Genet. Res., Camb. (2006), **87**, *pp.* 207–216.
doi:10.1017/S0016672306008172 Printed in the United Kingdom

Genetic mapping of allometric scaling laws

FEI LONG[1], YING QING CHEN[2], JAMES M. CHEVERUD[3] AND RONGLING WU[1]*
[1] *Department of Statistics, University of Florida, Gainesville, FL 32611, USA*
[2] *Program in Biostatistics, Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA*
[3] *Department of Anatomy and Neurobiology, Washington University School of Medicine, St Louis, MO 631110, USA*

(*Received 26 July 2005 and in revised form 4 December 2005*)

Summary

Many biological processes, from cellular metabolism to population dynamics, are characterized by particular allometric scaling relationships between rate and size (power laws). A statistical model for mapping specific quantitative trait loci (QTLs) that are responsible for allometric scaling laws has been developed. We present an improved model for allometric mapping of QTLs based on a more general allometry equation. This improved model includes two steps: (1) use model II regression analysis to estimate the parameters underlying universal allometric scaling laws, and (2) substitute the estimated allometric parameters in the mixture-based mapping model to obtain the estimation of QTL position and effects. This model has been validated by a real example for a mouse F_2 progeny, in which two QTLs were detected on different chromosomes that determine the allometric relationship between growth rate and body weight.

1. Introduction

Among a vast range of species from microorganisms to the largest mammals, many biological variables seem to bear a specific quarter-power scaling relationship to overall body size (Fig. 1*A*). For example, various biological times (e.g. lifespan and the time between heartbeats) scale with body mass to the 1/4 power, and resting metabolic rate scales with body mass to the 3/4 power (McMahon, 1973; Calder III, 1984; Schmidt-Nielsen, 1984; Enquist *et al*., 1998, 1999; Brown & West, 2000). Several attempts have been made recently to derive such general allometric scaling laws based on maximum efficiency (West *et al*., 1997, 1999*a*; Banavar *et al*., 1999), which has been regarded as the fundamental design principle for biological systems. Andresen *et al*. (2002) argued that maximum efficiency built on evolutionary (Bonner & Horn, 2000) as well as thermodynamic grounds (Dewey & Donne, 1998) may suffer from internal inconsistencies when it is used to explain scaling laws.

* Corresponding author. Tel: +1 (352) 3923806. Fax: +1 (352) 3928555. e-mail: rwu@stat.ufl.edu

Allometric scaling laws can be described mathematically by a power function $B=\alpha M^{\beta}$, where B is a biological variable, M is the body weight, α is the constant and β is the scaling exponent. Allometric relationships result from the regulation of scale and proportion in living organisms and are thought to have a genetic component (Wu *et al*., 2003). Wu *et al*. (2002) presented a basic statistical framework for mapping quantitative trait loci (QTLs) responsible for universal quarter-power scaling laws of structure and function with the entire body size. A key issue for QTL mapping of allometry is how to explain the genetic value of a biological variable in terms of that of body weight under the scaling relationship. Wu *et al*.'s model takes advantage of a linear relationship between B and M after the power function is log-transformed and, therefore, is statistically straightforward to derive. When more complicated power functions, for which the log-transformation cannot lead to a linear relationship, are needed to describe allometric scaling laws, a different model based on Taylor's approximation is developed (Ma *et al*., 2003).

Although Taylor's approximation can establish a general model for allometric mapping, its actual

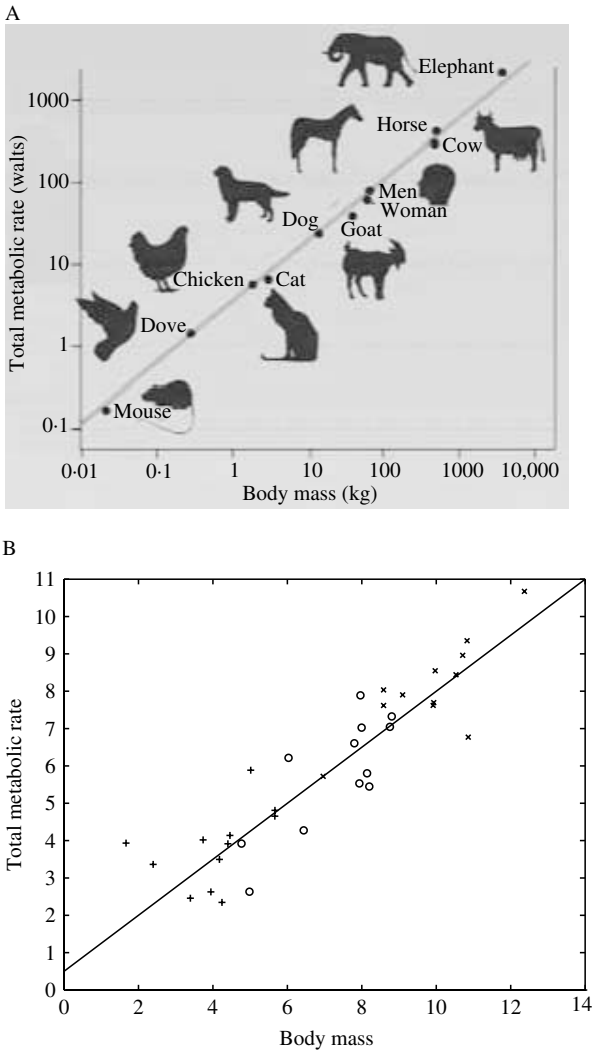

Fig. 1. Allometric scaling laws across different species (*A*) (downloaded from http://biology.unm.edu/jhbrown/Research/Scaling/Scaling.htm) and within species (*B*). Three different genotypes in (*B*) are assumed, coded by $\times$, $\bigcirc$ and $+$. The log-transformed genotypic values for the three genotypes are distributed along a straight line. If the coordinates of the three genotypic means do not overlap, this means that there exists a gene to affect the allometric law.

application presents a significant problem. The expression of the B mean as a function of the M mean depends on the order with which Taylor's series is expanded. Although the expansions of higher order can theoretically result in a more precise description of the B–M relationship than those of lower order, the former tend to be computationally more expensive than the latter. Because the precision of the model and computational efficiency are equally important, it would be difficult to make a compromise between these two aspects for a practical problem.

