Orthogonal contrast based models for quantitative genetic analysis in autotetraploid species

Luo, Zewei; Leach, Lindsey; Chen, Jing

DOI:
10.1111/nph.15284

Orthogonal Contrast Based Models for Quantitative Genetic Analysis in Autotetraploid Species

Jing Chen[1*], Fengjun Zhang[2*], Lin Wang[2], Lindsey Leach[1] and Zewei Luo[1,2@]

[1.] School of Biosciences, The University of Birmingham, Birmingham B15 2TT, UK

[2.] Institute of Biostatistics and Genetics, Fudan University, Shanghai 200433, China

Key words: Quantitative Genetic Models, Orthogonal Contrasts, Double Reduction, Tetrasomic Inheritance, Autotetraploids

* these authors contributed equally to the research

Word count total = 6496

Introduction (969); Materials and Methods (2076); Results (2065); Discussion (1356); Acknowledgements (30)

Number of figures = 1 (figure 1 to be published in colour)

Number of tables = 8

Supporting information: (2 figures, 5 methods, 1 note and 6 tables)

@ The corresponding authorship

Prof. Zewei Luo

School of Biosciences

The University of Birmingham

Edgbaston, Birmingham B15 2TT, United Kingdom

Tel: +44 121 414 5404

Fax: +44 121 414 5925

E-mail: z.luo@bham.ac.uk or zwluo@fudan.edu.cn

Summary (200 words)

- Dissecting the genetic architecture of quantitative traits is a crucial goal for efficient breeding of polyploid plants, including autotetraploid crop species, such as potato and coffee, and ornamentals such as rose. To meet this goal, a quantitative genetic model is needed to link the genetic effects of genes or genotypes at quantitative trait loci to the phenotype of quantitative traits.
- We present a statistically tractable quantitative genetic model for autotetraploids based on orthogonal contrast comparisons in the general linear model. The new methods are suitable for autotetraploid species with any population genetic structure and take full account of the essential features of autotetrasomic inheritance. The statistical properties of the new methods are explored and compared to an alternative method in the literature by simulation studies.
- We have shown how these methods can be applied for quantitative genetic analysis in autotetraploids by analysing trait phenotype data from an autotetraploid potato segregating population. Using trait segregation analysis, we showed that both highly heritable traits of flowering time and plant height were under the control of major QTL.
- The orthogonal model directly dissects genetic variance into independent components and gives consistent estimates of genetic effects provided that tetrasomic gene segregation is considered.

Key words: autotetraploids, double reduction, orthogonal, polyploid, potato, quantitative genetic model

Introduction

Polyploidy plays an important role in the evolution of eukaryotes, especially for flowering plants, all of which have undergone at least one round of polyploidization in their evolutionary history (Otto & Whitton, 2000; Jiao *et al.*, 2011). Between 30-80% of species are currently polyploids, while the rest exist as paleopolyploids (Wolfe, 2001), having undergone a gradual process of "diploidization" over evolutionary time. Many of the world's most important crop species are either autopolyploid, for example, the autotetraploid potato, coffee, and alfalfa, or allopolyploid, including wheat, oats and canola. Several economically important aquaculture animals are also autotetraloids, including Atlantic salmon and trout (Danzmann & Garbi, 2001; Vaughn *et al.*, 2007). Therefore, in order to address the global food security crisis, rigorous genetic analysis of autopolyploid species becomes a timely task.

Most biological characters important in organismal evolution and relevant to plant and animal breeding, such as reproductive isolation, yield, quality and resistance to biotic and abiotic stresses, are quantitative traits affected by genes at more than a single locus, as well as by environmental factors. Understanding the polygenic architecture underlying such quantitative traits is essential to enable their genetic improvement as part of effective plant or animal breeding programs. However, progress in quantitative genetic analysis in polyploid species lags far behind compared to that achieved in diploids for several major reasons.

Firstly, polyploids display a much more complicated pattern of gene segregation and recombination than diploids. For example, multiple alleles at individual loci of polyploids cause a substantially wider spectrum of genotypic segregation. In autopolyploids, multivalent pairing of homologous chromosomes during meiosis may result in the phenomenon of double reduction, in which identical alleles carried on sister chromatids enter into the same gamete, resulting in systematic allelic segregation distortion. Our studies (Luo *et al.*, 2006a) show that recombination frequency between a pair of loci can be as high as 75% under a tetrasomic model (compared to 50% in diploids) and that double reduction can occur at a frequency of 25%, showing the remarkable difference in the pattern of gene segregation and recombination between diploid and autopolyploid species. These factors have made polysomic genetic analysis one of the most challenging topics in theoretical and applied genetics since the pioneering works of quantitative geneticists such as Haldane, Mather and Fisher (Haldane, 1930; Mather, 1936; Fisher, 1947).

