption. A hypothesis posited by Cook and al-Torki (1975) suggests that the development of milking cultures occurred in relatively harsh environments, which were not suitable for agriculture. In order to survive in these environments, humans had to extract more calories from their environment, which includes domesticate animals. This hypothesis suggests less suitable environments, or environments that could support fewer individuals, should have greater frequencies of lactase persistence, the opposite of the hypothesis posited in the current work and the opposite of the effect of lactase persistence found in columns (1) and (2). An alternative view considers the added advantage of milk consumption. Peoples with greater frequencies of lactase persistence had an advantage in productivity. This advantage should have led to populations with a high frequency of lactase persistence dominating populations with a low frequency lactase persistence, resulting in an even distribution of lactase persistence and an even distribution in population density.[23] The analysis, however, considers the marginal, or added, benefit of lactase persistence holding constant other determinants of development in the precolonial period. The amount of time a country has practiced agriculture along with the effectiveness of agriculture are important determinants of development in the precolonial period. To account for these effects, column (3) includes the natural log of the millennia a country has practiced agriculture and the natural log of land productivity, which accounts for both the fraction of arable land and the average probability of cultivation within a country (Ashraf and Galor 2011; Michalopolous 2012; Putterman 2008; Ramankutty et al. 2002). With the inclusion of all agricultural controls in column (3), the effect of lactase persistence remains similar to the bivariate regression of column (1) and the inclusion of continent and region dummies in column (2). When favorable or harsh environments are held constant, the positive and significant effect of lactase persistence is unaltered.

Dairying is dependent upon the availability of domesticate animals, which provide additional benefits to population (e.g., labor, meat, and wool). In an attempt to control for these additional effects of domesticate animals that are independent of milk production and dairying cultures, column (4) includes a number of controls that intend to capture the alternative channels in which domesticate

exclusion of the indicator itself, does not alter the estimated relationship between the country-level frequency of lactase persistence and population density in 1500 CE.

[23] As is shown in Table 6, the country-level advantage of lactase persistence is a relatively recent phenomenon. Lactase persistence was most likely associated with pastoral societies and didn't result in denser populations until the widespread use of sedentary agriculture.

animals influence population density. The additional controls include the agricultural suitability of plow-positive and plow-negative crops, the suitability of land for pasture, and the ecological suitability of the tsetse fly. The inclusion of the plow agriculture variables intends to capture other non-dairying benefits of domesticate animals, particularly increases in productivity from the use of draft animals. For the same reasons, the suitability of land for pasture is also controlled for, from which the inclusion of this variable intends to capture other non-observed benefits of domesticate animals that are independent of the cultural adaptation of dairying.[24] Finally, the inclusion of tsetse suitability accounts for ecological channels that limit the presence of large domesticate animals (Alsan 2014).[25] As reported in column (4), the inclusion of all controls, which serve as proxies for the number and usage of domesticate animals, does not alter the previously found relationship between the frequency of lactase persistence and population density in 1500 CE.

As stated previously, the evolution of lactase persistence is dependent upon dairying cultures. It is possible, that cultures associated with dairying were also associated with other technological advantages. For this purpose, it is essential to control for differences in technology. The measure of technology in 1500 CE is the great circle distance from a technological frontier. Old World technological frontiers are London (U.K.), Paris (France), Cairo (Egypt), Fez (Morrocco), Constantinople (Turkey), and Peking (China). In addition to the distance to the technological frontier, I also include an indicator variable for whether or not a country was a part of the Roman Empire. Acemoglu et al. (2005) find that countries belonging to the Roman Empire had significantly higher levels of development between 1400 and 1600 CE, which encompasses the year of precolonial development of 1500 CE. Additionally, members of the Roman Empire shared a common government and culture, which provides a further control against a potentially omitted cultural variable associated with precolonial population density. Column (5) takes into account these measures of technological variation. The measures of technology have the expected signs and are statistically significant and the frequency of lactase persistence remains positive and significant at the 1% level and is similar in magnitude to the estimates of column (2), which excludes all controls except continental dummies.

As a final set of controls for the baseline regression model, I include measures for geographical variation. Byproducts of dairying culture, such as butter, cheese, and yogurt, are able to be stored and preserved, whereas milk spoils relatively quickly. These added goods could potentially be traded,

[24] The suitability of land for pasture uses the contemporary suitability, implying that this measure may also be influenced by 1500 CE population densities; therefore, as with the measures from the *Ethnographic Atlas*, the use of land suitable for pasture may not be controlling for *historic* differences in land for pasture but differences in population density in 1500 CE.

[25] The main measure of ecological suitability for the tsetse fly is the average across three strains–Fusca, Morsitans, and Palpalis. The inclusion of each suitability measure, either separately or together, does not alter the coefficient of the frequency of lactase persistence.

adding a benefit to dairying culture unrelated to the consumption of milk. As a potential control for this effect, I include the mean distance from a coast or navigable river.[26] As a further measure for climatic, or geographic differences, I also control for the absolute latitude of a country. The added geographical controls do not alter the significance or magnitude of the coefficient of interest. This is shown in column (6).

Column (7) comprises the regression model specified in equation (1). After including relevant controls, the frequency of lactase persistence is shown to have a positive relationship with population density in 1500 CE that is statistically significant at the 1% level. Figure 4 gives the orthoganolized plot of the estimation in column (7). The coefficient of column (7) gives that a one standard deviation increase in the frequency of lactase persistence is associated with a 36% increase in population densities in the year 1500. In other words, if Egypt's lactase persistence frequency was equal to that of Pakistan, everything else constant, Egypt's population density would increase from roughly 4 to 5.7 people per squared kilometer.

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. (2009a) and Alesina et al. (2003). The estimated effect of lactase persistence remains positive but is statistically indistinguishable from zero; however, due to the reduced sample and resulting loss in precision, I can not rule out that the coefficient of lactase persistence in column (8) is different in magnitude from the base estimation of column (7) that uses the extended sample (p=0.55) .

The estimates of Table 2 support the main hypothesis. Those societies which consumed milk had the advantage of an additional resource. Access to this additional resource is hypothesized to provide an added benefit, which holding constant other factors of development, is associated with denser populations in the precolonial era. Concerns, however, of reverse causality and omitted variable bias limit the causal interpretation of the strong statistical relationship found in Table 2.

Whether or not the relationship between the frequency of lactase persistence and precolonial population density is causative, depends in part upon the source of the cross-country differences in lactase persistence. Lactase persistence has arisen from cultural variation. The cultural cause of differences in lactase persistence creates an ambiguity in the exogeneity of the measure. In other words, did those cultures that adopted dairying have other unseen population advantages? An additional complication arises from the unmeasured benefits of cattle and other milk producing animals (e.g., draft labor, meat, etc.), which are likely correlated with the frequency of lactase persistence. In part, the baseline estimating equation addresses these potential sources of endogeneity by controlling for the suitability of

[26] Distance from the a technological frontier should also account for the potential effects of trade.

pasture and plow intensive crops as well as a number of controls for cultural similarity across countries. The next section will attempt to further alleviate concerns of bias through sample adjustments and the inclusion of possible omitted variables.[27]

3.2 Sensitivity Analysis

Sample Adjustments

The 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 3 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 four countries from the baseline sample; as a result, the significance and magnitude of the coefficient of interest are similar to those in column (7) of Table 2. Column (2) excludes countries in which less than 75% of the contemporary population is derived from the 1500 CE population. This results in the exclusion of 17 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 estimated coefficient is statistically significant and positive. Column (4) excludes countries with less than 95% of the current population derived from 1500 CE populations. This results in excluding 56 countries from the baseline sample. The estimate of the coefficient of interest, however, remains positive and is significant at the 1% level. The measurement error that results in the approximation of 1500 CE ethnic compositions does not appear to affect the 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.

While the baseline estimate for the effect of lactase persistence takes into account continent effects, I believe within continent estimations are necessary to further establish the relationship between milk consumption and population density in 1500 CE. Of particular importance is the role of Europe. Europe contained on average denser populations as well as higher frequencies of lactase persistence. This implies that the proposed effect of lactase persistence may be the byproduct of some unseen European advantage. Towards this end, Table 4 restricts the sample of the baseline regression model to

[27] A previous version of this paper also considered solar radiation as a potential instrument for the country-level frequency of lactase persistence.

each of the three continents that makeup the Old World. 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 larger in magnitude compared to 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 1969). Column (2) constricts the sample to countries within Africa alone. The coefficient of interest remains positive but becomes insignificantly different than zero. As with Africa, the coefficient of lactase persistence with Asia remains positive but is insignificantly different than zero at conventional levels. In column (4), I consider all countries outside of Europe. The coefficient of the frequency of lactase persistence is once again statistically significant at conventional levels and the magnitude only differs slightly from that given in the baseline estimate. The estimation of column (4) shows that Europe alone is not responsible for the significance of the coefficient of lactase persistence. Column (5) excludes Africa from the sample with little change in the coefficient of interest from the baseline estimate. Column (5) shows that within Eurasia milk consumption is associated with denser populations and development in the precolonial era. Table 4 provides substantial evidence that the effect of lactase persistence is not being driven by a European externality. This narrows the possibility of a spurious correlation and provides a better understanding of the role of lactase persistence in explaining variations in precolonial population density.

Omitted Variables

Tables 5 explores whether additional controls can make the effects of lactase persistence frequencies disappear. Column (2) includes a control for celiac disease; column (3) includes an additional genetic measure; column (4) includes additional environmental and geographic controls; column (5) controls for historic cultural variation; column (6) includes biogeographic controls; column (7) includes additional controls mentioned in columns (2)-(5); and column (8) includes all additional variables.

In addition to the digestion of milk, celiac disease, or the negative autoimmune response to gluten, a protein of grains, represents a dietary change initiated by the Neolithic Revolution. Celiac disease is the result of a genetically predisposed individual being exposed to gluten (Di Sabatino and Corazza 2009), and similar to lactase persistence, celiac disease has a greater prevalence amongst peoples of European descent (Fasano et al. 2003). Given the similarities in timing and location amongst celiac disease and lactase persistence, it may be possible that the two dietary changes are correlated, from which the presence of dairying may have lessened the need for grains thereby lessening the harm of celiac disease and allowing it to rise in frequency. This substitution between grains and dairy may also affect population density; therefore, I need to control for the frequency of celiac disease in order to more accurately measure the direct effect of milk consumption. In order to measure the frequency of celiac disease, I create a proxy based upon the frequency of genes associated with the disease. The HLA-DQ2.5 haplotype, or gene pair, has a strong association with celiac disease (Fallang et al. 2009). I therefore control for the frequency of this haplotype to account for the effect of celiac disease. This is shown in column (2) of Table 5. The inclusion of this variable does not significantly alter the relationship between the country-level frequency of lactase persistence and precolonial population density, implying the positive correlation between celiac disease and lactase persistence is not driving the positive relationship between lactase persistence and precolonial population density. Celiac disease itself has an insignificant effect on population density in 1500 CE.