In this article, we develop an ad-hoc model for genetic mapping of general allometric scaling laws, aimed at the simultaneous improvement of the precision of the model and computational efficiency. Our model includes two different steps. First, model II regression analysis based on a loss function is used to fit observed bivariate data using a known allometric equation (Ebert & Russell, 1994). According to allometric scaling laws, differences among different organisms can be described by the same power function. Thus, it is reasonable to assume that different genotypes at an underlying QTL share a common power function which can be described by the same set of parameters. Second, the estimated parameters for the power function are substituted in a QTL mapping framework built on the finite mixture model. Instead of the estimation of all parameters, we need to estimate only the parameters describing the QTL effects on the dependent variable (i.e. body weight) and QTL position. A reduced number of model parameters being estimated increases the precision of parameter estimation. We derived the observed information matrix to investigate the precision of our statistical model. A worked example for mouse body weight growth illustrates the usefulness of model. In this example, we have identified two QTLs on different mouse chromosomes that regulate the allometric scaling relationship between growth rate and body weight.

2. Linear model for allometry mapping

Assume that our mapping population is an F_2 progeny of size N founded by two inbred lines. In this F_2 population, we measure one biological variable, B, and body weight, M, for each individual and construct a genetic map based on polymorphic markers. Suppose these two measured traits B and M are related by a power equation,

$$B = \alpha M^{\beta}. \tag{1}$$

The estimation of the constant parameter α and exponential power β in this equation is obtained by a linear regression following the log-transformation, i.e.

$$\ln(B) = \alpha + \beta \ln(M)$$

or

$$y = \alpha + \beta x, \tag{2}$$

where $y = \ln(B)$ and $x = \ln(M)$. An interval mapping approach has been developed to locate QTLs that affect allometric scaling laws based on this log-transformed linear regression function (Wu *et al.*, 2002). We will modify this approach by making appropriate transformations to minimize the number of unknown parameters being estimated. An in Wu *et al.* (2002), two common mechanisms for trait correlations, pleiotropic and linkage, will be considered.

https://doi.org/10.1017/S0016672306008172 Published online by Cambridge University Press

(i) *Pleiotropic model*

The pleiotropic model proposes that a common QTL affects variation in both log-transformed traits x and y. At a putative QTL, the F_2 population can be sorted into three genotype groups, QQ, Qq and qq, coded by 2, 1 and 0, respectively. These three groups form three clusters in the coordinate of x and y. If the allometric change of trait y with respect to trait x, as described by equation (1), results from the pleiotropic effect of this QTL, then the three points, each representing a pair of the expected mean values of the two traits in one genotype group, should be significantly different from each other but should be on the same line described by the log-transformed allometry equation (2) (Fig. 1*B*). With such a linear relationship, the expected mean value of the transformed trait y can be predicted exactly from the mean value of the transformed trait x.

The differences in x or y among the three QTL-genotypic means reflect the magnitude of the genetic effects of the QTL on the corresponding trait. Denote a and d as the additive and dominant effects of the QTL on x. Thus, the phenotypic values of the two log-transformed traits for individual i are expressed by linear statistical models,

$$x_i = \mu + \xi_i a + \zeta_i d + e_i^x,$$
$$y_i = \alpha + \beta x_i + e_i^y = \alpha + \beta(\mu + \xi_i a + \zeta_i d) + \beta e_i^x + e_i^y,$$

where μ is the overall mean for x; ξ_i and ζ_i are the dummy variables indicating the QTL genotype of individual i, with ξ_i denoted as 1 for QQ, 0 for Qq and -1 for qq and ζ_i denoted as 1 for Qq and 0 for QQ or qq; $e_i^x \sim N(0, \sigma_x^2)$ and $e_i^y \sim N(0, \sigma_y^2)$ are the error terms for traits x and y, respectively, which are correlated among individuals with correlation coefficients R. Because the genetic effects are fixed effects, the variances for traits x and y and their correlation can be expressed, respectively, as

$$v_x^2 = \sigma_x^2,$$
$$v_y^2 = \beta^2\sigma_x^2 + \sigma_y^2,$$
$$\rho = \frac{\beta\sigma_x + R\sigma_y}{\sqrt{\beta^2\sigma_x^2 + \sigma_y^2}}.$$

Let μ_j^x and μ_j^y be the mean values of x and y for QTL genotypes j ($j = 2, 1, 0$), respectively. According to equation (2), the relationship between the means of the two transformed traits can be modelled by

$$\mu_j^y = \alpha + \beta\mu_j^x, \tag{3}$$

for QTL genotype j.

Suppose that the QTL is bracketed by two flanking markers M_1 and M_2 with recombination frequency of r. The recombination frequencies between M_1 and the QTL and between the QTL and M_2 are r_1 and r_2, respectively. The QTL position can be specified using r_1 or r_2. The conditional probability of a QTL genotype given each of the nine two-marker genotypes can be derived and expressed as a function of r, r_1 and r_2. We use $\pi_{j|i}$ to denote such a conditional probability for individual i to carry QTL genotype j.

A conventional allometry mapping model, advocated by Wu *et al.* (2002), estimates allometry coefficients, QTL effects, QTL position and residual (co)variance, arrayed by a unknown vector $\mathbf{\Omega} = (\mu, a, d, \alpha, \beta, r_1, \sigma_x^2, \sigma_y^2, R)^T$. The likelihood of the sample of bi-variate measurements can be expressed by a mixture model as

$$L(\mathbf{\Omega}|x, y) = \prod_{i=1}^{n}\sum_{j=0}^{2}\left[\pi_{j|i} f_j(x_i, y_i)\right] \tag{4}$$

where the two-dimensional normal density, $f_j(x_i, y_i)$, is expressed as

$$f_j(x_i, y_i) = \frac{1}{2\pi v_x v_y\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\right.$$
$$\left.\times\left[\left(\frac{x_i-\mu_j^x}{v_x}\right)^2 - 2\rho\frac{(x_i-\mu_j^x)(y_i-\mu_j^y)}{v_x v_y} + \left(\frac{y_i-\mu_j^y}{v_y}\right)^2\right]\right\}.$$

The maximum likelihood estimators (MLEs) of the unknown vector $\mathbf{\Omega}$ can be obtained by differentiating the log-likelihood function (equation (4)) with respect to each unknown, setting the derivatives equal to zero and solving the log-likelihood equation. By defining

$$\prod_{j|i} = \frac{\pi_{j|i} f_j(x_i, y_i)}{\sum_{j'=0}^{2}\left[\pi_{ij'} f_{j'}(x_i, y_i)\right]}, \tag{5}$$

which could be thought of as a posterior probability that individual i has QTL genotype j, the EM algorithm is implemented to solve the likelihood function. The posterior probabilities, $\prod_{j|i}$, are calculated for each individual and each QTL genotype in the E step and they are then used to obtain the MLEs of each parameter contained in $\mathbf{\Omega}$ in the M step which are derived from the log-likelihood equations. For detailed iterative EM steps, refer to Wu *et al.* (2002).