Secondly, the evolution of polyploid genomes is an extremely dynamic process compared to that of diploids, characterized by extensive genetic and epigenetic changes occurring in the nuclear

genome following polyploidization (Soltis & Soltis, 1995; Song *et al.*, 1995; Comai *et al.*, 2000; Adams & Wendel, 2005). Genome structure and function of polyploids may therefore differ markedly from that of their diploid relatives. This necessitates that breeding programs targeted at improving genetic performance of an autopolyploid species should ideally be conducted at the polyploid level rather than with its diploid counterparts.

The quantitative genetic model which links genetic effects of genes or genotypes at quantitative trait loci to the phenotype of quantitative traits is an essential basis for any quantitative genetic analysis. The theory and methods for modelling and analysing quantitative genetic effects have been well established and routinely practised in diploid species (Mather & Jinks, 1971; Falconer, 1989; Lynch & Walsh, 1998). In contrast, there are no methods currently available for modelling quantitative genetic effects in autotetraploids that take proper account of the complex features of autotetrasomic inheritance.

Early models for the quantitative genotypic effects at a single locus in randomly mating autotetraploid populations (Kempthorne, 1955; Kempthorne, 1957) were intractable for real data analysis because they involve a large number of genetic parameters. Li (1957) developed a simplified two-allele version of Kempthorne's model and proposed successive linear regression of genetic values of genotypes onto the corresponding frequencies in a tetraploid population under Hardy-Weinberg equilibrium (HWE). This model allowed genetic variance at a single locus to be represented by only four major components.

Mather and Jinks (1971) extended their concept of additive and dominance effects for quantitative genetic analysis in diploids to define these effects in tetraploids (Mather & Jinks, 1971). Analysis with any quantitative genetic model involves the distribution of genotypes at quantitative trait loci (QTL) in the population under study. In autotetraploids, this distribution depends on the coefficient of double reduction (Luo *et al.*, 2004). Killick (1971) therefore explored the influence of double reduction on Mather and Jink's additive-dominance model for autotetraploids. Nevertheless, all of the classical additive-dominance models developed either for diploids, autotetraploids, or more recently autohexaploids (van Geest *et al.*, 2017), share the undesirable property of correlation between estimates of different types of effects in the model (Li, 1957; Killick, 1971; Wright, 1979; Li *et al.*, 2010; van Geest *et al.*, 2017). This correlation structure may bias estimation of the model parameters and variance components of the genetic effects. Addressing this limitation, Cockerham (1954) pioneered in developing a quantitative genetic model for diploids based on the principle of orthogonal linear comparison, which enables phenotypic variation of a quantitative trait to be

partitioned in a way that ensures independence between different model effects, enabling direct dissection of genetic variance into independent components (Cockerham, 1954).

This paper presents a novel and statistically tractable tetrasomic quantitative genetic model based on orthogonal contrasts that are suitable for use with either natural or artificially created populations of autotetraploid species. The model represents the first example of quantitative genetic models for autotetraploids that take account of the essential features of tetrasomic inheritance, including double reduction, while retaining computational feasibility. The statistical properties of these new models are explored and compared with another method in the literature by computer simulation analyses. We have demonstrated their utility in quantitative genetic analyses of autotetraploid species by analysing trait phenotype data from an outbred segregating population of autotetraploid potato.

Materials and Methods

General one locus model

We first consider segregation of two alleles (*A* and *a*) at a single locus in an autotetraploid population. There are a total of 5 possible genotypes at the biallelic locus, namely *AAAA* (quadruplex), *AAAa* (triplex), *AAaa* (duplex), *Aaaa* (simplex) and *aaaa* (nulliplex). The *i*th genotype A_ia_{4-i}, is defined with a genotypic value of G_i and its frequency in the population is denoted by f_i , with i = 0, 1, …, 4 indicating the number of *A* alleles involved in the genotype, as shown in Table 1. In practice, there may be more than two QTL alleles. However, these may be grouped into the two classes of either increasing alleles or decreasing alleles, based on their effects on the trait phenotype. This effectively reduces the maximum number of possible genotypes at a locus down to 5 (++++, +++-,++--,+---,----) in any population, creating a tractable model.

We define here the genotypic effect for an individual through a regression model of allelic effects

$$G = \mu + x_1\theta_1 + x_2\theta_2 + x_3\theta_3 + x_4\theta_4 \qquad \text{Eqn 1}$$

where μ is the population mean, and θ_i (i = 1,…,4) are accordingly the monogenic, digenic, trigenic and quadrigenic genetic effects of the QTL, and x_i (i = 1, …, 4) are the corresponding genetic effect design variables. The monogenic effect will always be positive and represents the average effect caused by substituting allele *A* for allele *a* at the QTL. The digenic effect represents the average interaction effect between two alleles in a tetraploid genotype, denoted I_{Aa} in the biallelic model. The trigenic effect represents the average interaction effects among three alleles**.** Existence of a trigenic effect means that the interaction between two alleles differs according to the identity of the

third allele. In the model, there are two different three-way interactions, given by I_{AAa} and I_{Aaa}. The quadrigenic effect represents the average interaction effects among four alleles. Existence of quadrigenic effects means that the interaction between two alleles differs depending on the identity of the third and fourth alleles. In the biallelic model, there are three different four-way interactions, given by I_{AAAa}, I_{AAaa} and I_{Aaaa}. If there are no two-way, (three-way, four-way) allelic interactions, then the corresponding monogenic (trigenic, quadrigenic) genetics effects will be equal to zero. A more detailed explanation of the genetic effects is given in Supporting Information Method S1, Fig. S1 and Table S1.

