THEORETICAL BIOLOGY AND
MEDICAL MODELLING

RESEARCH   **Open Access**

A statistical model for mapping morphological shape

Guifang Fu[1,2], Arthur Berg[2], Kiranmoy Das[1,2], Jiahan Li[1,2], Runze Li[1,2], Rongling Wu[3,2,1*]

* Correspondence: rwu@hes.hmc.
psu.edu
[3]Center for Computational Biology,
Beijing Forestry University, Beijing
100083, China

Abstract

Background: Living things come in all shapes and sizes, from bacteria, plants, and animals to humans. Knowledge about the genetic mechanisms for biological shape has far-reaching implications for a range spectrum of scientific disciplines including anthropology, agriculture, developmental biology, evolution and biomedicine.

Results: We derived a statistical model for mapping specific genes or quantitative trait loci (QTLs) that control morphological shape. The model was formulated within the mixture framework, in which different types of shape are thought to result from genotypic discrepancies at a QTL. The EM algorithm was implemented to estimate QTL genotype-specific shapes based on a shape correspondence analysis. Computer simulation was used to investigate the statistical property of the model.

Conclusion: By identifying specific QTLs for morphological shape, the model developed will help to ask, disseminate and address many major integrative biological and genetic questions and challenges in the genetic control of biological shape and function.

Background

Morphological shape is one of the most conspicuous aspects of an organism's phenotype and provides an intricate link between biological structure and function in changing environments [1,2]. For this reason, comparing the anatomical and shape feature of organisms has been a central element of biology for centuries. Nowadays, attempts have been made to unlock the genetic secrets behind phenotypic differentiation in developmental shape [3], understand the origin and pattern of shape variation from a developmental perspective [4,5], and predict the adaptation of morphological shapes in a range of environmental conditions [6].

Three major advances in life and physical science during the last decades will make it possible to study shape variation and its biological underpinnings. First, DNA-based molecular markers allow the identification of quantitative trait loci (QTLs) and biochemical pathways that contribute to quantitatively inherited traits such as shape. In his seminal review, Tanksley [3] summarized some major discoveries of genes for fruit size and shape in tomato. In a long process of domestication, tremendous shape variation has occurred in tomato fruit from almost invariably round (wild or semiwild types) to round, oblate, pear-shaped, torpedo-shaped, and bell pepper-shaped (cultivated types). Some of the QTLs that cause these differences, namely *fw2.2*, *ovate*, and *sun*, have been cloned [7-9].
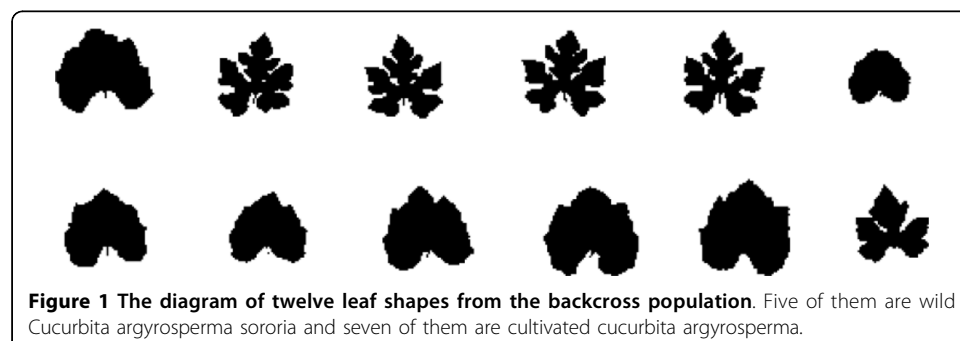
Second, digital technologies through computerized analyses and processing procedures can obtain a comprehensive representation of the involved objects, capable not only of representing most of the original information, but also of emphasizing their less redundant portions [10-15]. Third, statistical and computational technologies have well been developed for analyzing high-dimensional, large-scale, high-throughput data of high complexity [16,17]. With the development of missing data analysis, Lander and Botstein [18] have been able to pioneer an approach for dissecting complex quantitative traits into individual QTLs using genetic linkage maps constructed with molecular markers. There has been a vast wealth of literature in the development of QTL mapping models (see [19-25] among many others).

The motivation of this study is to develop a statistical and computational model for mapping specific QTLs that are responsible for differences in morphological shape. Historically, genetic mapping has been focused on the genetic control of a trait at a static point, ignoring the dynamic behavior and spatial properties of the trait. Now, by integrating the developmental principle of trait growth, a new genetic mapping approach, called functional mapping [26-28], can be used to study the dynamic control of genes in time course. The central idea of functional mapping is to connect the genetic control of a developmental trait at different time points through robust mathematical and statistical equations. Complementary to functional mapping, the model developed for shape mapping in this study links gene action with key morphometric parameters of a shape within a statistical framework. We will perform computer simulation to examine the statistical properties of the model.

Model

Genetic Design

We assume a backcross design although the model can be modified to accommodate any other mapping designs. Consider a backcross progeny population of size n, founded with two inbred lines that are sharply contrasting in leaf shape. Because of gene segregation, there is a range of variation in leaf shape among the backcross progeny. Such shape variation is illustrated in Fig. 1 by using leaf morphology in cucurbit plants [29]. To map the shape trait, the mapping population is typed for a panel of molecular markers from which a genetic linkage map covering the genome is constructed. The statistical approach for linkage analysis and map construction is reviewed in Wu et al. [30]. Assume that there are some specific QTLs responsible for the

Figure 1 The diagram of twelve leaf shapes from the backcross population. Five of them are wild *Cucurbita argyrosperma sororia* and seven of them are cultivated *cucurbita argyrosperma*.

biological shape. The approach being developed aims to detect and map such QTLs by capitalizing on knowledge about shape analysis and biological principles behind shape formation and variation.

