Beyond orchids and dandelions: Testing the 5HTT "risky" allele for evidence of phenotypic capacitance and frequency dependent selection

Dalton Conley*
New York University & NBER
Department of Sociology
6 Washington Square North Room 20
New York, NY 10003

Emily Rauscher
New York University
Department of Sociology
295 Lafayette Street, 4th Floor
New York, NY 10012

Mark L. Siegal
New York University
Center for Genomics and Systems Biology and the Department of Biology
12 Waverly Place
New York, NY 10003

*To whom correspondence should be addressed: 6 Washington Square North, Room 20; New York, N.Y. 10003. conley@nyu.edu.

ACKNOWLEDGEMENTS
This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). No direct support was received from grant P01-HD31921 for this analysis. This research was funded by the National Science Foundation's Alan T. Waterman Award, SES-0540543.

ABSTRACT

The persistence of behaviorally deleterious genes in the human population poses an interesting question for population genetics: If certain alleles at these loci are deleterious, why have they survived in the population? We consider evidence for phenotypic capacitance and/or frequency dependent selection for an allele that has been putatively shown to have negative associations with human behaviors (the "short" 5-HTT promoter region allele) yet which has persisted in human and non-human primate populations. Using National Longitudinal Study of Adolescent Health data, we compare sibling and twin variation in depression by 5-HTT genotype (specified in several ways) and investigate sibship-level cross-person gene-gene interactions. In support of the "orchid / dandelion" hypothesis, we find evidence that the short allele increases variation in phenotypes in response to environmental (or genetic) differences (i.e. acts as a perturbation of a phenotypic capacitor). Further, we also find some evidence that the effects of allelic variation at this locus are moderated by the genetic environment of the sibship unit (i.e. may be susceptible to frequency dependent selection). We discuss implications of these findings for genetic models in general, specifically with respect to stable unit treatment value assumption violations (i.e. non-independence of units of analysis).

INTRODUCTION

Over the last decade, a robust research tradition claiming to show that certain human alleles can lead to deleterious behavioral phenotypes such as anti-social behavior and depression has grown (1,2). For example, much debate has been conducted as to whether having a short-promoter variant of the serotonin transporter gene-linked promoter region (5HTTLPR) leads to a greater risk of depression—conditional on stressful life events. However, replication has been inconsistent (see, e.g., 3,4). If indeed such an effect—conditional or unconditional on environmental stress—withstands the rigors of replication, it poses an interesting question from the perspective of population genetics: If a certain allele at the 5-HTT promoter region locus is deleterious, why has it survived in the population?

Furthermore, genetic screening generally assumes that specific alleles affect mean levels of an outcome. However, it is equally important to understand variation in health characteristics. Certain alleles may promote greater *variation* rather than simply *more* depressive symptoms, for example.

In this study, we consider evidence for two potential explanations for the survival of the short 5-HTT allele, which has persisted in human and non-human primate populations despite negative associations with human behaviors – including depression. Below we provide some theoretical and empirical background, followed by details about our methods, data, results, and a discussion including some implications of these findings for genetic models, particularly with respect to stable unit treatment value assumption violations (i.e. non-independence of units of analysis).

THEORETICAL AND EMPIRICAL BACKGROUND

Evolution and Natural Selection

Variation in alleles in the human population can emerge by chance through mutation. Most mutations are deleterious (or lethal) or have no obvious effect on fitness. However, on the occasions when a new allele emerges that confers a fitness advantage, its frequency in the population is likely to increase thanks to the relative advantage of the allele for the survival and reproduction of the organism. Therefore, like other populations, we expect alleles associated with deleterious outcomes to slowly disappear from the human population to the extent that they impact reproductive fitness. One might reasonably speculate—but by no means assume—that an allele that conveys greater susceptibility to depression may also deleteriously influence reproductive fitness. (However, it could not matter for fitness or even have a positive effect through antagonist pleiotropy.)

Potential Mechanisms of Survival

Possible explanations for the survival of an apparently deleterious allele include that it may represent a relatively recent evolutionary change that has not yet been selected out; the context that makes the allele deleterious may have recently emerged; the persistence of this allele could reflect linkage disequilibrium with a beneficial allele at a nearby locus; heterozygous advantage; pleiotropy; negative frequency dependent selection; or phenotypic capacitance. We focus on two of these potential explanations – phenotypic capacitance and frequency dependent selection – but also address the argument that the allele could be a recent evolutionary change. More details about the remaining alternative explanations are provided in the supplemental section.

Evidence suggests that the risky allele mentioned above is not relatively recent in our evolutionary history. In fact, the repeat-containing, variation-prone 5HTTLPR dates back to the

ancestor of humans and simian primates, approximately 40 million years ago (6). Although the human short/long polymorphism does not appear to be shared with other hominoids (6), it does appear to have originated early in the history of *Homo sapiens*, in that short and long alleles occur at high frequency in human populations representing all major geographic areas (7).

Given the exclusion of novelty as an explanation, frequency dependent selection could be at play, namely that the effect of the alleles depends on their distribution in the population. A classic example of this is coloration in prey; by maintaining polymorphic coloration (apostatic selection), prey are able to collectively increase the search cost for predators (8). Another example comes from human immunology and the variation found in the major histocompatibility complex (MHC) involved in the recognition of foreign antigens (9). Along these lines, perhaps being prone to depression enhances fitness when there are lots of non-depressives around who lavish attention and resources on the individual with the "demanding" or sensitive allele.

Finally, it could be that the gene acts as a phenotypic capacitor. In the biological literature, the concept of phenotypic capacitance has been fairly well established over the last decade. A capacitor of phenotypic variation (or phenotypic capacitor) is a gene that suppresses phenotypic sensitivity when fully functional; when such a gene is knocked out or otherwise compromised, formerly neutral ("cryptic") genetic variation has an effect on the phenotype, yielding greater phenotypic variation among individuals. The classic example of a phenotypic capacitor is the chaperone Heat Shock Protein 90 (10,11); however, capacitance is expected to be a general feature of transcriptional and post-transcriptional regulatory networks (12, 13) and perhaps of neuronal networks as well (14).

Capacitance is essentially what a recent spate of literature has asserted in categorizing the polymorphisms studied at behaviorally implicated loci into "orchid" and "dandelion" alleles (see,

e.g., 15). Namely, the argument is that those with the risky alleles are actually more sensitive to context, more emotionally reactive. The result of this heightened sensitivity is that when nurturing conditions are ideal, the orchid alleles (such as the short allele at 5-HTTLPR) lead to more adaptive phenotypes (and to better outcomes as conventionally measured) (16,17). However, when environmental conditions are less than ideal (such as was the case in Caspi et al.'s "childhood maltreatment") the orchid allele leads to worse outcomes (18). Meanwhile, the long 5-HTT promoter, for example, is less sensitive to context (hence the moniker "dandelion" allele since these flowers seem to be able to grow in many ecosystems but are reputedly less striking than the more fragile orchids).

Belsky et al. review previous research and argue that the short 5-HTT allele increases sensitivity, not simply vulnerability, recalling Suomi's point that the only "weed species" of primates that live in diverse ecological environments – humans and rhesus macaques – are genetically distinctive because of the presence of the short 5-HTTLPR allele: "It seems unlikely that that which might afford these two species such an adaptive advantage would only be 'vulnerability genes' that predispose carriers to depression in the face of contextual stress" (19). However, there is conceptual ambiguity in this argument, in that environmental reactivity is equated with adaptive plasticity. That is, Suomi implicitly assumes that the environmental sensitivity of orchids reflects an ability to tune behavior to the demands of a particular environment, rather than a potentially maladaptive lack of behavioral robustness.