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 (see e.g., Ashraf and Galor 2013). Column (3) of Table 5 includes two additional genetic controls into the baseline estimation. These are genetic distance from the technological frontier, which is represented by Britain in 1500 CE, and the country-level genetic diversity in 1500 CE. Spolaore and Wacziarg (2009) argue that a smaller genetic distance (i.e., similar genotypes) eases the diffusion of technology, whereas Ashraf and Galor (2013) argue genetic diversity within a country has an effect on precolonial (and contemporary) economic development. The additional genetic controls do not alter the significance or magnitude of lactase persistence.[28] 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 (4) introduces a number of climate and environmental controls; these include an index for the transmission of malaria, the mean elevation, the mean ruggedness, the mean temperature, the mean precipitation, and the fraction of the country belonging to each Köppen-Geiger climate zone. An important 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 a historic presence of malaria. An additional hypothesis posed by Anderson and Vullo (1994) argues malarial environments may have selected for lactase deficiency. This is due to the

[28]The individual inclusion of either measure does not alter the coefficient of lactase persistence. Additionally, the use of the complete base sample given in Table 2 does not alter the results.

beneficial effects of malaria resistance from a diet lacking in riboflavin, which is supported by reduced consumption of riboflavin rich milk. This indicates that the relationship between lactase persistence and historic populations may be driven by the disease environment. It is therefore essential to control for differences in the disease environment. In addition to the ecological suitability of the tsetse fly, column (4) also controls for the index of malarial transmission from Kiszewski et al. (2004); while this is a contemporary measure, I have little evidence to believe it is an ineffective control variable.

Column (4) also controls for elevation and ruggedness, which is roughly defined as the average variation in elevation between cells within a country (Nunn and Puga 2012). For my concerns, ruggedness accounts for land variations that make farming difficult; and, therefore, may promote the use of animal husbandry, which increases the likelihood of milk consumption. The inclusion of these additional geographic controls into the baseline regression should alleviate any potential biases that may occur due to land conditions that increased pastoralism. An argument has also been put forward that extreme environments may contribute to variations in lactase persistence (Cook and al-Torki 1975). With this in mind, I control for extreme environments by including the percent of land within the tropics and the percent of land which is desert, as well as the percent of land within each Köppen-Geiger climate zone. The percent of land within the tropics, for my purposes, represents an environment in which resources are rich; consequently, there should be little need for dairying.[29] At the other extreme, deserts are poor in resources, implying a greater need for dairying. Additionally, I include the mean temperature and precipiation within a country to account for other unobserved environmental differences that may potentially be associated with population density or the frequency of lactase persistence. The inclusion of additional environmental, climate, and geographic controls in column (4) does not alter the relationship between lactase persistence and population density. The coefficient of lactase persistence remains positive, significant, and similar in magnitude to the baseline estimate with the revised sample of column (1). The relationship between dairying and historic populations is not the result of a simultaneous correlation with a previously omitted environment or climate.

Additional historic cultural and institutional control are considered in column (5), which takes into account differential state histories within the Old World as well as an index for historical economic development, which represents a range of development from purely nomadic societies to those with complex state organization, and the fraction of the diet derived from animal husbandry.[30] A major

[29] The percent within the tropics also accounts for tropical disease environments that may prevent the spread of cattle and populations.

[30] Both historic economic development and the fraction of the diet from animal husbandry are derived from George Murdock's *Ethnographic Atlas*; therefore, each measure is based on an ethnicity. Using this data on ethnicities and the contemporary ethnic composition of countries (Lewis 2009), Alesina et al. (2013) create country-level measures for many of the variables within the *Ethnographic Atlas*. These country-level measures are based on contemporary populations and not population compositions in 1500 CE; however, limited migration in the Old World likely results in roughly consistent ethnic compositions across time–e.g., the correlation between the 1500 CE measure of lactase persistence and

concern of endogeneity is tied to the cultural differences that led some societies to practice dairying. Favorable cultures that are associated with milk consumption, and therefore lactase persistence, may also be associated with other differences associated with economic development. The inclusion of state history, historic economic development, and the concentration of animal husbandry into the baseline regression is intended to control for unobserved cultural differences associated within societies of each country. The inclusion of these potentially omitted variables does not significantly alter the estimated effect of lactase persistence, which remains positive, statistically significant at the 1% level, and similar in magnitude to the baseline effect of column (1). The results of column (5) strengthen the proposed relationship between lactase persistence and population density in 1500 CE by accounting for relevant cultural differences across societies.

Domesticable animals were a necessary condition for the development of lactase persistence. But domesticable animals also provide population benefits, e.g., meat, labor, etc. Column (6) of Table 5 uses the number of potential domesticate animals as an additional proxy for the additional benefits conferred by domesticate animals as well as the number of potentially domesticate plants from Hibbs and Olsson (2004).[31] Due to space considerations, the baseline estimation with the sample based upon available biogeographic data is excluded. With the revised sample, the baseline estimate for the coefficient of lactase persistence is 2.23, which is significant at the 1% level. The inclusion of biogeographic controls to the baseline model does not influence the estimated relationship between the frequency of lactase persistence and population density in 1500 CE, providing further support the direct relationship between dairying, not the presence of domesticate animals, and population density.

Column (7) of Table 5 simultaneously introduces the potential omitted variables discussed in columns (2)-(5).[32] The inclusion of these potentially omitted variables does not affect the coefficient of lactase persistence, implying the relationship between lactase persistence and population density is robust to the inclusion of a large and theoretically important set of additional controls. Column (8) includes all additional controls with little effect on the magnitude of coefficient of interest, although statistical significance is lost from the reduction in sample size and resulting imprecision in estimation. The results of Table 5 suggest omitted variable bias is not likely.

In summary, the coefficient on lactase persistence remains relatively constant in magnitude and significance throughout the numerous empirical specifications performed. Throughout the sensitivity

the measure of lactase persistence using contemporary populations is roughly 1. A further concern of data from the *Ethnographic Atlas* is the time of collection occurs after the primary dependent variable of interest, population density in 1500 CE. As with the frequency of lactase persistence, variation in these measures may be driven by differences in historic population densities; therefore, the inclusion of the intensity of animal husbandry and economic complexity may not be exogenously accounting for differences in the specified variable.

[31] The land suitable for pasture, the intensity of plow agriculture, and the ecological suitability of the tsetse fly are also intended as controls for the additional benefits of domesticate animals.

[32] Biogeographic controls are excluded due to the sample truncation.

analysis, the coefficient of the frequency of lactase persistence remains highly significant and is rarely different in magnitude from the bivariate or baseline estimations (Column (7) of Table 2).[33] 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 a country's 1500 CE frequency of lactase persistence and population density in the same period.

3.3 Alternative Dependent Variables

Given the strong relationship between the frequency of lactase persistence and development in the precolonial era, it's of interest to see whether the beneficial effects of milk consumption are related to other measures of development. Of particular importance are income in the precolonial era, more historical levels of development, and the contemporary effect of lactase persistence.

Historic Population Density

The natural selection of lactase persistence is a relatively new adaptation resulting from the domestication of cattle and other milk producers. This idea is confirmed through a number of studies. Burger et al. (2007) find the gene variant associated with lactase persistence is absent in early Neolithic European populations. Plantinga et al. (2012) find that the frequency of lactase persistence was relatively low amongst Southwestern European populations in the late Neolithic Period, roughly 5,000 years before the present. Therefore, it is of interest to explore the effects of lactase persistence on more historic levels of development. Lactase persistence is a relatively new trait that should have more pronounced benefits in later periods. This is explored in Table 7.

Columns (1), (2), and (3) regress the natural log of population density in 1, 1000, and 1500 CE, respectively, for the sample of countries in which data is available for all years. Column (1) explores the effect of lactase persistence on population density in the first year CE.[34] In this very early period, the estimated coefficient of interest is positive but statistically insignificant, implying the significant, positive population effects of lactase persistence are realized in more recent periods of the Malthusian era. Column (2) looks at the association of lactase persistence and population density in 1000 CE, and again finds a positive but statistically insignificant significant coefficient. Column (3) repeats the baseline estimation of column (7) of Table 2 with the sample of countries that have data for population density in 1 and 1,000 CE. No significant difference in the estimated relationship is seen with the reduced

[33] The effect of lactase persistence remains positive, statistically significant, and consistent in magnitude when altering the sample across the differing variables of Table 5. The reduced consistent sample does not significantly alter the relationship between lactase persistence and precolonial population density.

[34] The frequency of lactase persistence is calculated with ethnic compositions for the year 1500. I have little reason to suspect the measure is invalid as a proxy for previous periods.

sample, and the positive and significant coefficient of the frequency of lactase persistence implies that the relationship between lactase persistence and population density is only recently realized.

As mentioned in Section 1.3, population density, which potentially serves as a proxy for the prevalence of infectious disease, may lead to the increased selection of lactase persistence, creating a spurious relationship in the previous estimations. To partially account for this difference in historical disease environments and the resulting reverse causality, columns (4) and (5) of Table 6 control for population density in earlier periods. A high level of serial correlation is found for population density. For the sample of Table 6, population density in 1500 CE has a correlation coefficient of 0.94 for population density in 1,000 CE and 0.82 for population density in 1 CE. Therefore, controlling for population density in earlier periods reduces the amount of variation in 1500 CE population density, thereby potentially reducing the magnitude of its relationship with differences in lactase persistence. In column (4), which repeats the baseline estimation while including population density in 1000 CE and with the sample of Table 6, the association between lactase persistence remains positive and statistically significant at conventional levels; although, the magnitude of the coefficient of interest is substantially lessened.[35] For column (5), which controls for population density in 1 CE in place of 1000 CE, a similar effect is observed: the coefficient of the frequency of lactase persistence is smaller in magnitude but remains positive and statistically significant. While the findings of columns (4) and (5) can not fully account for the potential of reverse causality, the estimates do suggest that the unobserved latent infectious disease environment is not the sole factor responsible for the significant relationship between 1500 CE population density and the frequency of lactase persistence.

