Research Article
E-Index for Differentiating Complex Dynamic Traits

Jiandong Qi,[1] Jianfeng Sun,[1] and Jianxin Wang[1,2]

[1]*School of Information, Beijing Forestry University, Beijing 100083, China*
[2]*Center for Computational Biology, Beijing Forestry University, Beijing 100083, China*

Correspondence should be addressed to Jianxin Wang; wangjx@bjfu.edu.cn

While it is a daunting challenge in current biology to understand how the underlying network of genes regulates complex dynamic traits, functional mapping, a tool for mapping quantitative trait loci (QTLs) and single nucleotide polymorphisms (SNPs), has been applied in a variety of cases to tackle this challenge. Though useful and powerful, functional mapping performs well only when one or more model parameters are clearly responsible for the developmental trajectory, typically being a logistic curve. Moreover, it does not work when the curves are more complex than that, especially when they are not monotonic. To overcome this inadaptability, we therefore propose a mathematical-biological concept and measurement, *E*-index (earliness-index), which cumulatively measures the earliness degree to which a variable (or a dynamic trait) increases or decreases its value. Theoretical proofs and simulation studies show that *E*-index is more general than functional mapping and can be applied to any complex dynamic traits, including those with logistic curves and those with nonmonotonic curves. Meanwhile, *E*-index vector is proposed as well to capture more subtle differences of developmental patterns.

1. Introduction

Whether there are different genes responsible for the formation of a trait and how these genes regulate the trait are of fundamental importance biologically, agriculturally, and/or medically. Quantitative traits, or characteristics varying in degree, can be attributed to the effects of genes and their environment [1]. Lander and Botstein [2] pioneered the systematic integration of molecular genetics and statistical methodologies to dissect quantitative traits to an individual genetic locus, well known as quantitative trait loci (QTLs). Since then, quantitative differences in mass, length, and so forth of the whole individual or an organ in their mature state are used to identify genes [3–8]. According to QTL mapping, individuals with different marker locus genotypes will have different mean values of a quantitative trait, if a QTL is linked to the marker locus.

It should be noted, however, that the single-valued traits are only a portion of the numerous traits, of which many others change with time or other independent variables and are the so-called complex traits. In fact, measurement values in the mature state provide much less information than the growth process leading to it [9]. For example, growth may be defined as quantitative changes in size, mass, or number, and the process is more biologically meaningful than the final state solely: the measurement value of an individual or an organ. The complex traits, which can be expressed as a functional or visually a curve, were thought to be *infinite-dimensional characters* in [10] or *function-valued traits* in [11]. Researchers made effort to study this problem extensively by biological, mathematical, or statistical means [12–15]. And other researchers tried to solve this problem by considering the complex trait (with a set of sampled values) as a bunch of simple traits [16–19]. However, if the values (a set of traits) of a complex trait are considered separately, the relationships between the values are lost or are too time-consuming to capture due to the large size of the residue covariance matrix. But with the eigenvalues of the matrix, the dimensions can be reduced greatly and it becomes feasible for genetic mapping of a large number of traits [20, 21]. These methods, however, do not take into account the developmental mechanisms that regulate trait formation and variation.

Ma et al. [22] proposed *functional mapping*, a statistical framework, for mapping QTL regulating dynamic trajectories
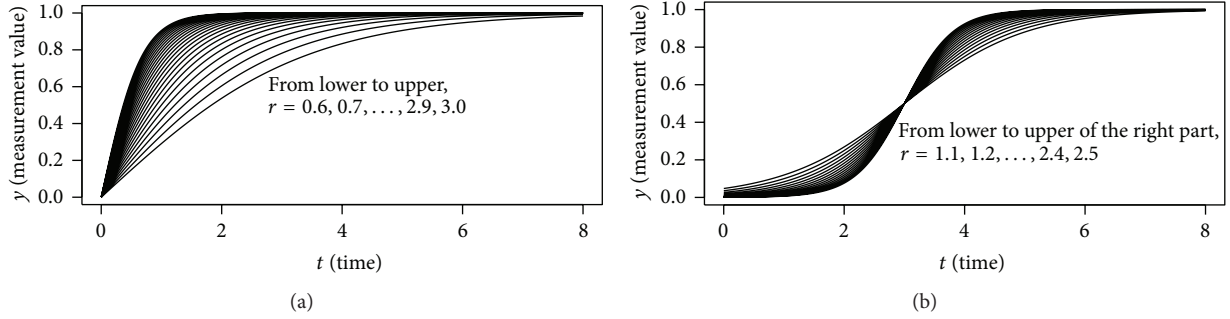
FIGURE 1: Growth curves with different parameter values. (a) is a bunch of curves described by (1). (b) is a bunch of curves described by (2).

of traits. Functional mapping is constructed on the basis of a biological law, as presented by West et al. [23], that the growth of many an organism follows a logistic curve due to the fundamental metabolic principles for allocating energy between maintaining current tissue and gaining biomass. By incorporating the logistic curve, functional mapping differentiates a complex dynamic trait by the parameters of the logistic function, instead of directly by the trait values, and thus makes computation less time-consuming and makes the results more biologically meaningful. Since functional mapping was proposed in [22], there has been a wealth of literature about its variations, improvements, and applications [24–29]. Up to date, functional mapping has successfully applied to associate high dimensions of SNPs with high dimensions of dynamic traits [30].

We now briefly review how functional mapping differentiates developmental trajectories. First of all, the growth process of an individual or an organ can be described by a growth curve, a function of a measurable variable against time. Theoretically, a growth curve may provide infinite amount of information, unlike a single measurement value in a mature state. For example, we consider two bunches of growth curves which are described by the following two equations and illustrated in Figure 1:

$$y = \frac{2}{1 + e^{-rt}} - 1, \tag{1}$$

$$y = \frac{1}{1 + e^{-r(t-3)}}. \tag{2}$$

In practice, discrete values of the developmental process are measured and collected, based on which functional mapping recovers the process by describing it with a curve which is determined by one or more parameters (r here indicating the growth rate). We can observe in Figure 1(a) that as the parameter r increases, the corresponding growth curve becomes steeper at the beginning part. And in Figure 1(b), a growth curve with greater r increases slower at the beginning of the left part and faster at the end of left part. The inclination of the right part is opposite to that of the left part. Therefore the parameter r may act as a characteristic value of the bunch to differentiate the curves and thus differentiate growth types or styles.

If all growth curves can be described by a function with one or more varying parameters, then we can employ
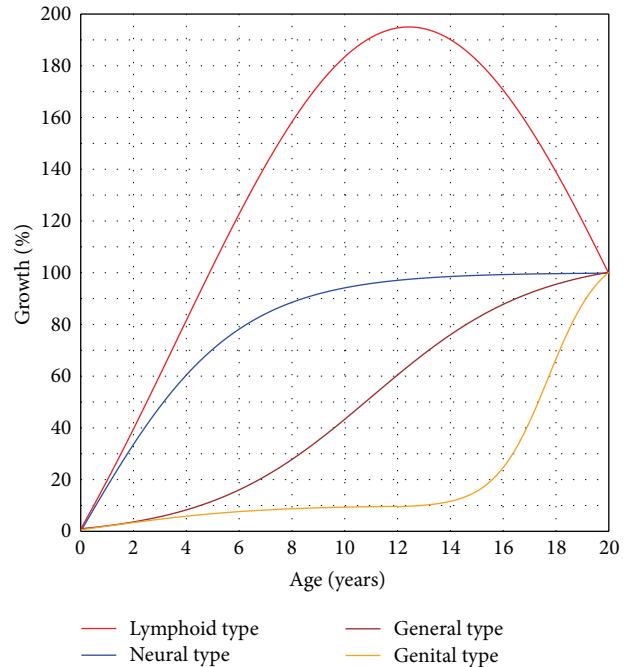

FIGURE 2: Four major types of growth curves of the organs and the body as a whole, from birth to 20 years [31]. (Later we will obtain their E-indices from upper to lower as 1.328, 0.802, 0.459, and 0.191, resp.)

these parameters to be the characteristic values, which is the essence of functional mapping. Unfortunately, no function is qualified for describing all growth types. Specifically, Figure 2 shows Scammon's classic illustration [31] of different growth types of human beings that are almost impossible to describe with a uniform function. Therefore, functional mapping fails to work with curves like the nonmonotonic lymphoid type in Figure 2.

