Title

The Origins and Maintenance of Female Genital Modification across Africa

Permalink

https://escholarship.org/uc/item/2vm5p0f2

Journal

Human Nature, 27(2)

ISSN

1045-6767

Authors

Ross, Cody T
Strimling, Pontus
Ericksen, Karen Paige
et al.

Publication Date

2016-06-01

DOI

10.1007/s12110-015-9244-5

Peer reviewed

Hum Nat
DOI 10.1007/s12110-015-9244-5

The Origins and Maintenance of Female Genital Modification across Africa

Bayesian Phylogenetic Modeling of Cultural Evolution under the Influence of Selection

Cody T. Ross[1] · **Pontus Strimling**[2] · **Karen Paige Ericksen**[3] · **Patrik Lindenfors**[2] · **Monique Borgerhoff Mulder**[4]

Abstract We present formal evolutionary models for the origins and persistence of the practice of Female Genital Modification (FGMo). We then test the implications of these models using normative cross-cultural data on FGMo in Africa and Bayesian phylogenetic methods that explicitly model adaptive evolution. Empirical evidence provides some support for the findings of our evolutionary models that the de novo origins of the FGMo practice should be associated with social stratification, and that social stratification should place selective pressures on the adoption of FGMo; these results, however, are tempered by the finding that FGMo has arisen in many cultures that have no social stratification, and that forces operating orthogonally to stratification appear to play a more important role in the cross-cultural distribution of FGMo. To explain these cases, one must consider cultural evolutionary explanations in conjunction with behavioral ecological ones. We conclude with a discussion of the implications of our study for policies designed to end the practice of FGMo.

Electronic supplementary material The online version of this article (doi:10.1007/s12110-015-9244-5) contains supplementary material, which is available to authorized users.

✉ Monique Borgerhoff Mulder
mborgerhoffmulder@ucdavis.edu

Cody T. Ross
ctross@ucdavis.edu

[1] Behavioral Sciences Program, Santa Fe Institute, Santa Fe, NM, USA

[2] Centre for Studies of Cultural Evolution, Stockholm University, Stockholm, Sweden

[3] Department of Psychology, University of California, Davis, Davis, CA, USA

[4] Department of Anthropology, University of California, Davis, Davis, CA, USA

Keywords Female circumcision · FGM · Cultural evolution · Social transmission · Marriage markets · Phylogenetic modeling

Introduction

In many parts of the world the prevailing culture requires that women undergo genital cutting/modification prior to their marriage. These operations typically occur between infancy and late puberty, and they may be extremely costly in terms of health, survival, and reproduction; these operations thus constitute a major challenge for evolutionary explanation. The analyses presented herein test competing hypotheses for the distribution of this cultural trait, with a particular focus on how the conditions that may have selected for the emergence of the trait (arguments for origins) might differ from those that account for its persistence (arguments for maintenance). We first present a model that demonstrates how conditions selecting for the origins of female genital modification (FGMo)[1] might differ from those that select for its persistence. In particular, we model how wealth and/or status differences within social groups might select for the emergence of this potentially costly trait, and how subsequent frequency-dependent forces might keep the trait in the population, despite its costs. We then use empirical data from Africa to test whether status differences and social stratification are associated with the origins of FGMo. Our expectation (derived from the model) is that stratification will be associated with the de novo origins of FGMo and will increase the likelihood of the intercultural transmission of FGMo. We use our results to discuss the adaptive significance of apparently costly cultural traits, and to discuss the importance of differentiating explanations for the origins and the maintenance of cultural traits.

Our study lies at the intersection of several literatures. First is the substantial literature on how marriage markets affect a range of cultural practices, including marriage payments (Anderson 2003; Bell and Song 1994; Borgerhoff Mulder 1995) and other household outcomes (Chiappori et al. 2002; Quisumbing et al. 2000). Many of these insights are inspired by Becker (1981). Second, we develop hypotheses on the basis of a wide-ranging research literature on female "circumcision" (Ericksen 1989; Shell-Duncan and Hernlund 2000; Silverman 2004). While contributors to this literature span many different disciplines, there is considerable agreement over the hypothesized functions of FGMo and related practices, though less agreement with regards to its specific geographical origin. Third, we draw

---

[1] Also known as Female Circumcision, Female Genital Mutilation (FGM), Female Genital Cutting (FGC), or a combination of these terms, such as FGM/C. The terminology one should use when discussing this practice is a matter of concern. We purposefully avoid using the term "mutilation" in the text of this paper because we feel that it is unduly value-laden. Likewise, we feel that it is wrong to distance the practice of female genital modification from male genital modification (circumcision) because such an action seems to validate one type of unnecessary, non-consensual removal of genital tissue (common in "Western" culture), while stigmatizing a similar practice in other cultures. We use the more neutral term FGMo to contextualize the practice within the wider anthropological scope of body modification.

indirectly on the insights of signaling (or handicap) theory, as developed in evolutionary biology and applied in anthropology (Grafen 1990; Smith et al. 2003), viewing FGMo as a signal demonstrating sexual fidelity and paternity certainty. Fourth, we use formalized evolutionary models to identify distinct hypotheses for the origins and maintenance of costly cultural traits, as these potentially differing dynamics are rarely separated in analysis (Borgerhoff Mulder et al. 2006). Fifth, a growing (albeit widely-dispersed) literature across academic venues and policy circles addresses the morality of, and challenges to, eradicating the practice of FGMo (Shell-Duncan and Hernlund 2000). We use our results to address some of the debates in this complex area.

Female Genital Modification

The prevalence of FGMo worldwide is unknown, but it is estimated that more than 125 million girls and women alive today have undergone some form of FGMo, with another 30 million at risk of being cut in the next decade (UNFPA-UNICEF 2013). It is most common in Africa, affecting up to 90% of women in Djibouti, Egypt, Eritrea, Mali, Sierra Leone, and Somalia, and more than 50% of women in other African countries, such as Benin, Ethiopia, Burkina Faso, Chad, Ethiopia, Gambia, Guinea Bissau, Kenya, Liberia, Nigeria, Sudan, and Togo. It occurs elsewhere across southern and central Africa and in the Middle East—for example, in Oman, Yemen, and the United Arab Emirates, as well as in some Asian countries, including Indonesia, Malaysia, Sri Lanka, and India (WHO 1998).

Little is known about the origins of the practice. Scholars have proposed a single origin in Egypt, on the basis of circumcised fifth century BC mummies (Little 2003), or ancient Meroe (Mackie 2000; Mackie and LeJeune 2009). In these highly stratified ancient empires, infibulation may have been practiced in the context of extreme resource inequality, with families cutting girls or women to signal their fidelity to wealthy, highly polygynous males. Mackie has hypothesized that the practice diffused across social strata and spread along female slave trade routes. Others suggest a multi-source origin, arguing that as FGMo spread out of its original core areas, it encountered and merged with preexisting practices associated with initiation rites for both males and females (Dorkenoo 1994). Strong evidence for either claim is nevertheless lacking. Although the practice sometimes spread with Islam, many non-Islamic groups practice FGMo.

Types of operations vary in severity, entailing the removal of the prepuce or hood of the clitoris (clitoridectomy), the clitoris and all or part of the labia minora (excision), or the complete clitoris, labia minora, and all or part of the labia majora, followed by a sewing together of the labia (infibulation, or Pharaonic circumcision).

The health costs (and benefits) of FGMo are hotly disputed (e.g., Shell-Duncan and Hernlund 2000). Estimates are likely to be biased by the complications being concealed where the practice is illegal, exaggerated by prejudice and by proponents of elimination strategies, and impacted by lack of good data. Furthermore, Western observers tend to stress pain, reduced sexual enjoyment, and medical complications, whereas (some) African scholars emphasize the cultural importance of the tradition; for example, a Kenyan woman commented that FGMo might be seen as "buying maturity with pain" (Davison 1996:60). Nevertheless, the evidence is clear that these operations, and particularly the more severe forms, can cause extensive short- and long-term medical complications, with implications for maternal health, pregnancy loss and stillbirth, primary and secondary fertility, and child survival (Mackie 2003; Shell-Duncan and Hernlund 2000; Banks et al. 2006), as well as for mental health (e.g., Whitehorn et al. 2002), and HIV risk (WHO 2010).

Most pertinent to this paper are arguments regarding the function of FGMo, or similar practices with likely negative effects for women's health and well-being. There are three types of explanations. The first, and by far the most predominant, concerns the marriageability of women. Though specifics differ markedly in different parts of the world, virginity, "purity," and sexual restraint before marriage are highly regarded in many societies that practice FGMo. FGMo operations are seen as a way of protecting sexual propriety, morality, and paternity (Ericksen 1989; Little 2003; UNFPA-UNICEF 2013), or demonstrating the obedience and respect required for marriageability (WHO 2010). Furthermore, FGMo is often associated with veiling, child betrothal, virginity testing at marriage, a transfer of sexual and reproductive rights to the groom and his family at marriage, and the legitimacy of subsequent births. FGMo is sometimes found in groups that lack strong chastity and fidelity expectations—for example, the Rendille women of Kenya are free to engage in premarital sex but must undergo FGMo at marriage (Mackie 2000). Nevertheless, scholars investigating FGMo across multiple sites typically stress the theme of safeguarding female sexual purity, enhancing marriage chances, and preserving family honor (Shell-Duncan and Hernlund 2000; UNFPA-UNICEF 2013; WHO 2010). For instance, uninfibulated women in Sudan during the early 1970s were widely considered to be like prostitutes (Hayes 1975).

A second set of explanations is more heterogeneous. They include protecting the health of a baby, increasing the likelihood of conception, reducing (or increasing) sexual pleasure, achieving an aesthetic ideal, or becoming a fully adult member of society; these more proximate considerations are also linked to marriageability, albeit indirectly. Particularly idiosyncratic notions include the belief that FGMo prevents child mortality [through avoiding fatal connection between baby's head and the clitoris during delivery (Myers et al. 1985)], that FGMo cured certain "female psychological disorders" in Victorian England (Little 2003), and that genital cutting is a functionless fad, as witnessed recently in Chad. Although FGMo in southern Chad may have been adopted in the nineteenth century to deter slave raiders from taking women, adolescent girls in the 1990s reportedly sought out the operation, often without parental knowledge, as something "fun, rebellious and cool" (Leonard 2000:190).

The third explanation stresses the importance of tradition, custom, and cultural identity. For instance, the very name of the Kipsigis of Kenya implies "we the circumcised," referring to the rebirth that is believed to occur at circumcision ceremonies, a central part of Kipsigis identity vis-a-vis other ethnic groups (Daniels 1970). Even more strategically, the Kenyan Mukogodo appear to have adopted circumcision ceremonies to hasten their transition in becoming Maasai (Cronk 2004). In a compilation of studies on attitudes toward FGMo, more than half the respondents reported that tradition was their primary reason for undergoing the operation (Warzazi 1986). Some authors emphasize only the importance of following the ancestors; others note strong social pressure to conform to the behavior of others, and still others explicitly recognize the force of current social convention.

As many researchers on FGMo recognize, there are intricate interdependencies between these hypotheses, even in cases where marriageability is not the explicitly-stated motivation for FGMo (for example, as in the Senegambian region [Shell-Duncan and Hernlund 2007]). It is extremely difficult to distinguish the more general motivations of social respectability and conformity from the motivation to be seen as a suitable wife (Mackie and LeJeune 2009). Furthermore, some of the more idiosyncratic beliefs underlying the need for FGMo may also serve as a way of encouraging conformity to social norms. Accordingly Mackie (2000)—emphasizing the critical role of social convention in maintaining particular practices—notes how the belief that an uncut woman will be unfaithful fixes the custom, even if the conditions that first gave rise to the trait have changed: "As soon as women believed that men would not marry an unmutilated woman, and men believed that an unmutilated would not be a faithful partner in marriage, the convention was locked in place" (Mackie 2000:264).

An Evolutionary Approach

Our evolutionary analysis of FGMo draws closely on the literatures reviewed above. Parents or other kin may choose to subject young female relatives to costly operations to enhance their marriageability. In other words, the trait may be sexually selected, functioning to enhance the access of females to favored mates. The intuition here is that in an initial non-FGMo population parents use FGMo to ensure and signal the virginity of their daughters at marriage. Insofar as women who have undergone FGMo observe more sexual restraint (whether as a consequence of the operation or of associated customs), FGMo also signals higher paternity certainty to a prospective groom. Where there is little difference in quality among prospective grooms, it is unlikely that parents would inflict this physiological cost on their daughters. However, where there is extensive competition among women to become the brides of particularly wealthy or powerful men, a costly trait such as FGMo could arise as a bargaining tool. The simple logic here parallels evolutionary arguments for the origin of another costly display—dowry, and its association with stratification (Dickemann 1979; Gaulin and Boster 1990; see also Fortunato et al. 2006)—and has in fact been marshaled as an explanation for the origins of FGMo (Mackie 2000). Some support comes from evidence that female puberty rituals are more elaborate and costly as social complexity increases, although this is based on a non-random sample drawn from a selection of foraging societies deemed to be only weakly stratified (Owens and Hayden 1997). Other support comes from comparative analysis that links public menarcheal ceremonies to strong fraternal interest groups, where it appears that powerful kin-based units are making explicit contractual negotiations over women (Ericksen 1989; Paige and Paige 1981).

As noted above, other functions may contribute to the practice of FGMo over time. For example, FGMo (at least when accompanied by ritual) can constitute a key rite of passage (Van Gennep 1960); it is also a cardinal symbol of ethnicity since groups that do not practice FGMo are sometimes seen as deficient or inferior by groups that do. There are also clear reasons why it is difficult to dislodge FGMo once it is in place as a marker of marriageability since a family that abjures FGMo for their daughters may fail to find husbands, thereby possibly reducing their daughters' contribution to the next biological and cultural generation. Evolutionary scientists refer to this as a *coordination game* (e.g., Efferson et al. 2015), and it can strongly influence the payoffs to FGMo in some contexts. Without a critical mass of FGMo-rejecters, there can be a strong disincentive for families to reject the practice; in addition, having increased payoffs in the mating market as a function of FGMo use might also disincentivize abandonment of the practice.

These arguments lead to the hypothesis that the *origins* of FGMo are likely to be linked to stratification, based on the rationale that where considerable variance among men in mate quality exists, and where men invest heavily in their offspring, parents will be willing to inflict possible physiological (and other) costs on their daughters in order to reap the benefits of marriage to a high-ranking man. Conversely, the custom might *persist* purely as a consequence of the high cost of not conforming, irrespective of the presence of high variance in wealth or status among men.

Modeling the Emergence and Persistence of FGMo

We begin methodologically by providing formal theoretical models for the emergence and persistence of FGMo. We then use phylogenetic models of adaptive evolution to test the implications of the theoretical models using cross-cultural data.

Assumptions and Notations

The following assumptions and notations are constant across all models:

a. Mating assessment is universal—that is, each woman agrees on a mating value for each man, and each man agrees on a mating value for each woman. A given woman's mating value is denoted f (and a given man's, m). The distributions for the mating values of women and men are d_f and d_m, respectively. Mating values are reflective of the integrated suite of traits of interest to the opposite sex (Schacht and Grote 2015).
b. The symbols r_f and r_m denote the functions that map the mating value of an individual woman or man to her or his rank in the mating market. These functions lead to a ranking of all women where the woman with the highest mating value gets rank $r_f = 1$, the woman with the second highest mating value gets rank $r_f = 2$, and so on. The woman with rank $r_f = 1$ is then assumed to marry the man with rank $r_m = 1$, and so on. Thus, if f_1 is the woman with the highest

mating value in the d_f distribution, then her rank is described as $r_f(f_1, d_f) = 1$; she will be paired with the male whose $r_m = 1$, the highest-ranked male from the d_m distribution.

c. The marriage value of a given person in d_f or d_m can be returned as a function, V_f or V_m, of rank, such that $V_m(r_m(m_1, d_m))$, for example, is the marriage value of the highest-ranked man. The marriage value function is order preserving but allows for a nonlinear relationship between rank order and marriage value; a unit change in rank may not significantly impact marriage value near the bottom of a ranking scale, but it might have a huge impact on marriage value near the top of a ranking scale—especially in stratified societies.
d. Everyone gets married.
e. FGMo comes at a constant cost, c.
f. FGMo gives a woman a higher mating value and, therefore, an opportunity to marry a more valuable man.
g. The sex ratio is balanced such that the number of women in the population, N_f, is equal to the number of men in the population, N_m.
h. There is a benefit, s, to having the same behavior as the rest of the group. Generally, s is a function of the proportion of women, α, that have the FGMo behavior. We define $s(1/N) = 0$, indicating that no social benefit will accrue to an individual who is the only one in a population to have a given trait. We conceive of this pressure as resulting from frequency-dependent dynamics, such as conformity bias (Henrich and Boyd 1998; Henrich and Gil-White 2001; McElreath et al. 2003; and see Discussion).
i. Baseline fitness across women is described as $\bar{w}$.

The General Model

We will consider two models. In the first, FGMo gives an assurance of virginity, and therefore, a constant increase in mating value, i, to every woman who has FGMo. In the second model, FGMo acts as a costly signal of general fitness. In this model, we treat the increase in mating value as a random variable, x, since different women can signal differential fitness in a nondeterministic way. Presumably, a woman will not know the effect of the signal before she undergoes the operation. Death, for example, is a possible, and not entirely uncommon, outcome of undergoing genital cutting (UNICEF 2013). Further, the signaling value of FGMo is likely to depend on the social context of the practice. For example, in the Kipsigis population studied by Borgerhoff Mulder in the 1980s, FGMo was often practiced in semi-public ceremonies and girls were socially judged on their stoicism and ability to withstand the pain of the cutting. Recovery from the clitorectomy is the subject of discussion and social scrutiny, with considerable gossip (Power 1998, 2000) over who recovered quickly and who needed to be taken to the dispensary or hospital. The ability to tolerate pain and recover quickly are presumably signals of quality and immunocompetence that are impossible to fake. In contrast, FGMo in Egypt (studied by Ericksen in the 1990s) was conducted in private by midwives; in such a context there is much less scope for FGMo to act as a costly signal. Figure 1 illustrates

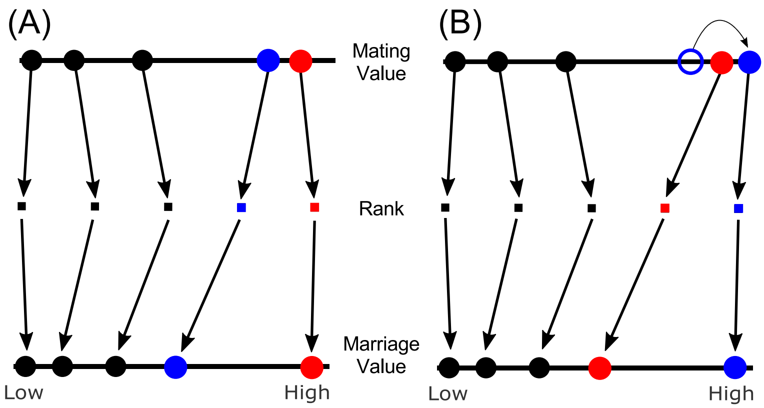


Fig. 1 A simplified sketch of the mathematical model described in this section. In frame (**a**), we observe a distribution of mating values (on the real number line) for five women. We have highlighted two individuals in red (highest-ranked) and blue (second highest-ranked). The arrows from Mating Value to Rank are indicative of the rank function r_f and map mating values to an ordered ranking. The arrows from Rank to Marriage Value are indicative of the value function V_m and map a woman's rank score on the marriage market to the marriage value of her husband (on the real number line). In frame (**b**), we imagine that the woman shown in blue has undergone FGMo and has thus added i or x to her mating value. This allows her to move from her previous position on the Mating Value scale (shown in the blue outlined circle) past the woman labeled in red. This in turn increases her rank and allows her access to the marriage value of the highest-ranked man. (Refer to online version for colors)

how the mathematical mechanisms outlined here function to produce the behavior of the model.

In both models, a woman who does not undergo FGMo has a payoff, P:

$$P = \bar{w} + V_m(r_f(f, d_f)) + s(1 - \alpha) \tag{1}$$

where α is the proportion of the population who undergo FGMo. Equation 1 says that the value she gets from marrying depends on the marriage value of the man she marries, which depends on her rank, $r_f(f, d_f)$, which in turn depends on her value, f, and the distribution of mating values across all females, d_f. The last term is the social payoff from having the same behavior as proportion $1 - \alpha$ of the population.