Shape Analysis

According to the definition of Kendall [31], "shape is all the geometrical information that remains when location, scale and rotational effects are filtered out from an object". Assume that each backcross progeny is measured for the leaf shape as shown in Fig. 1. For a given shape, I^i ($i = 1, ..., n$), described by a black and white image, it is gridded as an $L \times L$ matrix, where L is the number of pixels in the row and column. At each point in the matrix, we use 0 to denote the background (black) and 1 to denote the leaf (including an arbitrary shape of it) (white). The 1/0 value of the matrix is assumed to follow a Bernoulli distribution. All these n shapes, $T = \{I^1, I^2, ..., I^n\}$, need to be aligned, in order to minimize the interference caused by pose variations. This can be carried out by establishing a coordinate reference with respect to position, scale and rotation, commonly known as pose to which all shapes are aligned [10,12,14]. Denote the pose parameter for each shape I^i by $p^i = [a, b, h, \theta]^T$ where a and b correspond to x and y translations, h is the scaling parameter, and θ corresponds to rotation. The transformed image of I^i, based on the pose parameter p^i, is denoted by $\tilde{I}^i$, defined as

$$\tilde{I}^i(\tilde{x}, \tilde{y}) = I^i(x, y),$$

where

$$\begin{pmatrix} \tilde{x} \\ \tilde{y} \\ 1 \end{pmatrix} = T[p] \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & a \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} h & 0 & 0 \\ 0 & h & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\begin{pmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix},$$

which yields

$$\begin{cases} \tilde{x} &= a + hx\cos(\theta) - hy\sin(\theta), \\ \tilde{y} &= b + hy\cos(\theta) + hx\sin(\theta). \end{cases} \tag{1}$$

The translation matrix T [p] is the product of three matrices: a translation matrix M (a, b), a scaling matrix $H(h)$, and an in-plane rotation matrix $R(\theta)$. The transformation matrix $T[p]$ maps the coordinates $(x, y) \in R^2$ into coordinates $(\tilde{x}, \tilde{y}) \in R^2$, where $x, y = 1, ..., L$.

An effective strategy to jointly align the n binary images is to use a gradient descent to minimize the following energy function:

$$E = \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} \left\{ \frac{\iint_{\Omega} \left(\tilde{I}^i - \tilde{I}^j \right)^2 dA}{\iint_{\Omega} \left(\tilde{I}^i + \tilde{I}^j \right)^2 dA} \right\}, \tag{2}$$

where Ω denotes the image domain. Minimizing the energy function (2) is equivalent to simultaneously minimizing the difference between any pair of binary images in the training database. What we would like to estimate is the pose parameter p^i for each I^i.

The derivative respective to p^i of equation (2) is

$$\nabla_{p^i} E = 2 \Sigma_{j=1, j \neq i}^{n} \left\{ \frac{2 \iint_{\Omega} (\tilde{I}^i - \tilde{I}^j) \nabla_{p^i} \tilde{I}^i dA}{\iint_{\Omega} (\tilde{I}^i + \tilde{I}^j)^2 dA} \right.$$
$$\left. - \frac{2 \iint_{\Omega} (\tilde{I}^i - \tilde{I}^j)^2 dA \iint_{\Omega} \tilde{I}^i + \tilde{I}^j) \nabla_{p^i} \tilde{I}^i dA}{(\iint_{\Omega} (\tilde{I}^i + \tilde{I}^j)^2 dA)^2} \right\}, \tag{3}$$

$$\text{where } \nabla_{p^i} \tilde{I}^i = \left[\frac{\partial \tilde{I}^i}{\partial a}, \frac{\partial \tilde{I}^i}{\partial b}, \frac{\partial \tilde{I}^i}{\partial h}, \frac{\partial \tilde{I}^i}{\partial \theta} \right]^T.$$

By a chain rule and equation (1), we get

$$\frac{\partial \tilde{I}^i}{\partial a} = \frac{\partial \tilde{I}^i}{\partial \tilde{x}} = \frac{\partial I^i}{\partial x},$$

$$\frac{\partial \tilde{I}^i}{\partial b} = \frac{\partial \tilde{I}^i}{\partial \tilde{y}} = \frac{\partial I^i}{\partial y},$$

$$\frac{\partial \tilde{I}^i}{\partial h} = \frac{\partial I^i}{\partial x} \left(x\cos(\theta) - y\sin(\theta) \right)$$
$$+ \frac{\partial I^i}{\partial y} \left(y\cos(\theta) + x\sin(\theta) \right),$$

$$\frac{\partial \tilde{I}^i}{\partial \theta} = \frac{\partial I^i}{\partial x} \left(-hx\sin(\theta) - hy\cos(\theta) \right)$$
$$+ \frac{\partial I^i}{\partial y} \left(-hy\sin(\theta) + hx\cos(\theta) \right).$$

Hence, we can obtain the value of $\nabla_{p^i} E$ as long as p^i and $\tilde{I}^i$ are given in each iterative step. The steepest gradient algorithm is then used to minimize E in (2) and get the pose parameter p^i for each shape I^i. All the training shapes after the alignment procedure described above are obtained (see Fig. 2).

Statistical Model

After all the training shapes are aligned, a shape representation scheme needs to be chosen for T = { $\tilde{I}^1$, $\tilde{I}^2$, ..., $\tilde{I}^n$}., i.e., the transformed images, which now become continuous variables. The signed distance function was used as a shape descriptor to

Figure 2 Leaf shapes after alignment for leaf shapes shown in Fig. 1.
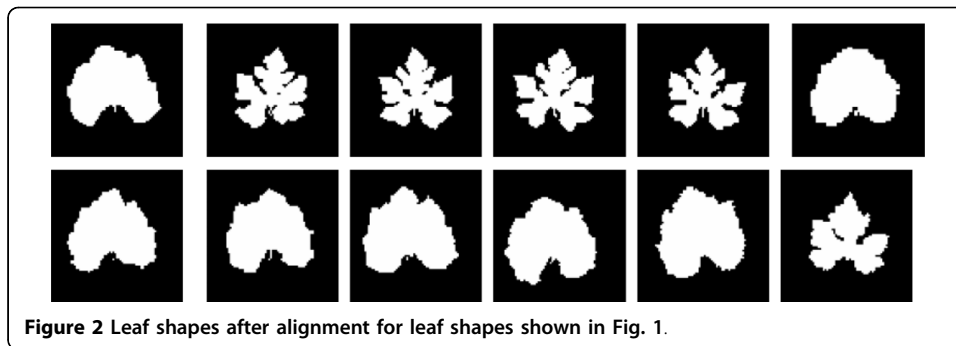
represent the contours of the shape. Each contour is embedded as the zero level set of a signed distance function with negative distances assigned to the inside and positive distances assigned to the outside. This technique yields n level sets functions $Y = \{Y_1, Y_2, \ldots Y_n\}$ corresponding to above n aligned training shapes. From the standpoint of QTL mapping, we treat $Y = \{Y_1, Y_2, \ldots, Y_n\}$ as the multiple phenotypic traits of n individuals. For a progeny i ($i = 1, 2, \ldots, n$), we have

$$Y_i = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1L} \\ y_{21} & y_{22} & \cdots & y_{2L} \\ \vdots & \vdots & \ddots \cdots & \vdots \\ y_{L1} & y_{L2} & \cdots & y_{LL} \end{pmatrix}. \tag{4}$$

Thus, each individual has a total of $m = L^2$ phenotypes.