The diversity of growth curves gives rise to a problem: how can we differentiate them with one or more characteristic values? An important characteristic value, E-index, will be proposed below and the rest of the paper is organized as follows. E-index is defined and its properties are discussed in Section 2. And in Section 3, a statistical framework for E-index is given and its effectiveness is validated through simulation studies. E-index vector is defined and validated in Section 4. And Section 5 concludes the paper.

2. E-Index's Definition and Properties

2.1. Concept and Definition. As is shown in Figure 1, growth and development may perform faster or slower, earlier or later, due to different types. And the earliness degree of growth that we are to define an index to measure may play an important role to evaluate a growth curve, both mathematically and biologically.

It is common sense that a growth curve is continuous and smooth, but in order to elucidate the concept of E-index intuitively and for simplicity, we design an imaginary scenario in which an individual gains part of its height instantaneously (though this is impossible), as shown in Figure 3. It takes each of the 6 individuals indicated in Figure 3 exactly 9 units of time (from 0 to 9) to gain 5 units of height (from 2 at the beginning to 7 at the end).

Take in Figure 3(c), for instance, first. The individual indicated by it keeps its original height 2 for the first 2 units of time, and then its height instantaneously increases by 4 units at the time point 2. After that, it keeps the height for 4 units of time, until it increases its height again by 1 unit at time point 6. Finally it keeps the height 7 to the end point.

Compared with (c), the individual indicated by (d) grows "later" since, at the "earlier" time point 2, it increases less, while, at the "later" time point 6, it increases more. Intuitively, the earliness degree of the individual indicated by (c) is more than that of the individual by (c). But we need to quantitatively measure the earliness degree to systematically reflect the difference and comparison. Obviously, two factors are to be considered: increased height and the time span from the time point when increasement occurs to the end time point, and therefore we use their product to represent the earliness degree.

On the basis of the analysis and discussion above, we are now able to calculate the earliness degree of individual (c) as the sum of the areas of two rectangles, one being 4 (increased height) by 7 (time span to the end) and the other being 1 (increased height) by 3 (time span). The area sum is $4 \times 7 + 3 \times 1 = 31$, which can be standardized to be $31/45 = 0.689$, by being divided by the area of the entire rectangle of 5 (total increased height) by 9 (whole time span from beginning to end).

Similarly, the earliness degree of (d) is calculated as $(1 \times 7 + 4 \times 3)/45 = 0.422$, which is much less than 0.689, the earliness degree of (c).

Using the same method we can calculate the earliness degree of the other 4 individuals, with those of (a) and (b) being trivially 1 and 0, respectively. But the cases of (e) and (f) are more complicated, since the height of (e), before reaching the end time point, has increased to a value 8, a greater value than that of the mature state, 7; and the height of (f) has decreased to a value 1, a smaller value than the beginning value 2. Nevertheless, the earliness degree of (e) can be calculated as $(6 \times 9 - 1 \times 2)/45 = 1.156$, a value greater than 1, and that of (f) as $(-1 \times 9 + 1 \times 6)/45 = -0.067$, a negative value.

We denote the quantitative earliness degree as E-index for short. Though imaginary and impossible in real life, the scenario presented in Figure 3 gives us a helpful intuition and clue to define the E-index rigidly.

To give the definition, we do not require a growth curve to be globally differentiable, but it is currently required to be piecewise differentiable, which as we will see later is not necessarily met. And we hence give the definition of E-index.

Definition 1. Suppose that the growth curve is a continuous function $f(t)$ defined on a closed interval $[a, b]$, $a < b$. $\{p_i\}_{i=0}^n$ is a sequence of points, $p_0 = a$, $p_n = b$, $0 \le i < j \le n$. Also suppose that $f(t)$ is differentiable on the open interval $(p_{i-1}, p_i), i = 1, 2, \ldots, n$, and that $f'(t)$ is its derivative function. Then we define the E-index of the growth curve as follows:

$$E_a^b(f) = \frac{1}{(b-a)(f(b) - f(a))} \sum_{i=1}^{n-1} \int_a^b f'(t)(b-t)\, dt. \quad (3)$$

It should be noted that $f'(t)$, the growth rate at time point t, is undefined on each inner split point. But this would not change the integration result, even if we set the growth rate at such point to be any value. For simplicity, we denote $E_a^b(f)$ as $E(f)$.

How early or how late the growth rate $f'(t)$ occurs is our key concern for growth, and the expression $f'(t)(b-t)$ in (3) quantifies the degree of earliness. The greater the product value $f'(t)(b-t)$ is, the earlier the growth or development occurs. In this sense, the E-index measures how early growth occurs in the whole process by accumulating the product along the time.

2.2. Properties of E-Index. From the definition above, we can derive several of E-index's properties which are to be discussed in the form of propositions. However, their proofs are all omitted since they can be found in calculus textbooks or related literature.

Proposition 2. *If the growth curve function $f(t)$ defined on the closed interval $[a, b]$ is strictly monotonically increasing and is globally differentiable, then its E-index can be calculated with the following equation:*

$$E(f) = \frac{1}{(b-a)(f(b) - f(a))} \left(\int_a^b f(t)\, dt \right.$$

$$\left. - f(a)(b-a) \right) = \frac{\int_a^b f(t)\, dt}{(b-a)(f(b) - f(a))} \quad (4)$$

$$- \frac{f(a)}{(f(b) - f(a))}.$$

Proposition 2 provides us an alternative approach to calculate the E-index and reveals to us the relations between the integrations along horizontal direction and along vertical direction.

Proposition 3. *The conclusion of Proposition 2 still holds if the growth curve function is still globally differentiable, but not necessarily monotonic.*

(a)

(b)

(c)

(d)

(e)

Earliness accumulated
Earliness lost
Earliness both accumulatedand lost

(f)