The Virginity Assurance Model

In the virginity assurance model, the payoff, P_{VA}, for undergoing FGMo is:

$$P_{\mathrm{VA}} = \bar{w} + V_m(r_f(f + i, d_f)) + s(\alpha) - c \tag{2}$$

Under this condition, FGMo will emerge when, for any woman, $P_{VA} - P > 0$, for $\alpha = 0$. That is, when:

$$V_m(r_f(f + i, d_f)) - V_m(r_f(f, d_f)) - s(1) - c > 0 \tag{3}$$

So for FGMo to emerge, we need at least one (or a combination) of the following: low social control or conformity bias, s; a high value of FGMo, i; or a large difference in the marriage value of men, V_m. We will not elaborate on how FGMo spreads through the population, nor will we discuss all possible equilibria that can appear. Rather, we focus on what is needed for an equilibrium in which every woman in the population undergoes FGMo. In the virginity assurance model, this occurs for cases such that when $\alpha = 1$, then $P_{\mathrm{VA}} - P > 0$ holds for all women in the population. That is, where

$$V_m(r_f(f+i, d_f)) - V_m(r_f(f, d_f)) + s(1) - c > 0 \tag{4}$$

Here two things have changed from the emergence condition in Eq. 3; the social effect has switched so that it now benefits FGMo, and also the value distribution of women, d_f, has changed since each woman has added i to her value. This means that when $\alpha = 1$ each woman is in the same position on the ranking scale as she would have been had no one undergone FGMo.

So Eq. 4 can be reduced to:

$$s(1) - c > 0 \tag{5}$$

Several interesting conclusions can be drawn from Eq. 5. Whereas the marriage value function for males, V_m, is critical in explaining the origins of FGMo, it has no effect on the *maintenance* of the trait at *fixation*. Another conclusion is that strong social pressure is required to obtain an equilibrium in which every woman has undergone FGMo. Finally, we can see that the difference in payoff between being a woman in a society in which every woman undergoes FGMo and a society in which no one does is strictly negative for each woman since the rankings are unchanged between the FGMo and non-FGMo equilibria, but all women pay a cost, c, in the FGMo equilibrium.

The Costly Signaling Model

Here we assume that FGMo functions as a general costly signal of fitness, rather than a signal of virginity. It is unreasonable to imagine that this effect is the same for each woman or child, or that the effect is known before the decision to commit FGMo is taken. Therefore, we model the mating value increase from having FGMo as a random variable, x. The payoff, P_{CS}, to undergoing FGMo is thus:

$$P_{CS} = \bar{w} + V_m(r_f(f+x, d_f)) + s(\alpha) - c \tag{6}$$

And FGMo emerges in a population if the expected value of P_{CS} exceeds P for at least one woman in the population when $\alpha = 0$. That is, when

$$E[V_m(r_f(f+x, d_f))] - V_m(r_f(f, d_f)) - s(1) - c > 0 \tag{7}$$

An equilibrium where everyone in the population undergoes FGMo is maintained if $E[P_{CS}] > P$ for every woman in the population when $\alpha = 1$. That is, when

$$E[V_m(r_f(f+x, d_f))] - V_m(r_f(f, d_f)) + s(1) - c > 0 \tag{8}$$

And here the difference from the virginity assurance model becomes clear. If we again look at a case in which a low-ranking woman in a non-FGMo society considers FGMo, she can (depending on the distributions of x, d_f and d_m) still have

$$E[V_m(r_f(f+x, d_f))] - V_m(r_f(f, d_f)) > 0 \tag{9}$$

Even though the probability of having a higher payoff after FGMo might be small, FGMo could have a high enough expected value to be worthwhile. This also means that there might be women who benefit from living in a society in which every woman undergoes FGMo, which was not the case in the virginity assurance model. On average though, each woman still gets c less payoff.

For a better understanding of our results, let's look closer at the implications of Eq. 9. For the woman with lowest mating value before FGMo, Eq. 9 can be expressed as:

$$\sum_{k=0}^{Nf} \Phi(k) V_m(N_f - k) - V_m(N_f) > 0 \tag{10}$$

where $\Phi(k)$ is the probability that undergoing FGMo gives her a higher value than k other women, and N_f is the number of women in the population. Note that as long as $\Phi(k)$ is not strictly zero for all k, then the left-hand side of Eq. 10 is strictly positive, although it is not necessarily greater than c. Thus, FGMo could function to increase a woman's mating value relative to other women, though this is not guaranteed generally.

To further investigate the implications of this model, we make the reducing assumptions that d_f and x are normally distributed random variables. In this case, all that matters to our model are the variances of the respective distributions; to understand the impact of variance in x and d_f, let us note that the probability that a realization from Normal(μ_1, σ^2) is greater than a realization from Normal(μ_2, σ^2) is just the probability that Normal($\mu_1 - \mu_2, 2\sigma^2$) > 0, which is determined by the difference between μ_1 and μ_2 and by σ. In our model, the difference between μ_1 and μ_2 is determined by the variance in the mating value of women, d_f, and σ is the variance of the random variable x.

An increasing variance of d_f decreases the chance that a low-ranking woman can move up on the scale as a function of FGMo adoption. An increase in the variance of x increases the chances that a low-ranking woman can move up on the mating value ranking scale. Finally, as V_m becomes an accelerating function of increasing rank, it increases the value for a woman of moving up on the mating value ranking scale.

Conclusions of the Model

In conclusion, our evolutionary models show that variance in male mating value should be associated with the origins of FGMo. The costly signaling model suggests that some women can benefit from FGMo at fixation, whereas under the virginity assurance model, no women benefit from FGMo at fixation. Further, we show that variance in male mating value is not strictly required for the maintenance of FGMo at fixation so long as social pressure is sufficient to maintain FGMo after it originates. Finally, we show that variance in male mating value may, however, factor into the maintenance of FGMo when the prevalence of FGMo is below fixation. This finding helps to explain the empirical observation by Efferson et al. (2015) that, in some cases, FGMo persists in populations at levels well below fixation—a situation that is unlikely to arise under a purely coordination-based model of FGMo persistence that does not also account for the effects of the differential value of males on the marriage market. We provide a more detailed discussion of our model-based findings in light of Efferson et al. (2015) in the Electronic Supplementary Materials, Section 9.

Methods

Adaptive Phylogenetic Analysis of Empirical Data

Data Sources and Sample Selection

To test whether variance in male value is associated with the distribution of FGMo across cultures, we use an extensive coding of data on female genital modification in Africa, the continent in which the custom is largely concentrated. The sample consists of 112 cultural clusters geographically distributed across Africa, drawn from Murdock's *Ethnographic Atlas* (Murdock 1969). For this sample, Ericksen (1989) examined all available ethnographic sources for evidence of FGMo. Following Ericksen (1989), clusters that contain more than one culture are represented by the culture with the best coverage on the topics being coded; in cases where more than one culture had good source materials, the choice was determined randomly. Interested readers are directed to Ericksen (1989) and Paige and Paige (1981) for detailed discussion of the included cultural clusters, sample construction, citations to primary sources, and ethnographic details; methodological details concerning the standard cross-cultural sample can be found in Murdock and White (1969).

Female genital modification, $FGMo$, was coded as present if the practice was mentioned as occurring by at least one ethnographer, either present or past, irrespective of how much detail was given; inter-coder reliability between three independent readers using English and non-English sources was 0.79 (Ericksen 1989). Recognizing that our analysis could be biased by considering FGMo as being practiced in cultural groups where the frequency of practice is actually quite low, we conduct a secondary analysis based on classifying FGMo as present only if contemporary FGMo prevalence is $> 50\%$ (Electronic Supplementary Materials, Section 8); the results of this analysis were qualitatively the same as those of our main analysis.

We use the existence of economic and/or social stratification as our measure of variation in male quality, based on the assumption that more stratified societies exhibit greater differences in male resource holdings, and greater power and prestige differences; stratification is coded using Variable 67 (class stratification) in Murdock (1957). This variable contains four categories:

1. Absence of significant class distinctions among freemen, ignoring variations in individual repute achieved through skill, valor, piety, or wisdom.
2. Wealth distinctions based either on the possession or distribution of property, present and socially important but not crystallized into distinct and hereditary social classes, or on hereditary aristocracy and lower class of ordinary commoners or freemen, where traditionally ascribed noble status is at least as decisive as control over scarce resources.

3. Dual or elite stratification in which an elite class derives its superior status from, and perpetuates it through, control over scarce resources, particularly land, and is therefore differentiated from a property-less class.
4. Complex stratification into social classes correlated in large measure with extensive differentiation of occupational statuses.

In our model, we use a binary variable for stratification in which categories 2, 3, and 4 are collapsed, yielding a variable that distinguishes cultural groups with wealth and class distinctions from cultural groups without such distinctions. For a small number of cultural groups, the *Ethnographic Atlas* had missing information; we investigated the primary ethnographic literature to obtain equivalent data for these groups. The Supplementary Materials (FGMoData.csv) contains our data and the relevant citations to supplementary primary sources. In total, we have data from 63 stratified cultural groups and 49 non-stratified groups; 44 groups practice FGMo and 68 do not.

Bayesian Phylogenetic Modeling of Adaptive Evolution

To model the effect of stratification on the evolution of FGMo, we utilize a Bayesian phylogenetic model of adaptive evolution (based on an Ornstein-Uhlenbeck process) in the spirit of Butler and King (2004). The methods advanced by Butler and King (2004) conceptualize evolution across a phylogeny as a function of both selective processes and drift. This approach thus constitutes critical progress in phylogenetic analysis insofar as it allows for adaptive hypotheses to be evaluated with phylogenetic models that explicitly include selection dynamics. The methods advanced by Butler and King (2004) allow us to (1) investigate the extent to which the likelihood of a cultural group practicing FGMo is conditioned on the state of that cultural group as stratified versus non-stratified and (2) contrast the strength of selection for FGMo based on stratification with the strength of drift (and selective forces operating orthogonally to stratification).

Standard tools for phylogenetic inference concerning the evolution of discrete traits (Pagel 1994; Pagel and Meade 2006; Ives and Garland 2010, 2014) represent an alternative analytic strategy. In the main text, we focus on the adaptive phylogenetic analysis because we find that the model dynamics more elegantly match the empirical processes we wish to understand. For thoroughness, we present the results from various other discrete trait phylogenetic models in the Electronic Supplementary Materials, Section 7; our findings are consistent across all approaches.

Butler and King (2004) developed a software package (OUCH, Ornstein-Uhlenbeck for Comparative Hypotheses) for adaptive phylogenetic analysis in the R software environment (R Core Team 2013). This software package, however, treats phylogenies as known data, deals poorly with parameter constraints, and relies on a maximum likelihood estimation procedure that fails to function reliably in many contexts, as detailed in Butler and King (2004).

To improve on the software introduced by Butler and King (2004), we wrote our own Bayesian implementation of the adaptive phylogenetic model using Hamilton Markov Chain Monte Carlo (HMC) simulation (Hoffman and Gelman 2014) in

C++, using the Stan 2.2.0 library (Stan Development Team 2013a). We provide code to run similar models in R, via the RStan interface. Our Bayesian implementation allows us to integrate over phylogenetic uncertainty, impose parameter constraints, and use prior information to identify parameters that are not necessarily identifiable under maximum likelihood estimation.

Following Butler and King (2004), we imagine trait evolution over a phylogenetic tree occurring as an Ornstein-Uhlenbeck process where a trait, X, evolves under a regime composed of both selection and drift such that:

$$\partial X(t) = \alpha[\beta(t) - X(t)]\partial t + \sigma \partial B(t) \tag{11}$$

where $\partial X(t)$ is the change in the character trait X over the course of a small increment of time, α is the strength of selection, $\beta(t)$ is the optimal trait value, and σ mediates the intensity of "white noise" fluctuations, $\partial B(t)$.

Below we describe the phylogeny used in our analysis and then outline the mathematical details of our statistical model, which uses the phylogenetic generalized linear model structure discussed by Ives and Helmus (2011) and Ives and Garland (2014)—with appropriate modifications to implement the adaptive process model introduced by Butler and King (2004).

A Phylogeny of Language Families Based on Lexical Similarity

We begin our phylogenetic analysis with a hierarchical clustering of the selected cultural groups ($N = 112$) in Africa (Murdock 1969) according to the linguistic divergences postulated in *Ethnologue* (Lewis 2009). Following Walker et al. (2012), we then utilize the estimated dates of lexical divergence produced by the ASJP (Automated Similarity Judgment Program) to define the prior expected proportionality of branch lengths (Holman et al. 2011). Neither the clustering of languages in *Ethnologue* nor the estimated times of language divergence produced by the ASJP are free of controversy (see peer commentary in Holman et al. 2011). There are two critical issues with the use of linguistic data to infer the splitting of cultural groups: first, linguistic evolution, as with any form of cultural evolution, is not necessarily treelike. There is, however, some treelike structure to linguistic and cultural evolution, and we believe accounting for this structure is better than ignoring it completely (Gray et al. 2010). And second, assumptions of constant rates of language evolution have been shown to be in conflict with empirical data (Gray et al. 2007). ASJP estimates of divergence dates, however, are not based on constant evolutionary rates, and are estimated using empirical calibration on the basis of historical, archaeological, and other evidence, as suggested by Gray et al. (2007).

Despite the fact that the phylogenetic tree used in our analysis is only of limited accuracy, in the Bayesian interpretation of the model fit in our analysis, phylogenetic information is utilized to construct prior beliefs concerning the expected covariance of the preferences for the FGMo trait across cultural groups resulting from shared ancestry. The fact that there is error in the ASJP estimates of divergence times is not necessarily a problem for our analysis. So long as the errors of estimated divergence times based on lexical similarity are roughly proportional across cultural groups, and do not vary as a function of stratification or FGMo prevalence, our inferences are unlikely to be biased (the actual calender dates of linguistic divergences do not matter for our analysis).

Furthermore, we know of no other published phylogenies that estimate the divergence times (branch lengths) of all African languages in a unified framework; the ASJP phylogeny is also derived under more theoretical and empirical rigor than standard Bayesian approaches to phylogenetic reconstruction that utilize a simple binary coding procedure of cognate classes from Swadesh lists. Future studies will surely improve inference concerning the nature and dating of the somewhat treelike, somewhat reticulated river-network-like cultural evolutionary pasts of extant human groups (Towner et al. 2012). When such data become available, our inferences herein should be reinvestigated.

Figure 2 displays our phylogeny with prior branch lengths scaled to the ASJP estimates. In the Electronic Supplementary Material in Section 3, we detail how exactly we bring the ASJP divergence times into our analysis, and how we model uncertainty in these values.

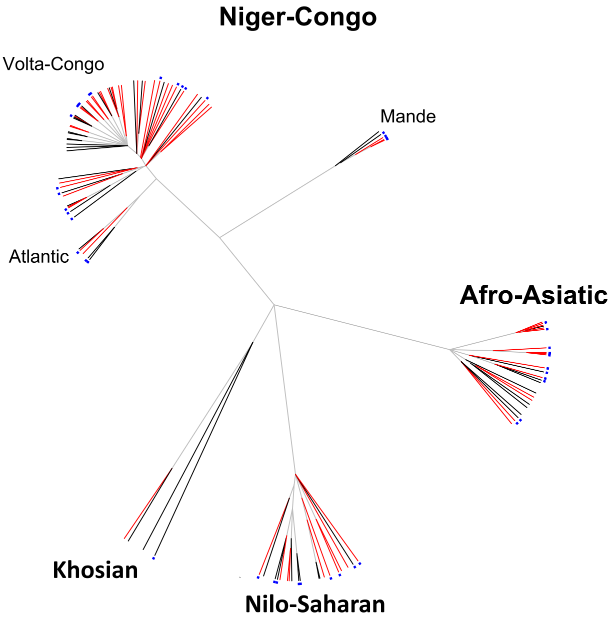


Fig. 2 The phylogeny of African languages used in this analysis. In this figure, branch lengths are presented as proportional to maximum likelihood AJSP divergence estimates; the actual phylogenetic tree implemented in the model, however, is a constrained random variable, allowing Bayesian integration over uncertainty in phylogenetic information. *Black* branches on the tree indicate non-stratified selection regimes. *Red* branches indicate stratified selection regimes. *Gray* branches indicate a mixture of stratified and non-stratified selection regimes. This mixture modeling allows for integration over uncertainty in deeper, nonterminal branches. The blue points on the edge of the phylogenetic tree indicate the presence of the practice of FGMO. (Refer to online version for colors)

Modeling Adaptive Evolution

At each MCMC (Markov Chain Monte Carlo) iteration, a random painting of divergence times onto the branches of the phylogenetic tree is proposed. We then standardize the phylogeny to the unit interval such that $t = 0$ is the time of the deepest divergence and $t = T = 1$ is the present. We then calculate two matrices, S and C, from the phylogeny. The S matrix is an N by N matrix (N =number of cultures included in this analysis=112) of the time points when cultural group n split from each and every other cultural group. The C matrix is an N by Γ matrix of the cut-points/transitions in epochs within a lineage, where $\kappa(n)$ is the total number of epochs in lineage n, and $\Gamma = \max(\kappa(n)) = 12$ is the maximum number of epochs in a cultural lineage observed in our data. A cultural lineage is defined as the path from the tip of the phylogenetic tree to its most basal node. The history of the n^{th} lineage is then a series of $\kappa(n)$ branch segments demarcated by epochs $0 = t_n^0 < t_n^1 < t_n^2 < \ldots < t_n^{\kappa(n)} = T$, where each epoch constitutes a single kind of selective regime, to be defined more thoroughly later (see Appendix A in Butler and King 2004).

Equation 11 describes a stationary, Gaussian, and Markovian process with well-defined moments; following, Butler and King (2004), we make a reducing assumption that in every lineage, evolution of a cultural trait, X, occurs along piecewise-constant selection regimes. Accordingly, the expected value of a trait evolving along a cultural lineage, n, can be defined as:

$$\mu_{[n]} = \mathrm{E}[X_n(T)|X_n(0) = \beta_{[n,0]}] = e^{-\alpha T}\beta_{[n,0]} + \sum_{\gamma=1}^{\kappa(n)} e^{-\alpha T}(e^{\alpha C_{[n,\gamma]}} - e^{\alpha C_{[n,\gamma-1]}})\beta_{[n,\gamma]} \tag{12}$$

where $\beta_{[n,0:\kappa(n)]}$ is a parameter vector "painted" with θ parameters. The first cell of this vector, $\beta_{[n,0]}$, is defined to be equal to θ_{anc}, the estimated trait value of the most basal node in the phylogeny (ancestral), and all other cells are populated with differing θ parameters that describe the hypothesized selective regime acting on the γ^{th} epoch in lineage n. The specific way that other θ parameters are painted on to the phylogenetic tree will be made clearer when we describe the exact models being compared in this study.

Regarding covariance, we assume that when $t < S_{[n,m]}$, lineages n and m evolved as a single group, and when $t > S_{[n,m]}$ the two lineages evolve independently. Accordingly, the covariance matrix $V_{[n,m]}$ can be defined as:

$$V_{[n,m]} = \mathrm{Cov}[X_n(T), X_m(T)|X_n(0) = X_m(0) = \beta_{[n,0]}] = \frac{\sigma^2}{2\alpha} e^{-2\alpha(T - S_{[n,m]})}(1 - e^{-2\alpha S_{[n,m]}}) \tag{13}$$

To complete the basic model definition, we define regularizing priors on the α and σ parameters, which concentrate prior probability density near zero. We considered both half-Gaussian and half-Cauchy (Gelman 2006) priors (see Electronic Supplementary Materials, Section 5 for more information on these priors). In our final analysis, we used the following half-Gaussian priors:

$$\alpha \sim \mathrm{Normal}(0, 5)T[0, \infty] \tag{14}$$

$$\sigma \sim \mathrm{Normal}(0, 5)T[0, \infty] \tag{15}$$

We use regularizing unit normal priors on each cell of the θ parameter vector. These priors can be understood as imposing the Bayesian corollary of Tikhonov regularization (Tikhonov and Arsenin 1977), or ridge regression (Hoerl and Kennard 1970), and aid in the identification of θ and α, which are not necessary well identified otherwise:

$$\theta \sim \mathrm{Normal}(0, 1) \tag{16}$$

We then model the data using a multivariate normal distribution parameterized to accept a Cholesky factor, L_V, from the decomposition of the variance-covariance matrix, V, in place of the variance-covariance matrix itself. This parameterization of the model improves the performance of the HMC estimation process for technical reasons that are outlined in the Stan manual (Stan Development Team 2013b). To link μ and L_V to the outcomes, we use a parameter vector, Ψ, which represents the strength of evolving social preferences for FGMo:

$$\Psi \sim \text{Multivariate Normal Cholesky}(\mu, L_V) \tag{17}$$

and then use the statement:

$$FGMo_{[n]} \sim \mathrm{Bernoulli}(\mathrm{Logistic}(\Psi_{[n]})) \tag{18}$$

to return the log probability of the data, $FGMo$, conditioned on the estimated social preferences for FGMo; $FGMo$ is a vector of binary indicators of the practice of FGMo.

Model Construction

In this paper, we fit two models to the data. In the first, we conceive of the evolutionary dynamics as an Ornstein-Uhlenbeck process with a single global optima for all cultural groups, an OU(1) model. This model serves as a null model, where stratification plays no role in the adaptive evolution of FGMo. In this case, we paint every branch of the phylogeny other than the most basal node, θ_{anc}, with a single parameter that represents a single global selection regime, θ_{gsr}.

In the second model, we conceive of the evolutionary dynamics as an Ornstein-Uhlenbeck process with separate optima for stratified and non-stratified cultural groups, an OU(2) model. In this case, we paint the phylogeny using two θ parameters, with one corresponding to a stratified selection regime, θ_{ssr}, and one corresponding to a non-stratified selection regime, θ_{nssr}.