Precolonial Income

Table 6 explores the relationship between lactase persistence and a proxy of income in 1500 CE. According the the Malthusian theory of Ashraf and Galor (2011), an increase in output will temporarily increase incomes within the economy, increasing demand for children, thereby permanently raising population while decreasing output per person, or income, to the initial level. If the relationship between the frequency of lactase persistence and population density is in concordance with the Malthusian theory, lactase persistence should not be related to income, as the benefit of lactase persistence has been translated into denser populations, not increased incomes.

Due to a lack of data, I use a measure of urbanization from Nunn and Qian (2011).[36] As shown in

[35]The p-value of the coefficient of the frequency of lactase persistence in column (4) is 0.058.

[36]Use of Maddison income estimates for the year 1500 results in a sample size of 25. As a consequence of this reduced sample, the estimated relationship between lactase persistence and population density becomes insignificant. Urbanization from Nunn and Qian (2011) is the fraction of a country's population living within a city of a population greater than 40,000.

Acemoglu et al. (2002), urbanization is strongly related to incomes in the Malthusian era; with this finding in mind, urbanization serves as a viable proxy for income. Column (1) reports the bivariate regression of urbanization in 1500 CE on the frequency of lactase persistence, which shows lactase persistence having an insignificant effect on urbanization, the proxy for income.[37] Consistent with the theory formalized by Ashraf and Galor (2011), the added benefit of lactase persistence did not lead to higher incomes in the Malthusian era. The baseline controls are included in column (2), resulting in a negative and statistically significant coefficient on the country-level frequency of lactase persistence. This negative relationship may be due to the structure of pastoral societies in relation to the proxy for income. Pastoral societies had the potential to develop great frequencies of lactase persistence but had little use for sedentary cities.[38] Columns (3) and (4) repeat the estimations of columns (1) and (2), replacing the urbanization rate in 1500 CE with urbanization in the early period of 1000 CE. The coefficient of the frequency of lactase persistence on the earlier measure of urbanization is lessened in magnitude but is consistent in the sign and significance.

As with population density in earlier periods, the fraction of the population living within a city may be a useful control to account for the effects of infectious disease. Due to relative large populations, high densities, and poor plumbing, cities serve as likely reservoirs for a number of infectious diseases. Therefore, countries with a greater fraction of the population living within a city experience greater exposure to infectious disease, which in turn, can potentially lead to a greater turnover of the population through the mechanisms mentioned previously. In order to account for this potential association with the infectious disease environment, columns (5)-(6) control for the urbanization rate in the baseline estimation.

Column (5) includes the urbanization rate in 1500 CE into the baseline estimation with no substantial effect on the coefficient of lactase persistence. Column (6) replicates column (5), replacing the urbanization rate in 1500 CE with the rate for 1000 CE. Again, the coefficient of interest remains similar to the baseline estimate. The use of urbanization as a control for the historic infectious disease environment does not alter the relationship between the frequency of lactase persistence and population density in 1500 CE. Providing support for a limited role of infectious disease in accelerating the selection of lactase persistence, thereby further lessening the concern of reverse causality.

[37] The 3 excluded countries for Table 7, which are missing data on urbanization, are Azerbaijan, Bahrain, and Georgia. The baseline relationship between lactase persistence and population density in 1500 CE is unaffected from the exclusion of these countries.

[38] The negative relationship between lactase persistence and urbanization is unaffected by the exclusion of the dummy for Roman heritage.

Modern Development

Given the strong connection between lactase persistence and historical development, it is of interest to explore the effects of lactase persistence on contemporary populations and incomes. Table 8 explores these relationship with the year 2000 as the measure of contemporary development and a contemporary measure for the country-level frequency of lactase persistence. The contemporary measure of lactase persistence is calculated with contemporary ethnic compositions from Alesina et al. (2003).

Column (1) regresses population density in 2000 CE on the contemporary frequency of lactase persistence as well as a number of contemporary controls and continent indicators.[39] The coefficient of interest in column (1) suggests lactase persistence remains positively associated with population density in the current era; however, this effect is insignificant at conventional levels. Considering income in column (2), lactase persistence is again shown to have an insignificant association with contemporary levels of development.

The country-level frequency of lactase persistence isn't directly related to contemporary economic outcomes. The effect of lactase persistence is captured in the Malthusian economy preceding the Industrial Revolution, not in the present economy. This coincides with the current natural selection of the lactase persistent trait. Due to the numerous dietary options and health interventions, the contemporary benefit of lactase persistence is likely to be negligible, and this insignificance is shown in the coefficients of Table 8.

4 Conclusion

Lactase persistence, which measures the ability to consume milk, is a recent genetic adaptation that has been naturally selected since the Neolithic Revolution. My main hypothesis is that the consumption of milk increases the carrying capacity of a fixed quantity of land by increasing both the quantity of calories produced and through qualitative improvements in diet from the high protein and other nutrients found within milk.

Towards this end, a statistically strong relationship is shown between the frequency of a country's 1500 CE population that is lactase persistent and the level of economic development in the precolonial era, measured by population density in 1500 CE. This relationship is robust, remaining consistent in magnitude and statistical significance through the addition of numerous controls and sample specifications. Importantly, the baseline estimating equation controls for a number of theoretically valid

[39]The contemporary controls include an ancestry adjusted measure of precolonial population densities, an ancestry adjusted measure for the millennia a country has practiced agriculture, ethnic fractionalization, land productivity (i.e., the first principle component of agricultural suitability and the percent of arable land), the suitability of land for plow-positive crops, plow-negative crops, and pasture, absolute latitude, and the average country-level distance to the coast or navigable river.

measures that account for agriculture, technology, geography, differing measures for the suitability of domesticate animals, and continent and region effects. Additional controls are also considered with little effect on the estimated relationship between the frequency of lactase persistence and population density in 1500 CE. Furthermore, the estimated relationship is robust to adjustments in the sample based upon migration and continent-level differences.

The analysis, while highly suggestive, is unable to truly establish a causal role for the frequency of lactase persistence in leading to differences in the population density in 1500 CE. Reverse causality remains plausible as more densely populated countries likely experienced a greater prevalence of infectious disease. This increased prevalence is associated with a rise in the mortality rate and, given the dynamics of the Malthusian economy, a rise in the birth rate, resulting in a greater turnover of the population and a faster selection of lactase persistence. In addition to concerns of reverse causality, the selection of lactase persistence is tied to the unobserved cultural practice of dairying, which may effect population density through other channels. However, there is little evidence for the proposed mechanism of reverse causality, and by controlling for a number of relevant, observable cultural differences, cultural variation does not appear to be a substantial source of bias.

Diamond states: "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, it alters this view by considering the role of the environment in shaping the frequency of genetic variants that may have long running effects on both health and economic outcomes.

References

Acemoglu, D., Johnson, S., & Robinson, J. (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review*, 91(5), 1369-1401.

Acemoglu, D., Johnson, S., & Robinson, J., (2002). Reversal of fortune: Geography and institutions in the making of the modern world income distribution. *Quarterly Journal of Economics*, 107, 12311294.

Acemoglu, D., Johnson, S., & Robinson, J., (2005). The rise of Europe: Atlantic trade, institutional change and economic growth. *American Economic Review*, 95(3), 546-579.

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., & Wacziarg, R. (2003). Fractionalization. *Journal of Economic Growth*, 8(2), 155-194.

Alesina, A., Giuliano, P., & Nunn, N. (2013). On the origins of gender roles: Women and the plough. *Quarterly Journal of Economics*, 128(2), 469-530.

Alsan, M. (2012). The effect of the tsetse fly on African development. Working Paper.

Anderson, B., & Vullo, C. (1994). Did malaria select for primary adult lactase deficiency?. *Gut*, 35(10), 1487-1489.

Ashraf, Q., & Galor, O. (2011a). Dynamics and stagnation in the Malthusian epoch. *American Economic Review*, 101(5), 2003-2041.

Ashraf, Q., & Galor, O. (2013). The "Out of Africa" hypothesis, human genetic diversity, and comparative economic development. American Economic Review, 103(1), 1-46.

Bersaglieri, T., Sabeti, P., Patterson, N., Vanderploeg, T., Schaffner, S., Drake, J., Rhodes, M., Reich, D., & Hirschhorn, J. (2004). Genetic signatures of strong recent positive selection at the lactase gene. *The American Journal of Human Genetics*, 74, 1111-20.

Bockstette, V., Chanda, A., & Putterman, L. (2002). States and markets: the advantage of an early start. *Journal of Economic Growth*, 8, 155-194.

Burger, J., Kirchner, M., Bramanti, B., Haak, W., & Thomas, M. (2007). Absence of the lactase-persistence-associated allele in early Neolithic Europeans. *Proceedings of the National Academy of Sciences*, 104, 3736.

Chanda, A., & Putterman, L. (2007). Early starts, reversals and catch-up in the process of economic development. *Scandinavian Journal of Economics*, 109(2), 387 413.

Clark, G. (2008). *A Farewell to Alms: A Brief Economic History of the World.* Princeton, NJ: Princeton University Press.

Coelho, M., Luiselli, D., Bertorelle, G., Lopes, A., Seixas, S., Destro-Bisol, G., & Rocha, J. (2005). Microsatellite variation and evolution of human lactase persistence. *Human Genetics*, 117, 329-339.

Cohen, M., & Armelagos, G. (1984). *Paleopathology at the Origins of Agriculture.* New York, NY: Academic Press.

Comin, D., Easterly, W., & Gong, E. (2010). Was the wealth of nations determined in 1000 BC?. *American Economic Journal - Macroeconomics*, 2(3), 65-97.

Cook, C.J. (forthcoming). Potatoes, milk, and the Old World population boom. Journal of Development Economics.

Cook, G., & al-Torki, M. (1975). High intestinal lactase concentrations in adult Arabs in Saudi Arabia. *British Medical Journal*, 135-136.

Cooper, M., & Spillman, W. (1917). *Human Food from an Acre of Staple Farm Products.* Washington, D.C.: US Department of Agriculture.

Copley, M., Berstan, R., Dudd, S., Docherty, G., Mukherjee, A., Straker, V., Payne, S., & Evershed, R. (2003). Direct chemical evidence for widespread dairying in prehistoric Britain. *Proceedings of the National Academy of Sciences*, 100, 1524.

Craig, O., Taylor, G., Mulville, J., Collins, M., & Parker Pearson, M. (2005). The identification of prehistoric dairying activities in the Western Isles of Scotland: An integrated biomolecular approach. *Journal of Archaeological Science*, 32, 91-103.

Di Sabatino, A., & Corazza, G. (2009). Coeliac disease. *Lancet*, 373, 14801493.

Diamond, J. (1997). *Guns, Germs, and Steel.* New York, NY: W.W. Norton & Company.