(ii) *Linked QTL model*

The allometric scaling of organisms may also be affected by two QTLs that are genetically linked on the same chromosome, one exerting an effect on trait x and the other on trait y. Under such a linked-QTL model, two putative QTLs may be located within the same marker interval or in different marker intervals. We consider a special linkage model in which each of the two linked QTL affects a different trait x and y.

https://doi.org/10.1017/S0016672306008172 Published online by Cambridge University Press

Consider two linked QTLs of which one ($\mathcal{P}$) affects trait x and the other ($\mathcal{Q}$) affects trait y. Let j_1 denote a QTL genotype PP, Pp and pp, coded as 2, 1 and 0, respectively. Similarly, we use j_2 to denote QTL genotypes QQ, Qq and qq, coded as 2, 1 and 0. If the two QTLs are located in different marker intervals, two different pairs of flanking markers should be used simultaneously to specify the likelihood of the data. Denote by $\mathbf{\Theta_1}$ the matrix for the conditional probability of QTL genotypes at $\mathcal{P}$ given the marker interval M_1–M_2. Similarly, the matrix for the conditional probability of QTL genotypes at $\mathcal{Q}$ given the marker interval N_1–N_2 is denoted by $\mathbf{\Theta_2}$. The conditional probability of joint QTL genotypes at $\mathcal{P}$ and $\mathcal{Q}$ given the two marker intervals can be expressed as

$$\mathbf{\Theta_1} \otimes \mathbf{\Theta_2},$$

where $\otimes$ is the Kronecker product. If two linked QTLs are located within the same marker interval, the conditional probability of joint QTL genotype given the marker genotypes is derived.

Under our linkage model, the additive and dominant effects of QTL $\mathcal{P}$ on x are denoted by a and d, whereas the additive and dominant effects of QTL $\mathcal{Q}$ and y are denoted as $\alpha+\beta a$ and $\alpha+\beta d$, respectively. The likelihood of the data under the linked-QTL model can be expressed as

$$L(\mathbf{\Omega})|x,y) = \prod_{i=1}^{n} \sum_{j_1=0}^{2} \sum_{j_2=0}^{2} \left[\pi_{j_1 j_2|i} f_{j_1 j_2}(x_i, y_i)\right] \tag{6}$$

where the vector $\mathbf{\Omega}$ contains the same unknowns as in the pleiotropic QTL model, except with one more parameters describing the position of the second QTL, $\pi_{j_1 j_2|i}$ is the conditional probability of the joint QTL genotype $j_1 j_2$ given a specific individual i that carries a known marker genotype and $f_{j_1 j_2}(x_i, y_i)$ is the joint normal distribution of x and y for a joint QTL genotype. Similarly, the EM algorithm can be used to estimate the unknowns under the linkage model.

(iii) *Precision analysis*

After the point estimates of parameters have been obtained by the EM algorithm, it is necessary to derive the variance–covariance matrix and evaluate the standard errors of the estimates. Because the EM algorithm does not automatically provide the estimates of the asymptotic variance–covariance matrix for parameters, an additional procedure has been developed to estimate this matrix (and thereby standard errors) when the EM algorithm is used (Louis, 1982; Meng & Rubin, 1991). Meng & Rubin (1991) proposed a so-called supplemented EM algorithm or SEM algorithm to estimate the asymptotic variance–covariance matrices. However, in this study, Louis' (1982) approach is used to calculate the standard errors for the MLEs of QTL parameters for allometry mapping.

(iv) *Hypothesis tests*

Several hypotheses about the QTL affecting quarter-power scalings of organisms can be formulated for our model. These hypotheses include: (i) there is a QTL in a linkage group that affects two traits, (ii) this significant QTL is pleiotropic with effect on both traits, or it actually presents two linked QTLs (one affecting each trait), and (iii) under a best-fitting model, the QTL detected affects the two traits in a quarter-power scaling. In addition, we can ask whether the normalization constant α is a characteristic of species or populations (Niklas, 1994). This can also be tested to find out whether the mapping population used conforms to a general scaling pattern. However, a recent survey by Niklas & Enquist (2001) suggested that all plants have similar allometric exponents and normalization constants and, therefore, comply with a single allometric formula.

The evidence for QTL on traits x and y can be tested by hypothesizing a single QTL versus no QTL on a linkage group, i.e.

$$\begin{cases} H_0\text{: } a=d=0 \\ H_1\text{: At least one of these equalities above does not hold.} \end{cases} \tag{7}$$

The H_0 states that no QTL affects trait x (the reduced model), whereas the H_1 proposes that there is such a QTL (the full model). The test statistic for testing the hypotheses (7) is calculated as the log-likelihood ratio of the reduced to the full model:

$$\mathrm{LR} = -2[\ln L(\widetilde{\mathbf{\Omega}}|x,y) - \ln L(\widehat{\mathbf{\Omega}}|x,y)], \tag{8}$$

where $\widetilde{\mathbf{\Omega}}$ and $\widehat{\mathbf{\Omega}}$ denote the MLEs of the unknown parameters under H_0 and H_1, respectively. Under the null hypothesis, the LR is asymptotically χ^2-distributed with two degrees of freedom. An empirical approach for determining the critical threshold is based on permutation tests. By repeatedly shuffling the relationships between marker genotypes and phenotypes, a series of maximum log-likelihood ratios are calculated, from the distribution of which the critical threshold is determined.

To test whether this detected QTL affects allometric scaling laws, we need to perform two additional tests for the significance of the exponential power β. The first is to test $\beta=0$ versus $\beta\neq 0$, which is associated with the pleiotropic effect of this QTL on trait y. The second is to test $\beta=k$ versus $\beta\neq k$, where k is a multiplier of a quarter such as 1/4 or 3/4. Only after

https://doi.org/10.1017/S0016672306008172 Published online by Cambridge University Press

both the null hypotheses above about x and y are rejected, is the underlying QTL suggested to be pleiotropic for allometric scaling laws.