Deep ancestral branches are not easily classifiable as stratified or non-stratified, but the phylogenetic tree contains information on the probability of stratification in deeper branches conditional on the state of the branch tips and the strength of selection for stratification along the branches. Accordingly, at each MCMC iteration, we estimate the probability of stratification in deep branches by using the Butler and King (2004) method to model the culture-group specific evolution of stratification with an OU(15) model (See Electronic Supplementary Materials, Sections 2 and 6 for details and model diagnostics). Following this, we model the evolution of FGMo, conditioned on the estimated state of stratification at every node in the phylogeny.

More formally, in the OU(2) model, at each MCMC iteration, we run an adaptive phylogenetic model on stratification using the observed binary data variable $Strat$. We model:

$$\Psi_{\text{strat}} \sim \text{Multivariate Normal Cholesky}(\mu_{\text{strat}}, L_{V_{\text{strat}}}) \tag{19}$$

where μ_{strat} and $L_{V_{\text{strat}}}$ are corollaries to μ and L_V and are derived from Equations 12 and 13 in the same way; Ψ_{strat} is a parameter vector representing the log odds that a given cultural group is stratified, and:

$$Strat_{[n]} \sim \text{Bernoulli}(\text{Logistic}(\Psi_{\text{strat}[n]})) \tag{20}$$

provides the log probability of the observed stratification data conditional on the proposed parameter values.

The β_{strat} matrix for the stratification model is painted with θ parameters such that the branches for each language family and subfamily have unique parameters (e.g., language family parameters are $\theta_{\text{NiloSaharan}}$, $\theta_{\text{NigerCongo}}$, etc., and language subfamily parameters are θ_{Omotic}, θ_{Mande}, etc.). The way in which these parameters are painted onto the branches is made clear in the Supplementary Stan code, using variables $GID2$ and $GID3$ from the Supplementary Data.

Under such a model it is straightforward to calculate the probability of stratification in any epoch in any lineage, $\Lambda_{[n,\gamma]}$, from Eq. 12 as:

$$\Lambda_{[n,\gamma]} = \text{Logistic}\left(e^{-\alpha T} \beta_{\text{strat}[n,0]} + \sum_{\gamma=1}^{\kappa(n)} e^{-\alpha T} \left(e^{\alpha C_{[n,\gamma]}} - e^{\alpha C_{[n,\gamma-1]}} \right) \beta_{\text{strat}[n,\gamma]} \right) \tag{21}$$

To model the evolution of FGMo as a function of stratification, we then paint the tips of the phylogeny (in the FGMo model) with parameters using observed data:

$$\beta_{[n,\kappa(n)]} = \begin{cases} \theta_{\text{ssr}}, & \text{if } Strat_{[n]} = 1 \\ \theta_{\text{nssr}}, & \text{if } Strat_{[n]} = 0 \end{cases} \tag{22}$$

and paint all nonterminal epochs (e.g., where $\gamma \neq \kappa(n)$) across all lineages (in the FGMo model) as a mixture of θ_{ssr} and θ_{nssr} with mixing proportions determined by Λ:

$$\beta_{[n,\gamma]} = \theta_{\text{ssr}} \Lambda_{[n,\gamma]} + \theta_{\text{nssr}}(1 - \Lambda_{[n,\gamma]}) \tag{23}$$

Software Environment

Each complete model was fit using the Stan 2.2.0 software, using the No-U-Turn sampler, a variant of Hamiltonian Monte Carlo (Stan Development Team 2013a). We ran multiple small chains to ensure that the models were well-defined and converged to similar posterior regions. Our final results are based on the results from a single long chain, run for 20,000 warm-up iterations and 400,000 sampling iterations, thinned at an interval of 20. We found that our model was well identified for all θ parameters, for α, and effectively identified for σ, although σ itself is not numerically well identified. Convergence diagnostics and model identification are discussed in detail in the Electronic Supplementary Materials, Sections 4 and 6.

All pre-processing of data and post-processing of MCMC samples was conducted using the R environment for statistical computing (R Core Team 2013).

Model Comparison

We compare models using the Watanabe-Akaike information criterion (WAIC) (Gelman et al. 2014), which is a more fully Bayesian generalization of the standard Akaike information theoretic criteria, AIC. Computed WAIC is defined as:

$$\text{WAIC} = -2(lppd - p_E) \tag{24}$$

The computed log pointwise posterior predictive density, $lppd$, is defined as:

$$lppd = \sum_{n=1}^{N} \log \left(\frac{1}{Q} \sum_{q=1}^{Q} Pr\left(FGMo_{[n]} | \Psi_{[n,q]}\right) \right) \tag{25}$$

where $q = 1 \ldots Q$ references the index of simulations from the posterior distribution. The effective number of parameters, p_E, is computed as:

$$p_E = \sum_{n=1}^{N} \text{Var}_{q=1}^{Q} \left(\log(Pr(FGMo_{[n]} | \Psi_{[n,q]})) \right) \tag{26}$$

where the symbol $\text{Var}_{q=1}^{Q}$ represents the function to calculate the sample variance over the posterior simulations.

Results

Analysis of the data with Ornstein-Uhlenbeck process models and information theoretic model comparison methods shows that while stratified cultural groups are on average slightly more likely to practice FGMo than non-stratified groups, there is little evidence that the stratification model provides a better fit for the data than the null model.

Table 1 presents the key parameters of our models, showing the posterior mean and medians estimates, as well as the central 95% posterior confidence intervals (95PCIs) from the fitted OU(1) and OU(2) models. Table 2 presents the results of model comparison using WAIC.

Table 1 Key parameter estimates from our models, showing relative support for α (selection based on stratification) and σ (drift and/or orthogonal selective forces) in explaining the distribution of FGMo in Africa

	Model	Mean	Median	2.5% PCI	97.5% PCI
θ_{anc}	OU(2)	−0.075	−0.078	−1.974	1.853
θ_{ssr}	OU(2)	−0.04	−0.042	−1.868	1.761
θ_{nssr}	OU(2)	−0.278	−0.289	−2.106	1.56
α	OU(2)	2.781	2.636	0.412	6.105
σ	OU(2)	6.891	6.712	2.234	12.502
θ_{anc}	OU(1)	−0.047	−0.039	−1.994	1.889
θ_{gsr}	OU(1)	−0.18	−0.212	−1.789	1.547
α	OU(1)	2.652	2.496	0.337	5.88
σ	OU(1)	7.404	7.102	2.701	14.023

The mean and median are point estimates of the posterior distribution; the 2.5% and 97.5% equal tail posterior confidence intervals (PCI) present the dispersion of the posterior distribution. The top set of parameter estimates is from the OU(2) model, and the bottom set of estimates is from the OU(1) model. The symbols prefixed by θ indicate the estimated optimal trait value (log odds) under the ancestral, global, stratified, or non-stratified selection regime, as indicated by the subscripts. We note that σ dominates α in both the OU(1) and OU(2) models, which is indicative that the distribution of FGMo in Africa is better explained by drift and/or selective forces operating orthogonally to stratification than by stratification itself. In the OU(2) model, however, we find that stratification relative to non-stratification is weakly, but positively, associated with elevated social pressure for FGMo

Inspection of the θ_{ssr} and θ_{nssr} parameter estimates from the OU(2) model in Table 1 shows a moderate difference in the mean estimated strength of social preferences for FGMo as a function of stratification. Converting these parameter estimates to the probability scale, we find that stratified cultural groups are about 6% more likely, on average, to practice FGMo than non-stratified cultural groups. Although the

Table 2 Results of formal model comparison using WAIC. WAIC, p_E, and *lppd* are defined in the text

Model	p_E	*lppd*	WAIC	dWAIC	wWAIC
OU(1)-Null	28.099	−28.409	113.017	0	0.848
OU(2)-FixedBranchTips	27.407	−30.821	116.458	3.441	0.152

dWAIC indicates the difference in WAIC between the best model and the second model, and wWAIC indicates the weight in probability that the specified model will make the best predictions on new data, conditional on the set of models being considered. We note that the OU(1) model outperforms the OU(2) model. This result indicates that the distribution of FGMo across our sample of African cultural groups can be most parsimoniously explained by a model with a single global optima, as opposed to a model with separate optima for stratified and non-stratified societies. However, these wWAIC values are very close, and the application of WAIC to these models is not completely justified given the relative strength of drift; as such, we argue that both models are important to consider. See Electronic Supplementary Materials, Section 5 for a discussion of methodological issues that arise with WAIC, as σ increases relative to α in adaptive phylogenetic models.

confidence intervals on these estimates are wide and largely overlap one another, the estimates of θ_{ssr} and θ_{nssr} are in the direction predicted by our evolutionary models.

Model comparison with WAIC, however, shows that the OU(1) model is preferred over the OU(2) model. So while increased stratification appears to be weakly associated with increased odds of FGMo, accounting for stratification does not improve predictions sufficiently to compensate for the increased model complexity in an information theoretic framework.

In conclusion, although we present evidence that social stratification places positive selection pressure on social preferences for FGMo, it is evident that σ (which accounts for drift and/or selection dynamics operating orthogonally to stratification) plays a more important role in explaining the cross-cultural distribution of FGMo. Future research is needed to disentangle the effects of drift from other possible selective drivers of FGMo.

Since our variables for both stratification and FGMo are binary, use of continuous measures of resource inequality (e.g., a Gini coefficient) or FGMo prevalence might yield more informative results. However, we do not as of yet have access to such data. Future comparative work in anthropology might benefit from using higher-resolution cultural data (where such data exist) or, even better, individual-level measures (e.g., Borgerhoff Mulder et al. 2009; Hill et al. 2011). For example, if we had cross-cultural, individual-level data on FGMo prevalence and male wealth, we could estimate FGMo prevalence and calculate a Gini coefficient on wealth, which would likely be more informative than the simple binary coding scheme used in this analysis. Likewise, as methods of estimating cultural-group divergences through language information become more refined, the effectiveness and accuracy of cultural phylogentic methods may improve substantially.

Discussion

This paper offers four principal contributions. First we develop explicit evolutionary models for the origin and maintenance of FGMo. In so doing we provide both a formal justification for an existing hypothesis that the practice is related to social stratification (Mackie 2000) and a novel set of predictions regarding the distinction between trait origins and maintenance (albeit also recognized by Mackie [2000]). Second, we consider two generative models for FGMo based on virginity assurance and costly signaling, and we find that both lead to similar macro-level model dynamics and could therefore be responsible for the emergence of FGMo when there is inequality in male mating or marriage value. Third, we fit an empirically grounded statistical model that includes the two main effects from the theoretical model: a selective effect based on the inequality in mate value (operationalized in the empirical analysis using stratification as a proxy) and a social transmission effect (operationalized in the empirical model using a term orthogonal to the proposed selective regime based on stratification). Finally, we provide code for a Bayesian implementation of the Ornstein-Uhlenbeck process model developed by Butler and King (2004).

We use mathematical models to show the importance of differentiating explanations for cultural trait origins from those for persistence; the importance of this distinction for explaining trait distributions has been made previously (Borgerhoff Mulder 2001) but has not, to our knowledge, been formally modeled or investigated with empirical data. The results of our adaptive phylogenetic analysis suggest that stratification is associated, though only weakly, with selection pressure on the uptake of FGMo. However, the fact that σ dominates α in explaining the evolutionary dynamics of the FGMo trait in Africa indicates that drift and/or selective forces operating orthogonally to social stratification play a very significant role in explaining the current cross-cultural distribution of the FGMo trait. This is entirely consistent with the view of earlier scholars who suggest that FGMo became decoupled from the signaling of marriageability and chastity as it spread into less stratified populations (Mackie 2000).

Historical records, albeit largely speculative, that place the origins of FGMo in ancient Egypt, or the strongly Egyptian-influenced and fabulously wealthy trading city of Meroe, are consistent with our model-based findings that variance in male mating value and social stratification should be causally linked to the de novo origins of FGMo, since ancient Egyptian society was marked by complex stratification (Murdock 1957).

Although our model-based results indicate that stratification is needed for the de novo origins of FGMo, once the trait has arisen in a single cultural group its adoption in any subsequent cultural groups does not require stratification. Once FGMo has arisen in a cultural group, and the system has equilibrated such that s—a measure of frequency-dependent or conformist biases (Boyd and Richerson 1985; Henrich and Gil-White 2001; Henrich and Boyd 1998)—functions to increase pressure on individuals to undergo FGMo, the genesis of FGMo in further groups can occur strictly as a function of s, irrespective of stratification. Thus, the dynamics of the de novo origins of FGMo differ from the subsequent dynamics of intercultural transmission (Ross et al. 2015). Though our adaptive phylogenetic model illustrates that the evolution of FGMo is not strictly dependent on stratification, the model still suggests that stratification increases the likelihood of intergroup transmission or adoption of FGMo.

Most intriguingly, our model shows how, over ecological time, ordinary adaptive processes can lead to the emergence of a trait that encourages individuals to accept huge fitness costs in their pursuit of mates. This, of course, is not unusual in the animal world (Chapman et al. 1995; Wigby and Chapman 2005). However, as a result of the biases that influence the spread of culturally transmitted traits (conformity and/or imitation of prestigious individuals), such traits can stabilize over evolutionary time and, because of the strength of cultural norms and social stigma against deviation, hold in place institutions that can, in this case, severely damage women. Our modeling therefore shows how ordinary adaptive processes, combined with cultural evolutionary feedbacks, can generate stable evolutionary outcome states in which sexual conflict is not resolved (Borgerhoff Mulder and Rauch 2009). Such outcome states might even set in place systems of gene-culture evolution in which genetic variations arise because of culturally specific sexual selection pressures (Ross and Richerson 2014). Futhermore, our modeling demonstrates the futility of thinking of cultural evolutionary and ordinary adaptive processes as alternative explanations since they are jointly required to explain the evolutionary dynamics of FGMo.

In our theoretical models, we focused mainly on differentiating the evolutionary forces responsible for the de novo origins of FGMo and its subsequent maintenance in a population. The results of our phylogenetic analysis, however, show that the effects of stratification are weak, and that characterizing the cross-cultural origins of FGMo requires more explicit treatment of the dynamics underlying intergroup transmission; these dynamics are investigated elsewhere in an empirical case study of the intergroup transmission of FGMo in the African diaspora and indigenous populations of Colombia (Ross et al. 2015).

The relationship between s and the intercultural transmission of FGMo is complex and will be structured by culturally and historically particular processes. For instance, imagine that migrants from a non-FGMo culture enter an FGMo cultural area. In such a case, s and conformist biases can now act to place selective pressure on the uptake of FGMo by these intercultural migrants, even if stratification was not present in the migrants' cultural group. Such an effect would be exacerbated if there were perceived prestige differences between the FGMo-practicing cultural group and the non-FGMo-practicing migrant group that cause migrants to copy the behaviors of the prestigious group (Richerson et al. 2015). This pattern would be likely if FGMo is associated with the prestigious class in stratified areas.

A similar pattern would be observed if adoption of the cultural traits of a dominant or prestigious cultural group is an essential step in successfully integrating into that culture; such dynamics might partially underlie the observation by Cronk (2004) that Kenyan Mukogodo appear to have adopted FGMo to hasten their transition to becoming Maasai, or the observation by Ericksen (1989) that the Fur appear to have adopted FGMo to marry with their nomadic, stock-owning, and more wealthy Zaghawa neighbors. Similarly, non-practicing displaced ethnic communities in Darfur (Sudan) have adopted FGMo when moving to cities, just as in the state of Khartoum, where migrants from West Africa now cut their daughters to gain acceptance in their new host community (DFID 2013). If successful intercultural migrants are emulated or held in high esteem, or if migration events involve a large portion of a cultural group, then frequency-dependent, conformist, and prestige biases might result in the propagation of FGMo back into the migrants' natal cultural group.

Different patterns of transmission are also possible. We know, for example, that during the intense military turmoil in the Great Lakes region of central Africa that refugee segments of Bantu-speaking Abaluhyia lineages, subsequently known as the Tiriki, adopted the age set organization of their Niolotic Terik neighbors. In return for asylum with the more military-prepared Terik, Tiriki elders accepted the full set of initiation rituals for their sons, including circumcision and seclusion, in a bid to obtain protection (Levine and Sangree 1962; Boyd et al. 1997).

In contrast, in many contexts the practice of FGMo may mark an ethnic boundary (McElreath et al. 2003) between groups that practice FGMo and those that do not. In such contexts, s could militate against adoption of the cultural traits of out-groups and work against the intercultural transmission of FGMo to neighboring populations. Case studies in which such socially or behaviorally marked ethnic boundaries function to prevent the transmission of information and behavior across groups are well described in the literature (Barth 1998; Van den Berghe 1987).

As such, we expect that the dynamics governing the transmission of FGMo and other costly traits will be historically and locally contingent based on the nature of intercultural contact. We suspect that the mechanisms underlying these intercultural dynamics are likely to include frequency-dependence or conformist learning biases—as outlined in our introductory review of how FGMo is embedded in deeply held societal conventions—as well as prestige-biased learning, whereby adopting FGMo may facilitate inclusion in a more powerful or prestigious population. Similarly, we note that the decline in FGMo is particularly striking in countries such as Kenya and Tanzania, where FGMo is less prevalent across constituent ethnic groups (UNFPA-UNICEF 2013), suggesting that abandonment is more tolerable, even practicable, with non-practitioners as neighbors. Future research should investigate in more detail the socioecological circumstances that aid in the intercultural transmission of FGMo and those that hinder transmission and hasten abandonment.

Our conclusions are generally consistent with those of Murdock (1959), who argues that the trait of FGMo—in his terminology, excision—is widely distributed across Africa because of cultural diffusion. For example, in East Africa, southern Nilotes and Bantu are thought to have adopted the custom from Cushitic neighbors, given that there is little evidence of FGMo among other Nilotes and the apparent Cushitic roots to the linguistic term used for the operation (Murray 1974). In West Africa too there is strong inferential evidence that FGMo was adopted by some of the neighbors of the FGMo-practicing Mande—for example, the Kissi and the Kran, but not the Kpelle, Guro, and Gbande—but the details of the adoption during this historic period of diffusion are unclear (Ericksen 1989). Furthermore, ethnographic reports (Stannus 1919) indicate that the Yao of northwestern Mozambique and southern Tanzania appear to have adopted female "circumcision" during their close collaboration with Swahili and Arab slave traders in the nineteenth century, although now it is largely dropped. We note, however, that these conclusions regarding the historical diffusion of the trait are based largely on the observation that FGMo appears in populations that do not necessarily share a common language, geography, and cultural history, and that our methods (like those of Towner et al. 2012) provide an improvement in inference.

Implications for Policy

Here we sidestep ethical arguments regarding whether or not FGMo should be abandoned, as well as the debates over who has the moral authority to take the lead in such initiatives, recognizing that this can only be locally adjudicated. As Shell-Duncan and Hernlund (2000:126) observe for the Rendille of Kenya, "awareness of the fact that female 'circumcision' is associated with adverse health consequences is widespread, yet the Rendille view the risks as worth taking in light of the implications for marriageability." There are clear social, psychological, and physical consequences to the practice of genital cutting, and these need to be weighed very carefully in each case. But, if abandonment is viewed as appropriate, the question of how to achieve such a goal remains. Many strategies are currently discussed, including medicalization of the practice (for example, the use of less extreme procedures in hospitals by specially trained practitioners), mass education campaigns, formal legislation with criminalization of operators and their clients, withdrawing of foreign aid programs, developing ritual alternatives, or simply relying on the processes of development and the erosion of tradition (Shell-Duncan and Hernlund 2000).

Many scholars argue that knowledge of the origins of FGMo will not contribute to a determination of its current significance; others maintain that understanding the historical roots of the tradition helps justify the persistence of the custom to disapproving outsiders, and still others insist that the origins and maintenance of the practice cannot be conceptually separated from development of strategies for its elimination (Mackie 2000). The present analysis, with its implications for the importance of frequency-dependent biases in the maintenance of FGMo, supports this third viewpoint. Specifically, it points to the potential value of programs that foster contracts within small communities whereby all parties make a pact not to send their daughters for the operation; our model suggests that such pacts may be especially effective in cultural groups who have acquired FGMo via social transmission, or in cultural groups where there no longer exists sufficient heterogeneity in the value of men on the marriage market to create the selective pressures that would eclipse the effects of such pacts. Mackie (1996) has made precisely this point, drawing an analogy between successful campaigns to eradicate footbinding in China through fostering of such local contracts. Furthermore, the most recent evaluation of the status of FGMo globally suggests that public declarations, or collective announcements of abandonment by village delegates, have for the most part been very successful in supporting change (UNFPA-UNICEF 2013), especially if the initiative is supported by locally respected leaders (WHO 2010).