Dunne, J., Evershed, R., Salque, M., Cramp, L., Bruni, S., Ryan, K., Biagetti, S., & di Lernia, S. (2012). First dairying in green Saharan Africa in the fifth millennium BC. *Nature*, 486, 390-394.

Engerman, S., & Sokoloff, K. (1997). Factor endowments, institutions, and differential paths of growth among New World economies: A view from economic historians of the United States. In Stephen Harber, ed., *How Latin America Fell Behind* (pp. 260-304). Stanford, CA: Stanford University Press.

Engerman, S., & Sokoloff, K. (2002). Factor endowments, inequality, and paths of development among New World economies. NBER Working Paper 9259. Cambridge, MA: National Bureau of Economic Research.

Evershed, R., Payne, S., Sherratt, A., Copley, M., Coolidge, J., Urem-Kotsu, D., Kostas Kotsakis, M., & others (2008). Earliest date for milk use in the Near East and southeastern Europe linked to cattle herding. *Nature*, 455, 528-531.

Fallang, L., Bergseng, E., Hotta, K., Berg-Larsen, A., Kim, C., & Sollid, L. (2010). Differences in the risk of celiac disease associated with HLA-DQ2.5 or HLA-DQ2.2 are related to sustained gluten antigen presentation. *Nature Immunology*, 10(10), 1096-1102.

Fasano, A., Berti, I., Gerarduzzi, T., Not, T., Colletti, R., & others. (2003). Prevalence of celiac disease in at-risk and not-at-risk groups in the United States: a large multicenter study. *Archives of Internal Medicine*, 163(3), 286-292.

Flatz, G., & Rotthauwe, H. (1973). Lactose nutrition and natural selection. *The Lancet*, 302, 76-77.

Galor, O., & Michalopoulos, S. (2012). Evolution and the growth process: Natural selection of entrepreneurial traits. *Journal of Economic Theory*, 147(2), 759-780.

Galor, O., & Moav, O. (2002). Natural selection and the origin of economic growth. *Quarterly Journal of Economics*, 117(4), 1133-1191.

Galor, O., & Moav, O. (2007). The Neolithic origins of contemporary variation in life expectancy. Brown University Department of Economics Working Paper 2007-14.

Gallup, J., Sachs, J., & Mellinger, A. (1999). Geography and economic development. CID Working Paper No. 1, March 1999.

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

Gonzalez-Galarza, F., Christmas, S., Middleton, D., & Jones, A. (2011). Allele frequency net: a database and online repository for immune gene frequencies in worldwide populations. *Nucleic Acids Research*, 39(Database Issue), D913-D1009.

Gueguen, L., & Pointillart, A. (2000). The bioavailability of dietary calcium. *Journal of the American College of Nutrition*, 19(2 Suppl.), 119S-136S.

Hartl, D. & Clark, A. (2006). *Principles of Population Genetics: Fourth Edition.* Sunderland, MA: Sinauer.

Hibbs, D., & Olsson, O. (2004). Geography, biogeography, and why some countries are rich and others are poor. *Proceedings of the National Academy of Sciences*, 101, 3715-20.

Heston, A., Summers, R., & Aten, B. (2012). Penn World Table Version 7.1. Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania.

Holden, C., & Mace, R. (1997). Phylogenetic analysis of the evolution of lactose digestion in adults. *Human Biology*, 69(5), 605-628.

Hoppe, C., Molgaard, C., & Michaelsen, K. (2006). Cow's milk and linear growth in industrialized and developing countries. *Annual Review of Nutrition*, 26, 131-173.

Ingram, C., Mulcare, C., Itan, Y., Thomas, M., & Swallow, D. (2009a). Lactose digestion and the evolutionary genetics of lactase persistence. *Human Genetics*, 124(6), 579-591.

Ingram, C., Raga, T., Tarekegn, A., Browning, S., Elamin, M., Bekele, E., Thomas, M., Weale, M., Bradman, N., & Swallow, D. (2009b). 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, 577-588.

Jain, A., Hsu, T., Freedman, R., & Chang, M. (1970). Demographic aspects of lactation and postpartum amenorrhea. *Demography*, 7, 255-271.

Kiszewski, A., Mellinger, A., Spielman, A., Malaney, P., Sachs, S., & Sachs, J. (2004). A global index representing the stability of malaria transmission. *The American Journal of Tropical Medicine and Hygiene*, 70(5), 486-498.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A., & Vishny, R. (1999). The quality of government. *Journal of Law, Economics and Organization*, 15, 222 279.

Lao, O., de Gruijter, J., van Duijin, K., Navarro, A., & Kayser, M. (2007). Signatures of positive selection in genes associated with human skin pigmentation as revealed from analyses of single nucleotide polymorphisms. *Annals of Human Genetics*, 71(3), 354-369.

Lewis, M.P. (2009). *Ethnologue: Languages of the World, Sixteenth Edition.* Dallax, TX: SIL International.

McEvedy, C., & Jones, R. (1976). *Atlas of World Population History.* New York, NY: Facts on File.

Meloni, T., Colombo, C., Ruggiu, G., Dessena, M., & Meloni, G. (1998). Primary lactase deficiency and past malarial endemicity in Sardinia. *Italian Journal of Gastroenterology and Hepatology*, 30(5), 490-493.

Michalopoulos, S. (2012). The origins of ethno-linguistic diversity. *American Economic Review*, 102(4), 1508-1539.

Mulcare, C. (2006). *The Evolution of the Lactase Persistence Phenotype.* London: University of London.

Murdock, G., & White, D. (1969). Standard cross-cultural sample. *Ethnology*, 8(4), 329-369.

Nielsen, R., Williamson, S., Kim, Y., Hubisz, M., Clark, A., & Bustamante, C. (2005). Genomic scans for selective sweeps using SNP data. *Genome Research*, 15, 1566-1575.

Nordhaus, W. (2006). Geography and macroeconomics: New data and new findings. PNAS, 103(10), 3510-3517.

Nunn, N. (2008). The long term Effects of Africa's slave trades. *Quarterly Journal of Economics*, 123(1), 139-176.

Nunn, N. (2009). The importance of history for economic development. *Annual Review of Economics*, 1(1), 65-92.

Nunn, N., & Puga, D. (2012). Ruggedness: The blessing of bad geography in Africa. *Review of Economics and Statistics*, 94(1), 20-36.

Nunn, N., & Qian, N. (2011). The potato's contribution to population and urbanization: Evidence from a historical experiment. *Quarterly Journal of Economics*, 126(2), 593-650.

Plantinga, T., Alonso, S., Izagirre, N., & others (2012). Low prevalence of lactase persistence in Neolithic South-West Europe. *European Journal of Human Genetics*, 20, 778-782.

Putterman, L.(2008). Agriculture, diffusion, and development: Ripple effects of the Neolithic Revolution. *Economica*, 75, 729-748.

Putterman, L. (2012). State Antiquity Index. http://econ.brown.edu/fac/louis_putterman

Putterman, L., & Trainor, C. (2006). Agricultural Transition Year Country Data Set. http://econ.brown.edu/fac/louis_putterman

Putterman, L., & Weil, D. (2010). Post-1500 population flows and the long-run determinants of economic growth and inequality. *Quarterly Journal of Economics*, 125(4), 1627-82.

Ramankutty, N., Foley, J., Norman, J., & McSweeney, K. (2002). The global distribution of cultivable lands: Current patterns and sensitivity to possible climate change. *Global Ecology and Biogeography*, 11(5), 377392.

Sabeti, P., Reich, D., Higgins, J., & others (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature*, 419, 832-837.

Simoons, F. (1969). Primary adult lactose intolerance and the milking habit: A problem in biological and cultural interrelations. I. Review of the medical research. *The American Journal of Digestive Diseases*, 14, 819-836.

Simoons, F. (1970). Primary adult lactose intolerance and the milking habit: A problem in biological and cultural interrelations. II. A culture historical hypothesis. *The American Journal of Digestive Diseases*, 15, 695-710.

Simoons, F. (1978). The geographic hypothesis and lactose malabsorption: A weighing of the evidence. *The American Journal of Digestive Diseases*, 23, 963-80.

Simoons, F. (2001). Persistence of lactase activity among Northern Europeans: A weighing of evidence for the calcium absorption hypothesis. *Ecology of Food and Nutrition*, 40(5), 397-469.

Spolaore, E., & Wacziarg, R. (2009). The diffusion of development. *Quarterly Journal of Economics*, 124(2), 469-592.

Tishkoff, S., Reed, F., Ranciaro, A., Voight, B., Babbitt, C., Silverman, J., Powell, K., Mortensen, H., Hirbo, J., Osman, M., & others (2006). Convergent adaptation of human lactase persistence in Africa and Europe. *Nature Genetics*, 39, 31-40.

Verardi, V., & Croux, C. (2009). Robust regression in Stata. *The Stata Journal*, 9(3), 439-453.

Vuorisalo, T., Arjamaa, O., Vasemagi, A., & Taavitsainen, J. (2012). High lactose tolerance in North Europeans: A result of migration, not in situ milk consumption. *Perspectives in Biology and Medicine*, 55(2), 163-174.

Wint, W., & Rogers, D. (2000). Predicted distributions of tsetse in Africa. DFID working paper.

World Health Organization (1998). The World Health Organization multinational study of breast-feeding and lactational amenorrhea. I. Description of infant feeding patterns and of the return of menses. *Fertility and Sterility*, 70(3), 448-460.

World Health Organization (2009). *Milk Fluoridation for the Prevention of Dental Carries.* Geneva: WHO Press.