Earliness accumulated
Earliness lost
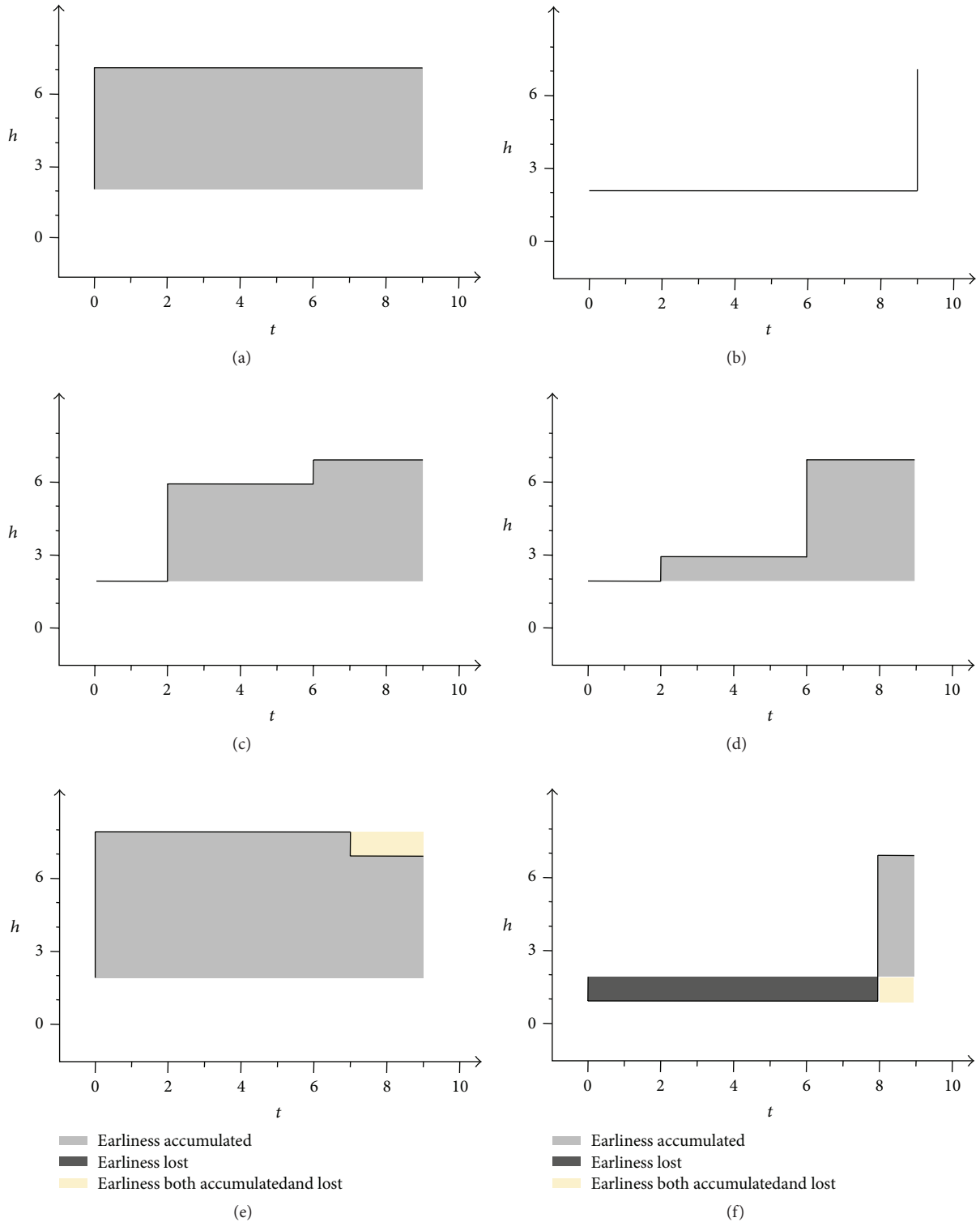Earliness both accumulatedand lost

FIGURE 3: An imaginary scenario about growth. The time interval of the growth is from 0 to 9, and the measurement values are all 2 at the beginning point and are all 7 at the end point.

We can illustrate the proof with Figure 4. Suppose there is only one inner extreme point (we can prove it similarly with more inner extreme points). The integration for the left part of the curve forms the red area, while that for the right part forms the green area which is negative. And their sum is the area between the curve and the horizontal line $y = f(a)$, which is the conclusion we wanted.

Proposition 4. *The conclusion of Proposition 3 still holds if the growth curve function is piecewise differentiable.*
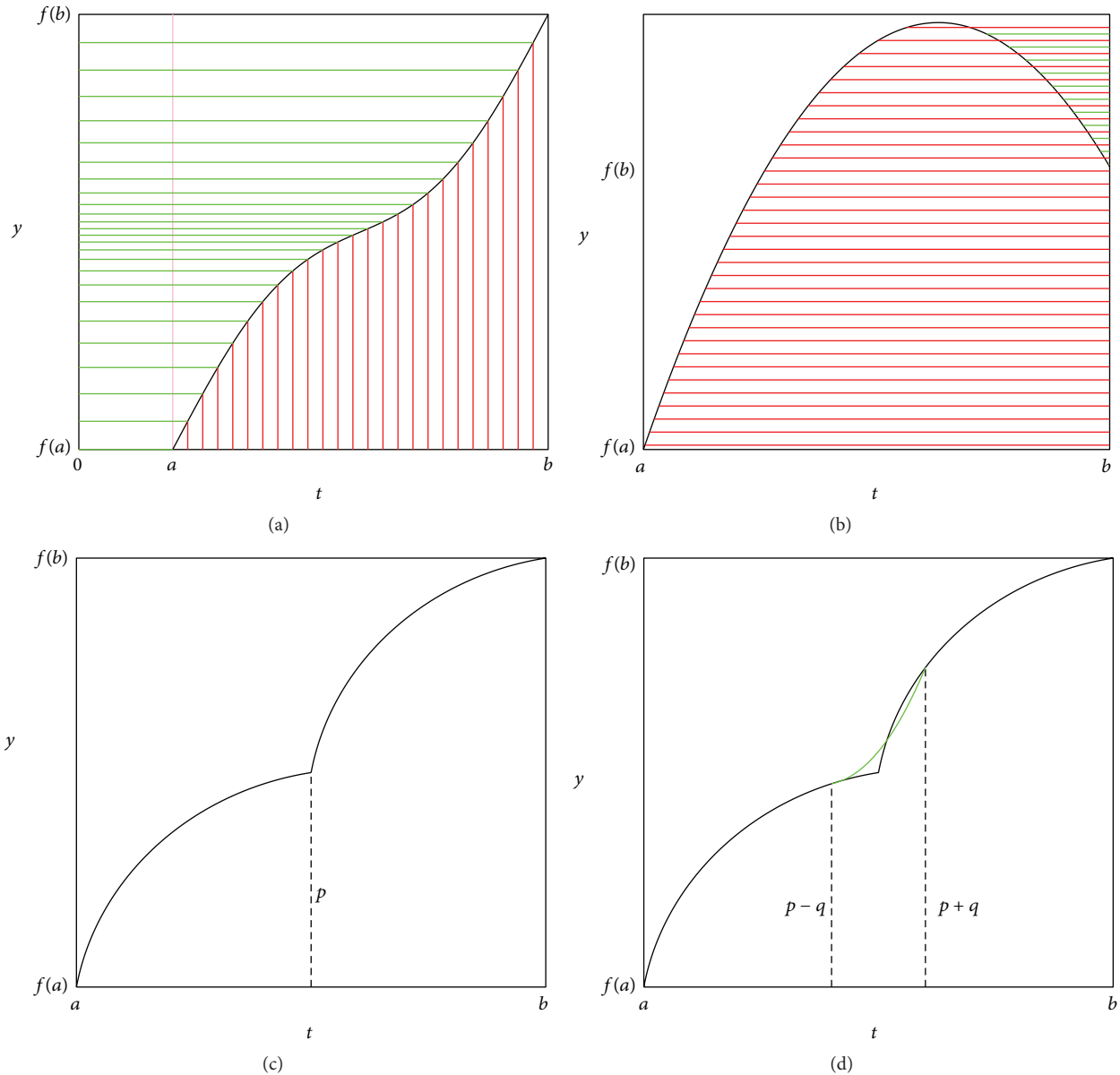
FIGURE 4: Illustrations for the proposition proofs. (a) is a growth curve strictly monotonically increasing. (b) is a nonmonotonic growth curve. (c) is a piecewise smooth growth curve. (d) is a resulting growth curve by smoothening that in (c) near the unsmooth internal point.

Propositions 2–4 indicate that we can calculate E-index with (4), no matter whether the growth curve is monotonic or not and no matter whether it is piecewise smooth or globally smooth. In fact, if the growth curve function is not differentiable, even not continuous, we are still able to calculate its E-index with (4), without changing the meaning of E-index.

Typically, the measurement value in the growth process is between the value at the beginning and that at the end. And we have still another proposition for this situation.