Regardless of whether or not orchid individuals make for weed species, the molecular mechanism underlying a pattern of greater/reduced environmental sensitivity based on polymorphic variation at a given locus would likely be one of phenotypic capacitance. That is, the long allele—resulting in higher expression levels of 5-HTT—would act as a capacitor (just as HSP90 does) by

muting the effects of variation in other genes (or of variation in the environment). Unlike HSP90, which primarily acts at the level of protein folding, high 5-HTT expression might suppress variation by working at the level of synaptic plasticity or other aspects of neural function (14). The result would be lower phenotypic variation among those with the long, high-expressing 5-HTT allele.

A Case Study: 5-HTT and Depression

In the present study we will test for the possibility that the so-called risky allele of 5-HTTLPR acts as a phenotypic capacitor and/or shows evidence of negative frequency dependent selection. First, we test whether the putative "orchid" allele appears to unleash greater phenotypic variation under conditions of genetic similarity and genetic variation. The typical approach to testing the orchid / dandelion hypothesis has been to interact genotype by some measure of environment such as parenting style or socio-economic status (see, e.g., 15). This approach is problematic on a number of fronts. First, due to the non-random distribution of alleles in the population (population stratification) it could be the case that it is not the genetic locus that mediates the degree of variance in outcomes observed but rather the environmental conditions with which it is associated. That is, a particular allele could be acting as proxy for ethnic background, region, religion or any number of other factors. Second, the alleles could be acting as proxy for unmeasured genetic differences, suggesting a capacitance effect but not necessarily at that locus. This could occur due to population stratification (as discussed above) or due to linkage disequilibrium, whereby the "true" capacitor is in linkage with the observed marker. Lastly, as a result of these issues, the typical approach fails to distinguish between phenotypic capacitors that suppress genetic variation and those that suppress environmental differences, which are more precisely termed "phenotypic stabilizers" (for a discussion of this distinction, see 20).

To deal with these issues, we take a novel approach: Namely, we compare identical twin sets that share the orchid alleles (the short promoter) with their counterpart twin sets that share the dandelion alleles (the long promoter). Under the orchid / dandelion hypothesis, it should be the case that the twin sets that have orchid alleles demonstrate greater differences in their measured phenotypic outcomes due to unmeasured environmental variation.

Due to population stratification, however, the orchid allele could be acting as a proxy for greater (or lesser) environmental variation. That is, results from the strategy above (comparing twin pairs from different families) could be driven by environmental rather than genetic differences. In other words, school, family income, or family closeness could hypothetically drive results. To address concern about population stratification, we also pursue a second strategy: We interact orchid alleles with birth weight differences between monozygotic twins. This approach takes advantage of birth weight differences between identical twin siblings as a random environmental (in utero) influence that is measured and exogenous to population stratification since birth weight differences themselves do not vary by this genotype (22 also see Table 1, here). If the effect of this measured difference in prenatal environment (and low birth weight itself has been associated with higher reactivity [see, e.g., 21, for a review]) also appears to be greater in the orchid sibships, then this would further bolster the argument that the gene is acting as phenotypic stabilizer.

Of course, in identical twin sets, all other genetic loci are held constant. Hence, we also examine these same relationships between unmeasured and measured (via birth weight) environmental differences in same-sex dizygotic twin sets (as well as same-sex, non-twin sibling sets). This introduces the further complication of within-pair genetic differences (in addition to environmental ones that MZ twins experience) and also allows for the possibility that the so-called orchid alleles are acting not just as stabilizers of environmental difference but also as phenotypic

capacitors of genetic variation. The importance of such capacitance to evolution has been suggested as far back as Waddington's (23) classic work on the revealing of cryptic variation as key to decanalization (a process by which a phenotype becomes less robust and genetic variation yields greater phenotypic variation), and it has attracted recent attention as a potential explanation of modern multifactorial illnesses, including psychiatric ones (24).

Second, since the family unit is the key institution in allocating attention and resources to children and adolescents, we look for a sibship-level gene-gene interaction as indicative of frequency dependent selection. Namely, we ask if the phenotype of an individual child depends not just on his/her allele at the aforementioned locus but if such an effect is conditional on the genotype of his/her siblings at that same locus (of course, cross-loci, cross-individual interactions could be at work, too, but to avoid ad hoc testing, we will constrain the present analysis to cross-sibling interaction effects at the same locus). In other words, it could be adaptive to have the putatively more emotionally reactive short 5-HTTLPR alleles when one is the only offspring to be homozygous for this allele among one's brood, thereby garnering more parental attention. However, if by luck of the draw, all offspring end up with the more demanding short alleles, the outcome is poorer for all. In other words, as with the classic prisoner's dilemma game, it is advantageous to have the short allele if you are the only one, but disadvantageous if you are not. Such a dynamic – negative frequency dependent selection – would likely lead to maintenance of polymorphism under a random mating situation (random with respect to this locus, that is).

Alternatively, positive frequency dependent selection could be at work: Having the short allele is advantageous when all offspring have it, but deleterious (through, say stigmatization) when only one offspring carries it. Such an equilibrium might arise thanks to parent-offspring competition: When all offspring are emotionally demanding, it pays off, since parents may be more

likely to invest in existing children at the expense of future ones. However, when only one is demanding and difficult, that child is stigmatized and disinvested vis-à-vis other, current or future siblings. This alternative scenario would lead to an unstable equilibrium; because it is deleterious when rare but beneficial when common, the allele may be very slow to appear in a population but move to fixation quickly once it appears in a frequency over a given threshold.

HYPOTHESES

To summarize our approach, we pose four hypotheses.

Orchid / Dandelion Hypotheses:

H1: Twin sets with the short 5-HTT ("orchid") alleles will demonstrate greater differences and variation in their measured depression outcomes due to unmeasured environmental variation.

H2: The effect of twin pair difference in prenatal environment (birth weight) on depression difference and variation will be greater in twins with the short 5-HTT alleles.

Frequency Dependent Selection Hypotheses:

H3: Holding twin depression constant, the relationship between individual 5-HTT genotype and depression will depend on the twin's genotype.

H4: The short 5-HTT allele will be advantageous (associated with lower depressive scores) if one's twin has no short alleles and disadvantageous otherwise (negative frequency dependent selection).

METHODS

The National Longitudinal Study of Adolescent Health (25) provides sequenced genotype data for five genes, including one putatively related to environmental sensitivity (5-HTT) with respect to depression. We focus on the third wave of panel data for sibling pairs, which surveyed respondents in 2001-2 when they were ages 18-26. Siblings of individuals identified as twins in the stratified (nationally representative) sample were added, yielding 64% of sibling pairs from the probability sample and 36% from convenience sampling. In other words, to increase the number of pairs, some siblings were added after the random sampling strategy. Sampling weights are therefore not available for genetic data. Buccal swabs were collected in wave 3 from 2612 of the 3139 eligible siblings from wave 1 (a compliance rate of 83%) for DNA sequencing at the Institute for Behavioral Genetics (26). Monozygosity was genetically confirmed, requiring complete matches on 11 "highly polymorphic, unlinked short tandem repeat (STR) markers: D1S1679, D2S1384, D3S1766, D4S1627, D6S1277, D7S1808, D8S1119, D9S301, D13S796, D15S652 and D20S481" and a sex chromosome identification marker (p.992 26).