Appendix

A. Frequency of Lactase Persistence for 1500 CE

Table A.1 Ethnic Lactase Persistence Frequencies

Ethnic Group	Country/Region	Frequency	Ethnic Group	Country/Region	Frequency
Arabs	JORDAN/Israel	0.23	Maasai	KENYA/Tanzania	0.80
Arabs	KUWAIT	0.53	Maori	NEW ZEALAND	0.36
Arabs	SAUDIARABIA	0.43	Mohajir	PAKISTAN	0.80
Ashkenazi Jews	ISRAEL	0.21	Mongols	CHINA	0.12
Asian	KUWAIT	0.42	Native Irish	IRELAND	0.96
Austrian	AUSTRIA	0.80	Nubians	SUDAN	0.33
Baloochi/Baluchistani	PAKISTAN	0.45	Nuer	SUDAN	0.28
Bantu	GABON/Uganda/Zambia	0.12	Pashtun	AFGHANISTAN/Pakistan	0.35
Bedouin	SAUDIARABIA	0.82	Polish	POLAND	0.63
Beja	SUDAN	0.84	Punjabi	PAKISTAN	0.44
british	UK	0.95	Rangi	TANZANIA	0.65
Burmese	MYANMAR	0.08	Roma	HUNGARY	0.44
Chinese (Taiwan)	TAIWAN	0.12	Russian	ESTONIA	0.43
Cypriots	CYPRUS	0.39	Saami	RUSSIA	0.52
Czech	CZECHOSLOVAKIA	0.82	San	BOTSWANA	0.09
Danes	DENMARK	0.96	Sandawe	TANZANIA	0.35
Dinka	SUDAN	0.27	Sardinians	ITALY	0.14
Diolas	SENEGAL	0.73	Sereres	SENEGAL	0.71
Dravidian	INDIA	0.13	Sicilians	ITALY	0.29
Egyptian	EGYPT	0.27	Sindhi	PAKISTAN	0.57
Estonian	ESTONIA	0.64	Somali	ETHIOPIA	0.24
Finn	FINLAND	0.84	Sotho	SOUTH AFRICA	0.35
Finns	RUSSIA	0.46	Spanish	SPAIN	0.66
French	FRANCE	0.71	Sri Lankans	SRI LANKA	0.28
German	GERMANY	0.85	Tajik	AFGHANISTAN	0.18
Greek	GREECE	0.55	Thai	THAILAND	0.02
Han	CHINA	0.08	Toucouleurs	SENEGAL	0.90
Hausa/Fulani/Wolof/Peuhls	NIGERIA/Senegal	0.64	Tswana	SOUTH AFRICA	0.17
Herero	SOUTH AFRICA/Namibia	0.04	Tuareg	NIGER	0.87
Hungarian	HUNGARY	0.63	Tunisian	TUNISIA	0.16
Hutu	RWANDA	0.42	Turks	TURKEY	0.31
Indians	INDIA	0.36	Tutsi	RWANDA	0.92
Indo-Aryan	INDIA	0.55	Twa	RWANDA	0.23
Iranian	IRAN	0.14	Udmurtian	RUSSIA	0.60
Iraqw/Fiome/Burunge	TANZANIA	0.60	Uzbek	AFGHANISTAN	0.00
Italians	ITALY	0.49	West-Siberian	RUSSIA	0.51
Jaali	SUDAN	0.51	Xhosa	SOUTH AFRICA	0.18
Japanese	JAPAN	0.28	Yemenites	SAUDIARABIA	0.53
Kazakh	CHINA	0.24	Yoruba	NIGERIA	0.17
Lebanese	LEBANON	0.22	Zulu	SOUTH AFRICA	0.14

Table A.2 Country Level Lactase Persistence Frequencies

Country	Frequency	Country	Frequency	Country	Frequency
Afghanistan	0.25	Guinea-Bissau	0.59	Nigeria	0.35
Albania	0.62	Hungary	0.63	Norway	0.96
Algeria	0.30	India	0.44	Oman	0.44
Angola	0.09	Indonesia	0.36	Pakistan	0.51
Armenia	0.31	Iran, Islamic Rep.	0.25	Philippines	0.36
Austria	0.79	Iraq	0.42	Poland	0.63
Azerbaijan	0.31	Ireland	0.96	Portugal	0.66
Bangladesh	0.55	Italy	0.48	Romania	0.46
Belarus	0.62	Japan	0.28	Rwanda	0.52
Belgium	0.80	Jordan	0.23	Saudi Arabia	0.47
Benin	0.20	Kazakhstan	0.23	Senegal	0.69
Bosnia and Herzegovina	0.57	Kenya	0.17	Serbia and Montenegro	0.48
Botswana	0.27	Korea, Rep.	0.28	Slovenia	0.63
Bulgaria	0.43	Kuwait	0.50	Somalia	0.23
Burkina Faso	0.65	Kyrgyz Republic	0.18	South Africa	0.27
Burundi	0.49	Lao PDR	0.02	Spain	0.66
Cambodia	0.35	Latvia	0.55	Sri Lanka	0.25
Cameroon	0.19	Lebanon	0.22	Sudan	0.41
China	0.08	Lesotho	0.32	Sweden	0.96
Congo, Dem. Rep.	0.12	Liberia	0.16	Switzerland	0.80
Congo, Rep.	0.12	Libya	0.41	Syrian Arab Republic	0.39
Croatia	0.62	Lithuania	0.61	Tajikistan	0.14
Czech Republic	0.76	Macedonia, FYR	0.56	Tanzania	0.14
Denmark	0.96	Madagascar	0.34	Thailand	0.05
Egypt, Arab Rep.	0.27	Malawi	0.12	Tunisia	0.16
Estonia	0.57	Malaysia	0.34	Turkey	0.34
Ethiopia	0.39	Mali	0.34	Turkmenistan	0.29
Finland	0.85	Moldova	0.42	Uganda	0.23
France	0.71	Mongolia	0.13	Ukraine	0.54
Gabon	0.12	Morocco	0.16	United Arab Emirates	0.37
Gambia, The	0.43	Mozambique	0.12	United Kingdom	0.95
Georgia	0.58	Myanmar	0.08	Uzbekistan	0.06
Germany	0.85	Namibia	0.09	Vietnam	0.35
Ghana	0.25	Nepal	0.50	Yemen, Rep.	0.53
Greece	0.55	Netherlands	0.85	Zambia	0.11
Guinea	0.39	Niger	0.56	Zimbabwe	0.15

B. Variable Definitions and Sources (Alphabetical Order)

Absolute Latitude

The absolute value of a country's representative latitude. Representative latitude is given by the centroid latitude of a country from *The World Factbook* (2011).

Average of Tsetse Suitability Index

An average measure of the ecological suitability of three species of tsetse fly: Fusca, Morsitans, and Palpalis. Data are from Wint and Rogers (2000).

Distance from Coast or Navigable River

The average distance in thousands of kilometers from an ice-free coast or navigable river. This variable is from the *Center for International Development*, which is derived from Gallup et al. (1999).

Distance to Technology Frontier

Technology frontiers are the two largest cities within each continent belonging to differing polities. For the Old World, the frontiers are London (UK), Paris (France), Cairo (Egypt), Fez (Morocco), Constantinople (Turkey), and Peking (China). Country-level distances, in thousands of kilometers, are calculated by the distance from a country's modern capital to the closest frontier. These data are from Ashraf and Galor (2011).

Fraction within Boreal

The fraction of a country within a Köppen-Geiger boreal climate. These data come from Gallup et al. (1999).

Fraction within Desert

The fraction of a country with sandy desert, dunes, rocky or lava flows. These data come from Nunn and Puga (2012).

Fraction within Dry Temperate

The fraction of a country within a Köppen-Geiger dry temperate climate. These data come from Gallup et al. (1999).

Fraction within Polar

The fraction of a country within a Köppen-Geiger polar climate. These data come from Gallup et al. (1999).

Fraction within Tropics

The fraction of a country with a Köppen-Geiger tropical climate. These data come from Nunn and Puga (2012).

Fraction within Subtropics

The fraction of a country with a Köppen-Geiger subtropical climate. These data come from Gallup et al. (1999).

Fraction within Wet Temperate

The fraction of a country within a Köppen-Geiger wet temperate climate. These data come from Gallup et al. (1999).

The Frequency of DQ 2.5

This variable is intended to be a proxy for the frequency of Celiac Disease. The data represent the fraction of a country's population containing the HLA-DQ 2.5 haplotype, or gene variant combination. The DQ 2.5 haplotype is the dual occurence of DQA1*0501 and DQB1*0201 genes. These data are given at the ethnic level, which are then matched to ethnic groups given by Ingram et al. (2009a) and aggregated to the country level by 1500 ethnic compositions. Data at the ethnic level can be found at allelefrequencies.net/hla6003a.asp (Gonzalez-Galarza et al. 2011).

The Frequency of Lactase Persistence in 1500 CE

Lactase persistence frequencies for Old World ethnicities are given by Ingram et al. (2009a). Ethnic data are then aggregated to the country level by matching ethnic groups from Ingram et al. (2009a) to compositions in Alesina et al. (2003) by language group similarities. This gives a contemporary, country-level measure for the frequency of lactase persistence. Contemporary ethnic compositions are modified by the inverse of the Putterman and Weil Migration Matrix (2010) to create representative ethnic compositions for the year 1500 CE.

GDP per capita in 2000

PPP converted GDP per capita in constant 2005 constant prices. Data come from the Penn World Table, version 7.1.

Genetic Distance from the U.K.

Genetic distance is a measure of genetic diversity between societies. This measure is calculated with the fixation index, or F_{ST}, from population genetics and measures the variation in gene frequencies across differing groups. F_{ST} scores are given for 42 indigenous populations; the data come from Cavalli-Sforza et al. (1994). The genetic distance measures are then aggregated to the country level by Spolaore and Wacziarg (2009), from which genetic distance to the UK is found for 206 countries. The UK is chosen as the technology frontier in 1500 CE. Genetic distance from this frontier is intended to convey difficulty in the diffusion of technology.

Genetic Diversity

Genetic diversity is the predicted country-level heterozygosity based on migratory distance from East Africa. These data represent the probability that two randomly selected individuals contain different gene variants at the same locus. The data are from Ashraf and Galor (2013).

Historic Intensity of Animal Husbandry

A measure from 0 to 1 percent that measures an ethnicity's historic dependency on animal husbandry. The measure is aggregated to the country-level by the contemporary ethnic composition of a country found within the Ethnologue (Lewis 2009). These data are from Alesina et al. (2013).

Historic Economic Development

An aggregated country-level index from 1 to 8 that represents historic economic development–nomadic or fully migratory, semi-nomadic, semi-sedentary, compact but temporary settlements, neighborhoods of dispersed family homes, separated hamlets forming a single community, compact and relatively permanent, complex settlements–for ethnicities within modern country borders. These data are from Alesina et al. (2013).

Land Productivity

Land productivity is the first principle component between a country's fraction of arable land and the country's suitability of agriculture. The fraction of arable land comes from the *World Development Indicators*. Suitability of agriculture is an index capturing soil and climate conditions favorable for agriculture. Suitability data are from Ramankutty et al. (2002) and aggregated to the country level by Michalopoulos (2012). Land productivity data are adopted from Ashraf and Galor (2011).

Malaria Ecology Index

The malaria ecology index takes into account differences in the environment and mosquito vectors that contribute to the spread of malaria. These data come from Kiszewski et al. (2004).

Mean Elevation

The average elevation in kilometers above sea-level. Data are from Nordhaus (2006) by way of Ashraf and Galor (2013).