For the backcross progeny population, there are always two different genotypes at each locus. The genotypes at a shape QTL, expressed as QQ (denoted as 1) and Qq (denoted as 2), cannot be observed directly but can be inferred from the markers that are linked to the QTL. For this reason, the basic statistical model for QTL mapping is based on a mixture model, in which each observation Y is assumed to have arisen from one of the two groups of QTL genotypes, each group being modeled from a density function (frequently a normal distribution is assumed). Thus, the population density function of Y is

$$f(Y \mid \omega, \phi, \eta) = \sum_{j=1}^{2} \omega_{j|i} f_j(Y_i \mid \phi_j, \eta), \tag{5}$$

where ω represents the mixture proportions ($\omega_{1|i}, \omega_{2|i}$), which are constrained to be nonnegative and sum to unity, ϕ_j is the expectation parameter specific to different QTL genotypes $j = 1, 2$, and η is the variance-covariance parameter common to all genotype groups, and $f_j(Y_i|\phi_j,\eta)$ is the probability density function for QTL genotype j. After images are transformed, Y_i can be assumed to follow a multivariable normal distribution, i.e.,

$$f_j(Y_i) = \frac{1}{(2\pi)^{m/2}|\Sigma|^{1/2}} \exp\left[-\left(Y_i - \mu_j\right)^T \Sigma^{-1} \left(Y_i - \mu_j\right)/2 \right], \tag{6}$$

with the expectation matrix of each QTL genotype expressed as

$$\mu_j = \begin{pmatrix} \mu^j_{11} & \mu^j_{12} & \cdots & \mu^j_{1L} \\ \mu^j_{21} & \mu^j_{22} & \cdots & \mu^j_{2L} \\ \vdots & \vdots & \ddots \cdots & \vdots \\ \mu^j_{L1} & \mu^j_{L2} & \cdots & \mu^j_{LL} \end{pmatrix}, \quad \text{for } j = 1, 2, \tag{7}$$

and $(m \times m)$ residual variance-covariance matrix of the variables Σ. If some patterns exist, we will use ϕ_j to model the mean structure of μ_j and η to model the covariance structure of Σ.

In order to simplify the problem, we use the most natural sampling strategy to utilize the $L \times L$ rectangular grid of the training shapes to generate $m = L \times L$ lexicographically ordered samples (where the columns of the matrix grid are sequentially stacked on top of one other to form one large row). Also, we assume that all the observations in the long row are independent among the progeny. Now, from equation (5), we get the likelihood function as

$$\begin{aligned} L(\gamma) &= \prod_{i=1}^{n} f(Y_i \mid \omega, \phi, \eta) \\ &= \prod_{i=1}^{n} \sum_{j=1}^{2} \omega_{j|i} f_j(Y_i \mid \phi_j, \eta) \\ &= \prod_{i=1}^{n} \sum_{j=1}^{2} \omega_{j|i} \frac{1}{(2\pi)^{m/2} |\Sigma|^{1/2}} \\ &= \exp[-(Y_i - \mu_j)^T \Sigma^{-1} (Y_i - \mu_j) / 2], \end{aligned} \tag{8}$$

where the mean matrix of QTL genotype j (μ_j) is modeled by parameter ϕ_j, and covariance matrix (Σ) modeled by parameter η.

Computational Algorithm

To obtain the maximum likelihood estimates (MLEs) of parameters in likelihood (8), we implement a standard EM algorithm. In the E step, we compute the posterior probability with which a backcross individual carries a QTL genotype j using

$$\Omega_{ij} = \frac{\omega_j f_j(Y_i | \phi_j, \eta)}{\sum_{l=1}^{2} \omega_l f_l(Y_i | \phi_l, \eta)}. \tag{9}$$

In the M step, we estimate the parameters using

$$\mu_{jk} = \frac{\sum_{i=1}^{n} \Omega_{ij} \gamma_{ik}}{\sum_{i=1}^{n} \Omega_{ij}}, \tag{10}$$

for $j = 1, 2$ and $k = 1, 2, ..., m$.

The EM steps are iterated between equations (9) and (10) until the estimates converge to stable values. It should be pointed out that the data set for shape analysis is highly sparse and high-dimensional. For example, if a shape is described by (256 ×

256) pixels, i.e., L = 256, then we will have m = 256^2 = 65, 536, and an ($n \times 65$, 536) matrix for the phenotypic observations. Several approaches will be developed to model the structure of the variance-covariance matrix. One of the simplest approaches is to use $\sigma = \frac{1}{2}\sqrt{2L^2}$. This choice is large enough to assure that various levels of differences lie well within a Gaussian distribution.