If a marker interval is detected to carry two QTLs each affecting a different trait, it is important to test whether the correlation between the two traits is due to the pleiotropic effect of the same QTL or the linkage between the two QTLs. Let two QTLs have positions symbolized by $p(1)$ for the QTL associated with trait x and $p(2)$ for the QTL associated with trait y. Whether or not these two QTLs are actually the same can be tested by formulating the hypotheses $p(1)=p(2)$ versus $p(1)\neq p(2)$. If the null hypothesis is accepted, this means that the pleiotropic effect of QTLs has a more important contribution to trait correlation than the linkage. But its rejection, i.e. the existence of the two QTLs, may result from two possibilities: (1) each QTL affects a different trait and (2) each QTL affects two traits simultaneously. These two possibilities can be further tested by formulating two alternative hypotheses. If the first possibility is true, the linkage is more important than the pleiotropy in affecting the trait correlation. If the second possibility is true, both the pleiotropy and linkage are important.

(v) *An improved mapping model*

In the previous sections, we described a conventional approach for allometry mapping. In this section, an improved model is proposed to increase the computational efficiency of allometry mapping. This improved model is based on a two-step estimation process. In step 1, the power parameters that govern allometric scaling laws are estimated by a regression model. In step 2, by substituting the estimated power parameters in the mapping model constructed by a mixture model, the effects and position of the underlying QTL are estimated using the EM algorithm. As shown in equations (1) and (2), the log-transformation makes two allometrically related traits linearly related and, at this time, the power parameters can be estimated, using a least squares approach, as

$$\breve{\beta}=\frac{n\sum_{i=1}^{n}x_iy_i-\left(\sum_{i=1}^{n}x_i\right)\left(\sum_{i=1}^{n}y_i\right)}{n\sum_{i=1}^{n}x_i^2-\left(\sum_{i=1}^{n}x_i\right)^2},$$

$$\breve{\alpha}=\frac{1}{n}\left(\sum_{i=1}^{n}y_i-\beta\sum_{i=1}^{n}x_i\right).$$

By viewing $\breve{\alpha}$ and $\breve{\beta}$ as the constants, we define a new variable

$$z_i=y_i-\breve{\beta}x_i-\breve{\alpha}, \tag{9}$$

which is normally distributed as $N(0,\ \sigma_z^2)$. It can be seen that x_i and z_i are independent. The joint distribution of x_i and z_i for the three different QTL genotypes can be written as

$$f_{\mathbf{2}}(x_i,z_i)=f_{\mathbf{2}}(x_i)*f(z_i)=\frac{1}{2\pi\sigma_x\sigma_z}\times\exp\left\{-\frac{1}{2}\left[\frac{(x_i-\mu-a)^2}{\sigma_x^2}+\frac{z_i^2}{\sigma_z^2}\right]\right\},$$

$$f_{\mathbf{1}}(x_i,z_i)=f_{\mathbf{1}}(x_i)*f(z_i)=\frac{1}{2\pi\sigma_x\sigma_z}\times\exp\left\{-\frac{1}{2}\left[\frac{(x_i-\mu-d)^2}{\sigma_x^2}+\frac{z_i^2}{\sigma_z^2}\right]\right\},$$

$$f_{\mathbf{0}}(x_i,z_i)=f_{\mathbf{0}}(x_i)*f(z_i)=\frac{1}{2\pi\sigma_x\sigma_z}\times\exp\left\{-\frac{1}{2}\left[\frac{(x_i-\mu+a)^2}{\sigma_x^2}+\frac{z_i^2}{\sigma_z^2}\right]\right\}.$$

The likelihood of the unknown parameters given the observed trait x and the newly defined variable z can be written under the pleiotropic (equation (4)) or linked QTL model (equation (6)). The maximization of this likelihood with respect to the unknown parameters leads to the following log-likelihood equations for the pleiotropic model, expressed as a function of the posterior probabilities (equation (5)):

$$\mu=\frac{1}{2}\left[\frac{\sum_{i=1}^{n}x_i\prod_{\mathbf{0}|i}}{\sum_{i=1}^{n}\prod_{\mathbf{0}|i}}+\frac{\sum_{i=1}^{n}x_i\prod_{\mathbf{2}|i}}{\sum_{i=1}^{n}\prod_{\mathbf{2}|i}}\right]$$

$$a=\frac{\sum_{i=1}^{n}x_i\prod_{\mathbf{1}|i}}{\sum_{i=1}^{n}\prod_{\mathbf{1}|i}}-\frac{1}{2}\left[\frac{\sum_{i=1}^{n}x_i\prod_{\mathbf{0}|i}}{\sum_{i=1}^{n}\prod_{\mathbf{0}|i}}+\frac{\sum_{i=1}^{n}x_i\prod_{\mathbf{2}|i}}{\sum_{i=1}^{n}\prod_{\mathbf{2}|i}}\right]$$

$$d=\frac{1}{2}\left[\frac{\sum_{i=1}^{n}x_i\prod_{\mathbf{0}|i}}{\sum_{i=1}^{n}\prod_{\mathbf{0}|i}}-\frac{\sum_{i=1}^{n}x_i\prod_{\mathbf{2}|i}}{\sum_{i=1}^{n}\prod_{\mathbf{2}|i}}\right]$$

$$\sigma_x^2=\frac{1}{n}\sum_{i=1}^{n}\left[\prod_{\mathbf{2}|i}(x_i-\mu-a)^2+\prod_{\mathbf{1}|i}(x_i-\mu-d)^2+\prod_{\mathbf{0}|i}(x_i+\mu+a)^2\right]$$

$$\sigma_z^2=\frac{1}{n}\sum_{i=1}^{n}(y_i-\breve{\alpha}-\breve{\beta}x_i)^2.$$

The EM algorithm is implemented to estimate these parameters. Compared with the traditional model, this improved model estimates fewer parameters and, thus, is expected to provide more power to detect allometric QTLs.

3. Non-linear model for allometry mapping

Although the simple allometry equation (1) has been used to model allometric scaling relationships, it is limited for the precise description of many important biological phenomena. For example, this equation, which forces two variables to pass through the origin, i.e. when $x=0$ then $y=0$, cannot describe the relationship between two developmentally asynchronous features, such as reproductive timing and body weight.

