LETTERS

# Pleiotropic scaling of gene effects and the 'cost of complexity'

Günter P. Wagner[1], Jane P. Kenney-Hunt[2], Mihaela Pavlicev[2], Joel R. Peck[3], David Waxman[3] & James M. Cheverud[2]

**As perceived by Darwin, evolutionary adaptation by the processes of mutation and selection is difficult to understand for complex features that are the product of numerous traits acting in concert, for example the eye or the apparatus of flight. Typically, mutations simultaneously affect multiple phenotypic characters. This phenomenon is known as pleiotropy. The impact of pleiotropy on evolution has for decades been the subject of formal analysis[1–6]. Some authors have suggested that pleiotropy can impede evolutionary progress (a so-called 'cost of complexity'[5]). The plausibility of various phenomena attributed to pleiotropy depends on how many traits are affected by each mutation and on our understanding of the correlation between the number of traits affected by each gene substitution and the size of mutational effects on individual traits. Here we show, by studying pleiotropy in mice with the use of quantitative trait loci (QTLs) affecting skeletal characters, that most QTLs affect a relatively small subset of traits and that a substitution at a QTL has an effect on each trait that increases with the total number of traits affected. This suggests that evolution of higher organisms does not suffer a 'cost of complexity' because most mutations affect few traits and the size of the effects does not decrease with pleiotropy.**

A set of quantitative traits in strains of inbred mice were chosen to represent a broad cross-section of linear skeletal traits from all major subsystems of the bony skeleton (see Supplementary Table 1). The strains used were initially selected for increased (LG/J) and reduced (SM/J) body size at 60 days *post partum*, respectively, and then inbred for more than 100 generations[7,8]. A QTL mapping analysis of 70 such phenotypic characters was performed on 1,040 mice of the $F_2$ generation that were derived from a cross between the inbred LG/J and SM/J lines. The number of scored markers used was 471, leading to an average distance between them of 3.98 centimorgans. A total of 102 autosomal QTLs were detected[9]. These QTLs exhibited pleiotropy, as assessed by a specific statistical test[10]. The QTLs identified by this test affected a variable number of characters, up to a maximum of 30, with a mean of 7.8 and a median of 6 (Fig. 1a). This means that in our data, 50% of QTLs affect fewer than 10% of the 70 characters measured. This result suggests that pleiotropic effects tend to be limited to subsets of the total phenotype rather than being widespread, as assumed in models of universal pleiotropy, in which all traits are affected by a mutation (see also Supplementary Fig. 1).

To compare the mutational effects of different QTLs on the affected traits, we calculated a standardized effect for each character by dividing the QTL effect by the trait's phenotypic standard deviation (see Supplementary Note 1). The standardized effect on trait $i$, denoted by $A_i$, is half the difference in means between homozygotes. The total effect, $T$, of a QTL is then defined as the euclidean distance spanned by all the single character effects (Supplementary Note 6):

$$T = \sqrt{\sum_{i=1}^{N} A_i^2}$$

In this equation, $N$ represents the degree of pleiotropy; that is, the number of characters affected by a QTL.

The distribution of the total mutational effect $T$ arising from the QTL data set is illustrated in Supplementary Fig. 2. The distribution is estimated to have a mean of 0.59 and a standard deviation of 0.334. We consider these estimates to be unbiased, because the well-known bias of QTL effects[11] is minimal in sample sizes above 500 as in our study. We use these data to test two well-known models of how the total effect depends on the degree of pleiotropy, the euclidean superposition model and the invariant total effect model.

The euclidean superposition model assumes that the expected squared effect of a mutation on a character $\langle m^2 \rangle$ is the same regardless how many other characters are also affected[4]. Taking the phenotypic standard deviation of a trait to be unity, the expected total effect of a mutation then is $\langle T \rangle = \sqrt{N}\sqrt{\langle m^2 \rangle}$. Thus, the total effect of a mutation is predicted to scale with the square root of the degree of pleiotropy $N$. This model was used by Rechenberg[2], for evolutionary optimization problems in engineering, and by Turelli[3], Wagner[4,12], Waxman and Peck[13] and others, in a population genetic context. Orr[5] and Wingreen *et al.*[6] proposed alternative models that we term invariant total effect models. These models assume that the probability distribution of total effect, $p(T)$, and hence $\langle T \rangle$ itself, is independent of the degree of pleiotropy. This assumption leads to progressively smaller mutational changes on each trait as the degree of pleiotropy increases and to slower rates of evolutionary adaptation with increased numbers of traits.