There is nevertheless huge variation within and between communities in the rate of abandonment (Shell-Duncan and Hernlund 2007), which begs further study. Recent observations by Efferson et al. (2015) that FGMo is not purely a coordination norm, in conjunction with the findings reported here, might help to explain this heterogeneity. Whereas community-based pacts might be effective in abating FGMo in contexts where FGMo is driven primarily by coordination problems (or other frequency- or prestige-based social norms), in cases where FGMo is maintained at intermediate levels by selective pressures related to heterogeneity in the value of men or ranking of women on the marriage market, abatement pacts are unlikely to be effective. This is because FGMo may not be maintained purely by coordination. Future empirical study of the relative cross-cultural effectiveness of FGMo abatement programs as a function of male wealth inequality in each cultural group may help to validate this hypothesis.

Acknowledgments CTR designed and tested the Bayesian models, conducted analysis, and wrote/edited the paper. PS designed the mathematical models and wrote/edited the paper. KE collected and compiled the ethnographic and linguistic information and edited the paper. PL conducted the alternative phylogenetic analysis and wrote/edited the paper. MBM conceived the study, reviewed the literature, and wrote/edited the paper.

We thank the UC Davis Behavioral Ecology and Cultural Evolution lab groups for helpful comments and critiques, Richard McElreath for feedback on the mathematical model, Richard McElreath and Andrew Gelman for providing code and advice on model comparison using WAIC, and the Stan Development Team for making Stan freely available and open source, and for providing impressive levels of software support and consulting. Mark Grote gave significant advice concerning model notation and provided statistical consulting that much improved our methodological framework. Charlie Nunn and Bruce Rannala gave helpful suggestions early in the analysis.

References

Anderson, S. (2003). Why dowry payments declined with modernization in Europe but are rising in India. *Journal of Political Economy*, *111*(2), 269–310.

Banks, E., Meirik, O., Farley, T., Akande, O., Bathija, H., & Ali, M. (2006). Female genital mutilation and obstetric outcome: WHO collaborative prospective study in six African countries. *Lancet*, *367*(9525), 1835–1841.

Barth, F. (1998). *Ethnic groups and boundaries: the social organization of culture difference*. Long Grove, IL: Waveland Press.

Becker, G.S. (1981). *A treatise on the family*. Cambridge: Harvard University Press.

Bell, D., & Song, S. (1994). Explaining the level of bridewealth. *Current Anthropology*, *35*(3), 311–316.

Borgerhoff Mulder, M. (1995). Bridewealth and its correlates: quantifying changes over time. *Current Anthropology*, *36*(4), 573–603.

Borgerhoff Mulder, M. (2001). Using phylogenetically based comparative methods in anthropology: more questions than answers. *Evolutionary Anthropology: Issues, News, and Reviews*, *10*(3), 99–111.

Borgerhoff Mulder, M., Bowles, S., Hertz, T., Bell, A., Beise, J., Clark, G., Fazzio, I., Gurven, M., Hill, K., Hooper, P.L., & et al. (2009). Intergenerational wealth transmission and the dynamics of inequality in small-scale societies. *Science*, *326*(5953), 682–688.

Borgerhoff Mulder, M., Nunn, C.L., & Towner, M.C. (2006). Cultural macroevolution and the transmission of traits. *Evolutionary Anthropology: Issues, News, and Reviews*, *15*(2), 52–64.

Borgerhoff Mulder, M., & Rauch, K.L. (2009). Sexual conflict in humans: variations and solutions. *Evolutionary Anthropology: Issues, News, and Reviews*, *18*(5), 201–214.

Boyd, R., Borgerhoff Mulder, M., Durham, W.H., & Richerson, P.J. (1997). Are cultural phylogenies possible? In P. Weingart, P. Richerson, S. Mitchell, & S. Maasen (Eds.), *Human by nature, between biology and the social sciences* (pp. 355–386). Mahwah, NJ: Erlbaum.

Boyd, R., & Richerson, P.J. (1985). *Culture and the evolutionary process*. Chicago: University of Chicago Press.

Butler, M.A., & King, A.A. (2004). Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *The American Naturalist*, *164*(6), 683–695.

Chapman, T., Liddle, L.F., Kalb, J.M., Wolfner, M.F., & Partridge, L. (1995). Cost of mating in *Drosophila melanogaster* females is mediated by male accessory gland products. *Nature*, *373*(6511), 241–244.

Chiappori, P.-A., Fortin, B., & Lacroix, G. (2002). Marriage market, divorce legislation, and household labor supply. *Journal of Political Economy*, *110*(1), 37–72.

Cronk, L. (2004). *From Mukogodo to Maasai: ethnicity and cultural change in Kenya*. Boulder: Westview Press.

Daniels, R.E. (1970). *By rites a man: a study of the societal and individual foundations of tribal identity among the Kipsigis of Kenya.* PhD thesis, University of Chicago.

Davison, J. (1996). *Voices from Mutira: changes in the lives of rural Gikuyo women, 1910-1995*. Boulder: Lynne Rienner.

DFID (2013). Business case: Sudan free of female genital cutting. Department for International Development: Sudan free of female genital cutting, UK. available online at iati.dfid.gov.uk/iati_documents/2850234.doc.

Dickemann, M. (1979). The ecology of mating systems in hypergynous dowry societies. *Social Science Information*, *18*(2), 163–195.

Dorkenoo, E. (1994). *Cutting the rose: female genital mutilation: the practice and its prevention*. London: Minority Rights Publications.

Efferson, C., Vogt, S., Elhadi, A., Ahmed, H.E.F., & Fehr, E. (2015). Female genital cutting is not a social coordination norm. *Science*, *349*(6255), 1446–1447.

Ericksen, K.P. (1989). Female genital mutilations in Africa. *Cross-Cultural Research*, *23*(1-4), 182–204.

Fortunato, L., Holden, C., & Mace, R. (2006). From bridewealth to dowry? *Human Nature*, *17*(4), 355–376.

Gaulin, S.J., & Boster, J.S. (1990). Dowry as female competition. *American Anthropologist*, *92*(4), 994–1005.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, *1*(3), 515–534.

Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, 1–20.

Grafen, A. (1990). Biological signals as handicaps. *Journal of Theoretical Biology*, *144*(4), 517–546.

Gray, R.D., Bryant, D., & Greenhill, S.J. (2010). On the shape and fabric of human history. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1559), 3923–3933.

Gray, R.D., Greenhill, S.J., & Ross, R.M. (2007). The pleasures and perils of Darwinizing culture (with phylogenies). *Biological Theory*, *2*, 360–375.

Hayes, R.O. (1975). Female genital mutilation, fertility control, women's roles, and the patrilineage in modern Sudan: a functional analysis. *American Ethnologist*, *2*(4), 617–633.

Henrich, J., & Boyd, R. (1998). The evolution of conformist transmission and the emergence of between-group differences. *Evolution and Human Behavior*, *19*(4), 215–241.

Henrich, J., & Gil-White, F.J. (2001). The evolution of prestige: freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evolution and Human Behavior*, *22*(3), 165–196.

Hill, K.R., Walker, R.S., Božičević, M., Eder, J., Headland, T., Hewlett, B., Hurtado, A.M., Marlowe, F., Wiessner, P., & Wood, B. (2011). Co-residence patterns in hunter-gatherer societies show unique human social structure. *Science*, *331*(6022), 1286–1289.

Hoerl, A.E., & Kennard, R.W. (1970). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55–67.

Hoffman, M.D., & Gelman, A. (2014). The no-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, *15*, 1351–1381.

Holman, E.W., Brown, C.H., Wichmann, S., Müller, A., Velupillai, V., Hammarström, H., Sauppe, S., Jung, H., Bakker, D., Brown, P., & et al. (2011). Automated dating of the world's language families based on lexical similarity. *Current Anthropology*, *52*(6), 841–875.

Ives, A.R., & Garland Jr., T. (2010). Phylogenetic logistic regression for binary dependent variables. *Systematic Biology*, *59*(1), 9–26.

Ives, A.R., & Garland Jr., T. (2014). Phylogenetic regression for binary dependent variables. In L.Z. Garamszegi (Ed.), *Modern phylogenetic comparative methods and their application in evolutionary biology* (pp. 231–261). Berlin: Springer.

Ives, A.R., & Helmus, M.R. (2011). Generalized linear mixed models for phylogenetic analyses of community structure. *Ecological Monographs*, *81*(3), 511–525.

Leonard, L. (2000). Adopting female "circumcision" in southern Chad: the experience of Myabe. In B. Shell-Duncan, & Y. Hernlund (Eds.), *Female "circumcision" in Africa: culture, controversy, and change* (pp. 167–192). Boulder: Lynne Rienner.

Levine, R.A., & Sangree, W.H. (1962). The diffusion of age-group organization in East Africa: a controlled comparison. *Africa*, *32*(02), 97–110.

Lewis, M.P. (2009). *Ethnologue: Languages of the world*, sixteenth edition. Dallas, TX: SIL International. Available online at www.ethnologue.com.

Little, C.M. (2003). Female genital circumcision: medical and cultural considerations. *Journal of Cultural Diversity*, *10*(1), 30–34.

Mackie, G. (1996). Ending footbinding and infibulation: a convention account. *American Sociological Review*, *61*(6), 999–1017.

Mackie, G. (2000). Female genital cutting: the beginning of the end. In B. Shell-Duncan, & Y. Hernlund (Eds.), *Female "circumcision" in Africa: culture, controversy, and change* (pp. 253–282). Boulder: Lynne Rienner.

Mackie, G. (2003). Female genital cutting: a harmless practice? *Medical Anthropology Quarterly*, *17*(2), 135–158.

Mackie, G., & LeJeune, J. (2009). Social dynamics of abandonment of harmful practices: a new look at the theory. Special Series on Social Norms and Harmful Practices, Innocenti Working Paper 6. Available online at http://www.unicef-irc.org/publications/pdf/iwp_2009_06.pdf.

McElreath, R., Boyd, R., & Richerson, P.J. (2003). Shared norms and the evolution of ethnic markers. *Current Anthropology*, *44*(1), 122–130.

Murdock, G.P. (1957). World ethnographic sample. *American Anthropologist*, *59*(4), 664–687.

Murdock, G.P. (1959). *Africa: its peoples and their culture history*. New York: McGraw Hill.

Murdock, G.P. (1969). *Ethnographic atlas*. Pittsburgh: University of Pittsburgh Press.

Murdock, G.P., & White, D.R. (1969). Standard cross-cultural sample. *Ethnology*, *6*, 329–369.

Murray, J.M. (1974). *The Kikuyu female circumcision controversy, with special reference to the Church Missionary Society's sphere of influence*. PhD thesis, UCLA.

Myers, R.A., Omorodion, F.I., Isenalumhe, A.E., & Akenzua, G.I. (1985). Circumcision: its nature and practice among some ethnic groups in southern Nigeria. *Social Science & Medicine*, *21*(5), 581–588.

Owens, D., & Hayden, B. (1997). Prehistoric rites of passage: a comparative study of transegalitarian hunter–gatherers. *Journal of Anthropological Archaeology*, *16*(2), 121–161.

Pagel, M. (1994). Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proceedings of the Royal Society of London. B: Biological Sciences*, *255*(1342), 37–45.

Pagel, M., & Meade, A. (2006). Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov Chain Monte Carlo. *The American Naturalist*, *167*(6), 808–825.

Paige, K., & Paige, J.M. (1981). *The politics of reproductive ritual*. Berkeley: University of California Press.

Power, C. (1998). Old wives' tales: the gossip hypothesis and the reliability of cheap signals. In J. R. Hurford, M. Studdert-Kennedy, & C. Knight (Eds.), *Approaches to the evolution of language* (pp. 111–129). Cambridge: Cambridge University Press.

Power, C. (2000). Secret language use at female initiation: bounding gossiping communities. In C. Knight, M. Studdert-Kennedy, & J.R. Hurford (Eds.), *The evolutionary emergence of language: social function and the origins of linguistic form* (pp. 81–98). Cambridge: Cambridge University Press.

Quisumbing, A.R., Maluccio, J.A., & et al. (2000). Intrahousehold allocation and gender relations: New empirical evidence from four developing countries. Washington, DC: International Food Policy Research Institute. Available online at http://ebrary.ifpri.org/cdm/ref/collection/p15738coll2/id/125398.

R Core Team (2013). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria.

Richerson, P., Baldini, R., Bell, A., Demps, K., Frost, K., Hillis, V., Mathew, S., Newton, E., Narr, N., Newson, L., Ross, C., Smaldino, P., Waring, T., & M. Zefferman (2015). Cultural group selection plays an essential role in explaining human cooperation: a sketch of the evidence. *Behavioral and Brain Sciences*, FirstView. doi:10.1017/S0140525X1400106X.

Ross, C.T., Joyas Campino, P., & Winterhalder, B. (2015). Cultural transmission and the interethnic transfer of female genital modification in the African diaspora and indigenous populations of Colombia. *Human Nature*, *26*(4). doi:10.1007/s12110-015-9234-7.

Ross, C.T., & Richerson, P.J. (2014). New frontiers in the study of human cultural and genetic evolution. *Current Opinion in Genetics & Development*, *29*, 103–109.

Schacht, R., & Grote, M. (2015). Partner choice decision making and the integration of multiple cues. *Evolution and Human Behavior*, *36*(6), 456–466.

Shell-Duncan, B., & Hernlund, Y. (2000). *Female circumcision in Africa: culture, controversy, and change*. Boulder: Lynne Rienner.

Shell-Duncan, B., & Hernlund, Y. (2007). Are there stages of change in the practice of female genital cutting? Qualitative research findings from Senegal and the Gambia. *African Journal of Reproductive Health*, *10*(2), 57–71.

Silverman, E.K. (2004). Anthropology and circumcision. *Annual Review of Anthropology*, *33*, 419–445.

Smith, E.A., Bird, R.B., & Bird, D.W. (2003). The benefits of costly signaling: Meriam turtle hunters. *Behavioral Ecology*, *14*(1), 116–126.

Stan Development Team (2013a). Stan: A C++ library for probability and sampling, version 2.0. Available at http://mc-stan.org/.

Stan Development Team (2013b). Stan modeling language users guide and reference manual, version 2.0. Available at http://mc-stan.org/.

Stannus, H. (1919). *The Wayao of Nyasaland*. Cambridge: Harvard African Studies III. Available online at http://catalog.hathitrust.org/Record/100366857.

Tikhonov, A., & Arsenin, V.Y. (1977). *Solutions of ill-posed problems*. New York: Wiley.

Towner, M.C., Grote, M.N., Venti, J., & Borgerhoff Mulder, M. (2012). Cultural macroevolution on neighbor graphs. *Human Nature*, *23*(3), 283–305.

UNFPA-UNICEF (2013). UNFPA-UNICEF joint evaluation of the UNFPA-UNICEF Joint Programme on Female Genital Mutilation/Cutting (FGM/C): Accelerating change, 2008–2013. https://www.unfpa.org/sites/default/files/admin-resource/FGM-report%2012_4_2013.pdf.

UNICEF. (2013). *Female genital mutilation/cutting: a statistical overview and exploration of the dynamics of change*. New York: UNICEF.

Van den Berghe, P.L. (1987). *The ethnic phenomenon*. Santa Barbara: ABC-CLIO.

Van Gennep, A. (1960). *The rites of passage*. Chicago: University of Chicago Press.

Walker, R.S., Wichmann, S., Mailund, T., & Atkisson, C.J. (2012). Cultural phylogenetics of the Tupi language family in lowland South America. *PloS One*, *7*(4), e35025.

Warzazi, H.E. (1986). Report of the working group on traditional practices affecting the health of women and children. UN Doc E/CN.4/1986/42.

Whitehorn, J., Ayonrinde, O., & Maingay, S. (2002). Female genital mutilation: cultural and psychological implications. *Sexual and Relationship Therapy*, *17*(2), 161–170.

WHO. (1998). *Female genital mutilation: an overview.* Available online at http://www.who.int/topics/femalegenitalmutilation/en/. Geneva: World Health Organization.

WHO. (2010). *Dynamics of decision-making and change in the practice of female genital mutilation in the Gambia and Senegal: social science policy brief*. Geneva: World Health Organization.

Wigby, S., & Chapman, T. (2005). Sex peptide causes mating costs in female *Drosophila melanogaster*. *Current Biology*, *15*(4), 316–321.

Cody Ross is a postdoctoral scholar at the Santa Fe Institute. His work focuses on the intersection of applied anthropology, human behavioral ecology, and cultural evolution.

Pontus Strimling is a research fellow at the Institute of Future Studies, Stockholm. He is also a coordinator of experiment-based research at the Institute for Analytical Sociology, Norrköping, and a part-time researcher at the Centre for Studies of Cultural Evolution, University of Stockholm. See http://pontusstrimling.com/.

Karen Paige Ericksen is professor emerita of psychology, University of California, Davis. Her research has focused on reproductive behavior, including infertility patterns in Africa, reproductive rituals, and FGM patterns in Africa and especially Egypt.

Patrik Lindenfors is an associate professor of zoological ecology at Stockholm University. He is currently focusing on the evolution of democracy and religion. See http://www.zoologi.su.se/research/lindenfors/.

Monique Borgerhoff Mulder is a professor of anthropology and is also in the Graduate Group in Ecology and the Center for Population Biology, UC Davis. She works on evolutionary aspects of demography, conservation, and economic development both in East Africa and crossculturally.

Human Nature

Supplementary Materials to: *The Origins and Maintenance of Female Genital Modification across Africa: Bayesian Phylogenetic Modeling of Cultural Evolution under the Influence of Selection*

Cody T. Ross · Pontus Strimling · Karen Ericksen · Patrik Lindenfors · Monique Borgerhoff Mulder

DOI 10.1007/s12110-015-9244-5

Contents

1 Contrasting Phylogenetic Modeling of Adaptive Evolution and Phylogenetic Correlation Methods

Basic phylogenetic correlation methods when applied to cultural trait data (using linguistic phylogenies) involve implicit assumptions that: 1) the cultural trait of interest coevolved with language (meaning the correlational structure of the focal cultural trait is assumed to be a scalar transform of the distance matrix derived

---

For questions, contact:
Cody T. Ross
Department of Anthropology
One Shields Avenue, University of California, Davis
Davis, CA 95616 USA
E-mail: ctross@ucdavis.edu

from linguistic data), and that 2) evolutionary change can be investigated without considering the effects of selection regimes operating on different branches of the tree, often—but not always—through assumption of a global Brownian motion process.

We find both of these assumptions problematic in the domain cultural phlyogenetic analysis. Namely, there is no reason to assume that the phylogeny of language and the phylogeny of a given cultural trait are the same, as independent cultural traits can have independent evolutionary histories [Borgerhoff Mulder et al., 2006]. The fact that linguistic data provides prior information about divergence times for the splitting of physical communities, does not rectify this concern—especially if selection for a cultural trait is strong as a function of a covariate. Additionally, the very hypotheses we hope to test with cross cultural data are often adaptive hypotheses—hypotheses that require an estimation of the evolutionary pressure exerted on a given cultural trait by socio-ecological conditions or another cultural trait.

We note that the mean vector used in our model (see Equation 12 in the main text) does not simply assume that there is a single global process which defines the expected value of a cultural trait, instead the expected value of a trait is defined by the strength of selection, and the postulated selective regimes operating on each epoch in each linage. Correlations in outcomes can thus be understood as arising from either similar selection regimes (through the definition of μ), or from shared cultural history (through the definition of V).

Next, we note that the variance-covariance matrix used in our model (see Equation 13 in the main text) is not simply a variance-covariance matrix derived from linguistic data—it is a variance-covariance matrix specified randomly at each MCMC iteration whose value is jointly conditional on: 1) a randomly generated matrix of culture group divergence times (based on the ASJP topology, with random local perturbations, see next subsection "Modeling Phylogenetic Uncertainty"), 2) the estimated strength of selection, α, acting on trait evolution over the phylogeny, and 3) the estimated strength of Brownian motion, σ.

The relative value of covariance in V across cultural groups, up to a factor, σ^2, is conditional only on the estimated divergence times of the cultural groups and the strength of selection acting on the evolving trait of interest, with covariance between cultural groups decreasing monotonically, as α increases and selective dynamics drive cultural groups to adopt locally optimal behaviors (See Figure 1).

[Fig. 1 about here.]

In cases with strong selection, error structure will not be determined by linguistic information; instead, error structure will tend towards assumptions of independence. Correlations in outcome realizations, however, will arise through the correlations introduced directly in the μ vector, via the θ parameters describing the selective regimes operating on each cultural group. If selection is strong and cultural groups share selective regimes, the values in their μ vectors will be positively correlated.

In cases where selection is very weak for a focal cultural trait (based on the selective regimes painted on to the phylogeny), then error structure tends towards that which would be expected from the linguistic divergence data and the estimated strength of Brownian motion fluctuations.

This model behavior makes the application of phylogenetic models of adaptive evolution to cultural data less problematic than the application phylogenetic correlation methods to such data, in that phylogenetic models of adaptive evolution can disentangle—to some extent—the effects of convergent selection and shared ancestry.

Furthermore, the selective regimes painted onto the phylogeny can even be based on the linguistic groups *themselves*. As such, the selective pressure exerted on an evolving cultural trait in specific language subgroups can even be estimated during model fitting. We use such methods in the OU(15) sub-model of the evolution of stratification conditional on language group, included in the OU(2) model of the evolution of FGMo conditional on stratification, detailed in the main text. Using such methods, the strength of the inferred association between an evolving cultural trait (stratification in this case) and language group, is estimated directly and allowed to vary across language clusters.

