Breeding racehorses: what price good genes?

Alastair J. Wilson* and Andrew Rambaut

Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3JT, UK

*Author for correspondence (alastair.wilson@ed.ac.uk).

Horse racing is a multi-million pound industry, in which genetic information is increasingly used to optimize breeding programmes. To maximize the probability of producing a successful offspring, the owner of a mare should mate her with a high-quality stallion. However, stallions with big reputations command higher stud fees and paying these is only a sensible strategy if, (i) there is a genetic variation for success on the racetrack and (ii) stud fees are an honest signal of a stallion’s genetic quality. Using data on thoroughbred racehorses, and lifetime earnings from prize money (LE) as a measure of success, we performed quantitative genetic analyses within an animal model framework to test these two conditions. Although LE is heritable ($V_h = 0.299 \pm 0.108$, $P_r = 0.002$), there is no genetic variance for stud fee and the genetic correlation between traits is therefore zero. This result is supported by an absence of any relationship between stud fees for currently active stallions and the predicted LE for their (hypothetical) offspring. Thus, while there are good genes to be bought, a stallion’s fees are not an honest signal of his genetic quality and are a poor predictor of a foal’s prize-winning potential.

Keywords: thoroughbred; *Equus caballus*; heritability; genetic correlation

1. INTRODUCTION

Thoroughbred horse racing is a multi-million pound international industry, in which there is an increasing emphasis on using genetic tools to optimize breeding programmes. Both quantitative and molecular approaches have revealed genetic variation for key racing performance traits (Gaffney & Cunningham 1988; Hill 1988; Harrison & Turrión-Gomez 2006), and accurately assessing the genetic potential of candidate parents may therefore be critical to breeding successful racehorses (Bailey 1998). In particular, the owner of a thoroughbred mare should mate her with a high-quality stallion in order to maximize the probability of producing successful progeny.

However, a breeder must also pay stud fees which increase with a stallion’s reputation, a reputation that might be based on his own racing career, or on the performances of his progeny to date (or both). Thus, by paying higher stud fees, breeders may assume that they are effectively buying better genes and thereby increasing the prize-winning potential of the resultant foal. However, this is only true if the stallion’s reputation, and hence nomination fee, is an honest signal of his genetic quality. This system has strong parallels in evolutionary biology, where theory predicts that honest signals of male genetic quality will be used by females to determine mate choice and influence levels of reproductive investment (e.g. Fisher 1958; Kokko et al. 2002; Maynard Smith & Harper 2003). Here, however, investment and mate choice are performed by proxy (i.e. the breeder rather than the mare), with a view to maximizing financial returns rather than evolutionary fitness. In this context, paying higher fees for a stallion with a bigger reputation may not make financial sense if fees are not an honest signal of genetic quality.

Here we perform quantitative genetic analyses on a pedigreed dataset of thoroughbred racehorses to address two questions. Firstly, we test whether there is genetic variation for success on the racetrack by analysing data on lifetime prize money earnings. Secondly, we ask whether stud fees are a useful indicator of a stallion’s genetic quality and hence its offspring’s prize-winning potential.

2. MATERIAL AND METHODS

(a) Data sources

Data relating to 554 currently (or recently) active stallions standing at stud in the UK and USA were collated from freely available sources at www.the-racehorse.com, http://www.britishhorseracing.com/ and http://breeding.bloodhorse.com. For each stallion we recorded stud farm, state (or country) and stallion nomination fee (SNF). For some stallions, fee records were available from multiple seasons (between 2001 and 2007). In such cases, we averaged over records for simplicity. Analyses using multiple records per individual were also performed, but they yielded quantitatively similar results and are not shown.

Lifetime prize money earned (LE) was recorded for each stallion, and also for each of his ancestors (both male and female) to a depth of four generations, from http://www.pedigreequery.com/. For each horse, the number of races run, and year and country of foaling were obtained from the same source. In total, we thus compiled a pedigreed database of 4476 horses foaled between 1922 and 2003, with stud fees for 554 active stallions and lifetime performance statistics for a much larger set of 2500 horses. All lifetime earnings and stud fees were converted to US dollars where reported in other currencies (based on exchange rates in the year of birth) and were log transformed prior to analysis. Data sources provide no guarantee as to the accuracy of the data, and while errors in either phenotypic values or pedigree links will cause downward bias in estimated genetic variance, this effect is expected to be minimal.

(b) Quantitative genetic analyses

Phenotypic variance for the traits of LE and SNF was partitioned into genetic and environmental variance components using univariate animal models (Kruuk 2004) solved by restricted maximum likelihood using the program ASReml (VSN International). For each trait $y_i$, the phenotype of individual $i$ ($y_i$) was modelled as

$$y_i = \mu + \text{fixed effects} + a_i + e_i$$

(2.1)

where $\mu$ is the overall population mean, $a_i$ is the additive genetic effect on individual $i$ assumed to be normally distributed with mean of zero and variance of $V_a$ (the additive genetic variance) and $e_i$ is a residual resulting from environmental effects with zero mean and variance $V_e$. The genetic variance $V_a$ is estimated from the variance-covariance matrix of additive genetic effects, which is equal to $A \sigma_a^2$, where $A$ is the additive numerator relationship matrix containing the individual elements $A_{ij} = 2 \theta_{ij}$, and $\theta_{ij}$ is the coefficient of coancestry between individuals $i$ and $j$ (obtained from the pedigree structure).