We first tested these predictions by considering the dependence of the total effect $T$ on the variable number of traits $N$ affected by a mutation (Fig. 1b). The total effect $T$ is strongly increasing with pleiotropy with a highly significant linear regression coefficient ($t = 24.05$, $P < 0.0001$) and a highly significant regression effect ($F_{1,100} = 578.7$, $P < 0.0001$). This result decisively rules out the invariant total effect models of pleiotropy. However, it is also clear that the data do not support the euclidean superposition model either, because that model predicts a square-root dependence rather than the near-linear dependence that is apparent in Fig. 1b (see Supplementary Note 2). The linear model has a higher $R^2$, suggesting a linear rather than a square-root dependence (linear regression adjusted $R^2 = 0.8512$, square-root regression adjusted $R^2 = 0.7935$; for more detail see Supplementary Note 2). We compared the fit of the two regression models by calculating their log-likelihood ratio, 3.638. This ratio indicates that the data are about 40-fold less likely under the square-root model than under the linear model. We thus conclude that the data are not only severely inconsistent with the invariant total effect

[1]Department of Ecology and Evolutionary Biology, Yale University, New Haven, Connecticut 06520-8106, USA. [2]Department of Anatomy and Neurobiology, Washington University, St Louis, Missouri 63110, USA. [3]Center for the Study of Evolution, School of Life Sciences, University of Sussex, Brighton BN1 9QG, UK.