Mean Precipitation

The average yearly precipitation in millimeters within a country between 1960 and 1990. Data are from Nordhaus (2006) by way of Ashraf and Galor (2013).

Mean Temperature

The average yearly temperature in Celsius within a country between 1960 and 1990. Data are from Nordhaus (2006) by way of Ashraf and Galor (2013).

Member of the Roman Empire

An indicator variable coded to one for countries with Roman heritage that were part of the Roman Empire but not belonging to the Ottoman Empire. These include Belgium, Britain, France, Italy, the Netherlands, Portugal, Spain, and Switzerland. These data are from Acemoglu et al. (2005).

Millennia of Agriculture

The millennia since the majority of a country's population adopted agriculture for subsistence. These data are from Putterman and Trainor (2006).

Number of Potential Domesticate Animals

The number of prehistoric, native animals that were a potential source of domestication within a country. These data are from Hibbs and Olsson (2004).

Number of Potential Domesticate Plants

The number of prehistoric, native plants that were a potential source of domestication within a country. These data are from Hibbs and Olsson (2004).

Population Density in 1, 1000, and 1500 CE

Population data for 1, 1000, and 1500 CE come from McEvedy and Jones (1978). Land area for each country is based on contemporary borders and is from the *World Development Indicators*. These data are adopted from Ashraf and Galor (2011).

Ruggedness

Ruggedness represents the average standard deviation of grid elevation within a country. These data come from Nunn and Puga (2012).

State History in 1500

This variable measures differing levels of societal formation in 50 year intervals from 1 CE to 1500 CE. The data are then aggregated to form an index for state history. These data come from Chanda and Putterman (2007).

Suitability of Land for Pasture

A suitability index, ranging from 0 to 1, for pasture within 5 arc-minute by 5 arc-minute grids, which is then averaged within modern country borders. This index takes into account climate, soil, and and terrain conditions necessary in developing grasslands. The raster data can be found at http://www.fao.org/geonetwork/srv/en/metadata.show?id=14167.

Suitability of Land for Plow-Negative Crops

The agricultural suitability for sorghum, maize, millet, roots, tubers, and tree crops. Data are from Alesina et al. (2013).

Suitability of Land for Plow-Positive Crops

The agricultural suitability for wheat, teff, barley, and rye. Data are from Alesina et al. (2013).

5 Tables and Figures

Table 1. Baseline Summary Statistics

Variable	N	Mean	Std. Dev.	Min	Max
Frequency of Lactase Persistence	108	0.4152	0.2388	0.0233	0.96
Europe	33	0.6816	0.1671	0.4246	0.96
Asia	37	0.3093	0.1505	0.0233	0.5783
Africa	38	0.287	0.1653	0.0853	0.6872
Population Density in 1500 CE	108	7.5214	9.2793	0.1429	45.9462
Millennia of Agriculture	108	5.4157	2.3988	1	10.5
ln Land Productivity	108	0.1244	1.2454	-4.3445	1.657
Percent of Arable Land	108	17.6569	14.5377	0.12	62.1
Agricultural Suitability	108	0.3696	0.2363	0.0034	0.8375
Suitability of Plow-Positive Crops	108	0.5881	0.3947	0	0.9977
Suitability of Plow-Negative Crops	108	0.4175	0.2352	0.0016	0.8519
Suitability of Land for Pasture	108	0.3551	0.2377	0.0222	0.82
Avg. of Tsetse Suitability Index	108	0.0811	0.1855	0	0.7094
Distance from Technology Frontier (in km)	108	2346.868	1669.689	1	6630.874
Member of Roman Empire	108	0.0741	0.2631	0	1
Absolute Latitude	108	29.7353	17.3115	0.5	64
Distance from Coast or River (in thousands of km)	108	0.3695	0.4516	0.0227	2.2917

Table 2. Baseline Estimates

	Dependent Variable: Log of Population Density in 1500 CE							
Sample:	Extended							Conservative
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Freq. of Lactase Persistence	2.4273***	2.4170***	2.1808***	2.6558***	2.1412***	2.2611***	1.5009***	0.9803
	(0.5087)	(0.6535)	(0.4840)	(0.6837)	(0.6811)	(0.6627)	(0.5046)	(0.8677)
ln Millennia of Agriculture			0.9778***				1.0127***	1.5038***
			(0.2587)				(0.3126)	(0.4617)
ln Land Productivity			0.6210***				0.5665***	0.4784***
			(0.0641)				(0.0578)	(0.1276)
Suit. of Land for Plow-Positive Crops				0.3693			-0.0225	-0.1088
				(0.5287)			(0.4051)	(0.4805)
Suit. of Land for Plow-Negative Crops				0.2443			-0.5443	-0.7444
				(0.7849)			(0.5483)	(0.7486)
Suit. of Land for Pasture				1.9449***			0.9578*	2.5840**
				(0.5640)			(0.4896)	(1.1728)
Avg. of Tsetse Suit. Index				1.1898*			0.2303	0.2738
				(0.6191)			(0.4414)	(0.4682)
ln Dist. to Technological Frontier					-0.1174**		-0.0692	-0.0107
					(0.0529)		(0.0636)	(0.0734)
Member of Roman Empire					0.5810*		0.8085***	0.5175
					(0.2971)		(0.2545)	(0.3811)
ln Absolute Latitude						-0.3744**	-0.1792	-0.3339
						(0.1545)	(0.1307)	(0.3092)
Dist. to Coast or River						-0.6086***	-0.5011***	-0.5718***
						(0.2281)	(0.1382)	(0.1391)
Region Effects:								
Continent Fixed Effects	N	Y	Y	Y	Y	Y	Y	Y
Scandinavia Dummy	N	Y	Y	Y	Y	Y	Y	Y
Obs.	108	108	108	108	108	108	108	49
R Sqr.	0.1958	0.3461	0.6522	0.4719	0.3829	0.4351	0.7696	0.8852

Summary: This table gives the base relationship between lactase persistence and population density in the year 1500 CE. According to the Malthusian theory, higher lactase persistence is associated with a transition to a higher production technology. This greater level of technology in the Malthusian era yielded larger populations. This relationship is established in the OLS estimated coefficients above. Col. (1) gives the bivariate estimate; col. (2) controls for continental effects; col. (3) controls for agriculture; col. (4) controls for domesticate animals; col. (5) controls for technology in 1500 CE; col. (6) controls for geography; and col. (7) includes all controls for the extended sample, while col. (8) restricts the sample to more conservative estimates for lactase persistence.

Notes: (i) The frequency of lactase persistence represents the fraction of a country in 1500 CE that is able to digest lactose, or milk. This frequency is calculated with the 1500-2000 migration matrix mentioned in Sec. 2.1.1. (ii) The sample of countries is restricted to only the Old World (i.e., Europe, Asia, and Africa). (iii) Continent dummies include indicator variables for Europe and Africa. (iv) The extended sample, columns (1)-(7), is composed by matching ethnic groups in Alesina et al. (2003) to those given by Ingram et al. (2009a) through language classifications; the conservative sample of Col. (8) represents direct matches between ethnic groups. (v) The natural log of land suitability is the first principle component of the percent of arable land and agricultural suitability. (vi) 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.

Table 3. Sample Truncation due to Migration

Dependent Variable: Log of Population Density in 1500 CE				
Fraction of Indigenous Population in 2000 CE:	> 50% (1)	> 75% (2)	> 85% (3)	> 95% (4)
Freq. of Lactase Persistence	1.5134***	1.4733***	1.4254**	1.6720**
	(0.5210)	(0.5415)	(0.5542)	(0.7709)
Controls:				
Baseline	Y	Y	Y	Y
Continent Fixed Effects	Y	Y	Y	Y
Scandinavia Dummy	Y	Y	Y	Y
Obs.	104	91	87	52
R Sqr.	0.7372	0.7453	0.7515	0.7842

Summary: This table shows that the positive relationship between the frequency of lactase persistence and precolonial population density is not influenced by potential bias in obtaining 1500 ethnic compositions from the inverse of the Putterman and Weil (2010) migration matrix. Columns (1)-(4) increasingly restrict the sample to countries with greater fractions indigenously derived since 1500 CE. This minimizes potential errors in measuring 1500 ethnic compositions.

Notes: (i) The baseline regression model given by Equation (1) and column (7) of Table 2 is used for all regressions of Table 3. (ii) The frequency of lactase persistence represents the fraction of a country in 1500 CE that is able to digest lactose, or milk. This frequency is calculated with the 1500-2000 migration matrix mentioned in Sec. 2.1.1 (iii) The sample of countries is restricted to the Old World (i.e., Europe, Asia, and Africa). (iv) Continent dummies include indicator variables for Europe and Africa. (v) The natural log of land suitability is the first principle component of the percent of arable land and agricultural suitability. (vi) 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.

Table 4. Within Continent Estimation

Dependent Variable: Log of Population Density in 1500 CE					
Continent:	Europe (1)	Africa (2)	Asia (3)	Africa + Asia (4)	Eurasia (5)
Freq. of Lactase Persistence	3.3379***	1.1040	1.8435	1.5925***	1.7763*
	(1.0792)	(0.7261)	(1.3867)	(0.5559)	(0.9536)
Controls:					
Baseline	Y	Y	Y	Y	Y
Continent Fixed Effects	–	–	–	Y	Y
Scandinavia Dummy	Y	N	N	N	Y
Obs.	33	38	37	75	70
R Sqr.	0.8866	0.8186	0.6971	0.7220	0.7334

Summary: This table shows that the positive relationship between the frequency of lactase persistence and precolonial population density remains within each Old World continent, as well as when Europe and Africa are excluded from the sample. Columns (1)-(3) perform the baseline regression within Europe, Africa, and Asia, respectively. Column (4) excludes European countries from the sample; this shows the effects of lactase persistence aren't driven solely by Europe. Column (5) shows that within Eurasia the ability to consume milk was associated with denser populations.

Notes: (i) The baseline regression model given by Equation (1) and column (7) of Table 2 is used for all regressions of Table 4. (ii) The frequency of lactase persistence represents the fraction of a country in 1500 CE that is able to digest lactose, or milk. This frequency is calculated with the 1500-2000 migration matrix mentioned in Sec. 2.1.1. (iii) Member countries of the Roman Empire are countries with Roman Heritage. This indicator variable is exclusive to European countries. (iv) The continent dummy in columns (4) and (5) is an indicator variables for Asia. (v) The natural log of land suitability is the first principle component of the percent of arable land and agricultural suitability. (vi) 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.