Proposition 5. *If the growth curve function $f(t)$ is defined on $[a, b]$ and $f(a) \leq f(t) \leq f(b)$, then $0 \leq E(f) \leq 1$.*

But $E(f)$ may get a value greater than 1 if, for some t, $f(t)$ is greater than $f(b)$ or even be a negative value if, for some t, $f(t)$ is less than $f(a)$.

Proposition 5 dictates the range of E-index, and we can design a growth curve function whose E-index is any designated value in the range.

3. Validating E-Index's Effectiveness

E-index can be easily calculated with integration operation stated in (4). However, is it as effective as the function parameters, say, r in (1) and (2), to differentiate growth curves? Or in

addition, is it able to differentiate the growth curves without a uniform function in Figure 2? Growth curves are usually formed by collecting successive measurements and finding a function (sometimes difficult to find) approximately fitting the data. But is E-index applicable in this situation? We will answer all these questions in the following subsections.

3.1. Contrasting E-Index and Function Parameters.

The function parameter r in the 2 bunches of growth curves can differentiate the curves, as is illustrated in Figure 1. We are trying to find out whether E-index is capable of doing so, and the results are shown in Figure 5.

For each of the 25 values of r in Figure 1(a), the E-index of the corresponding growth curve function is calculated with (4) as follows:

$$E(r) = E(f_r(\cdot)) = \frac{1}{8} \int_0^8 \left(\frac{2}{1 - e^{-rt}} \right) - 1. \qquad (5)$$

The relation between the values of the parameter r and the corresponding E-index values is plotted in Figure 5(a). It can be observed that as r increases, the E-index value increases accordingly, which implies that E-index is capable of differentiating growth curves as the parameter r that is related to growth rate. For the bunch of growth curves in Figure 1(b), we can obtain similar result illustrated in Figure 5(b).

3.2. E-Index Applied in Nonuniform Functions.

In some cases, we may use E-index as an equivalent of function parameters, to differentiate growth curves of a uniform type. In addition, we may continue to apply E-index to differentiate them without uniform function describing them. For instance, the 4 types of growth curves in Figure 2 are lymphoid, neural, general, and genital, respectively. Specifically, we can use the following 4 functions to precisely describe them:

$$f_{\text{Lymphoid}}(t) = 810e^{-(t-17.43)^2/153.1175} - 110.5,$$

$$0 \le t \le 20,$$

$$f_{\text{Neural}}(t) = \frac{200}{1 + e^{-0.35t}} - 100, \quad 0 \le t \le 20,$$

$$f_{\text{General}}(t) = \frac{108}{1 + e^{-0.32(t-11)}} - 2.1, \quad 0 \le t \le 20,$$

$$f_{\text{Genital}}(t) = \begin{cases} \dfrac{16}{1 + e^{-0.35(t-1)}} - 6, & 0 \le t \le 12, \\ \dfrac{100}{1 + e^{-(t-17.7)}} + 9.2, & 12 < t \le 20. \end{cases}$$

$$(6)$$

Applying (4) once again to the above functions, we will obtain the corresponding E-indices as 1.328, 0.802, 0.459, and 0.191 for the lymphoid, neural, general, and genital type of growth curves, respectively. This result is consistent with our observation and intuition: growth and development occur earliest for the lymphoid type, comparing to the other three types of growth curves; and the genitals grow and develop latest among the four types.

This example illustrated that E-index may, at least in some cases, differentiate growth curves even without uniform function describing them.

3.3. E-Index of Spline Interpolation.

In order to differentiate growth curves by function parameters, we have to assume the function type first and then calculate parameters making the function fit the successively collected measurement values best. The resulting parameters do not work well if the function does not fit the data well.

In fact, spline interpolation performs well to find a smooth function piecewise defined by polynomials. Unfortunately, splines are not uniform functions and therefore, function parameters do not work either for the case of splines. E-index, however, does work in this situation. Based on the successively collected measurements, we can define a smooth function to fit the data by spline interpolation and then calculate the E-index of the function. The resulting E-indices will provide help to differentiate the corresponding growth indicated by the collected measurements.

We will consider 2 growth curves. The first one is described by (1) with the parameter $r = 1$. Suppose that we do not have any knowledge of the curve type and all that we have is the function values of 5 interpolation points evenly dispersed in the time domain. A typical kind of spline, cubic spline, is calculated and compared to the original curve in Figure 5(c). The second growth curve is described by (2) with $r = 1$, and the derived cubic spline and the original curve are contrasted in Figure 5(d).

It is observed from Figures 5(c) and 5(d) that the spline fits the original growth curve well (and will fit it better with more interpolation points), which indicates that E-index works well even without knowledge of the growth curve type.

Next, spline is calculated for each of the functions in the bunch illustrated in Figure 1(a). For the same value of the parameter r, E-index of the original function and that of the spline are calculated, respectively. And the obtained E-index values are contrasted in Figure 5(e). Similarly, the E-index values are contrasted in Figure 5(f) for the original functions illustrated in Figure 1(b) and their splines. The results are encouraging, since E-index values of the splines are quite close to those of the original functions if the number of inner interpolation points is 5 or more (see Figures 5(e) and 5(f)).

3.4. Statistical Framework of E-Index.

How can we apply E-index to differentiate complex dynamic traits? We are typically given two genotypes A and B with m samples of A and n samples of B, each sample measured at T time points. And our purpose is to judge whether the genotypes significantly affect the phenotypes.

Suppose that the value vector of the ith sample of A is $v_i^A = (v_{i,1}^A, v_{i,2}^A, \ldots, v_{i,T}^A)$, $i = 1, 2, \ldots, m$, and that of the jth sample of B is $v_j^B = (v_{j,1}^B, v_{j,2}^B, \ldots, v_{j,T}^B)$, $j = 1, 2, \ldots, n$. The computation steps are as follows, using t-test of $(m + n - 2)$ degrees of freedom to discover the significance.