To accurately describe the scaling relationship between any biological traits, we need to extend the

https://doi.org/10.1017/S0016672306008172 Published online by Cambridge University Press

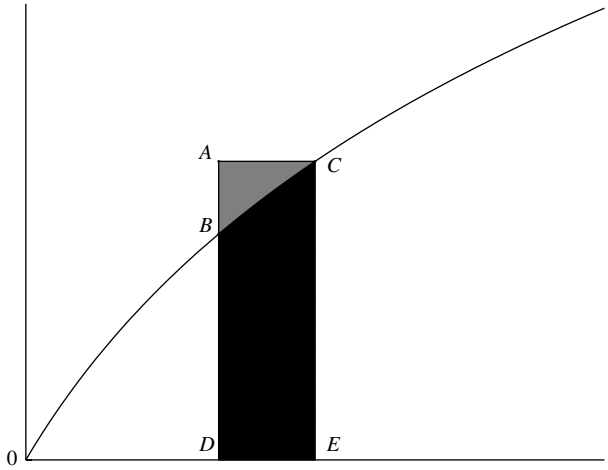

Fig. 2. The diagram for calculating the area under a curve using the loss function approach. Adapted from Ebert & Russell (1994).

simple allometry equation to accommodate general biological phenomena. We propose a two-step procedure to estimate the position and effects for a QTL that affects more general allometric scaling laws.

(i) *Step 1: Modelling general allometry equations*

A number of mathematical equations have been proposed by earlier biologists to describe general allometric scaling relationships. Among them, three representative allometry equations are:

$$y=\begin{cases}\alpha x^{\beta}+\gamma & \text{(Robb, 1929)}\\ \alpha(x-\gamma)^{\beta} & \text{(Reeve \& Huxley, 1945)}\\ \alpha(x-\gamma)^{\beta}+\delta & \text{(Lumer, 1937),}\end{cases} \quad (10)$$

where Lumer's four-parameter equation can be viewed as the most general. Ebert & Russell (1994) introduced a model II non-linear regression analysis to estimate the parameters contained in the allometry equations (10). Unlike model I regression that minimizes the squared distance between the coordinate of a data pair on the y-axis and the function, model II regression minimizes the area connecting the coordinates of the data pair and the function and thus deals with variation in both traits. Model II regression analysis assumes an equal error variance for both traits and is a special case of more general error-in-variables models (Seber & Wild, 1989).

Consider an individual i from a mapping population of size N. The observations of this individual for traits x_i and y_i present a point (A) with the coordinate (x_i, y_i) as shown in Fig. 2. To detect an allometry curve that has a minimum deviation to this coordinate, we define four more points B, C, D and E, whose coordinates are expressed as

$$B=\left[x_i, \alpha(x_i-\delta)^{\beta}+\gamma\right]$$
$$C=(z_i, y_i)$$
$$D=(x_i, 0)$$
$$E=(z_i, 0),$$

where

$$z_i=\begin{cases}\left(\frac{y_i-\gamma}{\alpha}\right)^{\frac{1}{\beta}} & \text{Robb equation}\\ \left(\frac{y_i}{\alpha}\right)^{\frac{1}{\beta}}+\gamma & \text{Reeve–Huxley equation}\\ \left(\frac{y_i-\gamma}{\alpha}\right)^{\frac{1}{\beta}}+\delta & \text{Lumer equation.}\end{cases}$$

The area of the rectangle $ADEC$ is

$$\mathbf{A}_i(ADEC)=y_i|z_i-x_i|.$$

The area under the curve, $BDEC$, is

$$\mathbf{A}_i(BDEC)$$
$$=\begin{cases}\int_{x_i}^{z_i}(\alpha x_i^{\beta}+\delta)dx_i & \text{Robb equation}\\ \int_{x_i}^{z_i}\alpha(x_i-\gamma)^{\beta}dx_i & \text{Reeve–Huxley equation}\\ \int_{x_i}^{z_i}\left[\alpha(x_i-\gamma)^{\beta}+\delta\right]dx_i & \text{Lumer equation}\end{cases}$$
$$=\begin{cases}\frac{\alpha}{\beta+1}\left|x_i^{\delta+1}-\left(\frac{y_i-\delta}{\alpha}\right)^{\frac{\beta+1}{\beta}}\right|+\delta\left|x_i-\left(\frac{y_i-\delta}{\alpha}\right)^{\frac{1}{\beta}}\right| & \text{Robb equation}\\ \frac{\alpha}{\beta+1}\left|(x_i-\gamma)^{\beta+1}-\left(\frac{y_i}{\alpha}\right)^{\frac{\beta+1}{\beta}}\right| & \text{Reeve–Huxley equation}\\ \frac{\alpha}{\beta+1}\left|(x_i-\gamma)^{\beta+1}-\left(\frac{y_i-\delta}{\alpha}\right)^{\frac{\beta+1}{\beta}}-\gamma\right|+\delta\left|x_i-\left(\frac{y_i-\delta}{\alpha}\right)^{\frac{\beta+1}{\beta}}-\gamma\right| & \text{Lumer equation.}\end{cases} \quad (11)$$

https://doi.org/10.1017/S0016672306008172 Published online by Cambridge University Press

We define the absolute value of the difference between $\mathbf{A}_i$ $(ADEC)$ and $\mathbf{A}_i$ $(BDEC)$, i.e. area

$$\mathbf{A}_i(ABC) = \mathbf{A}_i(ADEC) - \mathbf{A}_i(BDEC),$$

as the loss function for individual i. The loss function for all individuals is expressed as

$$\mathbf{A}(ABC) = \sum_{i=1}^{n} \mathbf{A}_i(ABC). \tag{12}$$

The estimates of α, β, γ and δ fitting the allometry equation can be obtained by minimizing the loss function defined by areas (12). It is impossible to derive their analytical solutions for this non-linear function. However, their numerical solutions can be obtained by using the simplex algorithm (Nelder & Mead, 1965). The advantage of the simplex algorithm is that it is derivative-free and easy to implement with current software, such as Matlab. A reason for caution with this algorithm is the possibility of obtaining local optimal solutions for the loss function (12). By carefully selecting the initial values, however, this problem can be the minimum if there exist global optimal solutions. If no global optimal solutions exist, we can take the minimum in the space of these parameters.

For a practical data set, it is essential to determine the best allometry equation that can describe the allometric scaling relationship. The criterion for the determination can be based on the values of loss function summed over all individuals. For the same value of the loss function, an allometry equation with fewer parameters is better than those with more parameters.