Table 5. Omitted Variables

Dependent Variable: Log of Population Density in 1500 CE								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Freq. of Lactase Persistence	1.8356***	1.8361***	1.5634**	2.2046**	2.5307***	1.9563**	2.7971***	1.8186
	(0.5736)	(0.6260)	(0.6037)	(0.8298)	(0.5568)	(0.7486)	(0.8207)	(1.0105)
Freq. of DQ 2.5 Haplotype		0.0059					3.8296	1.1120
		(2.3586)					(2.4147)	(3.6875)
Genetic Diversity			-23.8752***				-24.3579**	10.5604
			(8.8396)				(11.6406)	(26.6152)
ln Genetic Distance from UK in 1500 CE			-0.0446				0.0839	-0.0481
			(0.1139)				(0.0914)	(0.1503)
Malaria Ecology Index				-0.0406			-0.0136	-0.0054
				(0.0253)			(0.0273)	(0.0311)
Mean Elevation				-0.5877**			-0.1561	-1.1260
				(0.2428)			(0.2803)	(0.7528)
% in Tropics				-0.0009			0.0078	0.0206
				(0.0083)			(0.0072)	(0.0124)
% in Desert				-0.0290**			-0.0244*	-0.0439
				(0.0112)			(0.0128)	(0.0253)
% in Subtropics				-0.6562			-0.1946	-0.0284
				(0.5462)			(0.7351)	(0.6246)
% in Boreal				1.5086			-0.0633	5.7130**
				(1.4644)			(1.4658)	(2.1339)
% in Dry Temperate				-1.2198			-1.5501	0.9284
				(0.8349)			(0.9213)	(0.9574)
% in Wet Temperate				0.4361			-0.2712	3.9685**
				(1.0428)			(1.0331)	(1.7044)
% in Polar				3.3776			0.7687	5.8796
				(2.8109)			(2.7574)	(3.3079)
Mean Temperature				0.2643**			0.1934	0.3895
				(0.1189)			(0.1254)	(0.2178)
Temp. Sqr.				-0.0052*			-0.0044*	-0.0061
				(0.0026)			(0.0026)	(0.0041)
Mean Precipitation				0.8377*			0.9278	0.0257
				(0.4279)			(0.5599)	(1.4321)
Precip. Sqr.				-0.1164			-0.1520*	-0.0300
				(0.0706)			(0.0796)	(0.2030)
ln Ruggedness				0.1186			0.1065	0.2198
				(0.1151)			(0.1141)	(0.1975)
State History in 1500 CE					1.4432***		0.4200	0.2630
					(0.4041)		(0.5569)	(1.1527)
Historic Economic Complexity					0.0254		0.0314	0.8596**
					(0.0877)		(0.1252)	(0.2698)
Historic Intensity of Animal Husbandry					-1.7292**		-0.9900	1.9946
					(0.8191)		(1.2514)	(2.2302)
No. of Potential Domesticate Animals						0.1738*		0.0258
						(0.0922)		(0.1490)
No. of Potential Domesticate Plants						-0.0238		-0.0139
						(0.0332)		(0.0541)
Controls:								
Baseline Controls	Y	Y	Y	Y	Y	Y	Y	Y
Continent Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y
Scandinavia Dummy	Y	Y	Y	Y	Y	Y	Y	Y
N	70	70	70	70	70	46	70	46
R Sqr.	0.8314	0.8314	0.8590	0.9011	0.8808	0.8965	0.9312	0.9884

Summary: This table controls for potential omitted variables that may be causing the positive relationship between the frequency of lactase persistence and precolonial population density. The inclusion of additional, relevant controls does not alter the estimated relationship of interest. Column (1) adjusts the sample for countries that have data for all omitted variables; column (2) includes a control for celiac disease; column (3) includes additional genetic controls; column (4) includes additional geographic and climatic controls; column (5) includes additional historic cultural controls; column (6) includes biogeographic controls; column (7) includes additional controls from columns (2)-(5); and column (8) includes all additional variables.

Notes: (i) Baseline controls include the millennia a country has practiced agriculture, the first principle component of the fraction of arable land and an index for the suitability of agriculture, the suitability of land for plow-positive and plow-negative agriculture, the suitability of land to pastoralism, the ecological suitability of the tsetse fly, the great circle distance to a technology frontier, an indicator variable for whether a country has Roman Heritage, the absolute latitude of a country, and the mean distance to the coast or navigable river. (ii) The sample of countries is restricted to the Old World (i.e., Europe, Asia, and Africa). (iii) Continent fixed effects include indicator variables for Europe and Africa. (iv) 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.

Table 6. The Effect on, and Controlling for, More Historic Population Density

	Dependent Variable: Log of Population Density				
Year:	1 CE	1000 CE	1500 CE		
	(1)	(2)	(3)	(4)	(5)
Freq. of Lactase Persistence	0.8454	1.0171	1.1528**	0.3448*	0.6753**
	(0.6653)	(0.6445)	(0.5321)	(0.1791)	(0.2795)
Controls:					
ln Population Density in 1000 CE	–	–	–	Y	N
ln Population Density in 1 CE	–	–	–	N	Y
Baseline Controls	Y	Y	Y	Y	Y
Continent Fixed Effects	Y	Y	Y	Y	Y
Scandinavia Dummy	Y	Y	Y	Y	Y
Obs.	95	95	95	95	95
R Sqr.	0.7353	0.6681	0.7527	0.9560	0.8721

Summary: This table shows that the full effect of the frequency of lactase persistence on precolonial population density is realized in 1500 CE. Lactase persistence is a relatively new trait, which translated into denser populations in the year 1500 CE, but may not be associated with levels of development in previous periods. Additionally, controlling for earlier periods can potentially correct for reverse causality from the infectious disease environment. Column (1) regresses population density in 1 CE on the baseline estimating equation; column (2) considers population density in 1000 CE, while column (3) re-estimates the baseline regression while using the sample of countries that contain full data for the 3 periods considered. Columns (4) and (5) control for population density in 1000 and 1 CE, respectively.

Notes: (i) Baseline controls include the millennia a country has practiced agriculture, the first principle component of the fraction of arable land and an index for the suitability of agriculture, the historic use of plow agriculture, the suitability of land to pastoralism, the ecological suitability of the tsetse fly, the great circle distance to a technology frontier, an indicator variable for whether a country has Roman Heritage, the absolute latitude of a country, and the mean distance to the coast or navigable river. (ii) The sample of countries is restricted to the Old World (i.e., Europe, Asia, and Africa). (iii) Continent fixed effects include indicator variables for Europe and Africa. (iv) 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.

Table 7. The Effect on, and Controlling for, Urbanization

Dependent Variable:	Urbanization Rate in 1500 CE		Urbanization Rate in 1000 CE		Pop. Density in 1500 CE	
	(1)	(2)	(3)	(4)	(5)	(6)
Freq. of Lactase Persistence	-0.0136	-0.0739***	-0.0060	-0.0330	2.0437***	1.8641***
	(0.0152)	(0.0245)	(0.0147)	(0.0246)	(0.4931)	(0.4771)
Controls:						
Urbanization Rate in 1500 CE	–	–	–	–	Y	N
Urbanization Rate in 1000 CE	–	–	–	–	N	Y
Baseline	N	Y	N	Y	Y	Y
Continent Fixed Effects	N	Y	N	Y	Y	Y
Scandinavia Dummy	N	Y	N	Y	Y	Y
Obs.	105	105	105	105	105	105
R Sqr.	0.0065	0.2399	0.0012	0.1400	0.7858	0.7808

Summary: This table illustrates the Malthusian theory of Ashraf and Galor (2011) by showing that a higher frequency of lactase persistence, while being positively associated with population density, does not have a statistically positive relationship with historic income. Columns (1)-(2) show the frequency of lactase persistence is not positively related to the urbanization rate in 1500 CE, while columns (3)-(4) replicate the findings of columns (1) and (2) using the urbanization rate for 1000 CE, and columns (5)-(6) consider the urbanization rate as a proxy for the infectious disease environment to partially account for reverse causality.

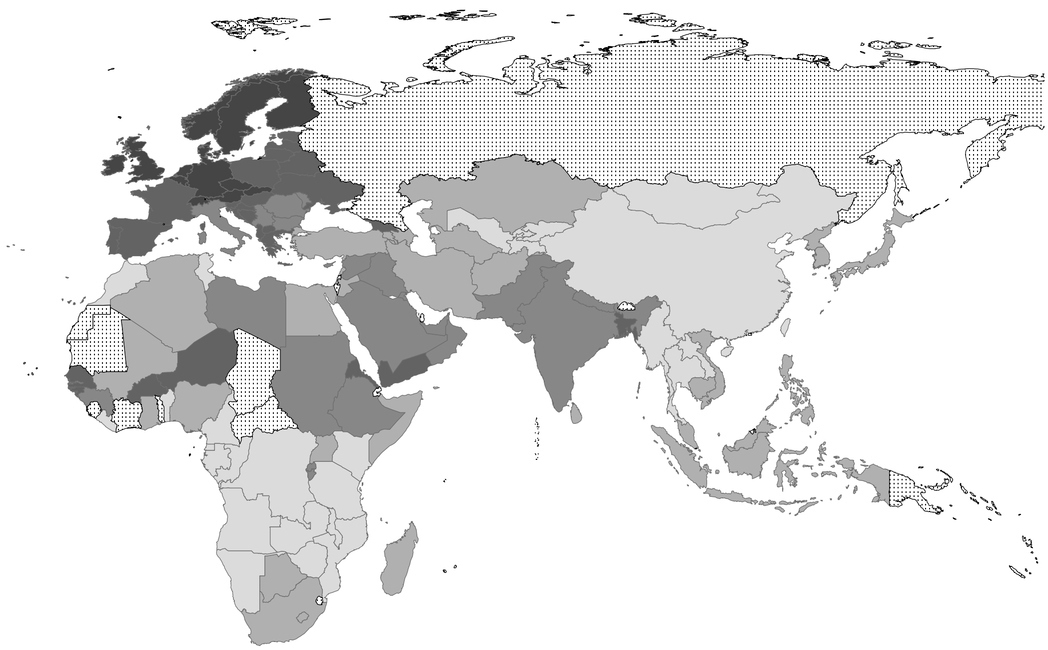
Notes: (i) Baseline controls include the millennia a country has practiced agriculture, the first principle component of the fraction of arable land and an index for the suitability of agriculture, the historic use of plow agriculture, the suitability of land to pastoralism, the ecological suitability of the tsetse fly, the great circle distance to a technology frontier, an indicator variable for whether a country has Roman Heritage, the absolute latitude of a country, and the mean distance to the coast or navigable river. (ii) The sample of countries is restricted to the Old World (i.e., Europe, Asia, and Africa). (iii) Continent fixed effects include indicator variables for Europe and Africa. (iv) 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.