Variation at the serotonin transporter gene locus (5-HTT) has been associated with propensity toward depression. Previous research has suggested that individuals with a short allele in the promoter region of 5-HTT have stronger depressive reactions to stressful life experiences (2) (however, c.f., 3). Consistent with the phenotypic stabilizer argument, however, the short allele at the serotonin transporter gene could increase environmental sensitivity rather than simply vulnerability. Rather than focus only on the "risky" allele, we specify 5-HTT genotype in several ways. The number of short alleles per individual and twin pair is measured to identify whether twin difference in depression is sensitive to each additional "orchid" allele. We also tested models where twins homozygous for the short allele, homozygous for the long allele, and heterozygotes (or one copy of each) are specified separately and compared to the other groups. Twinning itself, we should

note, is not affected by genotype at this locus. Namely, the overall prevalence of short (or long) alleles is not significantly different between dizygotic and monozygotic twins (see Table 1). There are some slight differences between singletons and twins, but they only reach marginal significance ($p<.10$).

Add Health data do not have information on parental genotype at this or any locus. Therefore we cannot rule out the possibility of population stratification (i.e. that the sibling's genotype is acting as proxy for the parental genotype and/or environmental conditions due to population stratification). With parental genotype controlled, we could be sure that the differences in sibling genotype are exogenous as a result of Mendelian randomization and segregation. However, to reduce concern about population stratification, we limit the sample to whites. Our sample of white twins includes over 100 identical twin pairs, nearly 130 fraternal pairs, and 900 full sibling pairs (although the exact sample size for each model depends on the number of pairs with complete genetic and depression data for both twins).

We sacrifice some power by identifying off within-twin pair differences. Thus, effects would have to be (almost) twice as large as models that use all individuals with genetic data in Add Health (these approaches also suffer from inflation of standard errors due to the non-independence of observations). So, while post-hoc power tests are discouraged in the literature (27, 28), the fact that previous studies report powerful impacts of these genes (and interaction effects) means that we should still be able to detect them with our reduced power, even if previous estimates are partially spurious. That said, acting as if we were designing an experiment to detect an effect size that is $\beta >= 0.15$, at $\alpha<.05$, with three predictors (other than the fixed effects), our minimum required sample size for a study with 0.80 power would be 76. All models meet this minimum. Meanwhile, our targeted, minimally detectable effect size of 0.15 is well within early reports for population studies of these genes on these

phenotypes. For example, Caspi et al. (1:853) report an MAOA-maltreatment interaction effect size of -0.36. Others report interaction effects ranging from -0.11 to -0.89 (2:388) and from -0.58 to -0.72 (29:599), nearly all larger than our targeted effect size.

Birth weight is reported by parents, measured in ounces, and logged. The average birth weight difference between twins is 8 ounces (half a pound), which is non-trivial and large enough to generate significant effects. Although this is retrospective data, when children are teens, parents typically remember birth weight well (e.g., one study reports an 85% accurate recall rate when children are teenagers) (30). Birth weight is reported in wave 1, 6 or 7 years prior to the outcome measure in wave 3, which reduces the possibility that birth weight is selectively remembered by parents depending on the depressive tendencies of their teenage children. Furthermore, regressions of birth weight differences on depression differences (or birth weight on depression at the individual level) show no significant relationships and correlation never exceeds 0.08. Depression is measured using nine items of the Center for Epidemiologic Studies-Depression Scale (CES-D). CES-D normally includes more items that were omitted from Wave 3. Therefore we also include the other six questions in Wave 3 about the frequency of depressive symptoms. The sum of responses for all items (listed in the supplemental section) indicates the frequency of depressive symptoms.

To test hypotheses 1 and 2, we first compare depression variance by genotype among twins and singleton siblings. Second, we regress twin pair depression difference and coefficient of variation (within-pair standard deviation over sibship mean) on genotype, specified several ways. To test hypotheses 3 and 4, we regress individual depression score on individual genotype, sibling genotype, and the interaction between the two – with and without controlling for sibling phenotype (depression). Regressions control for sex differences. In twin pair models, indicators for all-male pairs and opposite sex pairs are included, omitting all-female pairs as the suppressed category.

Before presenting results, we address a potential concern: Measures of variation in complex traits often show non-trivial dependencies on trait means (see, e.g., 31). That is, variation may be artifactually greater for twins with higher average depression scores, which could skew our results. We therefore examined the relationship between twin pair depression standard deviation and depression average by zygosity and 5-HTT genotype. Lowess curves show the locally weighted average for all points in each graph (Figure 1). That these curves are relatively linear suggests that the coefficient of variation is an appropriate measure for analyzing effects of genotype on depression variation. Although pairs are similarly distributed by genotype among MZ twins, DZ pairs with all long alleles tend to fall slightly below the Lowess curve, while DZ pairs with all short alleles more often fall above the curve, an observation that is consistent with the phenotypic capacitance hypothesis.

RESULTS

Since our primary interest is twin difference or variation, Table 1 provides means for the sibling pair average, standard deviation, and coefficient of variation by genotype for pairs in each sibling type. Average depression coefficient of variation is consistently highest among pairs who are both homozygous for the risky genotype. The only exception is singletons, where pairs who differ from each other at 5-HTT have the highest coefficient of variation. Consistent with the orchid / dandelion hypothesis, pairs homozygous for the short 5-HTT allele generally (excepting singletons) have the highest coefficient of variation in depression, followed by individuals with one and then two long alleles. The dandelions (with the long alleles) show the lowest sensitivity to environment – the only source of variation among identical twins. Simple differences in coefficient of variation are informative, but could reflect sex or other differences.

Orchid / Dandelion Hypotheses

Predicting twin difference and within-pair coefficient of variation among whites in Table 2, twin pair genotype shows a statistically significant relationship to depression variation. In support of H1, net of any differences due to sex, twins with two dandelion alleles show significantly lower depression differences. This difference is marginally significant ($p<.10$) among MZ and DZ twins. An indicator for twin pairs homozygous for the short allele yields no significant main effects.

Results are also consistent with the phenotypic capacitor explanation when predicting twin pair coefficient of variation (see Table 2). Pairs with two copies of the long serotonin transporter have lower variation, but this difference does not reach significance. Further supporting H1, pairs with two copies of the short, "orchid" allele have higher variation, but this difference is only marginally significant among DZ and all twins. There is stronger evidence of a linear relationship between number of short alleles and pair depressive variation. For each additional short allele, twin pairs have higher variation in depressive symptoms and this relationship is significant among DZ and all twins. Figure 2 depicts this significant relationship among all white twins (231 pairs). A twin pair's predicted coefficient of variation increases with each additional short "orchid" allele. Although the relationship is not significant among MZ twins, twins with additional "orchid" alleles consistently show greater variation in depression which offers some evidence for phenotypic capacitance.

Interacted with birth weight difference, regressions predicting twin depression *difference* offer little evidence to support H2. Only heterozygotes show any conditional effect on depression difference. With increasing birth weight differences, fraternal twins show significantly higher depression differences, but only if they each have one long and one short 5-HTT allele. The same

relationship holds when including all twins, but is insignificant among identical twins. Among identical twins, however, heterozygotes show marginally higher depression differences than others net of sex differences. Results thus contradict the heterozygote advantage argument, which would expect positive outcomes for both twins regardless of environment (e.g., birth weight).