(ii) *Step 2: The mapping process*

As pointed out above, the parameters, α, β, γ and δ, for more general allometry equations are regarded as universal parameters and they should be the same among different QTL genotypes. After these parameters have been estimated from model II non-linear regression by minimizing the loss function (12), we substitute these estimates, indexed by $\breve{\alpha}$, $\breve{\beta}$, $\breve{\gamma}$, and $\breve{\delta}$, to the likelihood functions described by equation (4) or (6). Before doing so, we take a simple change of more allometry equations for individual i:

$$\begin{cases} \ln(y_i - \breve{\gamma}) = \breve{\alpha} + \breve{\beta}\ln(x_i) & \text{Robb equation} \\ \ln(y_i) = \breve{\alpha} + \breve{\beta}\ln(x_i - \breve{\gamma}) & \text{Reeve–Huxley equation} \\ \ln(y_i - \breve{\delta}) = \breve{\alpha} + \breve{\beta}\ln(x_i - \breve{\gamma}) & \text{Lumer equation} \end{cases}$$

or

$$\begin{cases} \ln(y'_i) = \breve{\alpha} + \breve{\beta}\ln(x_i) & \text{Robb equation} \\ \ln(y_i) = \breve{\alpha} + \breve{\beta}\ln(x_i') & \text{Reeve–Huxley equation} \\ \ln(y'_i) = \breve{\alpha} + \breve{\beta}\ln(x_i') & \text{Lumer equation} \end{cases} \tag{13}$$

where ' ' denotes the transformation of the two traits. Because $\breve{\alpha}$, $\breve{\beta}$, $\breve{\gamma}$, and $\breve{\delta}$ can be regarded as known constants obtained from model II regression analysis, x_i' and y_i' can be calculated and display the same statistical distribution as raw data x_i and y_i. Thus, the new relationships described by equation (13) are identical to a linear log-transformed allometry equation (2). By using the improved model as described above, we estimate the remaining parameters $\mathbf{\Omega}_R = (\mu, a, d, \sigma_x^2, \sigma_z^2, r_1)^T$. The existence of the underlying QTL for more general allometry equations can be tested by formulating the hypotheses $\mathbf{H_0}$: $a = d = 0$ versus $\mathbf{H_1}$: at least one equality does not hold.

4. Results

The model proposed here is used to map age-dependent QTLs in a model system: the mouse. Cheverud *et al.* (1996) constructed a linkage map composed of 19 chromosomes based on 75 microsatellite markers in 535 $\mathbf{F_2}$ progeny population derived from two strains, Large and Small. The $\mathbf{F_2}$ hybrids were weighted at 10 weekly intervals starting at age 7 days. The raw weights were corrected for the effects of each covariate due to dam, litter size at birth, parity and sex. The growth rate at each time interval $[t, t+1]$ was calculated for each mouse my subtracting body weight at time t from body weight from time $t+1$. The mean growth rate across all the time intervals was then calculated for each mouse. Since we did not observe a marked trend that the variance increases with the mean, the model II non-linear regression is used to estimate the parameters for the equation that specifies the allometric relationship between growth rate and body weight (Niklas, 1994).

The four allometry equations (1) and (10) were used to fit the relationship between the growth rate and body weight. By comparing the values of loss function among these equations, Robb's equation was found to be the most parsimonious. Based on Robb's equation, we estimated the three parameters underlying the allometry equation $\tilde{\alpha} = 0{\cdot}024$, $\tilde{\beta} = 0{\cdot}68$ and $\tilde{\gamma} = -0{\cdot}016$. As shown by their sampling errors, these parameters estimates have reasonable precision (Table 1). It is interesting to find that the estimated β value is not significantly different from 0·75, which supports the three-quarter law for a biological process to scale with body weight (West *et al.*, 1997).

These estimated parameters from model II regression by minimizing the loss function (12) were substituted in the mapping model build by a log-transformed linear regression model as shown by equation (13). Through such a substitution, we only need to estimate the remaining parameters including the QTL position, QTL effects for body weight, and the residual (co)variance matrix between growth rate and body weight. We scan all the 19 chromosomes for

https://doi.org/10.1017/S0016672306008172 Published online by Cambridge University Press