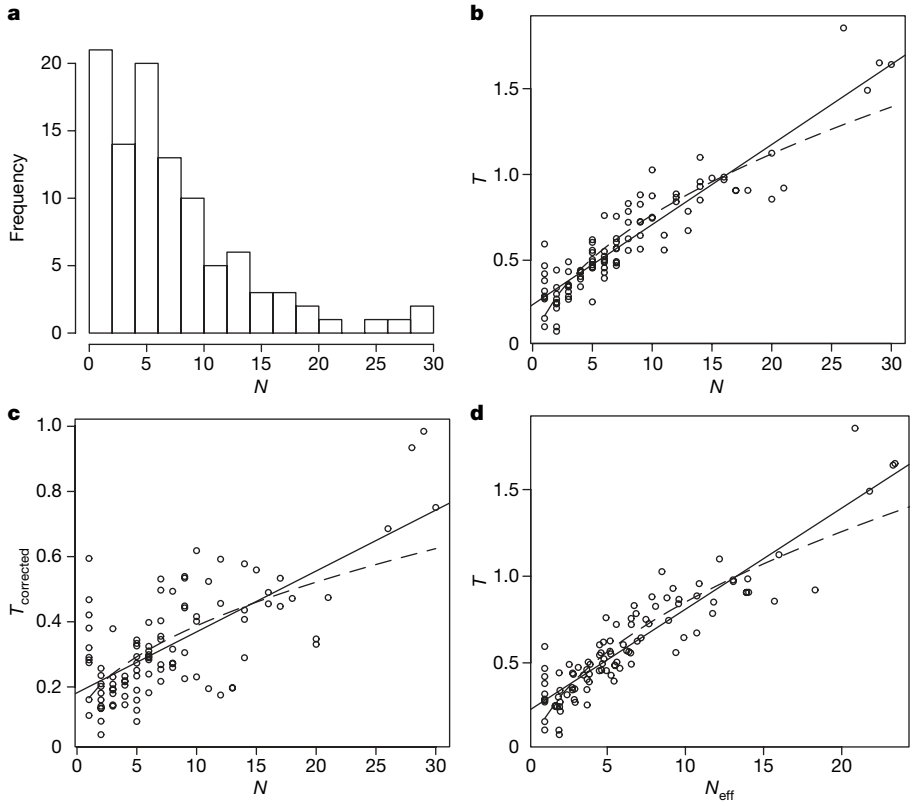


**Figure 1 | Distribution of QTL effects on 70 skeletal traits in the mouse.** **a**, The frequency distribution of the degree of pleiotropy of QTLs. Note that 50% of QTLs affect only up to six traits, but there is a set of QTLs affecting as many as 25–30 traits. **b**, Regression of uncorrected total effect $T$ on the number of traits $N$ affected by a QTL. Note the very tight relationship between $T$ and $N$. This result does not support the assumption that $T$ is independent of the degree of pleiotropy. It also suggests that $T$ increases more rapidly than predicted by the euclidean superposition model. Solid line, linear dependence; dashed line, square-root dependence. **c**, Regression of weight-corrected $T$ on the number of traits $N$. The residual effects represent the distribution of mutational effects better than the raw QTL effects, but there is still a strongly positive relationship between corrected $T$ and degree of pleiotropy. Solid line, linear dependence; dashed line, square-root dependence. **d**, To account for the correlations between the traits we calculated the 'effective number of traits' affected by a QTL, as described in the text. This correction reduces the maximal number of pleiotropic effects from 30 (**b**, **c**), to at most 24 effective traits. Even on this scale for pleiotropy the positive regression remains significant, suggesting that pleiotropy strongly influences the total phenotypic effect of a mutation. Solid line, linear dependence; dashed line, square-root dependence.

model but also suggest a stronger increase in the total effect with more traits affected than are predicted by the euclidean superposition model. The total effect of a mutation seems to be increasing more strongly with pleiotropy than predicted if we assume that the effect on each character remains the same.

One concern with this result is that a positive relationship between the number of traits affected by a QTL and the total effect could be caused by an artefact of QTL mapping. For instance, let us assume that all the QTLs have the same degree of pleiotropy and the same total effect on average. Let us further assume that for some QTLs the effects on a few traits are too small to be detected with our experimental design. This would lead to both an underestimate of the number of traits affected and an underestimate in the total effect, because the undetected effects are not added to the estimate of the total effect. In this way QTLs for which fewer effects are detected will also have systematically smaller total effects, leading to a positive regression of the total effect on the estimated pleiotropy (the number of traits affected). To assess whether our result in Fig. 1b can be explained as such an artefact, we calculated the predicted regression slope given the detection limit for QTL effects (see Supplementary Note 3). For our data this slope is predicted to be 0.0033 or less. The observed slope is about tenfold that predicted by this model: $0.047 \pm 0.0019$. We conclude that the result in Fig. 1b is unlikely to be caused by a detection artefact.

We considered two other ways in which the inferences about a positive relationship between pleiotropy and total effect could be flawed. First, the distribution of QTL effects revealed by this cross might not be an unbiased sample of all mutational effects, because the two lines used in this experiment were created by divergent selection for body weight. Alleles with a larger impact on body weight therefore had a greater chance of being fixed than alleles with smaller effects. The distribution of QTL effects might thus not reflect the true mutational distribution for skeletal traits but might over-represent alleles with large effects on components of overall body size. Second, the degree of pleiotropy might be overestimated because of correlations between traits. We address these two concerns below.

To test whether our conclusions are influenced by the fact that the strains were subject to directional selection on body weight rather than mutation accumulation, we calculated the partial regression of trait value on body weight and re-analysed the weight-corrected trait values. Plausibly, these values are more representative of the distribution of mutational effects. The results are shown in Fig. 1c and Supplementary Fig. 3. There is still a clear positive relationship between the total effect and the number of traits affected. The scatter is greater, but the regression coefficient is clearly positive and significant ($t = 9.887$, $P < 2 \times 10^{-16}$, $F_{1,100} = 97.74$, $P < 0.0001$).