2 Investigating the Coevolution of Stratification and Language

To investigate the relationship between stratification and language group, we fit an OU(1) model (a single global selection regime for stratification) and an OU(15) model (with unique optima as a function of language family and subfamily).

In the OU(15) model, the effect of selective forces associated with language group on the evolution of stratification can be estimated directly. In such a context, learning the language group of a culture contributes direct information about the strength of selective regimes operating on that cultural group during any epoch in the past, which in turn allows for the likelihood of that culture being stratified at any epoch to be calculated conditional on the model assumptions and data. Using linguistic information to analyze the evolution of stratification in such a context is methodologically appropriate, because the direct effect of each language group on the evolution of stratification is estimated in the model.

Table 1 displays the relevant parameter estimates from the OU(1) and OU(15) models. We observe increased likelihood of stratification in some language groups (Semetic, Berber, and Mande language groups), decreased likelihood of stratification in other groups (the Sudanic language groups), and a lack of relationship to stratification in other groups.

[Table 1 about here.]

Table 2 presents the results of model comparison. WAIC-based model comparison suggests that the larger OU(15) model with distinct evolutionary optima for language family and sub-family, generates better predictions than the OU(1) model, in spite of its increased parameter complexity. These results illustrate that the relationship between stratification and language group is reliable, and that our method of inferring the likelihood of stratification in non-terminal branches based on language group (as detailed in the main analysis) is well justified, and not based on assuming covariance between language group and stratification to a greater extent than is supported by the data.

In cultures where selective regimes based on language group are not associated with stratification, linguistic group conveys very little information about the likelihood of stratification to our main model linking stratification and the evolution of FGMo (See *Supplementary Table "StratificationPreferences.xlsx"* for estimates of the likelihood of stratification at deeper nodes of the phylogeny based on the OU(15) model).

[Table 2 about here.]

3 Modeling Phylogenetic Uncertainty

To account for phylogenetic uncertainty, we sample a new phylogenetic tree at each MCMC iteration. All sampled trees have the same gross topology, meaning that the hierarchical clustering of cultural groups is unchanged across MCMC iterations; however, the divergence times (S and C) of branches on the phylogeny differ across MCMC samples. Our model thus integrates over small scale perturbations in branch times, while still making use of the maximum likelihood tree structure based on ASJP data.

We begin with the C_{ML} matrix, which contains the maximum likelihood estimates of divergence times for each epoch in each lineage taken directly from the ASJP records [Holman et al., 2011, Supplementary Data A]. We fix two minor issues with ASJP dates: 1) there are four cases where the terminal divergence dates of a cultural group are not defined in the ASJP data—we model these splits as unknown parameters with uniform constrained support on the interval between the divergence time of the previous epoch and the present time, and 2) there are four cases where the ASJP dates for divergence time of a parent family in the linguistic hierarchy are younger than at least one of their daughters—we model these data by replacing the pathological ASJP date with a parameter that has uniform constrained support on the interval between the ASJP divergence times of the previous and subsequent epochs.

To proceed with the model, we then decompose the C_{ML} matrix into a two-dimensional array of values, D, which is composed of the unique divergence times in the phylogenetic tree, and a map, ID_D, that links the parameters in the D array back to their locations in the the C_{ML} matrix. D is of length $\varGamma$ and has a width equal to the number of unique parameters at each epoch. For example, in the first cell of D, the width is 1, since during the first epoch all cultural groups are considered to have been evolving as part of the same group, in the second cell the width is 4, since there are four unique divergence times (one for Khoe, one for Nilo-Saharan, one for Niger-Congo, and one for Afro-Asiatic); this pattern continues for each level of the hierarchical clustering of language groups. The variable ID_D indicates the locations where a parameter in D occurs in C_{ML}.

We then create an array of parameters, δ with direct correspondence to D in order to add random fluctuations to each value in D. Each cell in δ is given upper and lower constraints, M_x and M_n, respectively. A prior is specified each cell of δ such that:

$$\delta \sim Normal(0, M_x)T[M_n, M_x]$$

where $T[M_n, M_x]$ indicates the truncated likelihood operator, and the standard deviation set to equal M_x gives each δ parameter a weakly informative prior. We sum D and δ, and transform the results from years before present, as in ASJP, to years since the origin of the most basal node; we use the ID_D map to paint the resulting values back onto a parameter matrix, C_{Random}, which has the same structure as the original C_{ML} matrix. We then error check each proposed C_{Random} matrix to ensure that each and every lineage in the matrix is positive-ordered (ie. each cell in the lineage is positive and larger than the previous cell in the lineage). If any proposed matrix is not positive-ordered, the proposed MCMC sample is Metropolis rejected to maintain detailed balance. C_{Random} is standardized to the unit interval to become C, the parameter matrix used in the model description described in the main text. The S matrix is derived from C, by returning the value of C in lineage n_1 for the most recent epoch in which cultural group n_1 and n_2 were evolving as a single group.

The modeling described in this section and in the methods section of the text is described more directly in R/Stan computer code, also included in this supplement. Mx and Mn are set to 350 years and -350 years respectively, to give a 700 year window of uncertainty around each branch point in the phylogeny. See *Supplementary Table "CutTimesPosterior.xlsx"* for the ASJP dates, and the associated mean and 95PCIs of the posterior realizations of these dates generated by our model.

4 Testing the Basic Model with Simulated Data

The adaptive phylogenetic models used in this study are large and complicated. In this section, we briefly detail the output and results of our model when simulated data are analyzed. Testing a model with simulated data is an essential step in verifying that a model is capable of properly analyzing real data and returning parameter estimates that make sense in light of a given data set.

We begin by utilizing the empirical stratification data, and then simulating fourteen distinct FGMo data sets from the stratification data using a conditional Bernoulli probability generator with the probability parameter set to various values if $Stratification = 1$, and various other values if $Stratification = 0$ (See *Supplementary Table "SensitivityAnalysis.xlsx"*, for settings and results).

We fully analyzed each simulated data set with the main OU(1) and OU(2) models described in the main text. For each simulated data set and each model, we ran a single Markov chain for 5,000 warm-up iterations, and sampled the following 15,000 iterations thinning at an interval of 4.

The models fit in this sensitivity analysis used half-Cauchy priors on the positive constrained parameters; we later found half-Gassian priors to perform better. In spite of this, results from the simulation study show that our phylogenetic models are effective at recovering parameter values that make sense in light of the simulated data, and that the WAIC metric is effective in choosing the more complex model when stratification leads to divergence in FGMo frequency, and choosing the null model when stratification is not predictive of FGMo frequency.

5 Current Shortcomings in Our Formulation of Bayesian Phylogenetic Modeling of Adaptive Evolution

1) *Assumption of Piece-wise Constant Selection Regimes* - We find the assumption of piece-wise constant selection regimes to be quite troublesome. Simply because

we know a given cultural group to be stratified at the current time, does necessarily mean that this group has been stratified over the entire terminal epoch.

This assumption in our model gives the cultural groups with deep divergence times for their terminal epoches disproportionate weight in the estimation of model parameters; furthermore, because the terminal branches are fixed, the value of μ for a cultural group with deep divergence times is much more sensitive to the values of θ_{nssr} and θ_{ssr} than the value of μ for a cultural group with a more recent terminal divergence time, since θ_{nssr} and θ_{ssr} are mixed for the majority of this cultural group's evolution. Furthermore, this behavior undermines the effectiveness of WAIC, because WAIC is very sensitive to pointwise predictive density, and thus a single outlier with a deep divergence time for its terminal epoch can have a strong impact on WAIC.

It would be optimal to let the θ_{nssr} and θ_{ssr} parameters on the terminal branch segments mix as a function of time since present in order to better represent our uncertainty about the nature of the selection regime in the past. We know of no general, non-*ad hoc* methods for implementing such a function at this time, so new research is needed.

In our specific model formulation, an attractive option might be to use $\Lambda_{[n,\kappa(n)]}$ to mix θ_{nssr} and θ_{ssr} on the terminal branches, instead of simply painting them on to the terminal branches in a binary fashion using the observed data. We fit this 'random terminal branch tip model', in addition to the 'fixed terminal branch tip model'. The random terminal branch tip model model was preferred by WAIC (See Table 3) to the fixed branch tip model and the OU(1) model, possibly because the sensitivity of predictions to θ_{nssr} and θ_{ssr} in cultures with deep terminal divergences is reduced. This choice, however, throws away too much information, since we know that state of stratification in each cultural group to a greater degree of certainty than $\Lambda_{[n,\kappa(n)]}$ for the recent past.

[Table 3 about here.]

2) *Prior Specification* - Our model uses what we believe from theory and semi-systematic experimentation to be reasonable priors. However, no rigorous study has compared the performance of various priors on inference in models like ours. We find that both half-Gaussian ($Normal(0,5)T[0,\infty]$), and half-Cauchy ($Cauchy(0,1)T[0,\infty]$, or $Cauchy(0,0.1)T[0,\infty]$) priors led to similar results.

Our sensitivity analysis used half-Cauchy priors, but we later elected to use half-Gaussian priors in our final MCMC runs, since these priors mitigated some pathological behavior associated with the half-Cauchy parameterization.

3) *WAIC needs more careful thought* - WAIC is conditioned upon the pointwise evaluation of data and model parameters. When selection is strong, WAIC comparison in our model is properly defined because the model errors approach independence, and each datum can be seen as a unique point, for which a pointwise approach to predictive evaluation makes sense (in other words, the predictive fit of each cultural group gets equal weight in WAIC, because each cultural group is in fact an independent data point). However, as drift becomes important relative to selection, WAIC as we have defined it becomes somewhat tendentious, because WAIC still believes each datum to be a unique point useful for evaluating predictive information loss, even though each point is not completely independent.

However, to the extent that we care about our model predicting each point, our formulation of WAIC seems valid; on the other hand, one could argue that we should not care as much about our model predicting each point, but rather each language subfamily. We concede to such an argument, but lament that methods of model comparison useful for such contexts are still in development. Work on WAIC is still a very new field, and new computational formulations of WAIC may be useful for approximations to leave-one-group-out, instead of leave-one-point-out, cross validation. Such methods would be an improvement over our methods herein. WAIC appears to be well behaved in our sensitivity analysis, however, so we are reasonably confident that our WAIC results are valid.

6 Assessing Model Fit

We investigated model fit by plotting the traceplots and pairs plots of all model parameters. We include these plots for the main model parameters in each of the following models: OU(1)FGMo|Null, OU(2)FGMo|Stratification (Fixed Terminal Branch Tips for FGMo|Stratification), OU(2)FGMo|Stratification (Random Terminal Branch Tips for FGMo|Stratification), OU(1)Stratification| Null, OU(15)Stratification|Language(Fixed Terminal Branch Tips for FGMo|Stratification), and OU(15)Stratification|Language(Random Terminal Branch Tips for FGMo|Stratification).

6.1 Pairs Plots

[Fig. 2 about here.]

[Fig. 3 about here.]

[Fig. 4 about here.]

[Fig. 5 about here.]

[Fig. 6 about here.]

[Fig. 7 about here.]

6.2 Traceplots

[Fig. 8 about here.]

[Fig. 9 about here.]

[Fig. 10 about here.]

[Fig. 11 about here.]

[Fig. 12 about here.]

[Fig. 13 about here.]

Additionally, we used Stan's posterior summary statistics r_{hat} and effective sample size, n_{eff}, to assess model fit. These statistics are included in *Supplementary Table "StanSummaryOfResults.xlsx"*. All θ parameters were well estimated and clearly identified; however, σ, and to a lesser extent α, showed signs of poor identification in some models, especially under half-Cauchy rather than half-Gaussian priors. This behavior is not unexpected, due to the inherent issues with identification of α, θ, and σ in the Butler and King [2004] framework.

Our use of regularizing priors, however, improves identification of most model parameters over the standard Maximum Likelihood Estimation methods mentioned in Butler and King [2004]. All θ parameters are well behaved in every model, and we fail to see any serious identification issues between α and θ. We only note a minor issue that α can grow substantially when the θ parameters of a model approach zero in a given MCMC sample (this behavior is more common with half-Cauchy priors than with half-Gaussian priors). This behavior could be rectified by adopting stronger priors that place very low prior probability at values of $\alpha > 20$ as per the suggestion described in Butler and King [2004, Supplementary Appendix].

Half-Gaussian prior distributions appear to perform better than half-Cauchy priors in constraining α and σ. In our models, α never grew pathologically for more than a few MCMC iterations before returning to the area of high density (< 10), even with half-Cauchy priors, so both kinds of prior are deemed reasonable.

The σ parameter, however, shows the worst issues with identification across models, as it tends toward a random walk (under half-Cauchy priors) as it grows past a value of 8 or larger, especially in the OU(2)FGMo|Stratification model. This behavior likely results from two keys facts about our model: 1) in our model formulation σ and α are not perfectly orthogonal, so the 'correct' value of covariance in V, can always be specified by different combinations of α and σ, leading to a 'tug-of-war' between these parameters. Pragmatically, however, α is constrained heavily because it factors into the likelihood through both the mean vector and the variance-covariance matrix; as such, α and σ are fairly well identified from each other.

More importantly, 2) σ only factors into the overall model likelihood via its ability to modulate the variance and covariance in predicted values of Ψ; since $\lim_{x \to \infty} logistic(x) = 1$, and $\lim_{x \to -\infty} logistic(x) = 0$, once σ has grown sufficiently large, its evolution can diverge to an ever increasing random walk without leading to numeric differences in predictions or the likelihood of a model. While the exact numerical value of the σ parameter in such a context is not identifiable, lack of numerical identifiability does not lead to a non-identified model, nor does it hamper inference concerning the role σ plays in the evolutionary dynamics of the system. If σ is large enough such that it diverges to a random walk, then inferentially we know sigma to be large relative to selection pressure α, even if we cannot assign it a unique numerical value or range. Figure 14 illustrates how density on the unit interval—of logistic($\Psi_{[n]}$)—changes as a function of the value of σ. After σ exceeds approximately 8 or 10, probability density hardly changes as σ continues to increase on the linear scale. We note that σ is better constrained by half-Gaussian, than half-Cauchy priors.

[Fig. 14 about here.]

7 Alternative Phylogenetic Methods

In our paper, we elected to use and focus on the Butler and King methodology for phylogenetic modeling of adaptive evolution. Other approaches are also possible [see Pagel, 1994; Pagel and Meade, 2006]. We were keen on the ability of the Butler and King methodology to allow us to: 1) probabilistically model the state of stratification in deeper nodes of the phylogenetic tree, and to test if there are distinct evolutionary optima of the social preference for FMGo that depend on the state of stratification, and 2) partially disentangle the effects of selection and drift.

For thoroughness, we present the results of other standard methods for phylogenetic analysis of discrete traits in this supplement.

7.1 PGLS Methods

Phylogenetic signals were estimated using the D statistic for binary characters [Fritz and Purvis, 2010], implemented in the R package *caper* [Orme et al., 2011]. To investigate the relative importance of geographic and linguistic proximity, we used Mantel tests as implemented in the R-packages *cluster* [Maechler et al., 2013] and *vegan* [Oksanen et al., 2012], but see [Harmon and Glor, 2010]. For these tests, geographic proximity was measured using ordinary Euclidean distance, while phylogenetic distance was measured by counting the number of nodes between languages in the phylogeny. The model with highest Mantel r statistic was judged as the model providing the highest level of explanation. We checked for correlated evolution using the Pagel [1994] test of correlated (discrete) character evolution as implemented in *Mesquite* [Maddison and Maddison, 2011] and for the influence of stratification using Phylogenetic Generalized Least Squares (PGLS) as implemented in *caper* [Orme et al., 2011].

7.2 PGLS Results

There exists a significant phylogenetic signal in the binary variable, $FGMo$, (D=0.516; significance of phylogenetic signal, p = 0.001; significance of departure from Brownian motion p = 0.089). Interestingly, we found significant phylogenetic signals for both latitude (p<0.001) and longitude (p<0.001) indicating a marked correlation between the language phylogeny and the geographical location of the populations studied.

Utilizing Mantel tests, we attempted to tease apart the relative importance of geographic proximity and phylogenetic proximity. We tested four alternative predictor models: (1) linguistic proximity, (2) geographical proximity, (3) linguistic proximity with geographical proximity as a covariate, and (4) geographical proximity with linguistic proximity as a covariate. These tests indicated that geographical proximity with linguistic proximity as a covariate was the model with most explanatory power for all variables. However, due to the close correlation between linguistic and geographical proximity, as indicated by the strong phylogenetic signal for latitude, as well as the methodological problems with the Mantel tests

[Nunn et al., 2006; Dekker et al., 2007], we provide results of both phylogenetic and non-phylogenetic analyses of the hypothesis tests.

We failed to find evidence of a strong correlation between FGMo and stratification. Both non-phylogenetic, non-parametric Kruskal-Wallis tests (H = 1.270, p = 0.736), as well as Phylogenetic Generalized Least Squares tests (λ=0, p = 0.335), gave the same qualitative results.

7.3 PGLMM Methods

In addition to the simple models outlined above, we compared our main results to those produced using a phylogenetic implementation of a standard generalized linear mixed model (PGLMM) [Ives and Garland Jr, 2014], using a Bayesian estimation procedure.

As in Ives and Garland Jr [2014], we have:

$$\text{FGMo}_{[n]} \sim \text{Bernoulli}(P_{[n]}) \tag{1}$$

$$\text{logit}(P_{[n]}) = \beta_0 + \beta_1 \text{Strat}_{[n]} + \epsilon_{[n]} \tag{2}$$

$$\epsilon \sim \text{Multivariate Normal}((0, \ldots, 0), \sigma^2 S) \tag{3}$$

where S is a distance matrix with unit diagonal derived from the linguistic phylogeny.

In our model, we have priors:

$$\beta \sim \text{Normal}(0, 5) \tag{4}$$

$$\sigma^2 \sim \text{Cauchy}(0, 1)T[0, \infty] \tag{5}$$

See supplementary model code for implementation and further details.

7.4 PGLMM Results

Using the PGLMM model outlined in Ives and Garland Jr [2014] we find results that are qualitatively comparable to our main findings. The effect of stratification on the log odds of a population practicing FGMo is 0.49 (PCI95: -2.37, 4.52), which is positive, but non-significant.

8 FGMo Prevalence

Finally, since considering any cultural group as an FGMo practicing group if there is any normative use of FGMo (even at low frequencies) has the potential to inflate the size of the drift term, we collected data on contemporary FGMo prevalence, and reclassified groups as practicing FGMo only if contemporary prevalence was at least 50%. Because prevalence data by ethnicity is hard to come by, due to missing data, our revised model had N=100, rather than N=112, observations.

As in the main analysis, we find that σ dominates α in both the OU(1) and OU(2) models, and that the OU(1) model is preferred to the OU(2) by WAIC. See Tables 4 and 5.

[Table 4 about here.]

[Table 5 about here.]

9 Interpreting Efferson et al. [2015]

Following cautions raised in Efferson et al. [2015], we emphasize that our model does not assume that the social transmission bias arises purely from a coordination game; a wide range of social pressures, be they based on religious obligation [Gruenbaum, 2001], gender marking [Gruenbaum, 2001], coordination [Mackie, 2000], or cultural identity [Daniels, 1970], can all mutually interact to yield the net effects accounted for by $s(\alpha)$. Again following Efferson et al. [2015], our model formalizes the assumption that some males may hold considerable power in the mating market and explicitly considers the effects of providing these men with virginity assurance or costly signals of sexual fidelity, via the functions $V_m(r_f(f+i, d_f))$ or $V_m(r_f(f+x, d_f))$. Our model also integrates information on the costs, c, of FGMo, but we did not consider inter-individual variation in these costs in the main text.

These costs, however, are likely to be heterogeneous across individuals, and could possibly contribute to the empirical observation by Efferson et al. [2015] that, in some cases, FGMo occurs in populations at levels well below fixation. To see how, let us consider the behavior of the model when there are two classes of men, rich and poor, and the cost of undergoing FGMo depends on the wealth class of men to which a given woman is paired, with wives of the rich men paying cost c_r and wives of poor men paying the cost c_p—this assumption may be justified by noting that the costs of FGMo on women are often expressed via long-term medical complications, especially during pregnancy [Mackie, 2003; Shell-Duncan and Hernlund, 2000; Banks et al., 2006], and that males with increased financial resources might be able to better offset these costs by increasing the access of females to higher quality health care.