Hypothesis Tests

A hypothesis about the existence of a significant QTL that controls a morphological shape can be tested by calculating the log-likelihood ratio under the hypotheses:
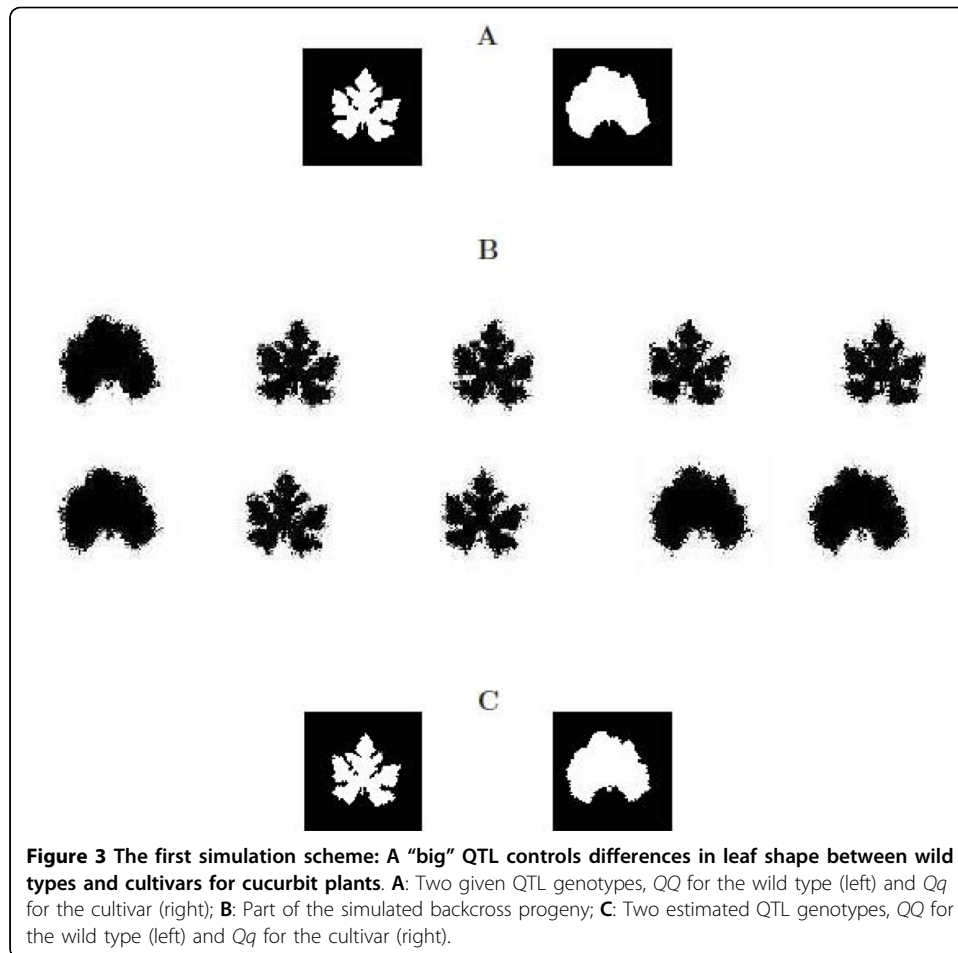
$$H_0 : \mu_1 = \mu_2 \text{ vs. } H_1 : \mu_1 \neq \mu_2. \tag{11}$$

As like an usual mapping approach, shape mapping has a problem of uncertain distribution for the log-likelihood test statistic. However, an empirical approach based on permutation tests, which does not rely on the distribution of log-likelihood ratios, can be used to determine the threshold for claiming the existence of a significant QTL.
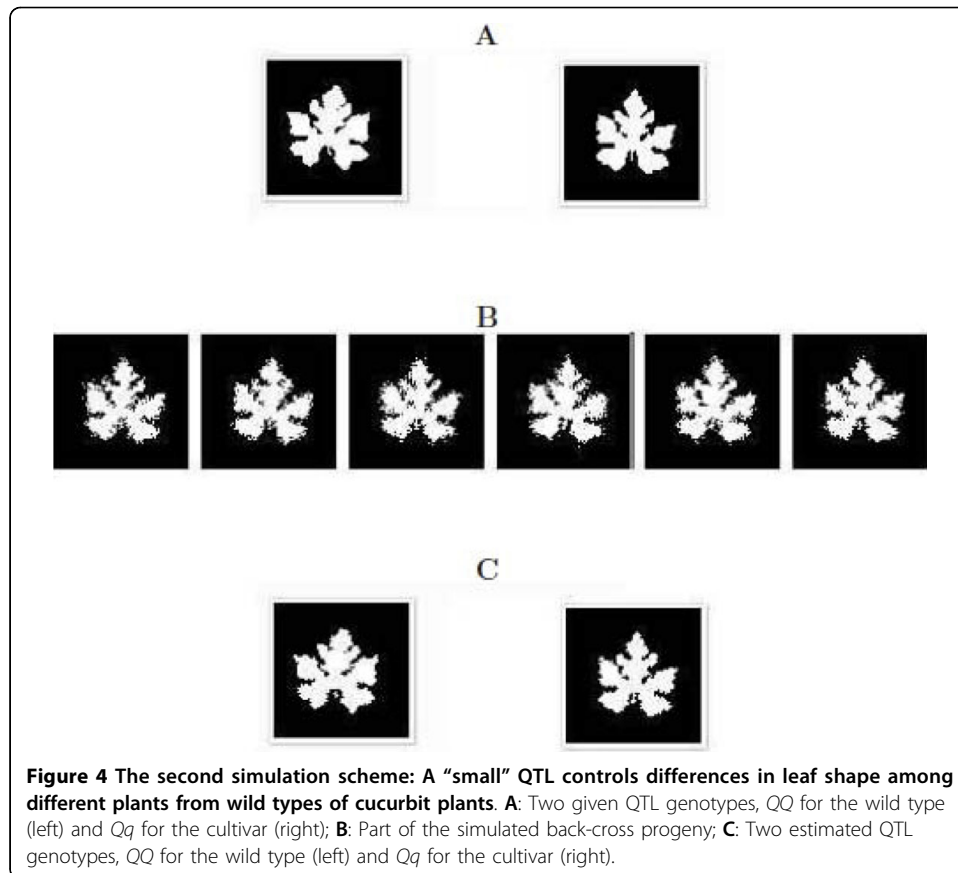
Computer Simulation

Cucurbit (*Cucurbita argyrosperm*) plants display tremendous variation in leaf shape between cultivars and wild types [29]. By mimicking leaf morphologies of this species, we performed simulation studies to examine the statistical behavior of our shape mapping model. A backcross population of 200 progeny was simulated for a linkage group with 11 equally spaced markers. A QTL that determines leaf shape is hypothesized on the third marker interval. The phenotypic values of the shape were simulated with a (75×75) dimension by $Y_i = \xi_i\mu_1 + (1-\xi_i)\mu_2 + e_i$, where μ_j is the mean shape matrix for QTL genotype j (j = 1, 2), ξ_i is the indicator variable defined as 1 and 0 if progeny i carries QTL genotype QQ (1) and qq (2), respectively, and e_i follows a multivariate normal distribution with mean vector zero and covariance matrix Σ. To simplify computing, we assumed that Σ is an identity matrix. We designed two simulation schemes to test our shape mapping algorithm.