When predicting depressive *variation*, identical twins show evidence supporting H2 in interactions between genotype and birth weight differences. For example, birth weight differences among MZ pairs with two long, "dandelion" alleles yield significantly lower variation differences than those with short alleles. For each additional short allele, MZ pairs show greater depressive variation with birth weight differences, although this only reaches marginal significance.

Frequency Dependent Selection Hypotheses

Frequency dependent selection predicts that genetic effects will depend on the genes carried by those around you. Controlling for sibling's phenotype, there is some evidence of this for depression. Table 3 lists coefficients for the frequency dependent selection analysis among white DZ twins. In the top model, controlling only for sibling phenotype, short 5-HTT alleles have no main effect on depression – whether indicating those with two short alleles or specifying the number of short alleles. The next model allows individual genotype effects to vary by sibling phenotype. Although depression is higher among those with two short alleles and increases with sibling depression, twins with two short alleles actually show a slight (but significant) decrease in depressive symptoms as their twin gets more depressed. In other words, genotype matters, but only conditional on sibling behavior. The detrimental effect of twin behavior is buffered among those with short alleles, which contradicts the vulnerability and sensitivity arguments. However, this buffering relationship does not account for sibling's genotype.

Offering evidence of frequency dependent selection (and supporting H3), Figure 3 and the final rows in Table 3 show that individual genotype effects depend on sibling genotype. Looking at the bottom rows in Table 3 that allow the effect of individual genotype to vary by sibling genotype, we find that fraternal twins with no short alleles show minimal change in depression by twin's genotype. However, a twin with two short alleles has high depression symptoms if their twin has no short alleles and lower depression scores otherwise. Having an "orchid" twin yields low depression scores regardless of individual genotype. However, having a "dandelion" twin makes individual genotype important. In other words, contradicting H4, an emotionally demanding or sensitive genotype can be detrimental unless those around you also have that same genotype, thereby presenting a unified front to parents. More broadly, the presence of such genotypes appears to attenuate individual genetic differences.

We originally hypothesized (H4) that from a parental investment point of view, it would be advantageous to have the putatively more emotionally reactive short 5-HTT alleles when one is the only offspring to be homozygous for this allele among one's brood, thereby garnering more parental attention. By contrast, if by luck of the draw, all offspring end up with the more demanding short alleles, the outcome would be poorer for all in this scenario – supporting *negative* frequency dependent selection (H4). In other words, as with the classic prisoner's dilemma game, we predicted it would be advantageous to have the short allele if you are the only one, but disadvantageous if you are not. Instead we found evidence of *positive* frequency dependent selection: when one has the reactive allele(s), it is more disadvantageous when one is the only one in the brood with this genotype. In other words, those who are alone among offspring with higher emotional reactivity may be stigmatized and isolated, which in turn, could lead to higher rates of depression symptomology.

Returning to the question of why these alleles persist in the human population, this preliminary evidence suggests the "orchid" alleles may only matter if others do not share them. In a social setting, the presence of others with orchid alleles appears to dampen their effect on individuals holding them. We note that this evidence is preliminary, since these data do not have parental genotype, and therefore we cannot rule out the possibility of population stratification (i.e. that the sibling's genotype is acting as proxy for the parental genotype and/or environmental conditions due to population stratification). With parental genotype controlled, we would be sure that the differences in sibling genotype are exogenous as a result of Mendelian randomization.

DISCUSSION

While we framed this paper as an investigation into the mystery of why a seemingly disadvantageous allele persists in the population, this was really a promise that the present analysis could not keep. We were limited by several facts. First, the phenotype we are examining is not one that is evolutionarily relevant—i.e. fitness—but rather a psychological state that may or may not be related in our hypothesized direction to reproductive success. Second, we are necessarily looking at a single generation whereas the arguments about fitness need to by dynastically tested. And finally, we cannot test alternative hypotheses such as linkage disequilibrium (i.e. hitchhiking), historical GE interaction (i.e. allele was adaptive in the ancestral environment but is no longer), and antagonistic pleiotropy (i.e. allele has positive effects on fitness despite negative effects on affect).

Despite these limitations, we believe the present study has made an important contribution by showing how to model genetic effects on variation (as opposed to mean effects), how to parse out whether these effects work via interaction with the external environment or internal cryptic genetic variation, and by testing cross-organism meta-genomic effects of allele combinations at a given

locus. So while limited by the lack of a replication dataset and a somewhat subjective behavioral phenotype measured with lots of error, we have shown a proof of concept in the methodological approach we take, hopefully opening up new pathways for biodemographic researchers who are examining effects of candidate genes (or genome wide genotypes) on social or behavioral outcomes (or physical outcomes for that matter).

Overall, results support hypotheses 1-3 and contradict hypothesis 4. In support of the "orchid / dandelion" or phenotypic capacitance hypothesis, we find some evidence that the short 5-HTT allele increases variation in phenotypes in response to environmental (or genetic) differences. Further, our results are also consistent with hypothesis 3 that the effects of allelic variation at this locus are moderated by the genetic environment of the sibship unit (i.e. would be susceptible to frequency dependent selection). However, contrary to hypothesis 4, results are consistent with positive rather than negative frequency dependent selection. It could be the case that the sibling-conditional phenotypic effects of genetic variation that we demonstrate are historically specific to the family arrangements and human capital investment patterns of 20th century U.S. society. If this were the case, the interaction between twin alleles to produce the phenotype would not necessarily imply negative frequency dependent selection over the evolutionary history of the 5-HTT promoter. So our results would be able to suggest balancing selection as a possibility but not prove it. However, given we find no support for our fourth hypothesis (and, in fact, significant results working in the opposite direction), this concern is moot. We do hope, however, that merely by estimating the model, we will lead other researchers to follow us to consider within-family, cross-individual allelic dynamics. We believe that cross-sibling allelic effects have other important implications in their own right.

Namely, the non-independence of sibling genotypes from each other is non-trivial for other models of molecular genetic effects, G x E estimation and heritability models. Specifically, if the behavioral phenotype of an individual is not just contingent on her/his own genotype but that of her/his siblings, then it suggests non-independence of the units of analysis (i.e. violation of the Stable Unit Treatment Value Assumption or SUTVA). This has implications for models that include molecular markers as covariates and for those that use variance decomposition methods to generate estimates, such as is done in classic heritability analysis. This non-independence could partly explain why studies trying to replicate evidence of candidate gene effects often fail. In fact, it may help explain why classic additive heritability estimates cannot be replicated (or even approached) by GWAS studies that regress phenotypes against all known polymorphic loci. That is, depending on how SUTVA is violated, it could result in attenuation bias in genome-wide marker regressions and/or overestimation of MZ-DZ differences in intra-class correlations for given phenotypes, which would overestimate heritability. This is a possibility that should be explored by future researchers with genome-wide data.

Finally, such effects have potential relevance to the debate over group selection and "extended phenotypes" to use the language of Dawkins (32). Namely, frequency dependent selection is a relatively uncontroversial way that group membership and dynamics result in certain alleles being favored. In our case of positive frequency dependent selection, such a dynamic could lead to group specialization where one population goes to fixation for a certain allele thanks to the emergent advantage of its presence in multiple members of the group while another population goes to elimination of the same allele holding all else constant. Among mobile organisms like humans, such a dynamic could lead to selection into groups by individuals, and provide a selective advantage based on group membership. This is much more plausible than other group selection arguments that

suffer from a mismatch between the principal (the group) and the agent (the individual) in driving allelic fitness.