To test whether our inference could be due to an overestimate of the degree of pleiotropy, we considered the hypothesis that the degree of pleiotropy is less than the number of traits as a result of correlations between the traits. To address this, we first tested whether the phenotypic covariance matrix of the traits affected by a QTL has fewer dimensions than the number of traits. As discussed in Supplementary Note 5, error variance has the effect of inflating the dimensionality of a covariance matrix. To correct for this effect, we implemented an adjusted bootstrap procedure that uses an estimated error threshold to test for the dimensionality of the covariance matrix (see Supplementary Methods). All QTLs except one affect sets of characters that have full dimensionality according to this criterion. Another way of assessing the effect of trait correlations on our result is to estimate an 'effective number of traits', $N_{\text{eff}}$. We propose to use the following simple equation (see Supplementary Note 4)

$$N_{\text{eff}} = N - \text{Var}(\lambda)$$

where $\text{Var}(\lambda)$ is the variance of the eigenvalues of the error-corrected correlation matrix. Figure 1d gives the result with QTL effects $T$ and the effective number of traits. Again there is a significant positive relationship (linear regression coefficient $t = 10.99$, $P < 0.0001$; $F_{1,100} = 548.7$, $P < 0.0001$). We therefore conclude that the total effect of a pleiotropic mutation increases with the degree of pleiotropy, even when correlations between traits are taken into account. Regressing weight-corrected mutational effects on effective trait number (Supplementary Fig. 4) still gives a positive relationship between mutation effects and degree of pleiotropy ($t = 8.974$, $P = 1.75 \times 10^{-14}$; $F_{1,100} = 94.96$, $P = 3.59 \times 10^{-16}$; $R^2 = 0.4819$) but does not discriminate between the linear and the square-root regression models.

The observed linear relationship between total effect size and the number of traits implies that the average mutational effect, per trait, is increasing with the square root of the degree of pleiotropy. Mutations with a high degree of pleiotropy have more substantial effects on each trait than mutations with a more limited degree of pleiotropy. This pattern is reminiscent of the decanalizing effects of major mutations[14] except that in this case the alleles are not pathological, as other large effect alleles may be, but are part of the natural variation of the species. In addition, this effect is not due to the release of hidden genetic variation, which is a generic feature of genes with epistasis[15]. We conclude that an increased degree of pleiotropy is accompanied by an increase in the overall phenotypic effects of mutations even among 'minor effect' alleles.

These findings affect predictions about the consequences of complexity on evolvability in two ways. First, the reason that Fisher's geometric model suggests a decrease in evolvability with increasing number of traits (complexity) is that his and all studies following his approach assume that each mutation potentially affects all traits ('universal pleiotropy'). Therefore with increasing complexity it becomes increasingly unlikely that all traits are affected by a mutation in a way that causes fitness to increase. However, the effects we detected in our study are not nearly as widely pleiotropic as assumed by the model of universal pleiotropy. QTL effects are more restricted to parts of the phenotype as suggested by the idea of variational modularity[16,17]. Why this is so is unclear, but there is increasing evidence that natural selection can change pleiotropy such that evolvability increases[18]. If, at any one time, only one or a few characters are maladapted, modularity increases evolvability[19,20]. The second factor that was cited as leading to a lower evolvability of complex organisms is the assumption of constant total effect[5]. This assumption was introduced to accommodate the fact that most mutations have small effects[5,6]. In contrast, the euclidean superposition model with universal pleiotropy predicts that the probability of small-effect mutations becomes very small. This is so because if many characters are affected by each mutation, then it would be unlikely that the total effect is small. The constant-total-effect model, however, has the consequence that the average effect per character decreases and thus the rate of response to directional selection also decreases, leading to another cost of complexity prediction. However, our data show that the total effects of mutation actually increase with pleiotropy. It therefore seems that in real organisms the combination of restricted rather than universal pleiotropy, and increasing total effects, could be seen as evolution's answer to the challenges of evolving complex organisms with random variation and selection.

## METHODS SUMMARY

The experimental population results from an intercross of inbred mouse strains LG/J and SM/J[21–23]. We measured 70 skeletal traits on 1,040 individuals of the $F_2$ generation (see Supplementary Table 1). QTLs were calculated by interval mapping[22,24], and correcting for multiple tests[25]. Tests for pleiotropy were performed[10]. The additive effect of an allele is $A = a/\text{s.d.}$, where $a$ is half the difference between the homozygotes. The total effect of a QTL is $T = \sqrt{\sum A_i^2}$.