The first scheme assumes that there exists a "big" QTL which triggers a tremendous effect on the difference in leaf shape of cucurbit plants between their cultivars and wild types. This QTL has two different genotypes, one, QQ, corresponding to the wild type shape (right) and the second, Qq, to the domesticated shape (left) (Figure 3A). The QTL genotypes are determined by the conditional probability of a QTL genotype, conditional upon the genotypes of the two markers that flank the QTL (see [30]). Part of the 200 progeny simulated with two assumed QTL genotypes were given in Figure 3B, in which some leaf shape looks more like the wild type, some more like the domesticated type, and the other is in between. The model described above was used to analyze the simulated data. The log-likelihood ratio test statistic calculated under hypotheses (11) is greater than the critical threshold for testing the existence of a QTL obtained from permutation tests, suggesting that two genotype-specific shapes for QQ and Qq were detected and identified. Figure 3B also illustrates the shapes of two detected QTL genotypes from the simulated data. As shown, the estimated shapes are similar to the true shapes for the two backcross QTL genotypes, suggesting that our model has great power to identify the QTL that control morphological shape.

Figure 3 The first simulation scheme: A "big" QTL controls differences in leaf shape between wild types and cultivars for cucurbit plants. **A**: Two given QTL genotypes, *QQ* for the wild type (left) and *Qq* for the cultivar (right); **B**: Part of the simulated backcross progeny; **C**: Two estimated QTL genotypes, *QQ* for the wild type (left) and *Qq* for the cultivar (right).

The second scheme simulated two QTLs that determine the differences of leaf shape among wild-type plants and domesticated plants, respectively. Compared to the "big" QTL assumed in the first scheme, these two QTLs are "small" because their two genotypes correspond to slightly different leaf shapes. Figures 4 and 5 provide the results about shape mapping for wild-type plants and domesticated plants, respectively. In the upper panel (A) of each figure, two original QTL genotypes are assumed, from which 200 backcross progeny were simulated with a range of leaf shape. The middle panel (B) gives part of the backcross. In the bottom panel (C), two genotypes were estimated using our algorithm. It can be seen that the model can well detect a QTL even if it has a small effect on morphological shape.

To show the fitness of our model, we put the estimated QTL genotypes on the simulated backcross population for the first (A) and second (B and C) simulation scheme (Fig. 6). The leaf shape of two QTL genotypes in each case well covers the simulated leaf shape, showing a good fitness of the mapping model. Also, we calculated the density functions for each simulated progeny and two QTL genotypes for each simulation scheme (Fig. 7). The "big" QTL displays two distinct modes of distribution (Fig. 7A), whereas there is a small difference in the density functions of two genotypes for each of two "small" QTLs (Fig. 7B,C). By comparing Fig. 1A with Fig. 7B and 7C, we can
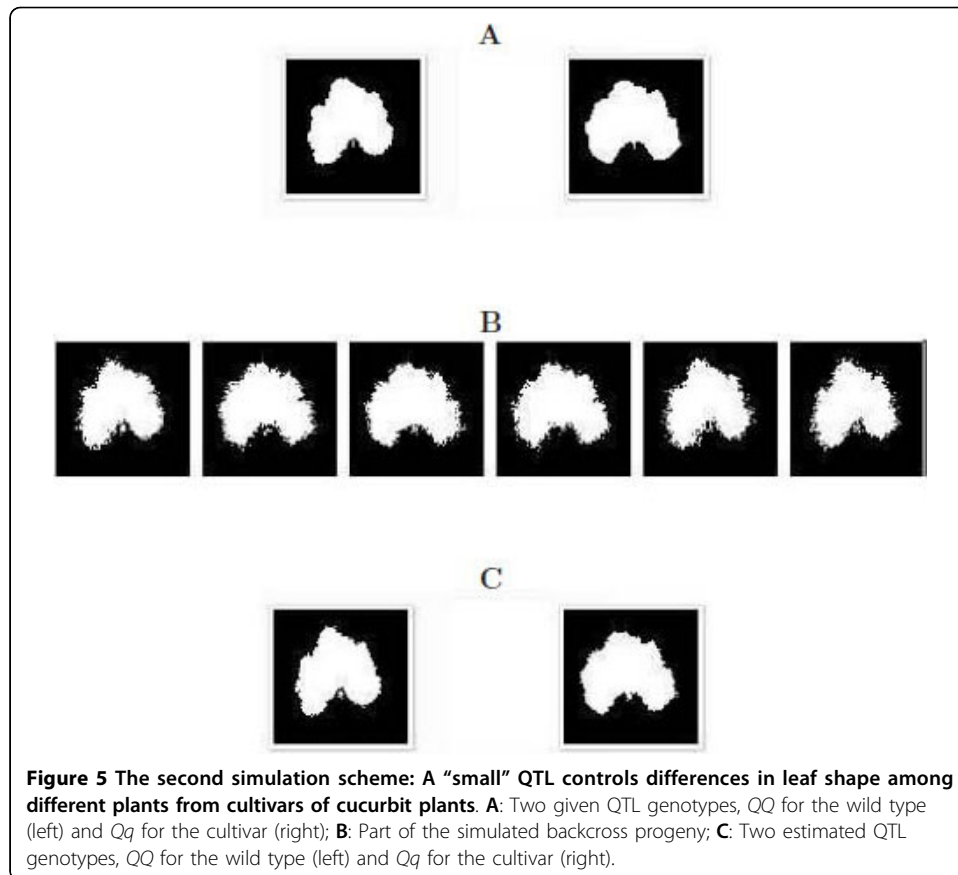
Figure 4 The second simulation scheme: A "small" QTL controls differences in leaf shape among different plants from wild types of cucurbit plants. **A**: Two given QTL genotypes, *QQ* for the wild type (left) and *Qq* for the cultivar (right); **B**: Part of the simulated back-cross progeny; **C**: Two estimated QTL genotypes, *QQ* for the wild type (left) and *Qq* for the cultivar (right).

obtain the basic information about how well different QTL genotypes are separated when QTLs exert different effects on leaf shape.