REFERENCES

1. Caspi, A. et al. 2002. "Role of Genotype in the Cycle of Violence in Maltreated Children." *Science* 297: 851-854.
2. Caspi, A. et al. 2003. "Influence of Life Stress on Depression: Moderation by a Polymorphism in the 5-HTT Gene." *Science* 297: 851-854.
3. Risch, N.; R. Herrell; T. Lehner; et al. 2009. A Meta-analysis: Interaction Between the Serotonin Transporter Gene (5-HTTLPR), Stressful Life Events, and Risk of Depression. *JAMA*. 301:2462-2471.
4. Caspi, A. et al. 2010. Genetic Sensitivity to the Environment: The Case of the Serotonin Transporter Gene and Its Implications for Studying Complex Diseases and Traits. *American Journal of Psychiatry* 167: 509-527.
5. Penke, L., J.J.A. Denissen, G.F. Miller. 2007. The Evolutionary Genetics of Personality. *European Journal of Personality* 21:549-587.
6. Lesch, K.P., J. Meyer, K. Glatz, G. Fliigge, A. Hinney, J. Hebebrand, S. M. Klauck, A. Poustka, F. Poustka 5, D. Bengel, R. Miissner, P. Riederer, and A. Heils. 1997. The 5-HT transporter gene-linked polymorphic region (5-HTTLPR) in evolutionary perspective: alternative biallelic variation in rhesus monkeys. *Journal of Neural Transmission*. 104:1259-1266.
7. Gelernter, J., J. F. Cubells, J. R. Kidd, A. J. Pakstis, and K. K. Kidd. 1999. Population studies of polymorphisms of the serotonin transporter protein gene. *American Journal of Medical Genetics.* 88: 61-66.
8. Clarke, B. 1962. Balanced polymorphism and the diversity of sympatric species. Pp. 47-70 in D. Nichols ed. Taxonomy and Geography. Systematics Association, Oxford.
9. Borghans, J.A.M., J. B. Beltman and R. J. Boer. 2007. MHC polymorphism under host-pathogen coevolution. *Immunogenetics*.55:732-739.
10. Rutherford S.L., S. Lindquist. 1998. Hsp90 as a capacitor for morphological evolution. *Nature* 396 (6709): 336–342.
11. Queitsch, C., T.A. Sangster and S. Lindquist. 2002. Hsp90 as a capacitor of phenotypic variation. *Nature*, v.417, pp. 618–24.
12. A. Bergman and M. L. Siegal. 2003. Evolutionary capacitance as a general feature of complex gene networks. *Nature* 424:549-552.
13. Sangster T.A., Lindquist S., Queitsch C. 2004. Under cover: causes, effects and implications of Hsp90-mediated genetic capacitance. *BioEssays* 26(4):348-62.
14. Niven, J.E. 2004. Channeling Evolution: Canalization and the nervous system. *PLoS Biology*. 2(1): e19.
15. Mitchell, C., D. Notterman, J. Brooks-Gunn, J. Hobcraft, I. Garfinkel, K. Jaeger, I. Kotenko, and S. McLanahan. 2010. The Role of Mother's Genes and Environment in Postpartum Depression. Paper presented at Integrating Genetics and the Social Sciences Boulder, Colorado June 2-3, 2010.
16. Boyce, W.T. and B.J. Ellis. 2005. "Biological Sensitivity to Context: I. An evolutionary-developmental theory of the origins and functions of stress reactivity." *Development and Psychopathology* 17: 271-301.
17. Ellis, B. J. and W. T. Boyce. 2008. Biological Sensitivity to Context. *Current Directions in Psychological Science.* 17: 183-187.

18. Obradovic, J., N.R. Bush, J. Stamperdahl, Nancy E. Adler, and W. Thomas Boyce. 2010. Biological Sensitivity to Context: The Interactive Effects of Stress Reactivity and Family Adversity on Socioemotional Behavior and School Readiness. *Child Development*. 81: 270-289.
19. Belsky, J., C. Jonassaint, M. Pluess, M. Stanton, B. Brummett, and R. Williams. 2009. Vulnerability Genes or Plasticity Genes? *Molecular Psychiatry* 14: 746-754.
20. Masel, J. and M. Siegal. 2009. Robustness: mechanisms and consequences. *Trends in Genetics*. 25: 395-403.
21. Rice, F., I. Jones, and A. Thapar. 2007. The impact of gestational stress and prenatal growth on emotional problems in offspring: A review. *Acta Psychiatrica Scandanavica*. 115:171-183.
22. Conley, D. and E. Rauscher. 2010. "Genetic Interactions with Prenatal Social Environment: Effects on Academic and Behavioral Outcomes." NBER Working Paper 16026. www.nber.org/papers/w16026.
23. Waddington, C. H. 1952. "Selection of the Basis for an Acquired Character." *Nature* 169: 278.
24. Gibson, G. 2009. Decanalization and the origin of complex disease. *Nature Reviews Genetics*. 10: 134-140.
25. Harris, K.M.. 2009. The National Longitudinal Study of Adolescent Health (Add Health), Waves I & II, 1994–1996; Wave III, 2001–2002 [machine-readable data file and documentation]. Chapel Hill, NC: Carolina Population Center, University of North Carolina at Chapel Hill.
26. Harris, K.M., C.T. Halpern, A. Smolen, and B.C. Haberstick. 2006. "The National Longitudinal Study of Adolescent Health (Add Health) Twin Data." *Twin Research and Human Genetics* 9, 6: 998-997.
27. Levine, Marc and Mary H.H. Ensom. 2001 "Post Hoc Power Analysis: An Idea Whose Time Has Passed?" *Pharmacotherapy* 21:405-9.
28. Hoenig, John M. and Dennis M. Heisey. 2001. "The Abuse of Power: The Pervasive Fallacy of Power Calculations for Data Analysis." *The American Statistician* 55(1):19-24.
29. Guo, Guang, Michael E. Roettger, and Tianji Cai. 2008a. "The Integration of Genetic Propensities into Social-Control Models of Delinquency and Violence among Male Youths." *American Sociological Review* 73(4):543-68.
30. Walton, K.A., L.J. Murray, A.M. Gallagher, G.W. Cran, M.J. Savage, C. Boreham. 2000. Parental Recall of Birthweight: A Good Proxy for Recorded Birthweight? *European Journal of Epidemiology* 16, 9: 793-796.
31. Levy S.F., M.L. Siegal. 2008. Network hubs buffer environmental variation in Saccharomyces cerevisiae. *PLoS Biol*. 2008 Nov 4;6(11):e264.
32. Dawkins, Richard. 1999. The Extended Phenotype: The Long Reach of the Gene. Oxford: Oxford University Press.
33. Bocquet-Appel, J.P. 2011. When the World's Population Took Off: The Springboard of the Neolithic Demographic Transition. *Science* 333(6042):560-1.
34. Homberg, J. R. and K.-P. Lesch. In press. "Looking on the Bright Side of Serotonin Transporter Gene Variation." *Biological Psychiatry*. http://biologicalpsychiatryjournal.com/content/1000856abs.