Although we limit the analytical exploration of our model in the main text to special cases describing the origins ($\alpha = 0$) and maintenance ($\alpha = 1$) of the FGMo cultural trait, our model provides a framework for understanding the dynamics leading to intermediate frequencies of FGMo, such that $0 < \alpha < 1$. Given the generality of our model, we can define the marriage value function for males, V_m, such that for some arbitrary fraction of males, $\hat{\alpha}$, $V_m(r_f(f+i, d_f)) - V_m(r_f(f, d_f)) > s(1-\hat{\alpha}) - s(\hat{\alpha}) + c_r$, and for $1-\hat{\alpha}$ of males, $V_m(r_f(f, d_f)) = V_m(r_f(f+i, d_f))$. This condition formalizes an empirical observation that in many populations the majority of male wealth values are similar, with a few outliers being responsible for the majority of the inequality [Borgerhoff Mulder et al., 2009], and leads to a situation where there is wealth-based incentive for the highest-ranked fraction of women, $\hat{\alpha}$, to practice FGMo, and there is no wealth-based incentive for the lowest-ranked $1-\hat{\alpha}$ fraction of women to practice FGMo, since there is no difference in marriage value as a function of undergoing FGMo, due to $V_m(r_f(f+i, d_f)) - V_m(r_f(f, d_f)) = 0$.

However, for the lowest ranked woman paired with one of the wealthy men, her marriage partner will be the same with or without her undergoing FGMo, so for her not to switch strategies, social pressures must be responsible; she will not switch strategies so long as $s(\hat{\alpha}) - s(1 - \hat{\alpha}) > c_r$. For the highest ranked female paired to one of the poor males, she will not undergo FGMo when $P - P_{VA} > 0$, which holds so long as $s(\hat{\alpha}) - s(1 - \hat{\alpha}) < c_p$; this condition implies that all lower-ranked women will also not engage in FGMo.

Because c_r, c_p, and s are arbitrary, we can always find parameter values to satisfy both critical conditions—specifically, $c_r < s(\hat{\alpha}) - s(1 - \hat{\alpha}) < c_p$—so our model is thus capable of producing a stable FGMo distribution for all values of $\hat{\alpha} \in [0, 1]$. Thus, while a pure coordination game may not be able to explain the frequency distribution of FGMo in some groups, as shown by Efferson et al. [2015], even if s was determined solely by a coordination game, our model demonstrates how consideration of FGMo in a mating market operating under under a system of mating payoffs coupled with coordination payoffs and variation in costs of FGMo could produce the empirical outcomes documented by Efferson et al. [2015]. Following this line of logic, and much previous emprical research, we argue that it may be more nuanced to argue that FGMo is not *purely* a social coordination norm, than to argue that it is not a coordination norm; depending on the parameters of the model, there could be very strong coordination norms driving FGMo, even when FGMo is practiced at intermediate levels.

10 Bibliography

E Banks, O Meirik, T Farley, O Akande, H Bathija, and M Ali. Female genital mutilation and obstetric outcome: WHO collaborative prospective study in six African countries. *Lancet*, 367(9525):1835–1841, 2006.

Monique Borgerhoff Mulder, Charles L Nunn, and Mary C Towner. Cultural macroevolution and the transmission of traits. *Evolutionary Anthropology: Issues, News, and Reviews*, 15(2):52–64, 2006.

Monique Borgerhoff Mulder, Samuel Bowles, Tom Hertz, Adrian Bell, Jan Beise, Greg Clark, Ila Fazzio, Michael Gurven, Kim Hill, Paul L Hooper, et al. Intergenerational wealth transmission and the dynamics of inequality in small-scale societies. *Science*, 326(5953):682–688, 2009.

Marguerite A Butler and Aaron A King. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *The American Naturalist*, 164(6): 683–695, 2004.

Robert E Daniels. *By rites a man: A study of the societal and individual foundations of tribal identity among the Kipsigis of Kenya.* PhD thesis, University of Chicago, 1970.

David Dekker, David Krackhardt, and Tom AB Snijders. Sensitivity of mrqap tests to collinearity and autocorrelation conditions. *Psychometrika*, 72(4):563–581, 2007.

Charles Efferson, Sonja Vogt, Amy Elhadi, Hilal El Fadil Ahmed, and Ernst Fehr. Female genital cutting is not a social coordination norm. *Science*, 349(6255): 1446–1447, 2015.

S. A. Fritz and A. Purvis. Selectivity in mammalian extinction risk and threat types: a new measure of phylogenetic signal strength in binary traits. *Conservation Biology*, 24(4):1042–1051, 2010.

Ellen Gruenbaum. *The female circumcision controversy: an anthropological perspective.* Philadelphia: University of Pennsylvania Press, 2001.

L. J. Harmon and R. E. Glor. Poor statistical performance of the Mantel test in phylogenetic comparative analyses. *Evolution*, 64:2173–2178, 2010.

Eric W Holman, Cecil H Brown, Søren Wichmann, André Müller, Viveka Velupillai, Harald Hammarström, Sebastian Sauppe, Hagen Jung, Dik Bakker, Pamela Brown, et al. Automated dating of the world's language families based on lexical similarity. *Current Anthropology*, 52(6):841–875, 2011.

Anthony R Ives and Theodore Garland Jr. Phylogenetic regression for binary dependent variables. In *Garamszegi, Lszl Zsolt (Ed.): Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology*, pages 231–261. Berlin: Springer, 2014.

Gerry Mackie. Female genital cutting: the beginning of the end. In *In Bettina Shell-Duncan and Ylva Hernlund (Eds.) Female "circumcision" in Africa: culture, controversy, and change*, pages 253–282. Boulder: Lynne Rienner, 2000.

Gerry Mackie. Female genital cutting: a harmless practice? *Medical Anthropology Quarterly*, 17(2):135–158, 2003.

W. P. Maddison and D. R. Maddison. *Mesquite: a modular system for evolutionary analysis. Version 2.75*, 2011.

Martin Maechler, Peter Rousseeuw, Anja Struyf, Mia Hubert, and Kurt Hornik. *cluster: Cluster Analysis Basics and Extensions*, 2013. R package version 1.14.4. For new features, see the 'Changelog' file (in the package source).

Charles L Nunn, Monique Borgerhoff Mulder, and Sasha Langley. Comparative methods for studying cultural trait evolution: A simulation study. *Cross-Cultural Research*, 40(2):177–209, 2006.

J Oksanen, F Guillaume Blanchet, R Kindt, P Legendre, PR Minchin, RB O'Hara, GL Simpson, P Solymos, M Henry, H Stevens, and H Wagner. *vegan: Community Ecology Package. R package version 2.0-5*, 2012.

C. D. L. Orme, R. P. Freckleton, G. H. Thomas, T. Petzoldt, S. A. Fritz, and N. J. B. Isaac. *caper: Comparative Analyses of Phylogenetics and Evolution in R. (Imperial College, London, UK) Version 0.4*, 2011.

Mark Pagel. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 255(1342):37–45, 1994.

Mark Pagel and Andrew Meade. Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov Chain Monte Carlo. *The American Naturalist*, 167(6):808–825, 2006.

Bettina Shell-Duncan and Ylva Hernlund. *Female Circumcision in Africa: Culture, Controversy, and Change.* Boulder: Lynne Rienner, 2000.

List of Figures

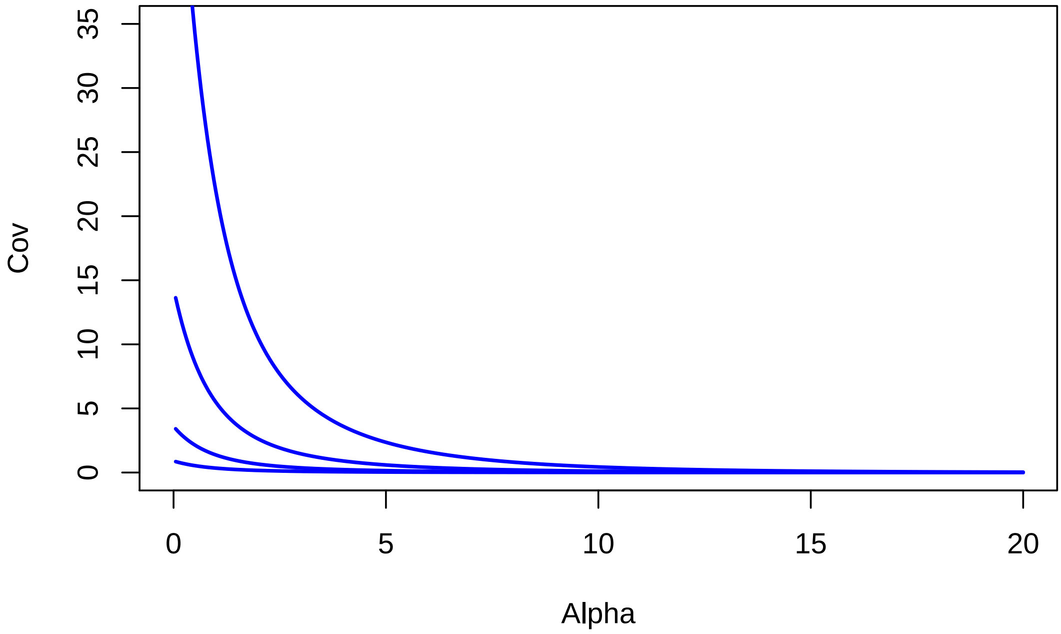


Fig. 1 Figure 1 displays a plot of the covariance (as estimated in Equation 13 of the main text) between two cultural groups who diverged at time $t = 0.9$ in the unit-scaled phylogeny. As α grows from $0 \to 20$ covariance declines exponentially toward zero. Each line is a contour for $\sigma^2 = \{1,4,16,64\}$ from lower-left to upper-right.

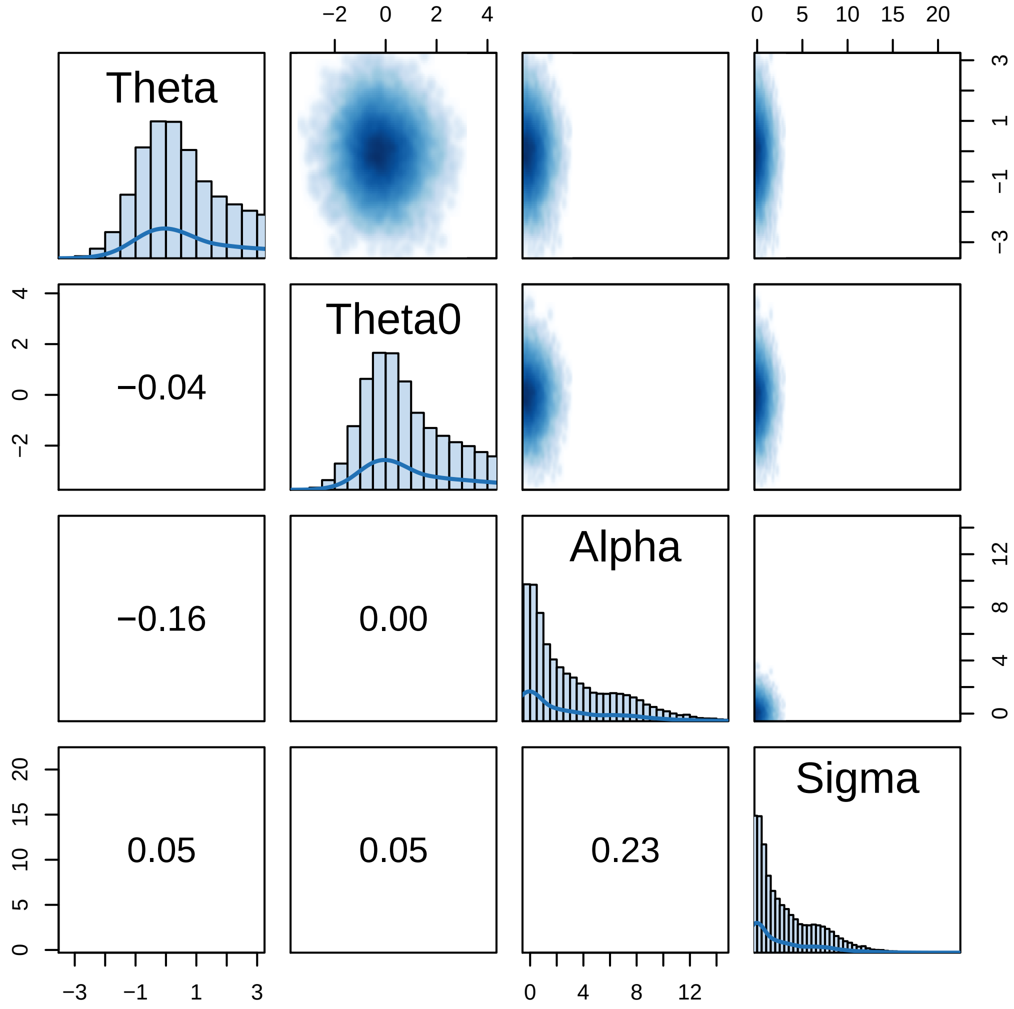


Fig. 2 Figure 2 displays the pairs plot for the OU(1)FGMo|Null model.

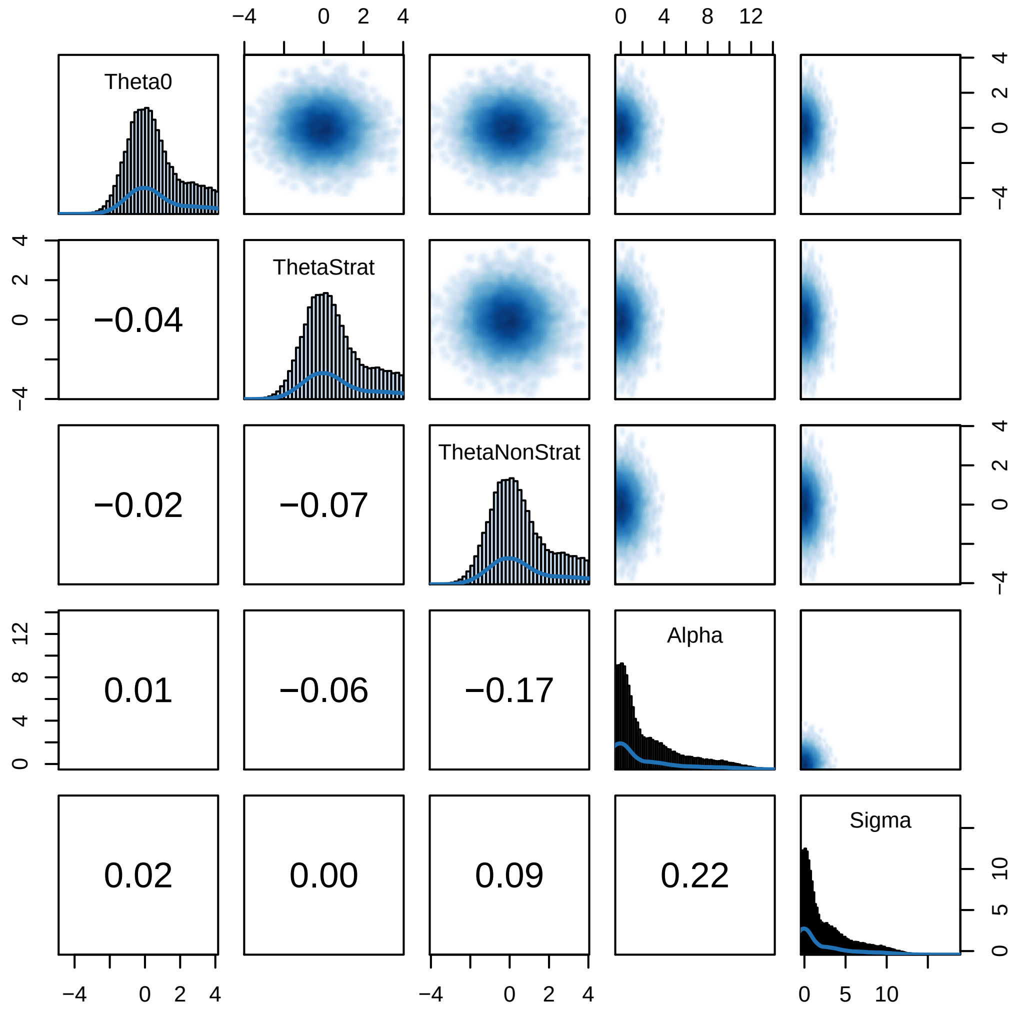


Fig. 3 Figure 3 displays the pairs plot for the OU(2)FGMo|Stratification model with fixed terminal branch tips for FGMo|Stratification.

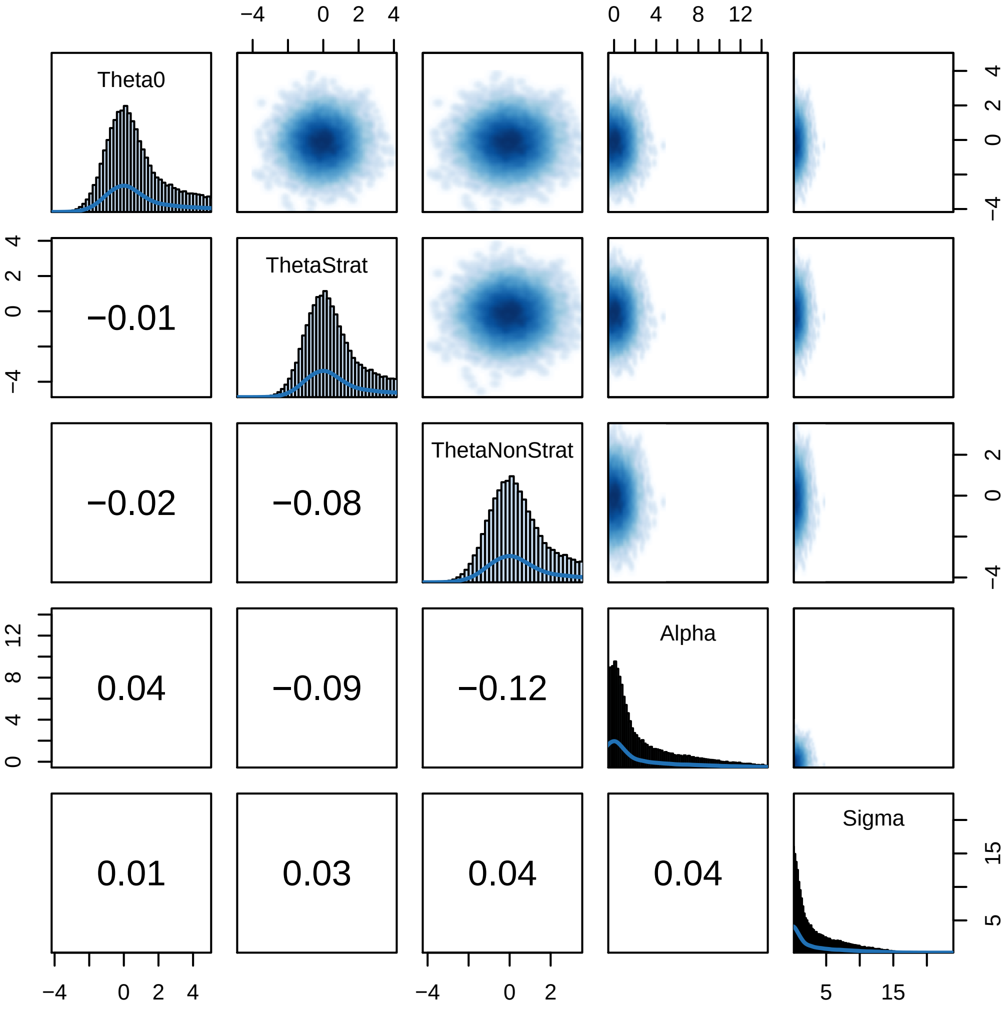


Fig. 4 Figure 4 displays the pairs plot for the OU(2)FGMo|Stratification model with random terminal branch tips for FGMo|Stratification.