Discussion

When specific genes that control morphological shape and physiological function are identified, we are in an excellent position to address fundamental questions related to growth, development, adaptation, domestication, and human health. In the past decades, the increasing availability of DNA-based markers has inspired our hope to map genes or quantitative trait loci (QTLs) for complex phenotypes [19-25]. However, only several studies have been alert to map so-called shape genes; a few successful examples are the positional cloning of genes for fruit shape in tomato [3,7-9]. These successes result from the fact that a major mutation occurs to determine shape difference. For many quantitatively inherited shape traits, genetic mapping will provide a powerful tool for characterizing QTLs affecting morphological shape. Klingenberg and colleagues [4,5] have developed quantitative genetic theory to estimate the heritability of shape by integrating geometric shape analysis. This theory was used to map specific QTLs for morphometric shapes in the mouse [32,33]. Airey et al. [34] used Procrustes superimposition to study shape differences in the cortical area map of inbred mice.
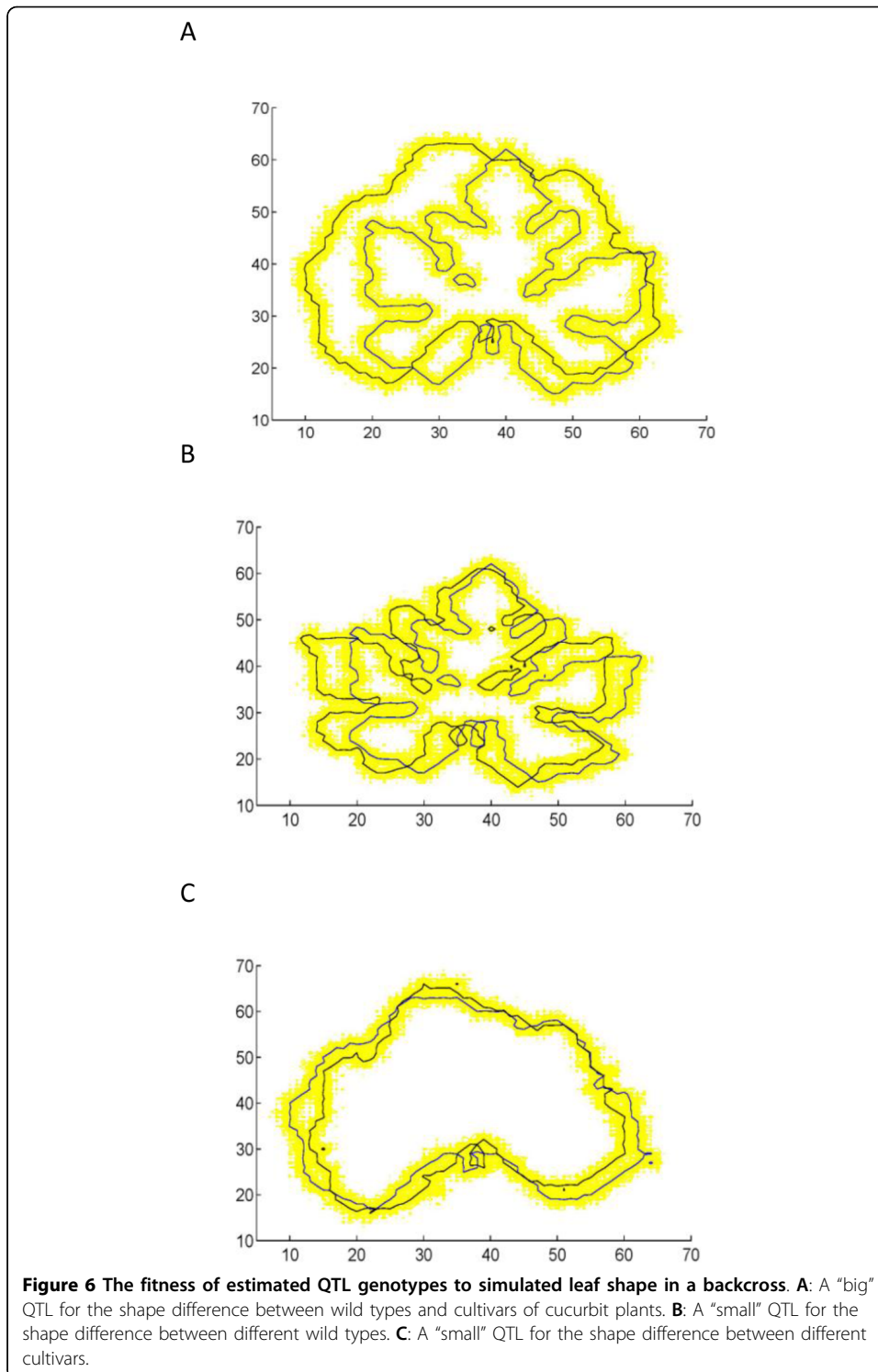
In this article, we present a new statistical model for mapping shape QTLs in a segregating population. The new model embeds shape analysis within a mixture model framework in which different types of morphological shape are defined for individual genotypes at a QTL. The model was solved using a traditional shape correspondence
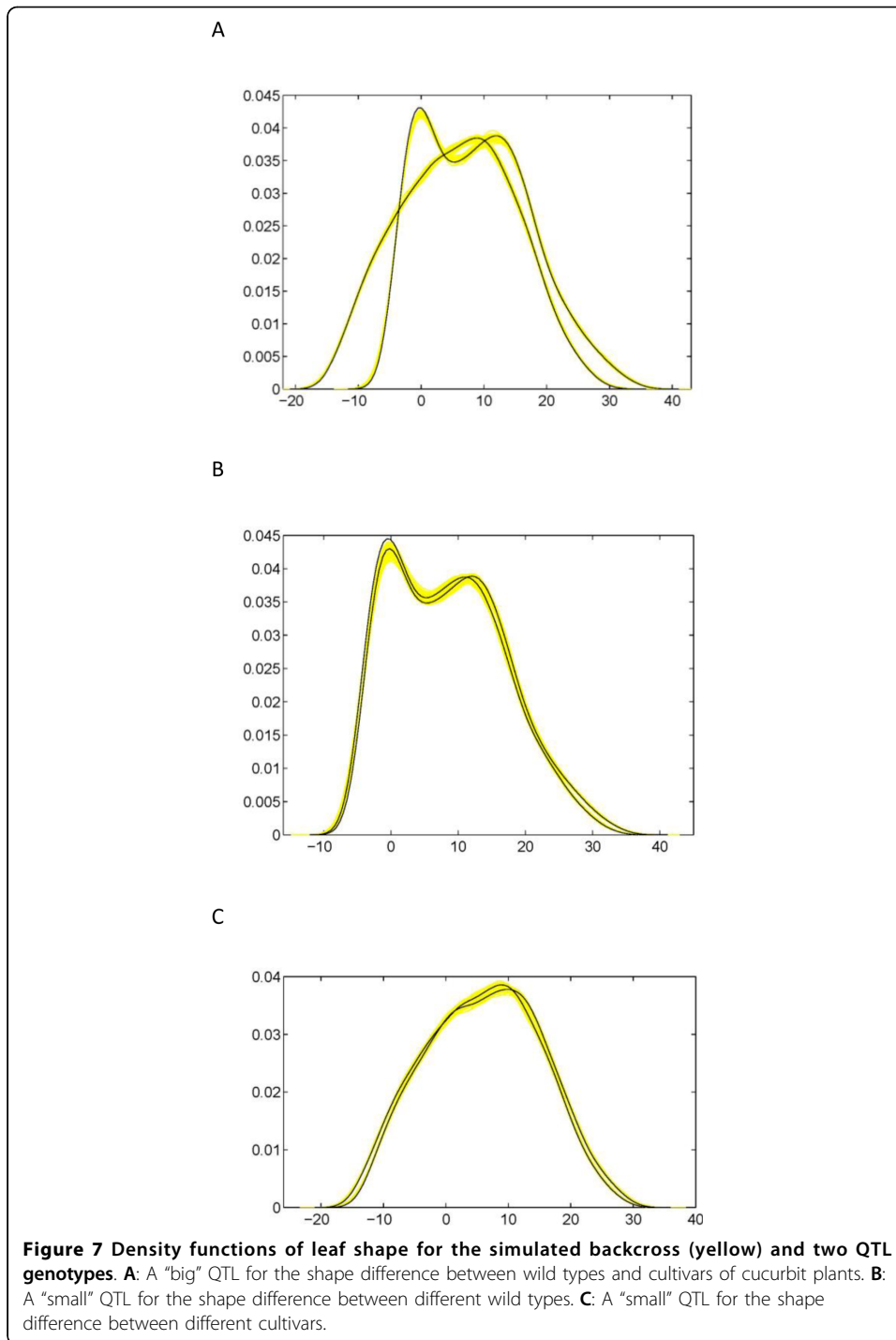
Figure 5 The second simulation scheme: A "small" QTL controls differences in leaf shape among different plants from cultivars of cucurbit plants. **A**: Two given QTL genotypes, *QQ* for the wild type (left) and *Qq* for the cultivar (right); **B**: Part of the simulated backcross progeny; **C**: Two estimated QTL genotypes, *QQ* for the wild type (left) and *Qq* for the cultivar (right).