Table 8. Contemporary Effects of Lactase Persistence

Dependent Variable:	ln Pop. Den. in 2000 CE	ln GDP per capita in 2000 CE
	(1)	(2)
Freq. of Lactase Persistence	0.3682	-0.1843
(Contemporary Ethnic Compositions)	(0.4924)	(0.5631)
ln Population Density, 1500 CE	0.3660***	0.2280**
(Ancestry Adjusted)	(0.0764)	(0.1131)
ln Millennia of Agriculture	0.2748	-0.5950
(Ancestry Adjusted)	(0.2514)	(0.3628)
ln Ethnic Fractionalization	-0.0538	-0.2924**
	(0.0878)	(0.1229)
ln Land Productivity	0.5168***	-0.4434***
	(0.1037)	(0.1053)
Suit. of Land for Plow-Positive Crops	-0.1006	0.6984
	(0.3625)	(0.4825)
Suit. of Land for Plow-Negative Crops	-0.7817*	-0.6923
	(0.4146)	(0.5195)
Suit. of Land for Pasture	0.0646	-0.5778
	(0.3372)	(0.5840)
Avg. of Tsetse Suit. Index	-1.2142***	-0.1137
	(0.4140)	(0.6764)
ln Absolute Latitude	-0.1355	-0.1264
	(0.1224)	(0.2721)
Dist. to Coast or River	-0.6836***	-0.5709*
	(0.2373)	(0.2884)
Continent Fixed Effects	Y	Y
Scandinavia Dummy	Y	Y
N	97	97
R Sqr.	0.7937	0.7451

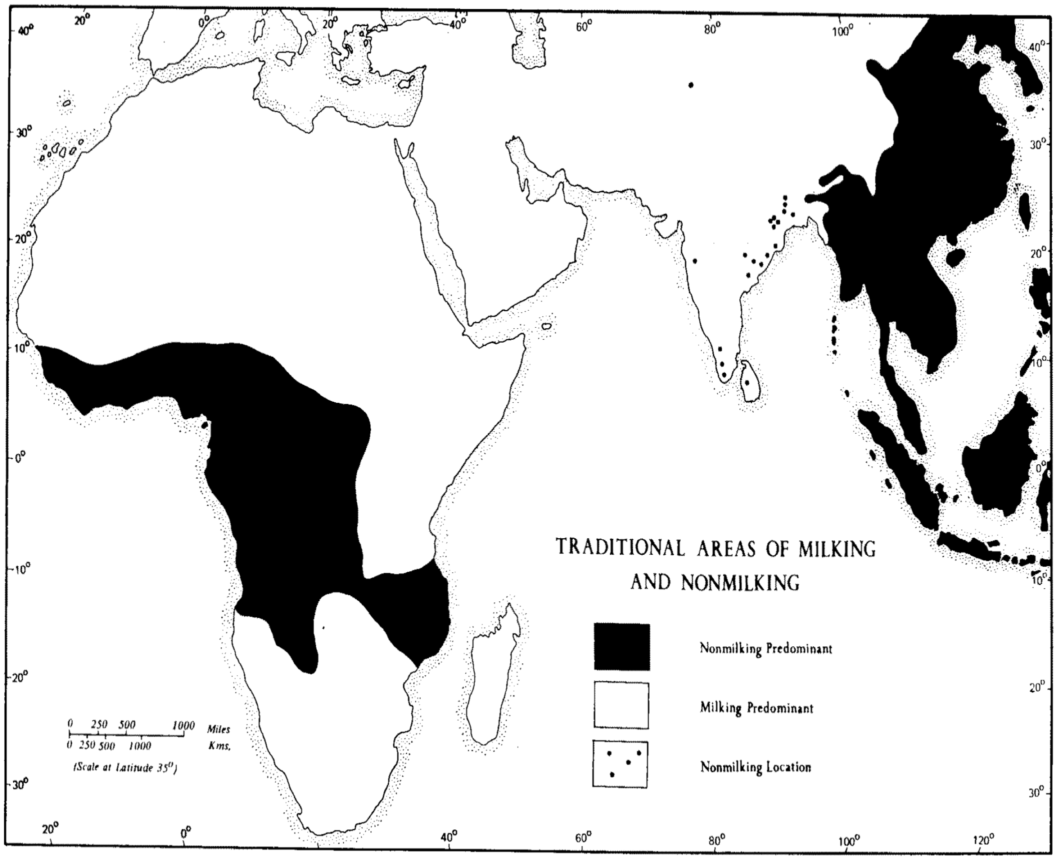
Summary: This table shows the effects of lactase persistence on contemporary population density and per capita GDP. Conditional on relevant controls, lactase persistence has no direct effect on contemporary economic development.

Notes: (i) The frequency of lactase persistence is based on contemporary ethnic compositions from Alesina et al. (2003). (ii) The sample of countries is restricted to the Old World (i.e., Europe, Asia, and Africa). (iii) Continent fixed effects include indicator variables for Europe and Africa. (iv) 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.



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. Country-Level 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)

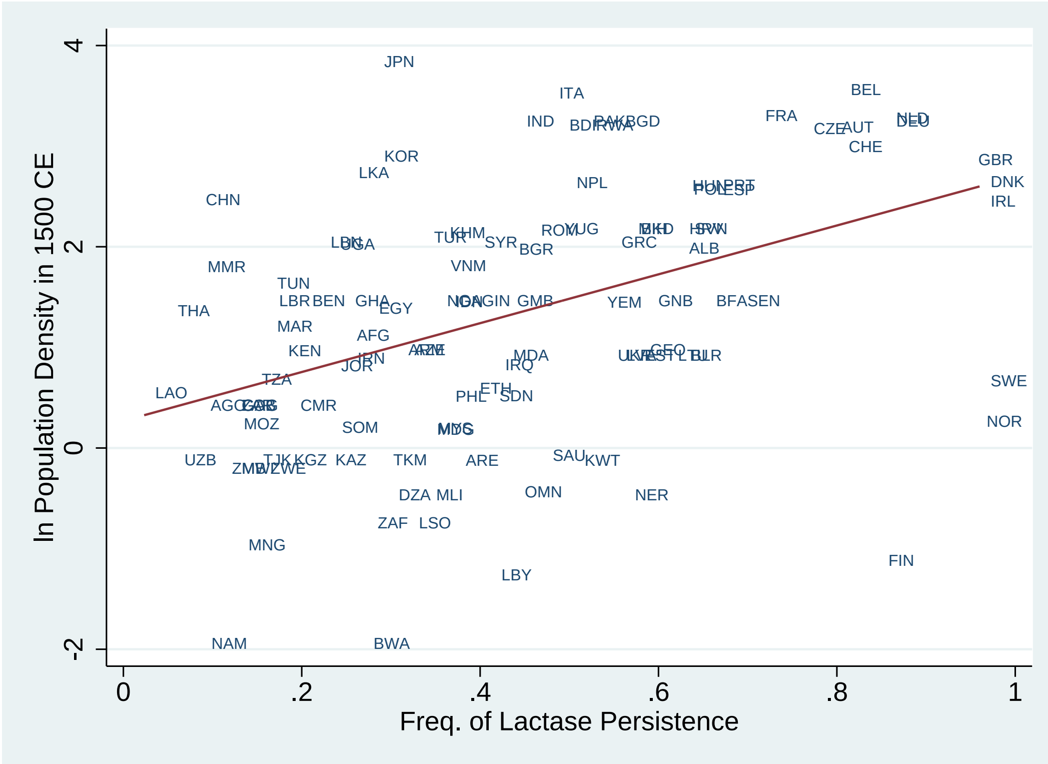


Figure 3. The Freq. of Lactase Persistence and the ln of Pop. Density in 1500 CE

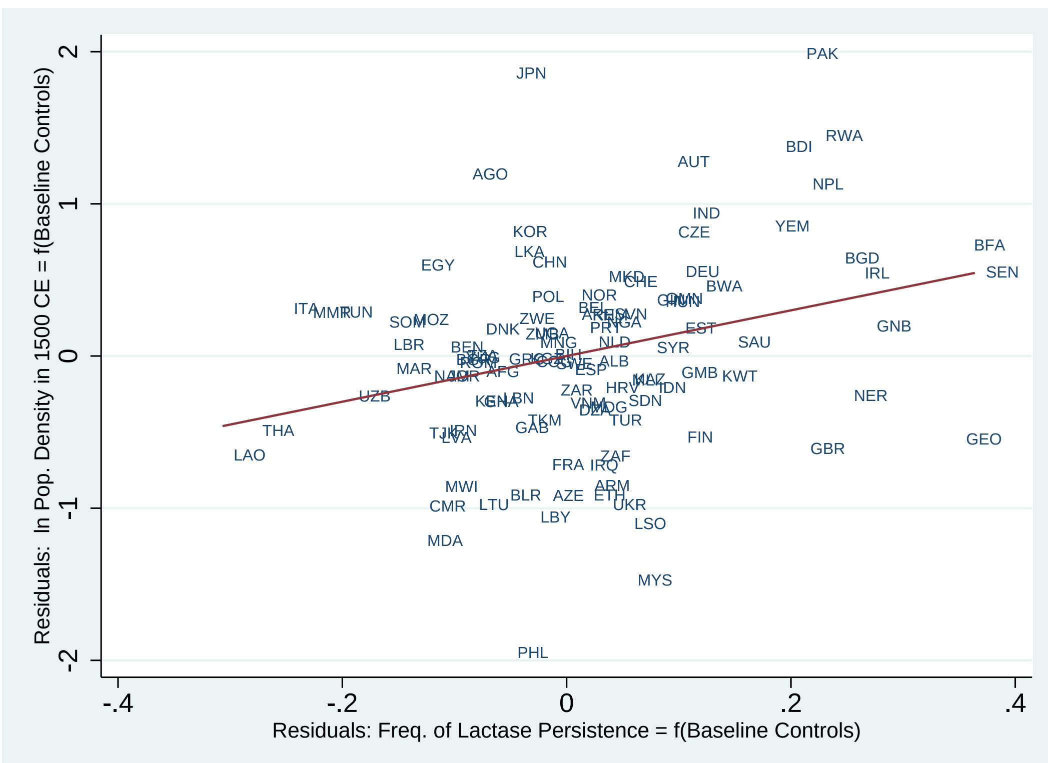


Figure 4. Orthogonalized Plot of Col. (7), Table 2