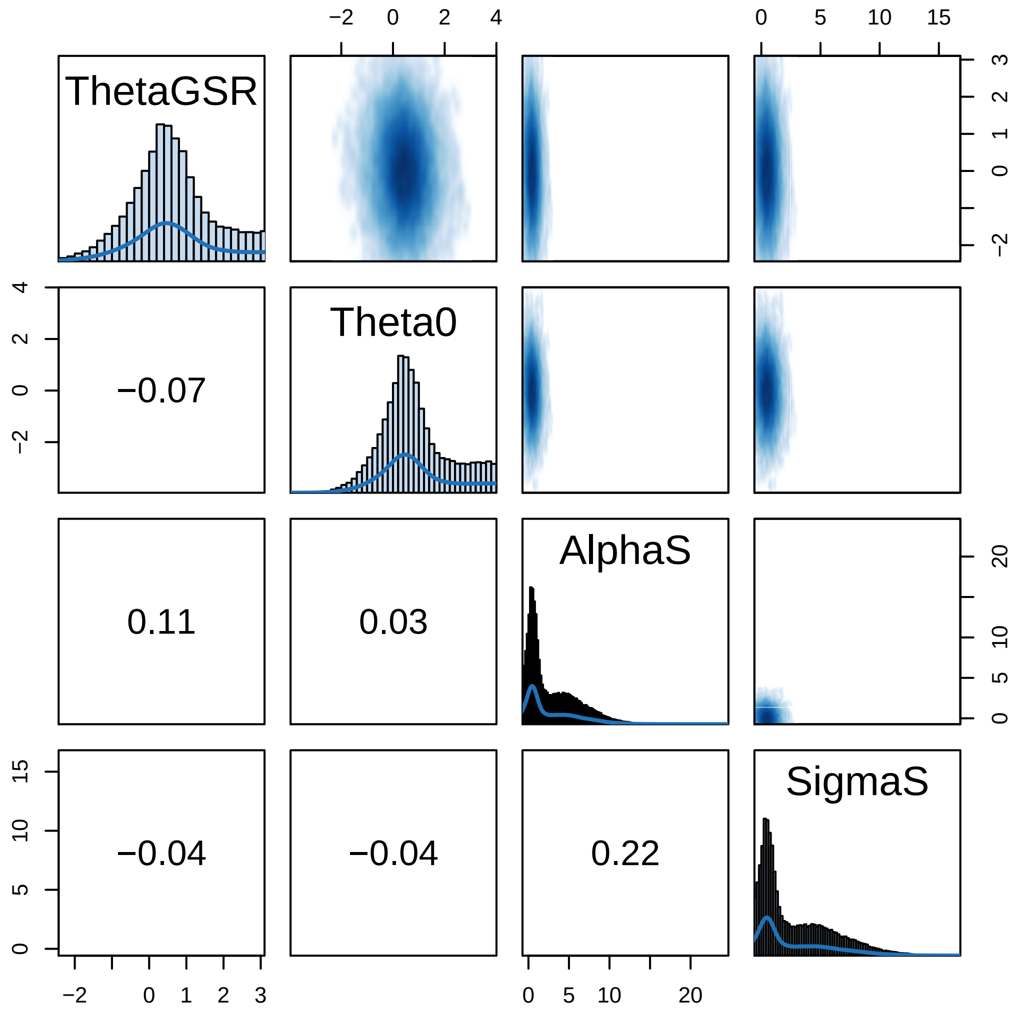


Fig. 5 Figure 5 displays the pairs plot for the OU(1)Stratification|Null model.

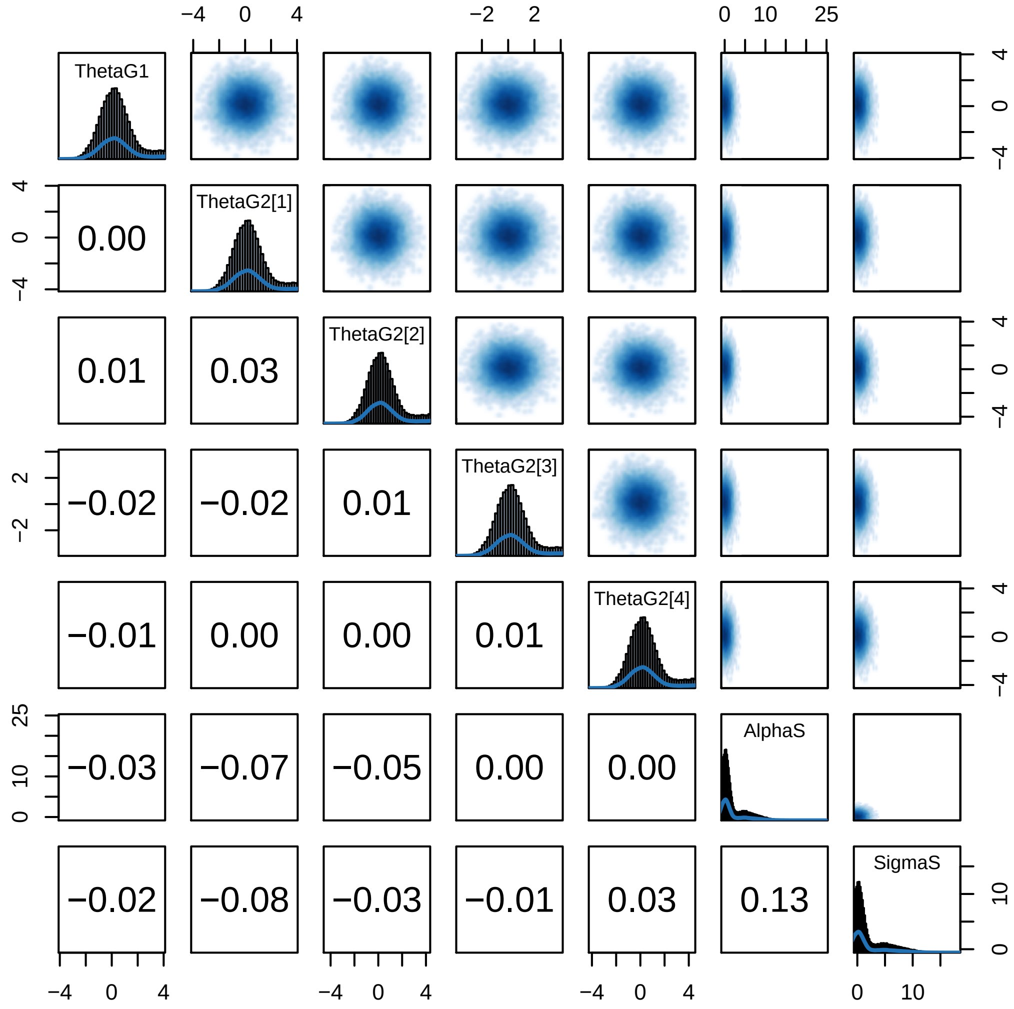


Fig. 6 Figure 6 displays the pairs plot for the OU(15)Stratification|Language model, with fixed terminal branch tips for FGMo|Stratification.

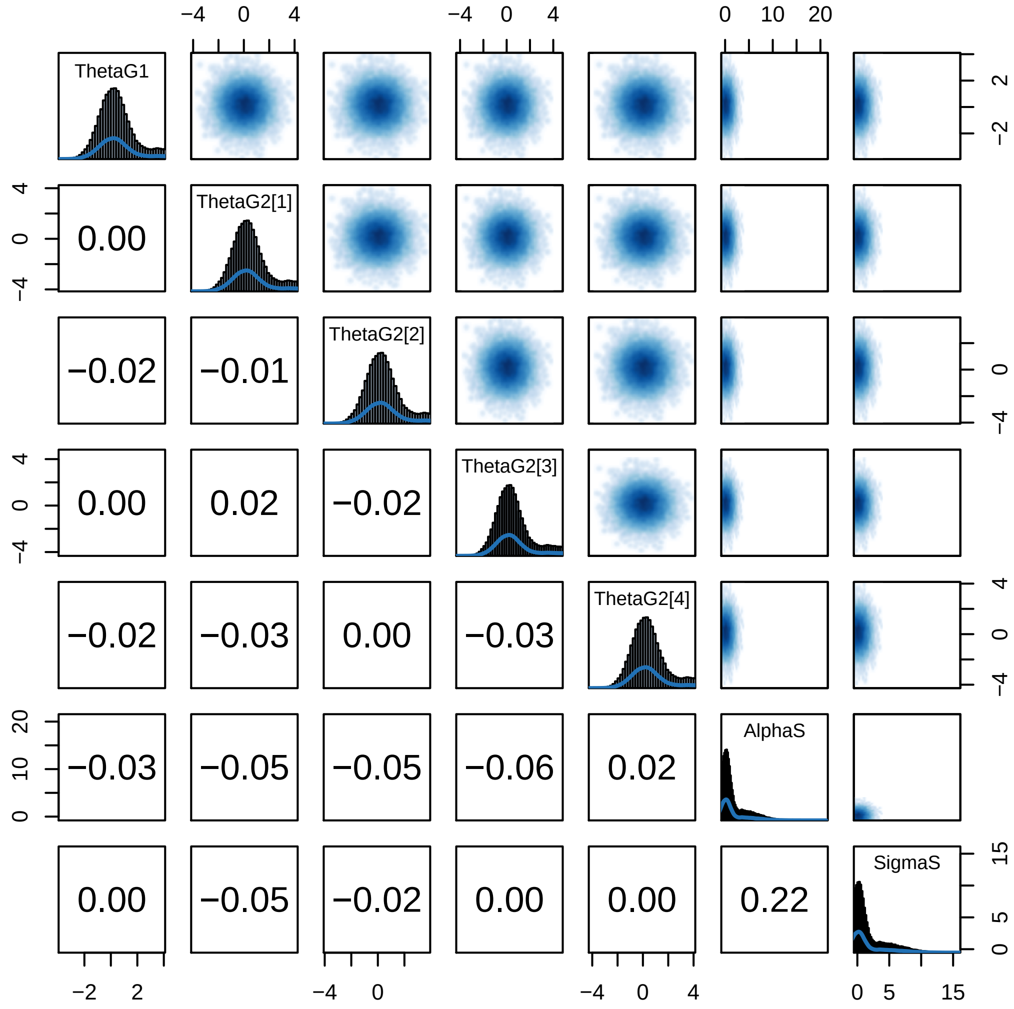


Fig. 7 Figure 7 displays the pairs plot for the OU(15)Stratification|Language model, with random terminal branch tips for FGMo|Stratification.

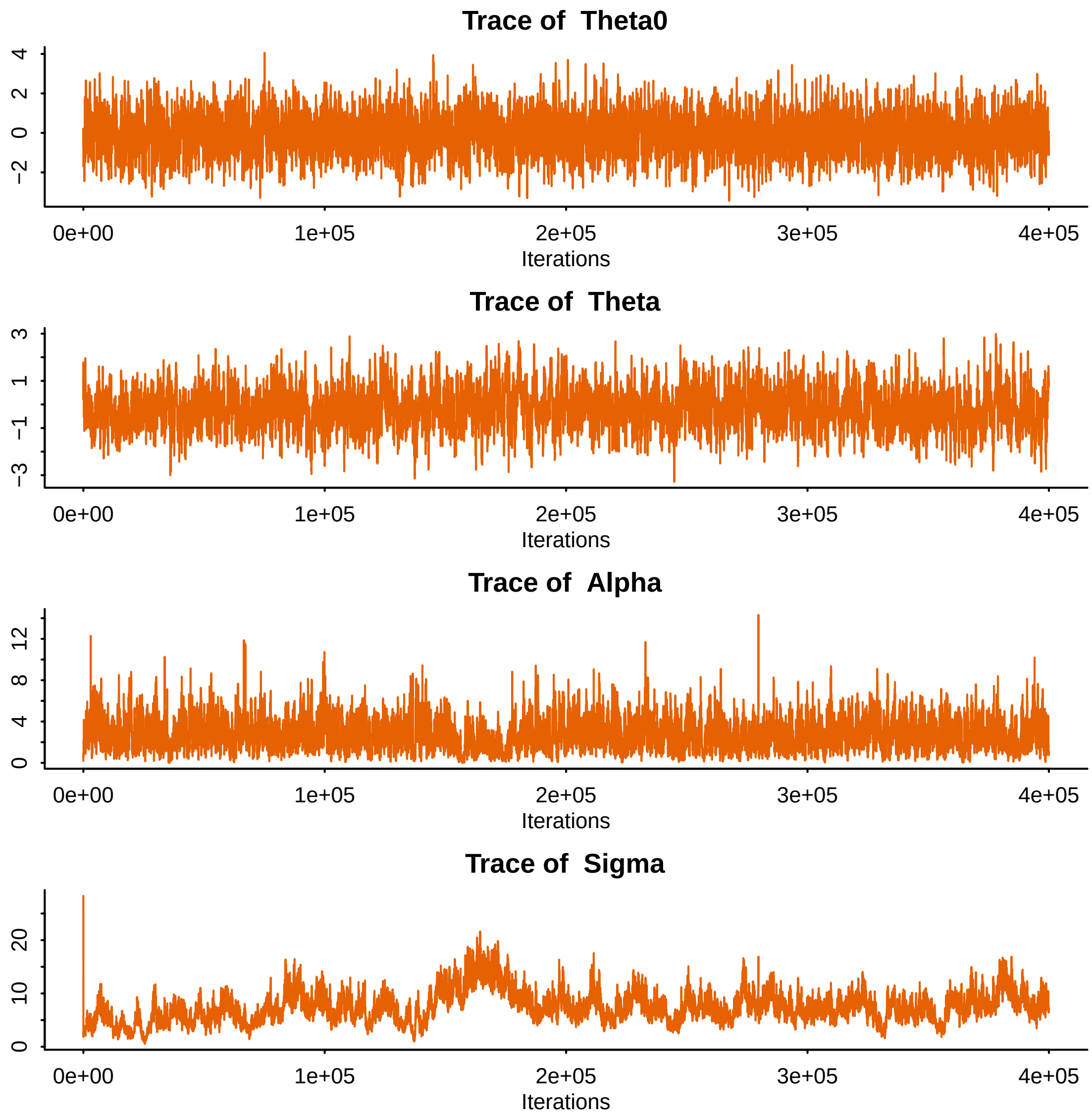


Fig. 8 Figure 8 displays the traceplot for the OU(1)FGMo|Null model.

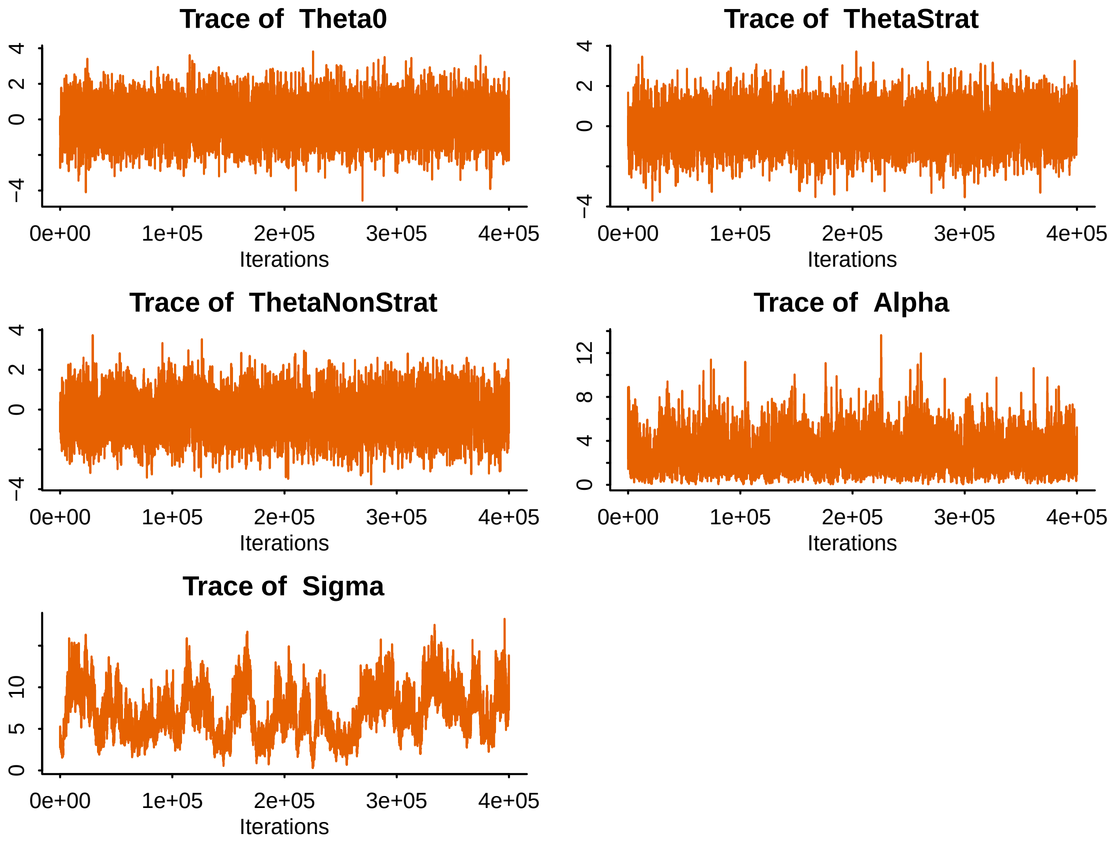


Fig. 9 Figure 9 displays the traceplot for the OU(2)FGMo|Stratification model with fixed terminal branch tips for FGMo|Stratification. Note that α and θ are stationary, well identified, and well mixed. On the other hand, we see that σ is well identified, but not as well mixed.

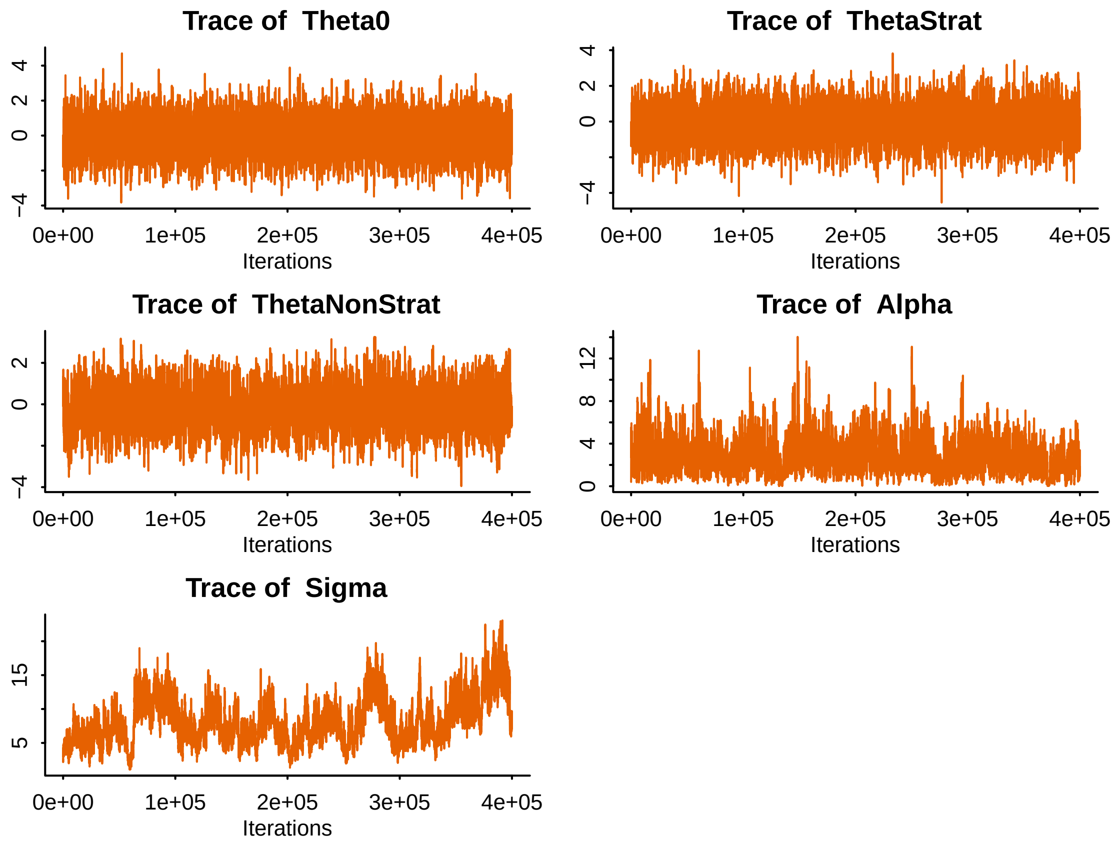
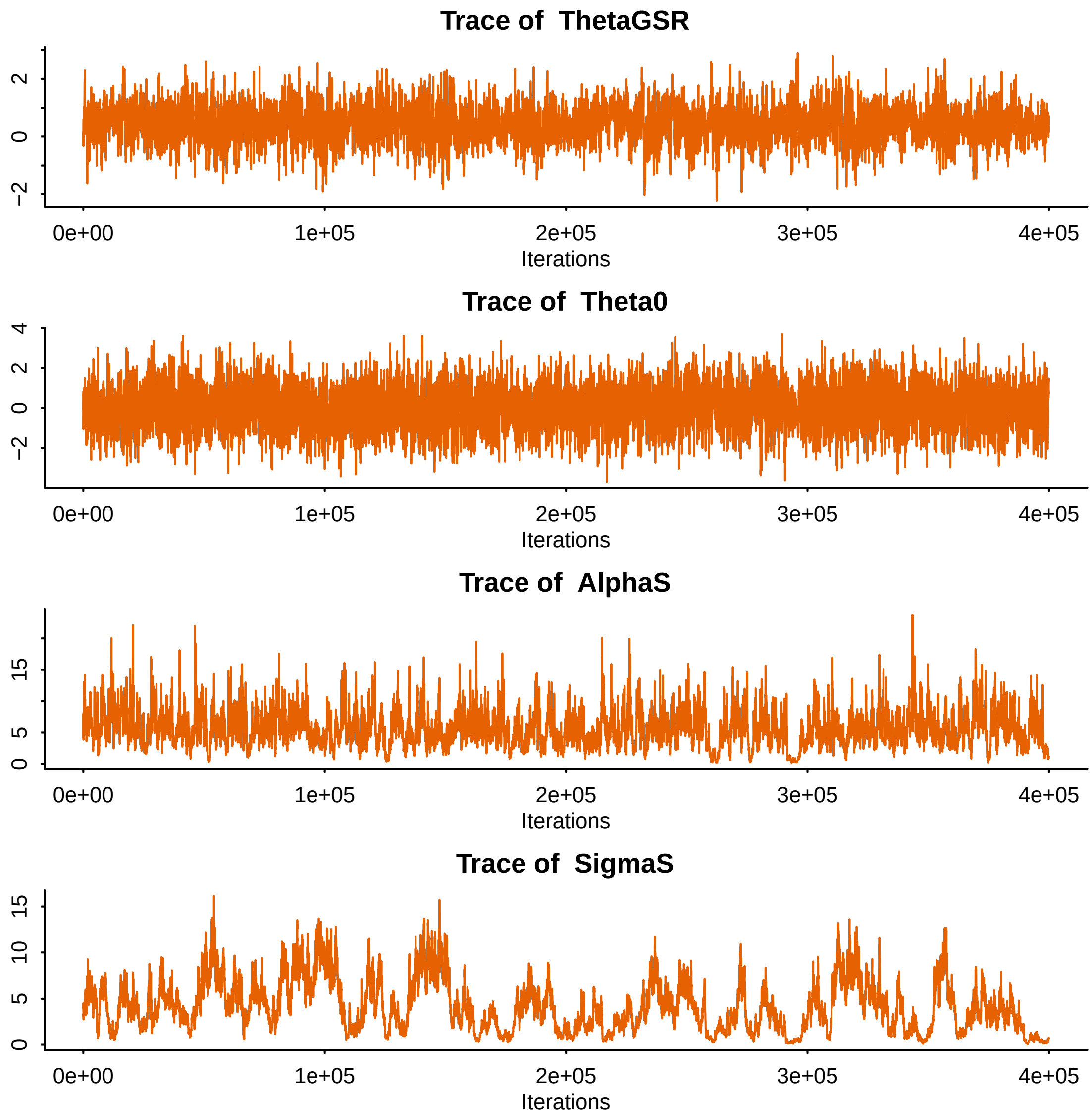


Fig. 10 Figure 10 displays the traceplot for the OU(2)FGMo|Stratification model with random terminal branch tips for FGMo|Stratification. Note that α and θ are stationary, well identified, and well mixed. On the other hand, we see that σ is not as well mixed, and shows some signs of poor identification during the tail end of the simulation.