(a) For each v_i^A and v_j^B, we can use spline interpolation to get continuous functions (curves) $f_i^A(t)$ and $f_j^B(t)$.
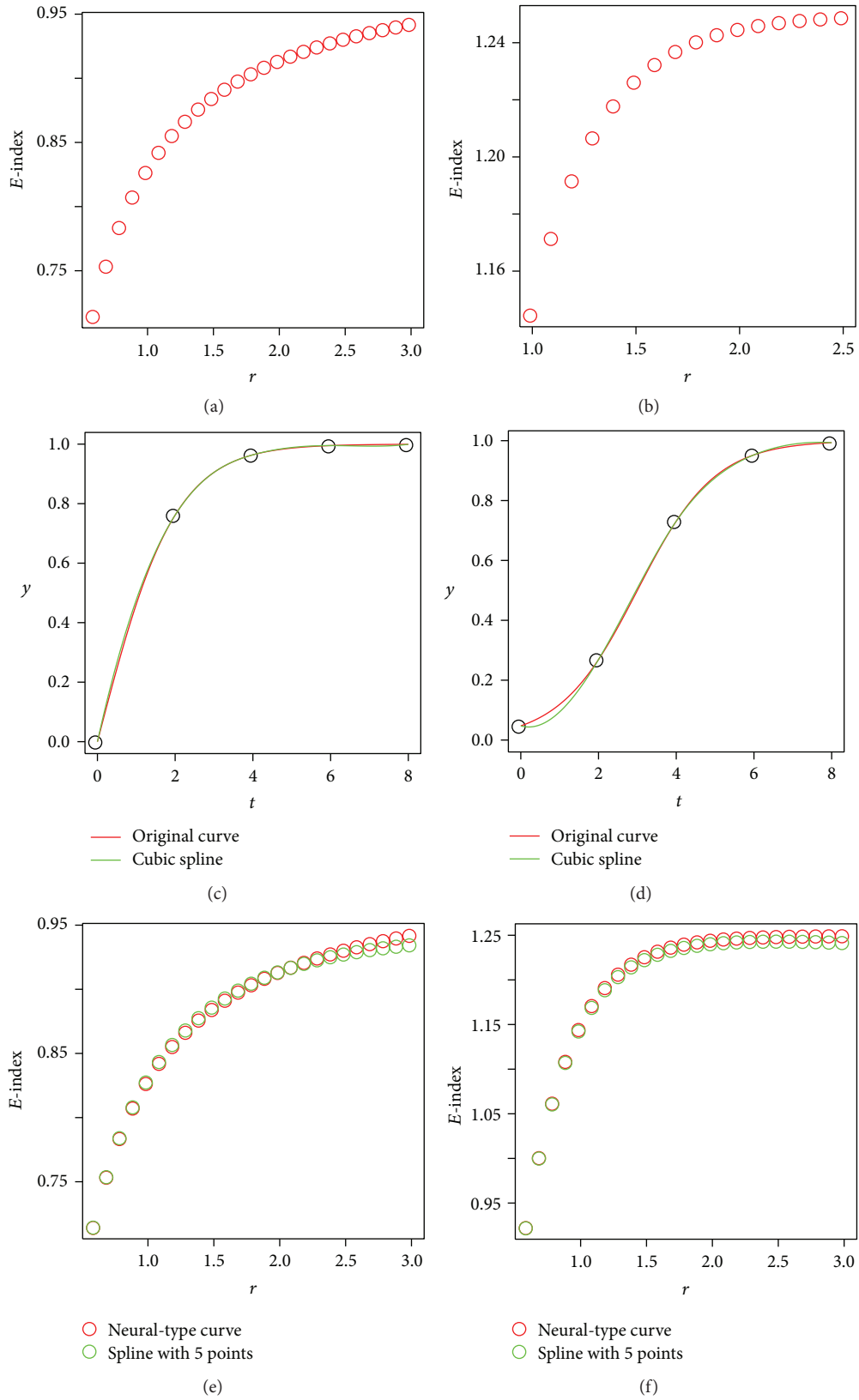
FIGURE 5: Simulations validating the effectiveness of E-index. (a) is E-indices of growth curves in Figure 1(a) plotted against parameter r. (b) is E-indices of growth curves in Figure 1(b) plotted against parameter r. (c) contrasts a growth curve in Figure 1(a) with $r = 1$ and its spline. (d) contrasts a growth curve in Figure 1(b) with $r = 1$ and its spline. (e) compares E-index values of the bunch of functions in Figure 1(a) and those of their splines. (f) compares E-index values of the bunch of functions in Figure 1(b) and those of their splines.

(b) Then (4) is employed to calculate each E-index of the curves, namely, E_i^A and E_j^B, respectively, for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$. And we denote that $E^A = (E_1^A, E_2^A, \ldots, E_m^A)$ and $E^B = (E_1^B, E_2^B, \ldots, E_n^B)$.

(c) After that we define a test statistic,

$$t = \frac{\overline{E_A} - \overline{E_B}}{\sqrt{s^2\,(1/m + 1/n)}}, \tag{7}$$

where $\overline{E_A}$ is the mean of the vector E^A and $\overline{E_B}$ is the mean of the vector E^B (here we suppose that $\overline{E_A} > \overline{E_B}$) and the common variance

$$s^2 = \frac{(m-1)\,s_A^2 + (n-1)\,s_B^2}{m+n-2}, \tag{8}$$

where s_A^2 is the sample variance of the vector E^A and s_B^2 is that of E^B.

(d) We can test the null hypothesis that the two groups of samples are not significantly different:

$$H_0: \overline{E_A} - \overline{E_B} = 0, \tag{9}$$

versus the alternative hypothesis that the two groups are significantly different:

$$H_1: \overline{E_A} - \overline{E_B} > 0. \tag{10}$$

H_0 will be rejected if

$$t > t_\alpha\,(m+n-2); \tag{11}$$

otherwise H_0 will be accepted, where t is the computation result of (7) and $t_\alpha(m+n-2)$ is the t-distribution value with the confidence level α and $(m + n - 2)$ degrees of freedom.

3.5. Applying E-Index. With the statistical framework of E-index in the previous subsection, we can now apply it to differentiate complex dynamic traits.

Two bunches of growth curves of genotypes A and B, respectively, are generated by simulation, and they, together with their E-indices, are illustrated in Figure 6.

The relative measurement value of each sample at the beginning time point 0 is 0 percent, and that at the ending time point 20 is 100 percent. Consequently, we are not able to differentiate them merely by the measurement value at the mature state and have to resort to the difference of developmental processes.

Intuitively, the two groups are far apart. But we fail to apply the functional mapping framework to differentiate the two groups, since there exist no parameters like r in (1) and (2) responsible for the curve shape. In addition, each curve is nonmonotonic, which functional mapping is not able to deal with.

But using the statistical framework given in Section 3.4, we get the standard deviations of the E-indices for genotypes A and B, $s_A = 0.0428$ and $s_B = 0.0597$, respectively. We have
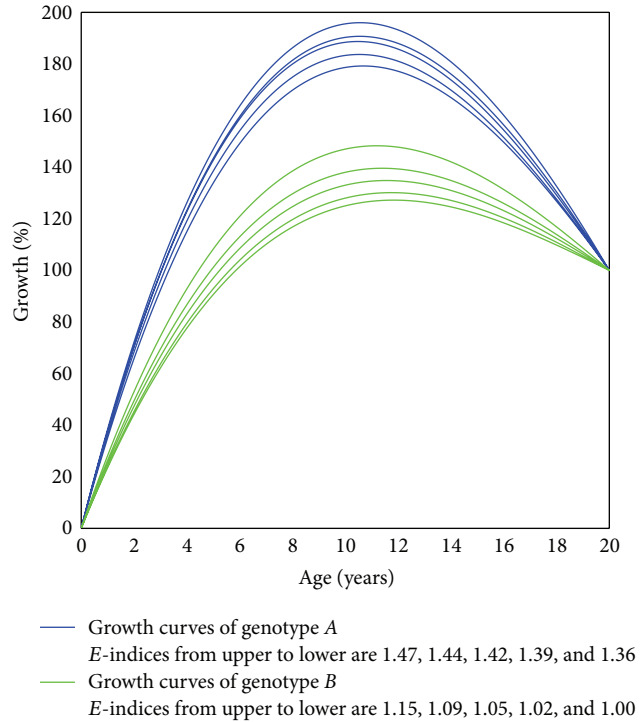

FIGURE 6: Simulated organ growth curves of different genotypes A and B with sample size 5 each.

— Growth curves of genotype A
E-indices from upper to lower are 1.47, 1.44, 1.42, 1.39, and 1.36
— Growth curves of genotype B
E-indices from upper to lower are 1.15, 1.09, 1.05, 1.02, and 1.00

by (8) the common sample variance $s^2 = 0.0027$ and by (7) the test statistic $t = 10.77$, much larger than $t_{0.01}(5+5-2) = 2.896$, which means that there is sufficient evidence to indicate that the genotypes are clearly responsible for the developmental processes of the organs.

4. E-Index Vector

As is mentioned earlier, a curve may theoretically provide infinite amount of information about growth. Though the E-index is sometimes capable of differentiating growth curves, it is after all only one characteristic value revealing one aspect of information. Therefore, it is natural for us to extend E-index into E-index vector.

4.1. Definition of E-Index Vector. Where and why is E-index insufficient to differentiate growth curves? An example from Figures 7(a), 7(c), and 7(e) will illustrate this.

Comparing the sizes of shade in Figures 7(a), 7(c), and 7(e), we will find that 3 totally different growth curves lead to the same E-index value (0.5). This is mainly due to the fact that the effect caused by the higher growth rate in Figure 7(b) or Figure 7(c) is counteracted by lower growth rate earlier or later.