TABLES AND FIGURES

Table 1: Twin Pair Means by 5-HTT Genotype – White Siblings

		MZ		DZ		Twins		Singletons		Full Sibs		Full Sibs (excl MZ)	
		Birth Wght	Depression	Birth Wght	Depression	Birth Wght	Depression	Birth Wght	Depression	Birth Wght	Depression	Birth Wght	Depression
HTT l-l	Mean	4.48	3.93	4.54	4.83	4.51	4.42	4.76	5.4	4.71	5.17	4.73	5.32
	Std Dev	0.16	3.55	0.21	4.69	0.19	4.21	0.19	4.96	0.21	4.81	0.2	4.92
	Range	4.16-4.77	0-15	4.16-4.89	0-27	4.16-4.89	0-27	4.16-5.20	0-28	4.16-5.20	0-28	4.16-5.20	0-28
	Variance	0.03	12.6	0.04	22	0.04	17.72	0.04	24.6	0.04	23.14	0.04	24.21
	Coeff of Var	0.04	0.90	0.05	0.97	0.04	0.95	0.04	0.92	0.04	0.93	0.04	0.92
	N	44	61	61	71	105	132	379	432	484	564	440	503
HTT l-s	Mean	4.48	5.56	4.52	5.24	4.5	5.38	4.74	5	4.68	5.1	4.71	5.04
	Std Dev	0.2	5.07	0.18	5.02	0.19	5.03	0.18	4.92	0.21	4.95	0.2	4.93
	Range	4.16-4.88	0-28	4.17-4.91	0-28	4.16-4.91	0-28	4.16-5.17	0-27	4.16-5.17	0-28	4.16-5.17	0-28
	Variance	0.04	25.7	0.03	25.2	0.04	25.3	0.03	24.21	0.04	24.5	0.04	24.3
	Coeff of Var	0.04	0.91	0.04	0.96	0.04	0.93	0.04	0.98	0.04	0.97	0.04	0.98
	N	86	108	104	136	190	244	546	667	736	911	650	803
HTT s-s	Mean	4.56	4.59	4.5	5.39	4.53	5.03	4.74	6.07	4.67	5.75	4.69	5.94
	Std Dev	0.17	4.28	0.17	4.71	0.17	4.52	0.2	6	0.21	5.6	0.21	5.77
	Range	4.22-4.88	0-15	4.19-4.84	0-17	4.19-4.88	0-17	4.20-5.23	0-31	4.19-5.23	0-31	4.19-5.23	0-31
	Variance	0.03	18.32	0.03	22.18	0.03	20.43	0.04	36	0.04	31.36	0.04	33.29
	Coeff of Var	0.04	0.93	0.04	0.87	0.04	0.90	0.04	0.99	0.04	0.97	0.04	0.97
	N	36	49	48	59	84	108	199	246	283	354	247	305
All	Mean	4.5	4.89	4.52	5.17	4.51	5.04	4.75	5.32	4.69	5.25	4.71	5.3
	Std Dev	0.18	4.55	0.19	4.85	0.19	4.72	0.19	5.15	0.21	5.04	0.21	5.1
	Range	4.16-4.88	0-28	4.16-4.91	0-28	4.16-4.91	0-28	4.16-5.23	0-31	4.16-5.23	0-31	4.16-5.23	0-31
	Variance	0.03	20.7	0.04	23.52	0.04	22.28	0.04	26.52	0.04	25.4	0.04	26.01
	Coeff of Var	0.04	0.93	0.04	0.94	0.04	0.94	0.04	0.97	0.04	0.96	0.04	0.96
	% l-l	0.27	0.28	0.29	0.27	0.28+†	0.27+†	0.34+	0.32+	0.32	0.31	0.33†	0.31†
	% s-s	0.22	0.22	0.23	0.22	0.22+	0.22+	0.18+	0.18+	0.19	0.19	0.18	0.19
	N	166	218	213	266	379	484	1124	1345	1503	1829	1337	1611

Birth weight is measured in natural log of ounces. Indicates marginal difference ($p<.10$) in prevalence of short 5-HTT allele between: DZ and Singletons (†); Twins and Singletons (+). These differences do not reach significance at $p<.05$ and other differences are insignificant.

Table 2: Twin Difference in Depression Scores – Genotype Coefficients controlling for Sex

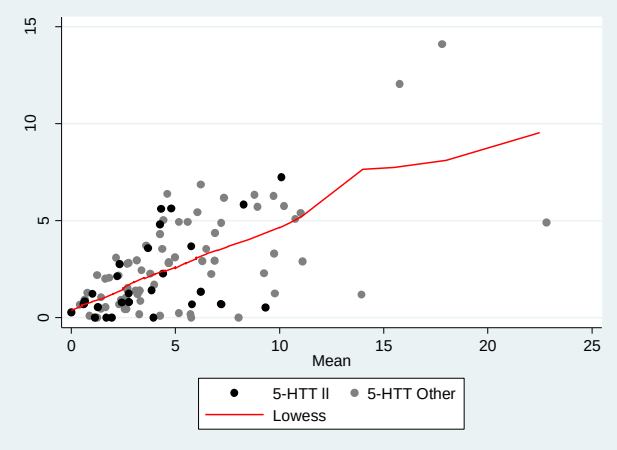
		Depression Score Difference			Depression Coeff of Var (σ/μ)			Z test difference MZ & DZ CoV coeffs
Homozygous for "Good" Alleles		MZ	DZ	Twins	MZ	DZ	Twins	
Main Effect	5-HTT ll	-1.269+	-1.833+	-1.542*	-0.022	-0.126	-0.087	0.693
		(.758)	(1.09)	(.656)	(.1)	(.112)	(.074)	
Interaction	5-HTT ll	-0.73	-1.618	-1.338	0.254	-0.041	0.07	-1.202
		(1.374)	(1.769)	(1.11)	(.168)	(.179)	(.123)	
	5-HTT ll * Birth Weight Diff	-7.25	-1.578	-2.177	-2.616*	-0.329	-1.077	1.418
		(9.432)	(11.11)	(7.085)	(1.154)	(1.127)	(.786)	
Homozygous for "Risky" Alleles								
Main Effect	5-HTT ss	-0.34	0.672	-0.056	0.125	0.247+	0.159+	0.719
		(.836)	(1.303)	(.748)	(.108)	(.131)	(.083)	
Interaction	5-HTT ss	-0.15	2.986	0.996	0.156	0.262	0.18	0.336
		(1.813)	(2.271)	(1.406)	(.222)	(.224)	(.153)	
	5-HTT ss * Birth Weight Diff	-1.193	-10.653	-4.578	0.403	0.475	0.521	0.032
		(15.12)	(13.376)	(9.739)	(1.852)	(1.32)	(1.057)	
Both Twins Heterozygotes								
Main Effect	5-HTT ls	1.273+	-1.215	-0.095	-0.069	-0.042	-0.067	0.211
		(.683)	(.888)	(.57)	(.09)	(.091)	(.064)	
Interaction	5-HTT ls	0.476	-4.054*	-1.755+	-0.349*	-0.18	-0.308*	0.744
		(1.242)	(1.638)	(.991)	(.152)	(.169)	(.108)	
	5-HTT ls * Birth Weight Diff	9.395	25.913*	16.428*	2.505*	1.202	2.018*	0.799
		(9.554)	(11.002)	(7.155)	(1.168)	(1.138)	(.782)	
Number of "Risky" Alleles								
Main Effect	5-HTT # s	0.2	0.451	0.305	0.026	0.069*	0.048*	0.921
		(.244)	(.338)	(.209)	(.032)	(.034)	(.023)	
Interaction	5-HTT # s	0.07	0.646	0.415	-0.038	0.074	0.033	1.389
		(.483)	(.552)	(.364)	(.059)	(.055)	(.04)	
	5-HTT # s * Birth Weight Diff	1.87	-2.261	-1.064	0.727+	-0.078	0.133	1.488
		(3.525)	(3.29)	(2.314)	(.43)	(.328)	(.254)	