In addition to the population mean $\mu$, further fixed effects were included to account for known influences on phenotype. Statistical significance of fixed effects was assessed using conditional Wald $F$ statistics. For LE, fixed effects included were the number of races started, sex, year of birth (modelled as a quadratic) and country of birth (as a factor). For SNF, state (or country if outside the USA) was included as a fixed effect while the stud farm was included as an additional random effect (with a corresponding variance component, $V_{\text{STUD}}$ consequently...
3. RESULTS
Animal model analysis indicated a heritable component of variation for lifetime prize earnings (LE; table 1). Additive genetic variance was found to be statistically significant based on comparison to a reduced model \( \left( \sigma_A^2 = 0.299 \pm 0.108 ; \chi^2 = 10.02, \Pr = 0.002 \right) \), with a corresponding heritability, \( h^2 = 0.095 \pm 0.034 \). Inclusion of parental and grandparental effects did not result in significantly improved likelihoods and we therefore do not present results from these more complex models.

In contrast to LE, there was no evidence of additive genetic variance for SNF. \( \sigma_A^2 \) was estimated as \(-0.015 \pm 0.052\), and fixed at zero when the model was constrained to biologically meaningful parameter space (i.e. variance components greater than or equal to zero). Substantial differences were found in stallion nomination fees among farms (univariate analysis of SNF; \( \sigma^2 = 0.4176 \pm 0.077 \)). This effect accounted for 44% of the phenotypic variance in SNF and was statistically significant based on a likelihood ratio test comparison with a reduced model \( \left( \chi^2_1 = 130, \Pr < 0.001 \right) \).

The phenotypic correlation between the log_\text{e}-transformed traits was positive but marginally non-significant \( r_{LE,SNF} = 0.074, \Pr = 0.074 \), while in the absence of additive genetic variance for SNF (a conclusion also supported by the bivariate animal model; results not shown) the genetic correlation between LE and SNF is defined as zero. Further support for the absence of a genetic relationship between LE and SNF is provided by model predictions of progeny lifetime earnings for currently active stallions (figure 1). Among stallions of varying genetic merit for LE, predicted lifetime progeny earnings for hypothetical female offspring varied from approximately $57 500 to $140 000. However, the slope of a linear regression of predicted lifetime earnings on SNF (performed on the dollar scale), though positive, was small \((0.020 \pm 0.011)\) and insignificant \( (F_{1,438} = 2.98, \Pr = 0.085) \). It should also be noted that since predicted lifetime progeny earnings incorporate estimates of genetic merit they are not independent among relatives, and statistical testing in this way is also inherently anti-conservative (Hadfield in press).

4. DISCUSSION
Our analysis shows that there is heritable variation for lifetime prize money earned. While environmental effects (including, for example, training, choice of races entered, jockeys employed and stochastic injury events) clearly have a much greater influence on lifetime earnings (accounting for 91.5% of the variance), LE is a heritable trait that could respond to appropriate selection. Thus, there is variation in genetic quality, and it should be possible for a breeder to pick a stallion for his good genes.

While there are good genes to be bought, it does not appear to be the case that you get what you pay for. Rather than having any underlying genetic basis, our analyses show that the phenotypic association between fees and lifetime earnings arises from environmental, not genetic, effects. Thus, it seems probable that breeders who can provide the best environment and expertise for training race horses are also likely to be those best able to afford the highest stud fees. However, there is certainly no support for the hypothesized positive genetic correlation between stallion nomination fees and lifetime prize earnings. Although power to directly estimate genetic parameters for SNF is limited (discussed further below), our conclusion is also supported by the regression of predicted progeny earnings on sire stud fee. While a regression line with slope of 1 would mean an expected financial return that was equal to the investment, here the slope is close to zero (with an expected return of $0.02 for each additional dollar spent on fees) and not significant.

It is important to note that prize money is by no means the only route to financial return in the racing industry (e.g. stallions may earn considerably more at stud than on the track) and it would also be wrong to
suggest that making money is the sole reason for owning or breeding racehorses. Consequently, it could be readily argued that maximizing expected lifetime prize earnings may not be the most important objective for a racehorse breeder. Nevertheless, in the absence of genetic variance for stallion nomination fees, it is not possible for SNF to show genetic covariance with other performance traits that might be more direct targets of selection. That is to say, the fees are not expected to correlate well with genetic merit for any other trait.

As a partial caveat to this conclusion, we acknowledge that sample size is relatively small for SNF and that this may limit power to detect low, but non-zero, levels of $V_A$ (at least by comparison with LE). Nevertheless, based on the unconstrained animal model we determined an upper (95%) confidence limit of just 0.09 for $V_A$ (approx. 8% of the phenotypic variance as estimated using a model containing fixed effects only). Furthermore, our conclusion is based on available stud fee data from 2001 to 2007 and it is possible that $V_A$ for SNF varies over time. For example, genotype-by-environment interactions are ubiquitous in nature and mean that genetic (co)variances change over time in temporally heterogeneous environments (Wilson et al. 2006). To the extent that the ‘economic environment’ in which stud fees are determined may also be variable, there is no reason to preclude similar effects here.

In summary, our results show that genetic variance exists for lifetime prize earnings as well as more specific performance traits measured on the track (e.g. Gaffney & Cunningham 1988). However, if the goal is to maximize lifetime prize winnings, then it seems clear that stud fees are not an honest signal of a stallion’s genetic quality.

We thank J. Hadfield and the participants of the second Wild Animal Model Biannual Meeting (Gotland, 2007) for useful discussion leading to this manuscript, and two anonymous referees for their constructive comments on the manuscript. A.J.W. is supported by the Natural Environment Research Council and A.R. is supported by the Royal Society.


Table 1. Univariate animal model analysis of log_{e} transformed lifetime prize earnings (LE) in thoroughbred racehorses. (Phenotypic variance ($V_P$) was partitioned into additive genetic ($V_A$) and residual environmental ($V_R$) components. The corresponding ratios of these components to phenotypic variance are indicated. Fixed effects included in the model are also expected effect in males relative to females.)