analysis approach and EM algorithm. The advantage of shape mapping lies in its capacity to quantify subtle differences in any corner of a morphological shape and detect specific QTLs that contribute to these differences. Results from simulation studies suggest that the model has reasonably high power to detect a QTL that control shape difference. Even with a modest sample size (200), the model is able to discern the effect of a QTL with a small effect on morphological shape. The model can be easily extended to model epistatic interactions on morphological shape by including more components in the mixture model.

The model will be needed to be modified for integrating developmental events and their consequences into ontogenetic trajectories of shape. Modern biological studies display an increasing interest in understanding shape variation in ontogenetic processes that bring about differentiation at an adult stage [35-37]. In a longitudinal study of radiographs of the Denver Growth Study, Bulygina et al. [37] investigated the morphological development of individual differences in the anterior neurocranium, face, and basicranium. The modified model can map the QTLs that cause variation in shape developmental trajectories.

In biology, a cell or organ fulfill certain biological functions through its shape. Shape is thought to govern the extent and pattern of energy, matter and signal transduction through the surface and inner structure of the biological object. For this reason, an understanding of biological curvature and texture has received a surge of interest in structural biology. The new model can be extended to map the QTLs that determine a

Figure 6 The fitness of estimated QTL genotypes to simulated leaf shape in a backcross. **A**: A "big" QTL for the shape difference between wild types and cultivars of cucurbit plants. **B**: A "small" QTL for the shape difference between different wild types. **C**: A "small" QTL for the shape difference between different cultivars.

three-dimensional (3D) shape and texture of a biological object. Vision technologies have been developed to estimate the 3D shape of an object from 2D image data without information about its texture (albedo), its pose and the illumination environment [38,39]. These technologies include a 3D morphable model (3DMM) that represents the 3D shapes and textures as a linear combination of shapes and textures principal

Figure 7 Density functions of leaf shape for the simulated backcross (yellow) and two QTL genotypes. **A**: A "big" QTL for the shape difference between wild types and cultivars of cucurbit plants. **B**: A "small" QTL for the shape difference between different wild types. **C**: A "small" QTL for the shape difference between different cultivars.

components, a stochastic Newton optimization algorithm that ts the 3DMM to a single facial image, thereby estimating the 3D shape, the texture and the imaging conditions, and a multi-features fitting algorithm that uses not only the pixel intensity but also other image cues such as the edges and the specular highlights. Statistical models can be developed to map QTLs that control the 3D shape and texture of a biological object

with image data. A series of hypothesis tests about the genetic control of topological features (such as stepness and ridgeness) and texture of a shape will be formulated.

Author details
[1]Department of Statistics, Pennsylvania State University, University Park, PA 16802, USA. [2]Center for Statistical Genetics, Pennsylvania State University, Hershey, PA 10733, USA. [3]Center for Computational Biology, Beijing Forestry University, Beijing 100083, China.

Authors' contributions
GF derived the model and performed simulation studies. AB, KD, and JL participated in simulation studies. RL participated in the design of the study. RW conceived of the study, coordinated the design and simulation studies, and wrote the manuscript. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