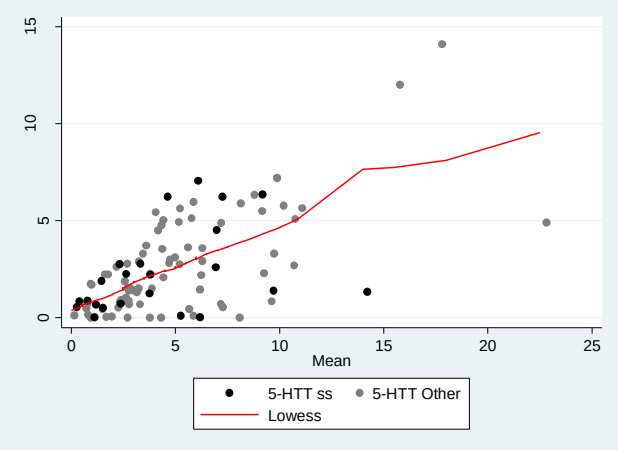
Table 3: Expression of depression phenotype is dependent on own genotype, twin phenotype and twin genotype, providing tentative evidence for possibility of frequency dependent selection. Data are for white DZ Twins – standard errors adjusted for family clustering; sex controlled in all models; Sample size ranges from 263 to 265 depending on the gene and model.

	DEPRESSION				
	Sib Pheno	Indiv	Sib	Indiv*Sib	Indiv*Sib Pheno
Sib Phenotype alone	0.11				
HTT s-s	0.12	0.41			
HTT - # s alleles	0.12	0.39			
HTT s-s	0.18 *	1.64			-0.24 *
HTT - # s alleles	0.28 *	1.24 *			-0.16 †
HTT s-s	0.12	0.70	-0.74		
HTT - # s alleles	0.13 †	0.90	-0.96		
HTT s-s	0.12	1.38	-0.12	-1.73	
HTT - # s alleles	0.12 †	1.79 *	-0.05	-0.95 †	
HTT s-s		1.41	0.12	-2.09	
HTT - # s alleles		1.84 *	0.17	-1.11 *	

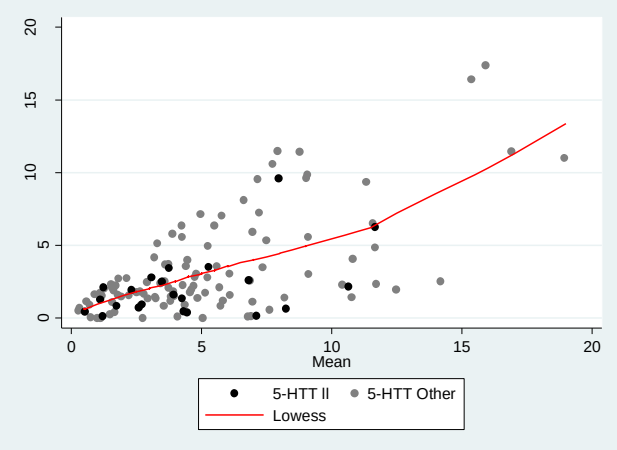
Figure 1: Twin Pair Depression Standard Deviation vs. Mean by Genotype

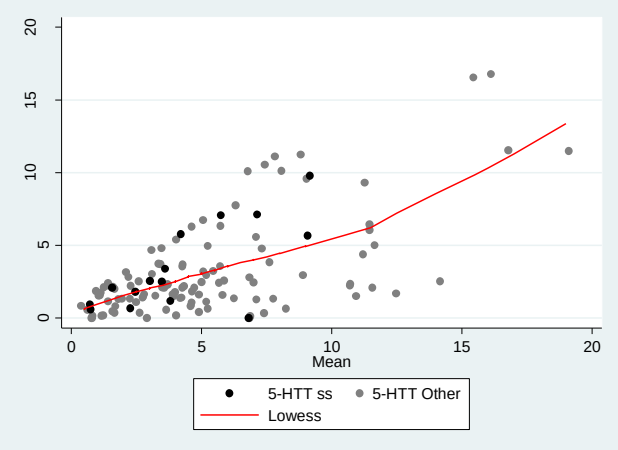
White MZ Twins



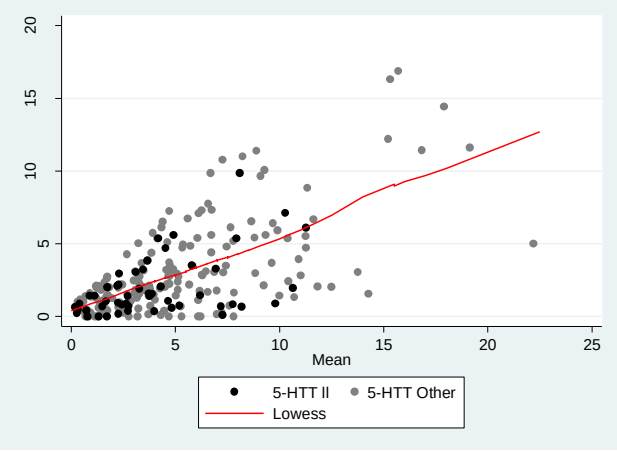


White DZ Twins





All White Twins



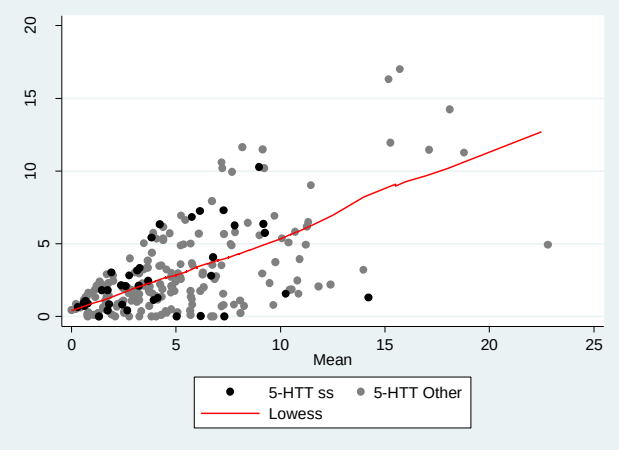


Figure 1 notes: Each figure plots the twin pair standard deviation of depression against the twin pair mean depression. The top two graphs show MZ twin pairs, followed by DZ, and all twin pairs. Plots on the left compare pairs in which both twins are homozygous for the long 5-HTT alleles to all other twin pairs. Plots on the right compare pairs homozygous for the short 5-HTT alleles to all other pairs. The line follows a Lowess curve, the locally weighted average for all points in each graph. The graphs show a linear relationship between twin pair mean and standard deviation, reducing concern that variance is dependent on the mean and suggesting that the coefficient of variation is an appropriate measure to analyze effects of genotype on depression variation.