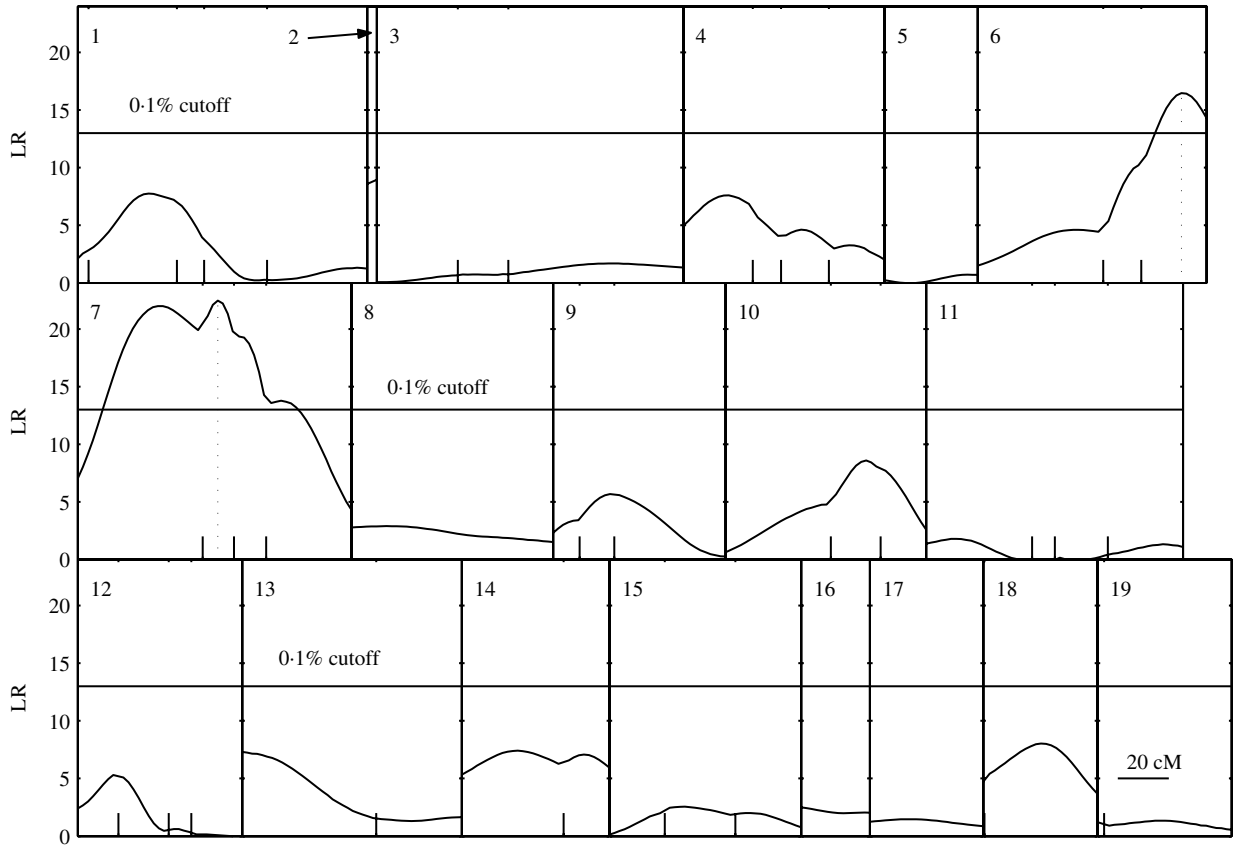

Fig. 3. The profiles of the log-likelihood ratios (LRs) between the full and reduced (no QTL) model for allometric scaling of mean growth rate to body weight across the entire genome using the linkage map constructed from microsatellite markers (Cheverud *et al*., 1996). The genomic positions corresponding to the peaks of the curves are the maximum likelihood estimators of the QTL positions. The genome-wide threshold values for claiming the existence of a QTL are shown as the horizontal lines. Tick marks on the *x*-axis represent the positions of markers on the linkage group, the names of which are given by Cheverud *et al*. (1996).

Table 1. *Estimates of genetic effects at two detected QTLs on chromosome 6 and 7 and the residual* (*co*) *variance matrix between two allometrically related traits x and y in the* F_2 *mouse progeny*

Parameter	Chromosome 6	Chromosome 7
$\hat{\mu}$	3·91	3·90
$\hat{a}$	0·35	0·25
$\hat{d}$	0·19	0·26
$\hat{\sigma}_x^2$	0·006	0·006
$\hat{\sigma}_y^2$	0·007	0·007
$\hat{R}$	0·62	0·62
$\hat{\alpha}$	0·024	
$\hat{\beta}$	0·68	
$\hat{\gamma}$	−0·02	

The maximum likelihood estimates (MLE) of parameters are symbolized by hats, and the estimates by model II regression are symbolized by breves.

the existence of QTLs affecting the allometric scaling relationship between growth rate and body weight. Fig. 3 gives the profile of the log-likelihood ratio (LR) test statistics for claiming the existence of QTLs across the entire mouse genome. There are two peaks for the LR profile: one (16·47) between markers *D6Nds5* and *D6Mit15* on chromosome 6 and the other (22·46) between *D7Nds1* and *D7Mit17* on chromosome 7. These two LR values are well beyond the genome-wide critical threshold (13·1) at the 0·001 significance level determined on the basis of permutation tests. These tests suggest the existence of two QTLs that are located at the positions corresponding to the peaks of the profile.

The genetic effects of these two QTLs on body weight and other model parameters were estimated in Table 1. In general, the estimates of these parameters in this mouse example from our model exhibit good precision. The QTL detected on chromosome 6 affects body weight in a partially dominant manner, whereas the QTL on chromosome 7 displays a strong dominant or overdominant effect on body weight. Our detection is broadly consistent with simple interval mapping analysis of the same material by Cheverud *et al*. (1996). Of 16 chromosomes observed to carry the QTLs for body weight, the QTLs on chromosomes 6

https://doi.org/10.1017/S0016672306008172 Published online by Cambridge University Press

and 7 were detected with larger LOD scores, explaining larger percentages of the observed variation, than those on the other chromosomes.

5. Discussion

Even though understanding the regulation of allometry would have broad implications for furthering our knowledge of developmental ontogeny, regeneration, population growth and evolutionary processes (West *et al*., 1997, 1999*a*, *b*), it is unfortunate that no general genetic model exists to mechanistically explain scaling laws. R. Wu and co-workers have, for the first time, incorporated allometric rules in a QTL mapping framework (Wu *et al*., 2002; Ma *et al*., 2003). Their models were validated by a real example from forest trees in which a QTL was detected to govern the allometric relationship of third-year stem height with third-year stem biomass. The result suggested that the QTL detected from the model is one specifically responsible for the allometric scaling. In this article, we proposed an improved model for mapping specific QTLs that are responsible for allometric scaling laws based on more general allometry equations.

Compared with previous models (Wu *et al*., 2002; Ma *et al*., 2003), our model is advantageous in several respects. First, it significantly reduces the number of parameters to be estimated, thus increasing computational efficiency. The derivation of our model is based on the fundamental principle for allometric scaling with which a particular biological variable (B) scales as one-quarter or three-quarters of body weight (M) across an incredible range of species (Fig. 1*A*). Reduced from interspecific to intraspecific allometric scaling, these quarter-power laws can be seen across different genotypes (Fig. 1*B*). If there is a particular QTL that affects allometric scaling laws, the means of different QTL genotypes should have significantly different coordinates for traits B and M but they should be located on a common straight line. As a result of this, we can first estimate the allometry parameters based on all genotypes using a conventional statistical method, such as least squares analysis, and then substitute these allometry parameters in a maximum-likelihood-based QTL mapping framework built by a mixture model. With the estimates of allometry parameters, we further make a simple transformation to remove the covariance between the response variable and body weight. A reduced number of the parameters about the QTL effects, QTL position and residual variances between the two traits are estimated by implementing the EM algorithm.

Second, our model take advantage of the log-linear property of the power equation as used in Wu *et al*. (2002). For more general allometry equations (10), this property is not applicable, in which Taylor series of different orders were used to approximate the genotypic means of the response variable based on body weight (Ma *et al*., 2003). Whereas a lower-order approximation may lead to system errors, a higher-order approximation demands substantial computational load. The model proposed in this article does not rely on Taylor's approximation. Third, our model divides the whole estimation procedure into two steps and, therefore, can be readily extended to any complicated allometry equations involving multivariate variables.