To correct for the effects of selection on body size during the generation of inbred lines, we included body weight as a covariate in the course of interval mapping[24]. We determined the true dimensionality by comparing the eigenvalues with the thresholds derived from known trait-specific measurement errors. For each QTL we constructed the phenotypic covariance matrix for the traits affected, and determined its eigenvalues and eigenvectors. The error threshold was estimated as the projection of the error variance onto the respective eigenvector. The significance of the eigenvalue is estimated by a bootstrap procedure.

We further estimated the 'effective number of traits' by accounting for correlations between traits. The measured covariance matrix was corrected for measurement error by subtracting the error variances from the diagonal. From the corrected covariance matrix we calculated the eigenvalues of the correlation matrix. The variance of these eigenvalues was subtracted from the number of traits to obtain the effective dimensionality.

To compare the fit of the data between linear and square-root regression we calculated the respective log-likelihood of the model from the regression residuals, assuming a normal distribution of residuals. We used the log-likelihood ratio to estimate the ratio between model fits. We consider a log-likelihood ratio of less than –3.0 to be 'significantly' better support for the particular model.

**Full Methods** and any associated references are available in the online version of the paper at www.nature.com/nature.

**Received 23 August 2007; accepted 28 January 2008.**

1. Fisher, R. A. *The Genetical Theory of Natural Selection* (Clarendon, Oxford, 1930).
2. Rechenberg, I. *Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution* (Fromman-Holzboog, Stuttgart, 1973).
3. Turelli, M. Effects of pleiotropy on predictions concerning mutation-selection balance for polygenic traits. *Genetics* **111,** 165–195 (1985).
4. Wagner, G. P. The influence of variation and developmental constraints on the rate of multivariate phenotypic evolution. *J. Evol. Biol.* **1,** 45–66 (1988).
5. Orr, H. A. Adaptation and the cost of complexity. *Evolution Int. J. Org. Evolution* **54,** 13–20 (2000).
6. Wingreen, N. S., Miller, J. & Cox, E. C. Scaling of mutational effects in models of pleiotropy. *Genetics* **164,** 1221–1228 (2003).
7. Chai, C. K. Analysis of quantitative inheritance of body size in mice II: gene action and segregation. *Genetics* **41,** 165–178 (1956).
8. Eppig, J. T., Bult, C. J., Kadin, J. A., Richardson, J. E. & Blake, J. A. and the members of the Mouse Genome Database Group. The mouse genome data base (MGD): from genes to mice – a community resource for mouse biology. *Nucleic Acids Res.* **33,** D471–D475 (2005).
9. Kenney-Hunt, J. P. *et al.* Pleiotropic patterns of quantitative trait loci for seventy murine skeletal traits. *Genetics.* (in the press).
10. Knott, S. A. & Haley, C. S. Multitrait least squares for quantitative trait loci detection. *Genetics* **156,** 899–911 (2000).
11. Xu, S. Theoretical basis of the Beavis effect. *Genetics* **165,** 2259–2268 (2003).
12. Wagner, G. P. Multivariate mutation-selection balance with constrained pleiotropic effects. *Genetics* **122,** 223–234 (1989).
13. Waxman, D. & Peck, J. R. Pleiotropy and preservation of perfection. *Science* **279,** 1210–1213 (1998).
14. Waddington, C. H. *The Strategy of Genes* (Macmillan, New York, 1957).
15. Hermisson, J. & Wagner, G. P. The population genetic theory of hidden variation and genetic robustness. *Genetics* **168,** 2271–2284 (2004).
16. Wagner, G. P. & Altenberg, L. Complex adaptation and the evolution of evolvability. *Evolution Int. J. Org. Evolution* **50,** 967–976 (1996).
17. Hansen, T. F. Is modularity necessary for evolvability? Remarks on the relationship between pleiotropy and evolvability. *Biosystems* **69,** 83–94 (2003).
18. Wagner, G. P., Pavlicev, M. & Cheverud, J. M. The road to modularity. *Nature Rev. Genet.* **8,** 921–931 (2007).
19. Welch, J. J. & Waxman, D. Modularity and the cost of complexity. *Evolution Int. J. Org. Evolution* **57,** 1723–1734 (2003).
20. Martin, G. & Lenormand, T. A general multivariate extension of Fisher's geometrical model and the distribution of mutation fitness effects across species. *Evolution Int. J. Org. Evolution* **60,** 893–907 (2006).
21. Cheverud, J. M. *et al.* Quantitative trait loci for murine growth. *Genetics* **142,** 1305–1319 (1996).
22. Vaughn, T. T. *et al.* Mapping quantitative trait loci for murine growth: a closer look at genetic architecture. *Genet. Res.* **74,** 313–322 (1999).
23. Cheverud, J. M. *et al.* Genetic architecture of adiposity in the cross of LG/J and SM/J inbred mice. *Mamm. Genome* **12,** 3–12 (2001).
24. Haley, C. S. & Knott, S. A. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69,** 315–324 (1992).
25. Cheverud, J. M. A simple correction for multiple comparisons in interval mapping genome scans. *Heredity* **87,** 52–58 (2001).

**Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.

**Acknowledgements** We thank A. Pyle, A. Kondrashov and B. Walsh for suggestions that have improved this manuscript. We thank the members of the Wagner and Cheverud laboratories for critical discussion. J.P. and D.W. thank members of the evolution group at Sussex. J.M.C. is funded by the National Institutes of Health and the National Science Foundation (NSF), G.P.W. acknowledges funding from the NSF, the Humboldt Foundation and the John Templeton Foundation, M.P. is funded by the Austrian Science Foundation (FWF) Fellowship, and the work of D.W. was supported by the Leverhulme Trust.

**Author Contributions** G.P.W. conceived this study, participated in the statistical analysis and wrote the manuscript. J.P.K.-H. collected the morphological data and performed the QTL analysis. M.P. did the statistical analyses. J.M.C. was responsible for generating the mouse populations and the genotype data used in the original mapping and advised on the pleiotropic scaling analysis. J.P. and D.W. performed a theoretical analysis of the scaling of trait effects with pleiotropy. All authors participated in the preparation of the manuscript.

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to G.P.W. (gunter.wagner@yale.edu) or J.M.C. (cheverud@pcg.wustl.edu).

## METHODS

**Experimental populations.** The experimental population results from an intercross of inbred mouse strains LG/J and SM/J[21–23]. LG/J and SM/J were selected for large and small body weight at 60 days of age, respectively[22,24]. In this study we combine the data from the two replications of an intercross protocol (intercross I and II) comprising a total of 1,040 $F_2$ mice. For details on markers and map resolution see ref. 9. Phenotypic traits comprise 70 skeletal measurements representing the cranium, axial and appendicular skeleton, as well as body weight at 10 weeks and at necropsy. For the list of measurements and their respective repeatability measures see Supplementary Information. For details on measuring techniques see ref. 9. Extreme outliers were eliminated to avoid biasing the data. The effects of dam, litter size, experimental block, sex, age at necropsy, and intercross were removed to reduce non-genetic variance, increasing the detectability of QTLs[9,21].

**QTL analysis.** QTL map positions and effects were calculated by interval mapping[22,24]. Genotypes were imputed every 1 centimorgan along each chromosome by using the genotypic scores of the flanking markers and the recombination rate[24]. The QTLs were positioned where there was the least probability ($P$) of the genotype–phenotype correlation occurring by chance. To account for multiple testing, overall significance level ($P = 0.05$) was derived by correcting the single-locus significance level for the number of tests by using a modified Bonferroni correction[25]. Formal tests for pleiotropy were performed, as proposed in ref. 10 and described in ref. 26.

The additive effect of an allele is defined as $A = a$/s.d. To determine the total effect of a QTL we calculate the euclidean distance of the genotype from the origin in the phenotype space by $T = \sqrt{\sum A_i^2}$.