Figure 2: 5-HTT and Phenotypic Capacitance; Depression score coefficient of variation among white twins by number of short 5-HTT alleles compared to twins with 0 short 5-HTT alleles, controlling for sex. The relationship is significant ($p<.05$). Sample size is 231 twin pairs.

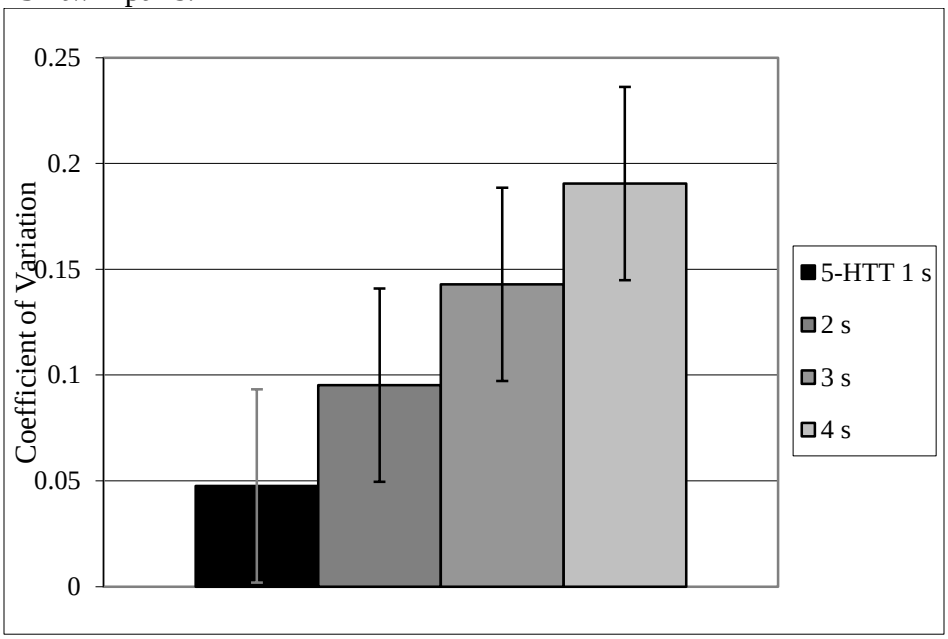
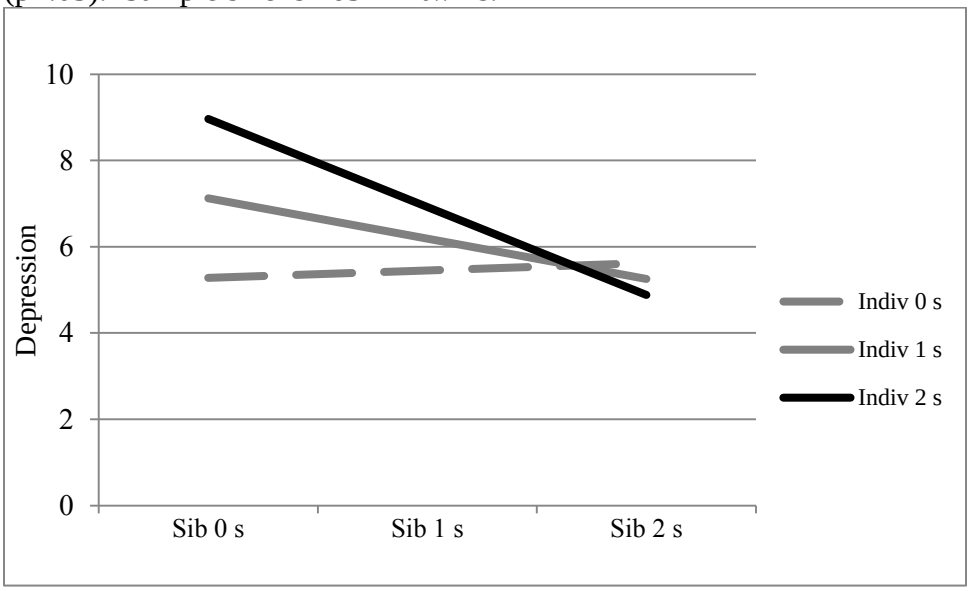


Figure 3: Preliminary Evidence for Frequency Dependent Selection: Phenotype of 5-HTTLPR Varies by Sib Genotype. Depression by number of individual and sibling short 5-HTT alleles among white DZ twins, controlling for sex. The relationship is significant ($p<.05$). Sample size is 265 DZ twins.



SUPPLEMENTARY MATERIAL

Alternative Explanations for the Survival of an Apparently Deleterious Allele

We cannot address whether a contextual change altered the importance of the 5-HTT allele but, having survived until the agricultural revolution, the resulting population increase may have allowed the allele to increase in frequency and survive to today. The agricultural revolution occurred relatively recently in evolutionary terms – approximately 12,000 years ago (33). If this or another relatively recent environmental change altered the importance of the allele, it may not have had time to be selected out of the population.

Alternatively, it could be that 5HTTLPR is near enough in the genome to other sites under sufficiently strong positive selection that purifying selection of the putatively deleterious allele is not possible due to linkage disequilibrium. Linkage disequilibrium occurs when alleles at certain loci "tag along" with others, making certain genotypes more common than random distribution would predict. In this case, the short 5-HTT allele would have to be close enough to an allele – so advantageous that it swamped any disadvantages of the short 5-HTT allele – to ride its coattails to survival in the population.

A third possibility to explain the apparent persistence of deleterious variants is that heterozygotes could display more hearty phenotypes than either homozygous form (a form of over-dominance, in which the heterozygous phenotype does not fall between the phenotypes of each homozygous genotype). However, evidence shows that heterozygotes tend to fall phenotypically between homozygotes on most measured phenotypes (see, e.g., 1). Of course, it could be the case that, as with malaria resistance given by the sickle cell trait, heterozygous advantage manifests in a completely different, as of yet unmeasured, phenotype.

This leads to a fourth, related possibility: pleiotropy—when a single gene influences multiple phenotypes. Some of these so-called risky alleles may have compensating fitness-enhancing effects on other phenotypes yet to be documented (see, e.g., 34). And we have to keep in mind that psychologically deleterious does not equate to reproductively deleterious.

This paper does not investigate evidence for the four potential explanations discussed above. We limit our analysis to two potential explanations – phenotypic capacitance and frequency dependent selection. Future research investigating these alternative explanations may need to take advantage of the increasing availability of genome-wide data, ideally among a large sample of twins.

List of Depression Index Items:
How often was each of the following things true during the past seven days?

1. You were bothered by things that usually don't bother you.
2. You could not shake off the blues, even with help from your family and your friends.
3. You felt that you were just as good as other people.
4. You had trouble keeping your mind on what you were doing.
5. You were depressed.
6. You were too tired to do things.
7. You enjoyed life.
8. You were sad.
9. You felt that people disliked you.
10. In the past 12 months, how often have you laughed a lot?
11. In the past 12 months, how often have you cried a lot?
12. How satisfied are you with your life as a whole?
13. Do you agree or disagree that you have many good qualities?
14. Do you agree or disagree that you have a lot to be proud of?
15. Do you agree or disagree that you like yourself just the way you are?

All items are coded on a scale from 0 to 3 so that higher scores represent more depressive symptoms.