References
1. Ricklefs RE, Miles DB: **Ecological and evolutionary inferences from morphology: an ecological perspective.** *Ecological morphology* Univ. of Chicago Press, ChicagoWainwright PC, Reilly SM 1994, 13-41.
2. Reich PB: **Body size, geometry, longevity and metabolism: do plant leaves behave like a animal bodies?** *Trends Ecol Evol* 2001, **16**:674-680.
3. Tanksley SD: **The genetic, developmental, and molecular bases of fruit size and shape variation in tomato.** *Plant cell* 2004, **16**:S181-S189.
4. Klingenberg CP, Leamy LJ: **Quantitative genetics of geometric shape in the mouse mandible.** *Evolution* 2001, **55**:2342-2352.
5. Klingenberg CP: **Quantitative genetics of geometric shape: heritability and the pitfalls of the univariate approach.** *Evolution* 2001, **57**:191-195.
6. Tsukaya H: **Leaf shape: genetic controls and environmental factors.** *Intl J Dev Biol* 2005, **49**:547-555.
7. Frary A, Nesbitt TC, Grandillo S, Knaap E, Cong B, Liu J, Meller J, Elber R, Alpert KB, Tanksley SD: **fw2.2: A quantitative trait locus key to the evolution of tomato fruit size.** *Science* 2000, **289**:85-88.
8. Liu J, Van Eck J, Cong B, Tanksley SD: **A new class of regulatory genes underlying the cause of pear-shaped tomato fruit.** *Proc Natl Acad Sci USA* 2002, **99**:13302-13306.
9. Xiao H, Jiang N, Schaffner E, Stockinger EJ, van der Knaap E: **A retrotransposonmediated gene duplication underlies morphological variation in tomato fruit.** *Science* 2008, **319**:1527-1530.
10. Bookstein FL: *The Measurement of Biological Shape and Shape Change* Springer- Verlag, New York 1978.
11. Monteiro LR, Diniz-Filho JA, dos Reis SF, Araujo ED: **Geometric estimates of heritability in biological shape.** *Evolution* 2002, **56**:563-572.
12. Adams DC, Rohlf FJ, Slice DE: **Geometric morphoetrics: ten years of progress following the "revolution".** *Ital J Zool* 2004, **71**:5-16.
13. Bernal B: **Size and shape analysis of human molars: Comparing traditional and geometric morphometric techniques.** *J Comp Hum Biol* 2007, **58**:279-296.
14. Stegmann MB, Gomez DD: *A Brief Introduction to Statistical Shape Analysis* Informatics and Mathematical Modelling, Technical University of Denmark, DTU 2002.
15. Basri R, Costa L, Geiger D, Jacobs D: **Determining the similarity of de- formable shapes.** *Vision Res* 1998, **38**:2365-2385.
16. Dempster AP, Laird NM, Rubin DB: **Maximum likelihood from incomplete data via the EM algorithm.** *J Roy Stat Soc Ser B* 1977, **39**:1-38.
17. Tsai A, Wells W, Warfield S, Willsky A: **An EM algorithm for shape classification based on level sets.** *Med Image Anal* 2005, **9**:491-502.
18. Lander ES, Botstein D: **Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps.** *Genetics* 1989, **121**:185-199.
19. Zeng Z-B: **Precision mapping of quantitative trait loci.** *Genetics* 1994, **136**:1457-1468.
20. Jansen RC, Stam P: **High resolution mapping of quantitative traits into multiple loci via interval mapping.** *Genetics* 1994, **136**:1447-1455.
21. Xu S, Atchley W: **A random model approach to interval mapping of quantitative trait loci.** *Genetics* 1995, **141**:1189-1197.
22. Lynch M, Walsh B: *Genetics and Analysis of Quantitative Traits* Sinauer Associates, Sunderland, MA 1998.
23. Broman KW, Speed TP: **A model selection approach for the identification of quantitative trait loci in experimental crosses (with discussion).** *J Roy Stat Soc Ser B* 2002, **64**:641-656.
24. Zou F, Fine JP, Hu J, Lin DY: **An efficient resampling method for assessing genome-wide statistical significance in mapping quantitative loci.** *Genetics* 2004, **168**:2307-2316.
25. Yi N, Yandell BS, Churchill GA, Allison DB, Eisen EJ, Pomp D: **Bayesian model selection for genome-wide epistatic quantitative trait loci analysis.** *Genetics* 2005, **170**:1333-1344.

26. Ma C-X, Casella G, Wu RL: **Functional mapping of quantitative trait loci under-lying the character process: A theoretical framework.** *Genetics* 2002, **161**:1751-1762.
27. Wu RL, Ma C-X, Lou Y-X, Casella G: **Molecular dissection of allometry, ontogeny and plasticity: A genomic view of developmental biology.** *BioScience* 2003, **53**:1041-1047.
28. Wu RL, Lin M: **Functional mapping How to study the genetic architecture of dynamic complex traits.** *Nat Rev Genet* 2006, **7**:229-237.
29. Schlichting CD, Pigliucci M: *Phenotypic Evolution: A Norm Reaction Perspective* Sinauer Associates, Sunderland, MA 1998.
30. Wu RL, Ma C-X, Casella G: *Statistical Genetics of Quantitative Traits: Linkage, Maps, and QTL* Springer-Verlag, New York 2007.
31. Dryden IL, Mardia KV: *Statistical Shape Analysis* John Wiley & Sons, New York 1998.
32. Leamy LJ, Klingenberg CP, Sherratt E, Wolf JB, Cheverud JM: **A search for quantitative trait loci exhibiting imprinting effects on mouse mandible size and shape.** *Heredity* 2008, **101**:518-526.
33. Klingenberg CP, Leamy LJ, Cheverud JM: **Integration and modularity of quantitative trait locus effects on geometric shape in the mouse mandible.** *Genetics* 2004, **166**:1909-1921.
34. Airey DC, Wu F, Guan M, Collins CE: **Geometric morphometrics defines shape differences in the cortical area map of C57BL/6J and DBA/2J inbred mice.** *BMC Neurosci* 2006, **7**:63.
35. Vioarsdottir US, O'Higgins P, Stringer C: **A geometric morphometric study of regional differences in the ontogeny of the modern human facial skeleton.** *J Anat* 2002, **201**:211-229.
36. Quillevere F, Debat V, Auray J-C: **Ontogenetic and evolutionary patterns of shape dierentiation during the initial diversication of paleocene acarininids (***planktonic foraminifera***).** *Paleobiology* 2002, **28**:435-448.
37. Bulygina E, Mitteroecker P, Aiello L: **Ontogeny of facial dimorphism and patterns of individual development within one human population.** *Am J Phys Anthrop* 2006, **131**:432-443.
38. Romdhani S, Vetter T: **Estimating 3D shape and texture using pixel intensity, edges, specular highlights, texture constraints and a prior.** *IEEE Computer Soc Conf Computer Vision Pattern Recog* 2005, **2**:986-993.
39. Romdhani S, Ho J, Kriegman DJ: **Face recognition using 3-D models: Pose and illumination.** *Proc IEEE* 2006, **94**:1977-1999.