**Correction for the effect of selection.** The experimental population used is an intercross of two inbred strains previously selected for low (SM/J) or high (LG/J) body weight at 60 days of age. The correlated response to this selection regime is manifested in the genotypic values of the traits. However, the pleiotropy scaling models tested here make assumptions about the mutational effect distribution, not the mutations fixed by a selective process. The observed distribution of allelic effects might therefore be biased towards alleles with large effects on body size, the target of selection. To correct for the effect of selection, we included the weight measured at nine weeks of age (closest available to 60 days) as a covariate in the multiple regression of traits in the course of interval mapping[24] (see above). The model thus estimates the genetic effects at the locus that accumulated in these lineages independently of the selected trait.

**Testing for matrix rank.** Measurement error for each trait is stochastically independent and therefore inevitably leads to a covariance matrix of full rank. If we were to take the rank of the covariance matrix as an estimate of the dimensionality of the phenotype, we could thus overestimate the pleiotropic range of a QTL. Hence we determined the true dimensionality of the trait distribution. We tested whether the eigenvalues were larger than zero as a result of the dimensionality of the underlying variables or as a result of error variance by using information about the repeatability of the measurements[9]. Measurement error variance was derived from trait-specific repeatability measures ($r^2$). Each $r^2$ is a coefficient of determination for the regression of mean value of all measurements taken on an individual onto a single repeated measurement. The error variance $E_j^2$ is obtained by multiplying this value by the total variance of the trait across all individuals.

For each QTL we constructed the phenotypic covariance matrix for the traits affected by that locus. Then we determined the eigenvalues and eigenvectors of this matrix. The expected variance due to measurement error for each eigenvector was estimated as

$$t_i = \sum_{j=1}^{n} v_{i,j}^2 E_j^2$$

that is, the projection of the amount of error variance onto the respective eigenvector.

The significance of each eigenvalue was estimated by a bootstrap procedure on the individuals (1,000 iterations). At each iteration we constructed the covariance matrix and determined its eigenvalues and eigenvectors. For each eigenvector of the original covariance matrix we projected the amount of variance in the bootstrap sample. The distribution of the projected bootstrap variances was compared with a threshold value derived from the trait-specific error variances. If at least 95% of the projected bootstrapped variances are larger than the threshold, we consider the corresponding eigenvalue to be significantly larger than measurement error. This procedure yielded the full rank for all pleiotropic ranges, so the number of traits affected ($N$) was used.

**Effective number of traits.** Because the phenotypic traits are correlated, we estimated the 'effective number of traits' by correcting the number of traits for their correlations. Highly correlated traits contribute little unique variance. The variance–covariance matrix was corrected for error variances by subtracting the error variances from the diagonal elements of the measured covariance matrix. Next, the correlation matrix was calculated from the corrected covariance matrix. Subsequently, the eigenvalues of the correlation matrix were determined and the variance of eigenvalues was calculated. The eigenvalue variance was then subtracted from the number of traits to derive the effective number of dimensions:

$$N_{\text{eff}} = N - \text{Var}(\lambda)$$

This function returns the value of 1.0 if all the traits are perfectly correlated and a value of $N$ if no correlation exists among the traits.

**Log-likelihood ratio calculation to compare regression models.** To determine whether a linear or square-root regression model better represents the data, we calculated the log-likelihood of each model from the residuals of the regression. We assumed that the residuals are normally distributed around the regression expectation. Hence the log-likelihood of a regression model M is

$$L(M) \propto \Pr(\text{Data}|M) = \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi V}} \exp\left[-\frac{s_j^2}{2V}\right]$$

$$\log L(M) = -\frac{1}{V}\sum_{j=1}^{n} s_j^2$$

where $s_j$ is the residual of the $j$th data point, and $V$ is the variance of the residuals for the model with the better fit. The log-likelihood ratio is then $\log L(M_1) - \log L(M_2)$. The factor

$$f = \exp[\log - \text{likelihood ratio}] = \frac{\Pr(\text{Data}|M_1)}{\Pr(\text{Data}|M_2)}$$

indicates the extent to which the data are less likely on the assumption of one model versus the other. If this factor is less than 0.05 it means that the data are less than 5% as likely with model 1 as with model 2. We consider this to 'significantly' better support for model 2 than for model 1.

26. Ehrich, T. H. *et al.* Pleiotropic effects on mandibular morphology I: developmental morphological integration and differential dominance. *J. Exp. Zool. B* **296**, 58–78 (2003).