This example indicates that E-index does not work in some cases to differentiate growth curves. Consequently, we have to move forward for a more sophisticated tool. Naturally we will extend E-index into E-index vector.

Definition 6. Suppose that $f(t)$ describing a growth curve is continuous on a closed interval $[a, b]$, $a < b$. And suppose
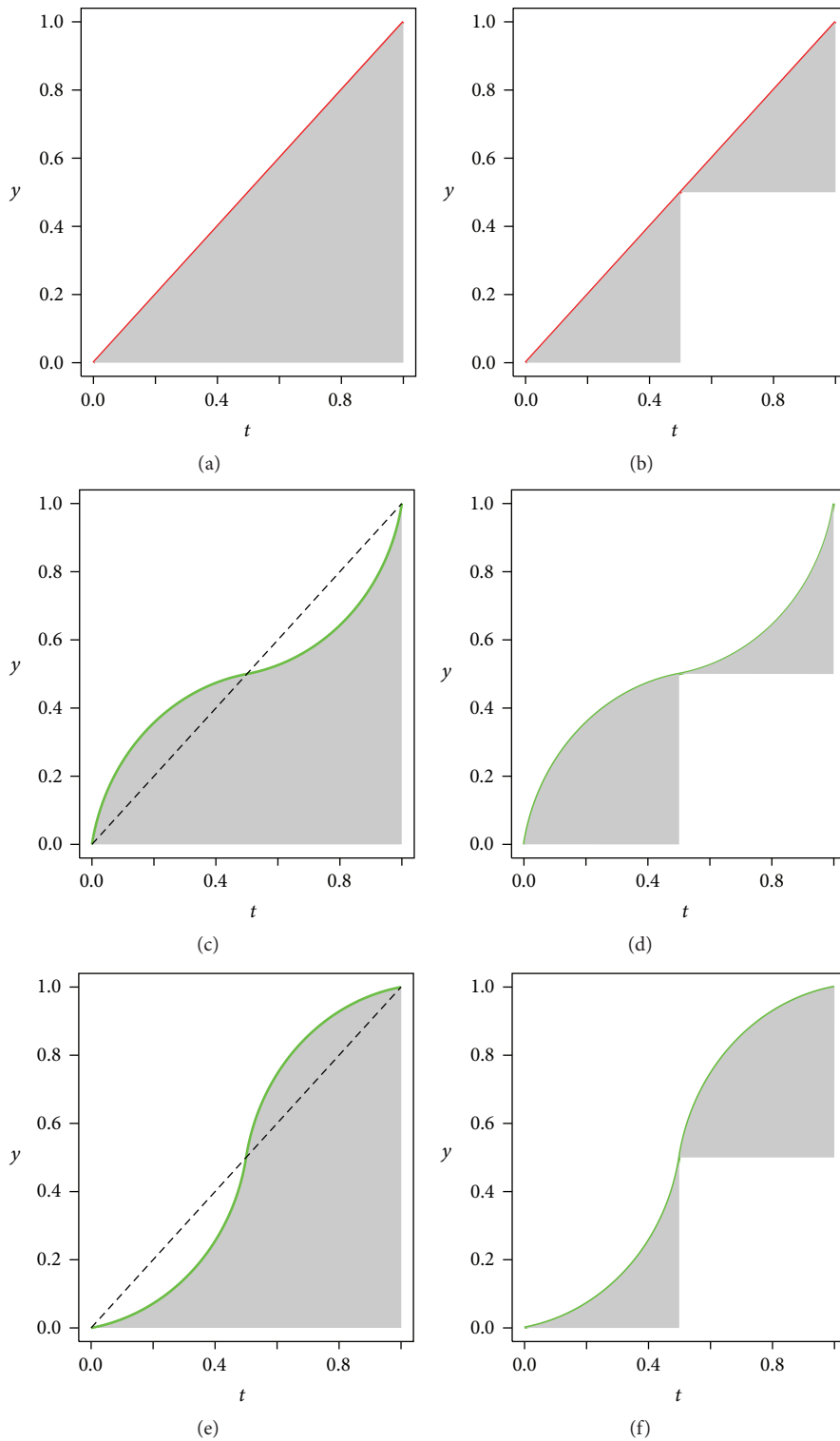
FIGURE 7: The limitation of E-index and the definition of E-index vector. (a) and (b) illustrate E-index and E-index vector for constant growth rate, respectively. (c) and (d) illustrate E-index and E-index vector for higher growth rate in earlier half and lower growth rate in later half. (e) and (f) illustrate E-index and E-index vector for lower growth rate in earlier half and higher growth rate in later half. E-indices of the curves in (a), (c), and (e) are of the same value 1. E-index vectors for the curves in (b), (d), and (e) are (0.5, 0.5), (0.716, 0.284), and (0.284, 0.716), respectively.

that $Q = \{p_i\}_{i=0}^n$ is a prescribed sequence of points, $p_0 = a$, $p_n = b$, $p_i < p_j$, $0 \le i < j \le n$. Then the *E-index vector* of $f(t)$ according to Q is defined as follows:

$$V_Q(f) = \left(nE_{p_0}^{p_1}(f), nE_{p_1}^{p_2}(f), \ldots, nE_{p_{n-1}}^{p_n}(f)\right). \tag{12}$$

$V_Q(f)$ is denoted as $V(f)$ for simplicity and $(nE_{p_0}^{p_1}(f),$ $nE_{p_1}^{p_2}(f), \ldots, nE_{p_{n-1}}^{p_n}(f))$ as $(E_1(f), E_2(f), \ldots, E_n(f))$.

The *E*-index vectors of the growth curves in Figures 7(b), 7(d), and 7(f) are calculated with (12). And the three resulting *E*-index vectors, (0.5, 0.5), (0.716, 0.284), and (0.284, 0.716), are apparently different, as is also illustrated with shaded areas in Figure 7. But how can we evaluate this difference quantitatively? In order to answer this question, we above all give another definition as the following.

Definition 7. Suppose $f(t)$ and $g(t)$ are two functions defined on a closed interval, describing two growth curves with the same prescribed sequence of points $Q = \{p_i\}_{i=0}^n$. Then the *growth dissimilarity* between $f(t)$ and $g(t)$ is defined as follows:

$$D(f, g) = \left[\sum_{i=1}^n \left(E_i(f) - E_i(g)\right)^2\right]^{1/2}. \tag{13}$$

It is easy to prove that growth dissimilarity satisfies distance axioms; that is,

$$D(f, f) = 0,$$
$$D(f, g) = D(g, f), \tag{14}$$
$$D(f, g) + D(g, h) \ge D(f, h).$$

And we hence have transformed problems about growth curves into problems about vectors which will help to analyze the relation between different growth curves in Figure 7. Denote the functions describing the growth curves in Figures 7(b), 7(d), and 7(f) as $f_1(t)$, $f_2(t)$, and $f_3(t)$, respectively. According to (13), their dissimilarities are calculated and listed as follows:

$$D(f_1, f_2) = 0.355,$$
$$D(f_1, f_3) = 0.355, \tag{15}$$
$$D(f_2, f_3) = 0.611.$$

The results above show that, in the growth perspective of earlier and later halves, $f_2(t)$ is more similar to $f_1(t)$ than it is to $f_3(t)$, which is consistent with what is observed in Figure 7.

Different weights can be designated to differently important phases of growth according to specific problems. So the growth dissimilarity defined in (13) can be accordingly redefined as the following equation with W_i denoted as the ith weight:

$$D(f, g) = \left[\sum_{i=1}^n W_i \left(V_i(f) - V_i(g)\right)^2\right]^{1/2}. \tag{16}$$

Equation (16) enables *E*-index vector to help differentiate two growth curves. What is more important, however, is to differentiate a set of growth curves or to divide them into groups or clusters, which will be discussed in the next subsection.