Trace of ThetaGSR

Trace of Theta0

Trace of AlphaS

Trace of SigmaS

Fig. 11 Figure 11 displays the traceplot for the OU(1)Stratification|Null model.

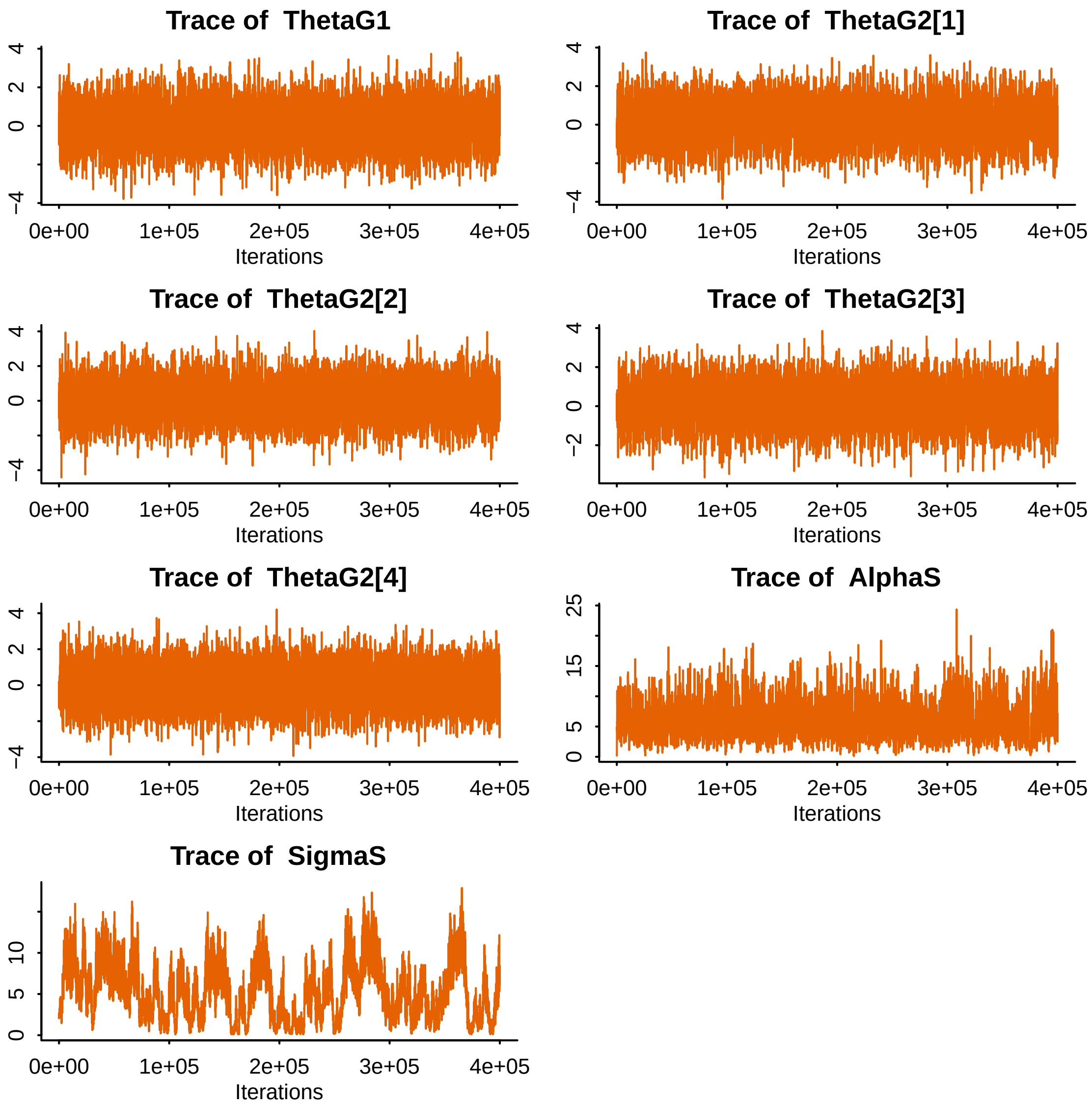


Fig. 12 Figure 12 displays the traceplot for the OU(15)Stratification|Language model with fixed terminal branch tips for FGMo|Stratification.

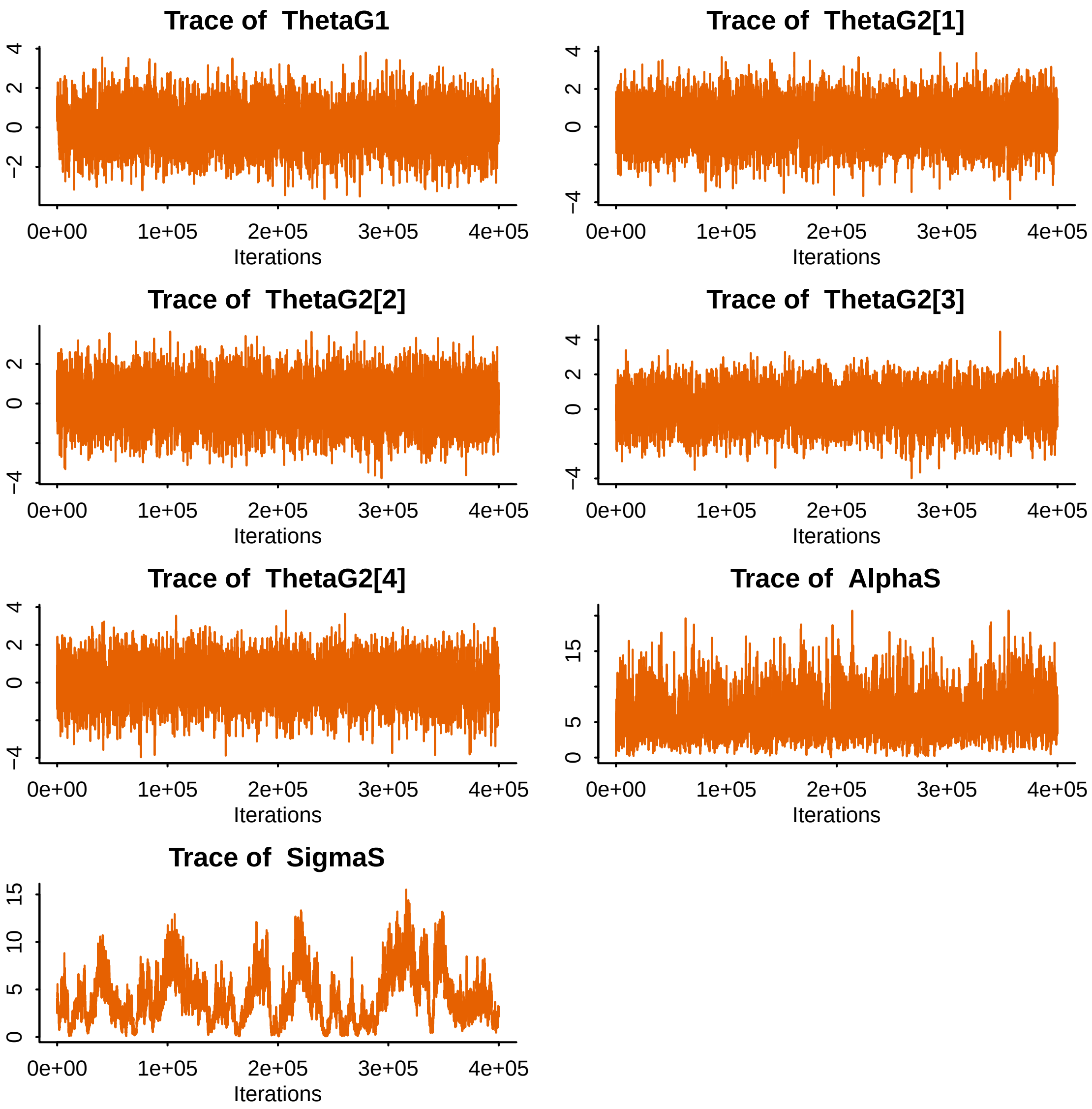
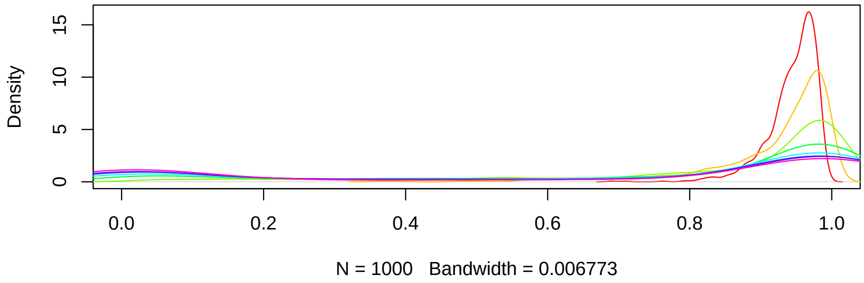
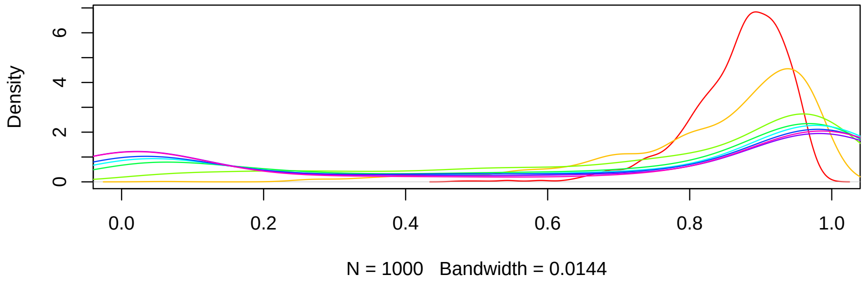


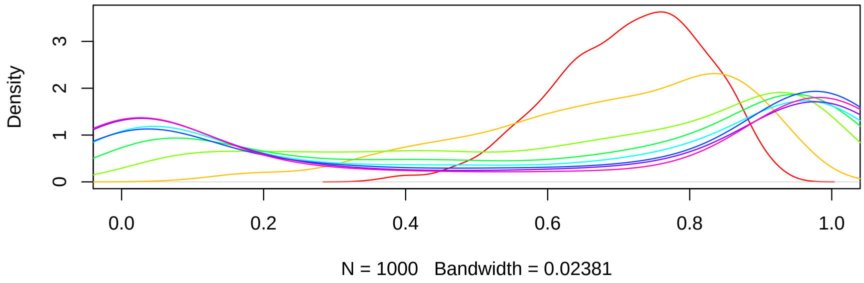
Fig. 13 Figure 13 displays the traceplot for the OU(15)Stratification|Language model with random terminal branch tips for FGMo|Stratification.



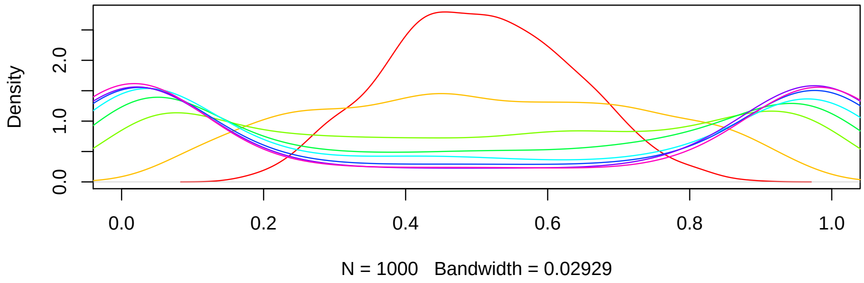
(a) $\mu = 3$



(b) $\mu = 2$



(c) $\mu = 1$



(d) $\mu = 0$

Fig. 14 Frames (a) through (d) illustrate how probability density on logistic($\Psi_{[n]}$) changes as a function of the value of σ, using simulated values for μ and V. In each subfigure, the mean value for $\Psi_{[n]}$ is held constant and σ increases from $\{1, 2, 4, 8, 10, 12, 14\}$, as red changes to magenta. Once σ has grown past approximately 8 on the linear scale, further growth fails to numerically alter probability density over logistic($\Psi_{[n]}$. These plots illustrate that while σ is not numerically identified, our overall model is well identified. We qualitatively see that σ is estimated to be large relative to α.

List of Tables

Table 1 Parameter estimates of the OU(1) and OU(15) submodels for the evolution of stratification conditional on language group. We note increased likelihood of stratification in the Semetic, Berber, and Mande language groups, and decreased likelihood of stratification in the Sudanic language groups. In other language groups, however, there appears to be no strong association between language group and stratification. We note that the parameters α and σ are of similar magnitude and fairly large, reflecting the fact that there exist signals of both selection and drift in the evolution of stratification conditional on language group.

Model	Parameter	Language	Mean	Median	2.50%	97.50%
OU(1)	θ_{GSR}	Africa	0.4352	0.4342	-0.777	1.5983
OU(1)	θ_{anc}	Ancestral	0.0284	0.0244	-1.9426	1.9886
OU(1)	α_S	-	5.6144	5.266	0.9623	12.0886
OU(1)	σ_S	-	4.1018	3.6508	0.3677	10.3519
OU(15)	θ_{anc}	Ancestral	0.034	0.037	-1.92	1.983
OU(15)	$\theta_{Fam[1]}$	AfroAsiatic	0.149	0.146	-1.754	2.042
OU(15)	$\theta_{Fam[2]}$	NigerCongo	0.049	0.051	-1.922	1.983
OU(15)	$\theta_{Fam[3]}$	NiloSaharan	0.086	0.096	-1.872	1.983
OU(15)	$\theta_{Fam[4]}$	Khosian	-0.016	-0.008	-1.976	1.913
OU(15)	$\theta_{SubFam[1]}$	Berber	0.335	0.336	-1.555	2.196
OU(15)	$\theta_{SubFam[2]}$	Chadic	-0.032	-0.041	-1.975	1.9
OU(15)	$\theta_{SubFam[3]}$	Cushitic	-0.085	-0.086	-2.01	1.854
OU(15)	$\theta_{SubFam[4]}$	Omotic	0.014	0.012	-1.94	2
OU(15)	$\theta_{SubFam[5]}$	Semitic	0.438	0.449	-1.493	2.343
OU(15)	$\theta_{SubFam[6]}$	Khosian	-0.001	-0.009	-1.952	1.95
OU(15)	$\theta_{SubFam[7]}$	AtlanticCongo	0.094	0.115	-1.513	1.594
OU(15)	$\theta_{SubFam[8]}$	Mande	0.302	0.302	-1.564	2.143
OU(15)	$\theta_{SubFam[9]}$	CentralSudanic	-0.294	-0.294	-2.176	1.618
OU(15)	$\theta_{SubFam[10]}$	EasternSudanic	-0.222	-0.22	-2.069	1.65
OU(15)	$\theta_{SubFam[11]}$	Fur	0.136	0.133	-1.802	2.087
OU(15)	α_S	-	5.655	5.262	1.338	11.99
OU(15)	σ_S	-	5.116	4.851	0.318	11.77

Table 2 Results of formal model comparison using WAIC. WAIC, p_E, and lppd are defined in the main text. The symbol dWAIC indicates the difference in WAIC between the best model and the subsequent models, and the symbol wWAIC indicates the weight in probability that the specified model will make the best predictions on new data, conditional on the set of models being considered. We see that WAIC prefers an adaptive model of stratification with distinct optima based on language family, to an OU(1) model unconditional on language family. We give further discussion about the relevance of fixed versus random branch tips later in this supplement.

Model	$\mathbf{p_E}$	**lppd**	**WAIC**	**dWAIC**	**wWAIC**
OU(15)-FixedBranchTips	23.657	-49.742	146.798	0	0.786
OU(1)-Null	19.988	-55.331	150.639	3.841	0.115
OU(15)-RandomBranchTips	20.376	-55.103	150.958	4.16	0.098

Table 3 Results of formal model comparison of the FGMo|Stratification model using WAIC. We see that WAIC prefers an adaptive model of FGMo based on stratification (with random branch tips for stratification), to an OU(1) model. We also note that the random branch tips model outperforms the the fixed branch tips model for predicting FGMo|Stratification, because the sensitivity of predictions to θ_{nssr} and θ_{ssr} in cultures with deep terminal divergences is reduced. Although the performance of the OU(2) model with random branch tips relative to the OU(1) appears to suggest that stratification plays an important role in the adaptive evolution of FGMo, inspection of the model's parameter estimates shows only very small effects, with σ still dominating the evolutionary dynamics (See *Supplementary Table "StanSummaryOfResults.xlsx"*).

Model	$\mathbf{p_E}$	**lppd**	**WAIC**	**dWAIC**	**wWAIC**
OU(2)-RandomBranchTips	27.486	-26.748	108.468	0	0.892
OU(1)-Null	28.099	-28.41	113.017	4.55	0.092
OU(2)-FixedBranchTips	27.407	-30.822	116.458	7.991	0.016

Table 4 Key parameter estimates from our revised model, showing relative support for α (selection based on stratification) and σ (drift and/or orthogonal selective forces) in explaining the distribution of FGMo in Africa. The mean and median are point estimates of the posterior distribution; the 2.5 and 97.5 percent equal tail posterior confidence intervals (PCI) present the dispersion of the posterior distribution. The top block of parameter estimates are from the OU(2) model, and the bottom block of estimates are from the OU(1) model. The symbols prefixed by θ indicate the estimated optimal trait value (log odds) under the global, stratified, or non-stratified selection regime, as indicated by the subscripts. We note that σ dominates α in both the OU(1) and OU(2) models, which is indicative that the distribution of FGMo in Africa is better explained by drift and/or selective forces operating orthogonally to stratification, than by stratification itself. In the the OU(2) model however, we find that stratification relative to non-stratification is weakly, but positively, associated with elevated social pressure for FGMo.

	Model	**Mean**	**Median**	**2.5% PCI**	**97.5% PCI**
θ_{ssr}	OU(2)	-0.245	-0.217	-2.22	1.523
θ_{nssr}	OU(2)	-0.601	-0.6	-2.732	1.551
α	OU(2)	1.299	1.148	0.095	3.326
σ	OU(2)	4.547	4.379	1.848	8.448
θ_{gsr}	OU(1)	-0.654	-0.693	-2.57	1.411
α	OU(1)	1.339	1.173	0.075	3.475
σ	OU(1)	5.784	5.397	2.576	10.95

Table 5 Results of formal model comparison on the revised model using WAIC. WAIC, p_E, and lppd are defined in the text. The symbol dWAIC indicates the difference in WAIC between the best model and the second model, and the symbol wWAIC indicates the weight in probability that the specified model will make the best predictions on new data, conditional on the set of models being considered. We note that the OU(1) model outperforms the OU(2) model. This result indicates that the distribution of FGMo across our sample of African cultural groups can be most parsimoniously explained by a drift model with a single global optima, as opposed to a model with separate optima for stratified and non-stratified societies. However, these wWAIC values are very close, and the application of WAIC to these models is not completely justified given the relative strength of drift; as such, we argue that both models are important to consider.

Model	$\mathbf{p_E}$	**lppd**	**WAIC**	**dWAIC**	**wWAIC**
OU(1)-Null	18.75	-19.34	76.18	0	0.91
OU(2)-FixedBranchTips	17.67	-22.71	80.76	4.59	0.09