Our model has been tested by an example of a mouse F_2 progeny. Using this model, we detected two QTLs that govern allometric scaling laws between growth rate and body weight. These two QTLs detected on chromosomes 6 and 7 were in agreement with the results from interval mapping of single growth traits (Cheverud *et al*., 1996). It has been shown that these two chromosomes are more likely to carry more significant QTLs than other chromosomes.

There are several areas in which our model can be modified. First, for model II non-linear regression used to estimate the allometry parameters (Ebert & Russell, 1994), we assume that error variances for traits x and y are equal. Although this may be reasonable in certain situations, incorporation of specified error variances of x and y is necessary for general allometry issues. Second, to better characterize the genetic architecture of allometric scaling laws, we should include modelling and analysis of epistatically interacting QTLs. Growing evidence has been observed for the role of epistasis in organ development (Cheverud, 2000; Wolf *et al*., 2000). Third, for simplicity, the example used to test our model deals with the allometric scaling between mean growth rate and body weight. However, growth rate is an age-dependent trait. The integration of age-specific allometric relationships in the QTL mapping framework will provide great insights into the genetic mechanisms for the developmental control of allometric scaling laws.

We thank three anonymous referees for their constructive comments on this manuscript. This work is supported by NIH grant DK055736 to J.M.C. and a grant from the National Science Foundation of China (09 95671) and NSF grant (0540745) to R.W.

References

Andresen, B., Shiner, J. S. & Uehlinger, D. E. (2002). Allometric scaling and maximum efficiency in physiological eigen time. *Proceedings of the National Academy of Sciences of the USA* **99**, 5822–5824.

Banavar, J. R., Maritan, A. & Rinaldo, A. (1999). Size and form in efficient transportation networks. *Nature* **399**, 130–132.

Bonner, J. T. & Horn, H. S. (2000). Allometry and natural selection. In *Scaling in Biology* (ed. J. H. Brown & G. B.

West), pp. 25–35. Oxford: Santa Fe Institute and Oxford University Press.
Brown, J. H. & West, G. B. (eds.) (2000). *Scaling in Biology*. Oxford: Oxford University Press.
Calder III, W. A. (1984). *Size, Function and Life History*. Cambridge, MA: Harvard University Press.
Cheverud, J. M. (2000). Detecting epistasis among quantitative trait loci. In *Epistasis and the Evolutionary Process* (ed. J. B. Wolf, E. D. Brodie III & M. J. Wade), pp. 58–71. New York: Oxford University Press.
Cheverud, J. M., Routman, E. J., Duarte, F. A. M., van Swinderen, B., Cothran, K. & Perel, C. (1996). Quantitative trait loci for murine growth. *Genetics* **142**, 1305–1319.
Churchill, G. A. & Doerge, R. W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics* **138**, 963–971.
Dewey, T. G. & Donne, M. D. (1998). Non-equilibrium thermodynamics of molecular evolution. *Journal of Theoretical Biology* **193**, 593–599.
Ebert, T. A. & Russell, M. P. (1994). Allometry and Model II nonlinear regression. *Journal of Theoretical Biology* **168**, 367–372.
Enquist, B. J., Brown, J. H. & West, G. B. (1998). Allometric scaling of plant energetics and population density. *Nature* **395**, 163–165.
Enquist, B. J., West, G. B., Charnov, E. L. & Brown, J. H. (1999). Allometric scaling of production and life-history variation in vascular plants. *Nature* **401**, 907–911.
Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society Series B* **44**, 226–233.
Lumer, H. (1937). The consequences of sigmoid growth for relative growth functions. *Growth* **1**, 140–154.
Ma, C. X., Casella, G., Littell, R. C., Khuri, A. I. & Wu, R. L. (2003). Exponential mapping of quantitative trait loci governing allometric relationships in organisms. *Journal of Mathematical Biology* **47**, 313–324.
McMahon, T. (1973). Size and shape in biology. *Science* **179**, 1201–1204.
Meng, X. & Rubin, D. (1992). Using EM to obtain asymptotic variance-covariance matrices: The SEM algorithm. *Journal of the American Statistical Association* **86**, 899–909.
Nelder, J. A. & Mead, R. (1965). A simplex method for function minimization. *Computer Journal* **7**, 308–313.
Niklas, K. L. (1994). *Plant Allometry: The Scaling of Form and Process*. Chicago, IL: University of Chicago.
Niklas, K. J. & Enquist, B. J. (2001). Invariant scaling relationships for interspecific plant biomass production rates and body size. *Proceedings of the National Academy of Sciences of the USA* **98**, 2922–2927.
Reeve, E. C. R. & Huxley, J. S. (1945). Some problems in the study of allometric growth. In *Essays on 'Growth and Form' presented to D'Arcy Wentworth Thompson* (ed. W. E. Le Gros Clark & P. B. Medawar), pp. 121–156. Oxford: Clarendon Press.
Robb, R. C. (1929). On the nature of heredity size-limitation. II. The growth of parts in relation to the whole. *British Journal of Experimental Biology* **6**, 311–324.
Schmidt-Nielsen, K. (1984). *Why Is Animal Size So Important*? Cambridge: Cambridge University Press.
Seber, G. A. F. & Wild, C. J. (1989). *Nonlinear Regression*. New York: Wiley.
West, G. B., Brown, J. H. & Enquist, B. J. (1997). A general model for the origin of allometric scaling laws in biology. *Science* **276**, 122–126.
West, G. B., Brown, J. H. & Enquist, B. J. (1999*a*). The fourth dimension of life: fractal geometry and allometric scaling of organisms. *Science* **284**, 1677–1679.
West, G. B., Brown, J. H. & Enquist, B. J. (1999*b*). A general model for the structure and allometry of plant vascular systems. *Nature* **400**, 664–667.
Wolf, J. B., Brodie III, E. D. & Wade, M. J. (2000). *Epistasis and the Evolutionary Process*. Oxford: Oxford University Press.
Wu, R. L., Ma, C. X., Littel, R. C. & Casella, G. (2002). A statistical model for the genetic origin of allometric scaling laws in biology. *Journal of Theoretical Biology* **219**, 121–135.
Wu, R. L., Ma, C. X., Lou, X. Y. & Casella, G. (2003). Molecular dissection of allometry, ontogeny, and plasticity: a genomic view of developmental biology. *BioScience* **53**, 1041–1047.