4.2. Grouping or Clustering Growth Curves by E-Index Vector. More and more growth traits are available and they can be described by growth curves. Studies [12] indicate that growth traits are powerful to identify genes some of which cannot be identified by traits only in one time point.

In order to identify genes with growth traits, we are required to divide into groups all growth curves that are as similar as possible in the same group while being as dissimilar as possible in different groups. But a common situation we are encountering is that it is difficult for us to obtain reasonable groups.

E-index, *E*-index vector, and the growth dissimilarity definition based on these two concepts may help us to group or cluster the growth curves.

With the *E*-index vectors, we can define describing rules for a curve group. Take the growth curves in Figures 7(b), 7(d), and 7(e), for instance. We define the rule describing the first group as "the first component of the *E*-index vector is greater than 0.6 and the second less than 0.4." And the second rule is defined as "the first component is less than 0.6 and the second greater 0.4." These two rules describe and define two groups, with growth curve (d) in the first group and growth curves (b) and (e) in the second one.

Though describing rules are capable of grouping growth curves, human experts are involved in prescribing the rules, and thus the rules, consequently the grouping results, may be different from person to person.

Unlike the grouping technique with describing rules, the clustering technique, k-mean algorithm, and its variations are almost automatic. It is a recursive algorithm with k randomly selected centers. To cluster the growth curves, the growth dissimilarities between each growth curve and each center are calculated with (16), and a group corresponding to a center will include all the growth curves nearer to its center than to the other centers. This process continues until the inclusion of each group keeps unchanged.

We simulated the k-mean algorithm by randomly generating 60 growth curves and dividing them into 5 clusters by the algorithm, according to *E*-index vector definition in (12) and growth dissimilarity definition in (16). In Figure 8, the primitive growth curves and the resulting groups are displayed. It can be seen from Figure 8 that each group represents a distinct growth style.

5. Conclusions

In order to generalize functional mapping and overcome the shortages of it, *E*-index and *E*-index vector are proposed in this paper, respectively, by means of measuring earliness degree of growth or development in the overall process and in a growth phase. We summarize their features as follows.
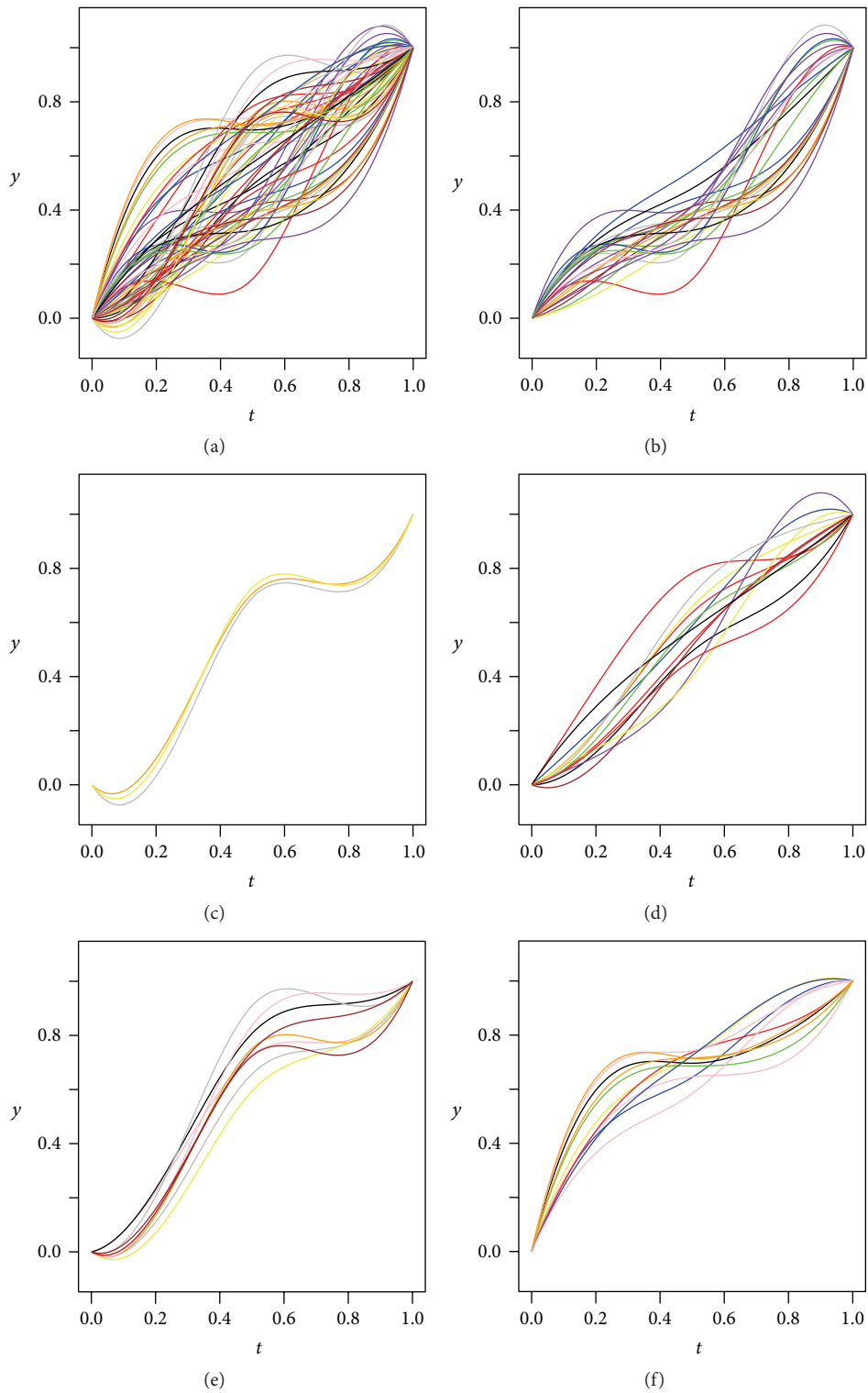
FIGURE 8: Clustering growth curves based on their E-index vectors. (a) illustrates 60 randomly generated growth curves, with a random color designated to each of them. (b)~(f) are the resulting groups after applying k-mean algorithm, with a group in each figure. Each growth curve in (b)~(f) retains its shape, position, and color from its original in (a).

(i) E-index is capable of differentiating growth curves (such as logistic curves) as function parameters are in the applications of functional mapping. Like functional mapping, E-index is good at differentiating growth trajectories with the same values of mature state, which traditional QTL takes as the same. In this sense, E-index generalized functional mapping.

(ii) E-index is sometimes unavailable according to its primitive definition given in Definition 1, due to strict restrictions. But it is always available for a growth curve and easier to calculate from another perspective stated in Proposition 4. Moreover, measuring the earliness degree of growth or development, E-index is as biologically meaningful as the important curve parameters employed in functional mapping.

(iii) A function globally and thoroughly describing the process of growth is unnecessary for calculating E-index. In fact, a cubic spline (as employed in [32]) approximates that well. Furthermore, E-index can be applied in any period of the developmental process; on the contrary, functional mapping can only be applied to the whole process in order to get suitable parameters.

(iv) Being a key and general characteristic value though, E-index provides limited information. As an extension of E-index, E-index vector is focused on the growth in different phases, the number of which may vary, and the time spans for the same vector may not be of equal length, according to the application background and requirements.

(v) By extracting the growth information in a curve and forming a vector, we can use well developed techniques for analysis, such as describing rules and k-mean algorithm.

(vi) E-index vector helps us reveal detailed characteristics in growth curves. It may be looked on as a microscope employed to observe a desired level of growth detail.

Competing Interests

The authors declare that there are no competing interests regarding the publication of this paper.

Acknowledgments

References

[1] T. F. C. MacKay, E. A. Stone, and J. F. Ayroles, "The genetics of quantitative traits: challenges and prospects," *Nature Reviews Genetics*, vol. 10, no. 8, pp. 565–577, 2009.

[2] E. S. Lander and S. Botstein, "Mapping mendelian factors underlying quantitative traits using RFLP linkage maps," *Genetics*, vol. 121, no. 1, pp. 185–199, 1989.

[3] C. Neuschl, G. A. Brockmann, and S. A. Knott, "Multiple-trait QTL mapping for body and organ weights in a cross between NMRI8 and DBA/2 mice," *Genetical Research*, vol. 89, no. 1, pp. 47–59, 2007.

[4] L. J. Leamy, D. Pomp, E. J. Eisen, and J. M. Cheverud, "Pleiotropy of quantitative trait loci for organ weights and limb bone lengths in mice," *Physiological Genomics*, vol. 2002, no. 10, pp. 21–29, 2002.

[5] C. Fan, Y. Xing, H. Mao et al., "GS3, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein," *Theoretical and Applied Genetics*, vol. 112, no. 6, pp. 1164–1171, 2006.

[6] G. Shao, S. Tang, J. Luo et al., "Mapping of qGL7-2, a grain length QTL on chromosome 7 of rice," *Journal of Genetics and Genomics*, vol. 37, no. 8, pp. 523–531, 2010.

[7] P. Ramya, A. Chaubal, K. Kulkarni et al., "QTL mapping of 1000-kernel weight, kernel length, and kernel width in bread wheat (*Triticum aestivum* L.)," *Journal of Applied Genetics*, vol. 51, no. 4, pp. 421–429, 2010.

[8] X. Y. Wan, J. M. Wan, L. Jiang et al., "QTL analysis for rice grain length and fine mapping of an identified QTL with stable and major effects," *Theoretical and Applied Genetics*, vol. 112, no. 7, pp. 1258–1270, 2006.

[9] R. L. Wu and M. Lin, "Opinion: Functional mapping—how to map and study the genetic architecture of dynamic complex traits," *Nature Reviews Genetics*, vol. 7, no. 3, pp. 229–237, 2006.

[10] M. Kirkpatrick and N. Heckman, "A quantitative genetic model for growth, shape, reaction norms, and other infinite-dimensional characters," *Journal of Mathematical Biology*, vol. 27, no. 4, pp. 429–450, 1989.

[11] S. D. Pletcher and C. J. Geyer, "The genetic analysis of age-dependent traits: modeling the character process," *Genetics*, vol. 153, no. 2, pp. 825–835, 1999.

[12] R.-L. Wu, M.-X. Wang, and M.-R. Huang, "Quantitative genetics of yield breeding for *Populus* short rotation culture. I. Dynamics of genetic control and selection model of yield traits," *Canadian Journal of Forest Research*, vol. 22, no. 2, pp. 175–182, 1992.

[13] W. R. Atchley and J. Zhu, "Developmental quantitative genetics, conditional epigenetic variability and growth in mice," *Genetics*, vol. 147, no. 2, pp. 765–776, 1997.

[14] J. M. Cheverud, J. J. Rutledge, and W. R. Atchley, "Quantitative genetics of development: genetic correlations among age-specific trait values and the evolution of ontogeny," *Evolution*, vol. 37, no. 5, pp. 895–905, 1983.

[15] W. R. Atchley, "Ontogeny, timing of development, and genetic variance-covariances structure," *The American Naturalist*, vol. 123, no. 4, pp. 519–540, 1984.

[16] W.-R. Wu, W.-M. Li, D.-Z. Tang, H.-R. Lu, and A. J. Worland, "Time-related mapping of quantitative trait loci underlying tiller number in rice," *Genetics*, vol. 151, no. 1, pp. 297–303, 1999.

[17] J. M. Cheverud, E. J. Routman, F. A. M. Duarte, B. Van Swinderen, K. Cothran, and C. Perel, "Quantitative trait loci for murine growth," *Genetics*, vol. 142, no. 4, pp. 1305–1319, 1996.

[18] D. Verhaegen, C. Plomion, J.-M. Gion, M. Poitel, P. Costa, and A. Kremer, "Quantitative trait dissection analysis in Eucalyptus using RAPD markers. 1. Detection of QTL in interspecific hybrid progeny, stability of QTL expression across different

ages," *Theoretical and Applied Genetics*, vol. 95, no. 4, pp. 597–608, 1997.

[19] L. C. Emebiri, M. E. Devey, A. C. Matheson, and M. U. Slee, "Age-related changes in the expression of QTLs for growth in radiata pine seedlings," *Theoretical and Applied Genetics*, vol. 97, no. 7, pp. 1053–1061, 1998.

[20] B. Mangin, P. Thoquet, and N. Grimsley, "Pleiotropic QTL analysis," *Biometrics*, vol. 54, no. 1, pp. 88–99, 1998.

[21] A. B. Korol, Y. I. Ronin, A. M. Itskovich, J. Peng, and E. Nevo, "Enhanced efficiency of quantitative trait loci mapping analysis based on multivariate complexes of quantitative traits," *Genetics*, vol. 157, no. 4, pp. 1789–1803, 2001.

[22] C. X. Ma, G. Casella, and R. L. Wu, "Functional mapping of quantitative trait loci underlying the character process: a theoretical framework," *Theoretical and Applied Genetics*, vol. 97, no. 7, pp. 1053–1061, 1998.

[23] G. B. West, J. H. Brown, and B. J. Enquist, "A general model for ontogenetic growth," *Nature*, vol. 413, no. 6856, pp. 628–631, 2001.

[24] R. L. Wu, C.-X. Ma, X.-Y. Lou, and G. Casella, "Molecular dissection of allometry, ontogeny, and plasticity: a genomic view of developmental biology," *BioScience*, vol. 53, no. 11, pp. 1041–1047, 2003.

[25] R. L. Wu, C.-X. Ma, M. Lin, and G. Casella, "A general framework for analyzing the genetic architecture of developmental characteristics," *Genetics*, vol. 166, no. 3, pp. 1541–1551, 2004.

[26] R. Wu, C.-X. Ma, R. C. Littell, and G. Casella, "A statistical model for the genetic origin of allometric scaling laws in biology," *Journal of Theoretical Biology*, vol. 219, no. 1, pp. 121–135, 2002.

[27] R. L. Wu, C.-X. Ma, M. Lin, Z. Wang, and G. Casella, "Functional mapping of quantitative trait loci underlying growth trajectories using a transform-both-sides logistic model," *Biometrics*, vol. 60, no. 3, pp. 729–738, 2004.

[28] R. L. Wu, Z. H. Wang, W. Zhao, and J. M. Cheverud, "A mechanistic model for genetic machinery of ontogenetic growth," *Genetics*, vol. 168, no. 4, pp. 2383–2394, 2004.

[29] X. Zhao, C. Tong, X. Pang et al., "Functional mapping of ontogeny in flowering plants," *Briefings in Bioinformatics*, vol. 13, no. 3, pp. 317–328, 2012.

[30] L. Jiang, J. Liu, X. Zhu et al., "2HiGWAS: a unifying high-dimensional platform to infer the global genetic architecture of trait development," *Briefings in Bioinformatics*, vol. 16, no. 6, pp. 905–911, 2015.

[31] L. Bogin, *Patterns of Human Growth*, Cambridge University Press, 2nd edition, 1999.

[32] M. Liu, X. Li, R. Fan, X. Liu, and J. Wang, "A systematic analysis of candidate genes associated with nicotine addiction," *BioMed Research International*, vol. 2015, Article ID 313709, 9 pages, 2015.