Estimation of genetic effects in the one locus model

In a natural autotetraploid population, genotypic frequencies vary across different loci in the genome and are usually not in Hardy-Weinberg equilibrium (Luo *et al.*, 2000). Orthogonal contrasts provide a way to partition genetic variance into independent components (Zeng *et al.*, 2005). We propose here general orthogonal scales w_{ij} for the genetic effects of genotype *i* for the *j*th contrast (*j* = 1,2,...,4), corresponding to monogenic, digenic, trigenic and quadrigenic genetic effects. The orthogonal scales are summarized in Table 1 and must satisfy a number of requirements to ensure that the comparisons are orthogonal, *i.e.* uncorrelated, as follows.

$\sum_{i=0}^{4} f_i = 1$; and

1). For monogenic effects $w_{i1}\left(i = 0,1,...,4\right)$

$$\begin{cases} \sum_{k=0}^{4} w_{k1} f_k = 0 \\ w_{i1} = w_{01} + i \qquad (i = 1,\mathrm{K}\ ,4) \end{cases}$$

2). For digenic effects $w_{i2}\left(i = 0,1,...,4\right)$

$$\begin{cases} w_{(i+1)2} - 2w_{i2} + w_{(i-1)2} = 1 \qquad (i = 1,2,3) \\ \sum_{k=0}^{4} w_{k2} f_k = 0 \\ \sum_{k=0}^{4} w_{k2} w_{k1} f_k = 0 \end{cases}$$

3). For trigenic effects, $w_{i3}\left(i = 0,1,...,4\right)$

$$\begin{cases} w_{(i+1)3} - 3w_{i3} + 3w_{(i-1)3} - w_{(i-2)3} = 1 \qquad \left(i = 2,3\right) \\ \sum_{k=0}^{4} w_{k3} f_k = 0 \\ \sum_{k=0}^{4} w_{k3} w_{k1} f_k = 0 \\ \sum_{k=0}^{4} w_{k3} w_{k2} f_k = 0 \end{cases}$$

4). For quadrigenic effects, $w_{i4}\,(i=0,1,...,4)$

$$\begin{cases} w_{44}-4w_{34}+6w_{24}-4w_{14}+w_{04}=1 \\ \sum_{k=0}^{4} w_{k4} f_k = 0 \\ \sum_{k=0}^{4} w_{k4} w_{k1} f_k = 0 \\ \sum_{k=0}^{4} w_{k4} w_{k2} f_k = 0 \\ \sum_{k=0}^{4} w_{k4} w_{k3} f_k = 0 \end{cases}$$

The above 1) - 4) ensure the key statistical properties of the orthogonal model as shown by Eqn (1). Firstly, $\sum_{i=0}^{4} w_{ij} f_i = 0$ for $j=1,...,4$ ensures the statistical definition of w_{ij} as contrast scales, which in turn define the design variables x_i in Eqn (1). Secondly, $\sum_{i=0}^{4} w_{ij} w_{ik} f_i = 0$ for $1 \le j \ne k \le 4$ ensures the orthogonality between the contrast scales w_{ij} and w_{ik} ($i=0,...,4$; $1 \le j \ne k \le 4$). The orthogonal scales calculated as above are then used to derive the genetic effect design variables in Eqn (1) as below

$$x_j = \begin{cases} w_{4j} & if\ G\ is\ AAAA \\ w_{3j} & if\ G\ is\ AAAa \\ w_{2j} & if\ G\ is\ AAaa \\ w_{1j} & if\ G\ is\ Aaaa \\ w_{0j} & if\ G\ is\ aaaa \end{cases} \qquad (j = 1, 2,\ldots,4)$$

We can then express the orthogonal model for the QTL effects at locus A in a matrix form given by

$$G_A = \begin{bmatrix} G_4 \\ G_3 \\ G_2 \\ G_1 \\ G_0 \end{bmatrix} = S_A E_A = \begin{bmatrix} 1 & w_{41} & w_{42} & w_{43} & w_{44} \\ 1 & w_{31} & w_{32} & w_{33} & w_{34} \\ 1 & w_{21} & w_{22} & w_{23} & w_{24} \\ 1 & w_{11} & w_{12} & w_{13} & w_{14} \\ 1 & w_{01} & w_{02} & w_{03} & w_{04} \end{bmatrix} \begin{bmatrix} \mu \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{bmatrix} \qquad \text{Eqn 2}$$

where S_A is the genetic effects design matrix and E_A is the genetic effects of the QTL genotypes, which can be calculated from

$$E_A = S_A^{-1} G_A \qquad \text{Eqn 3}$$

Accordingly, the five QTL genotypic values can be specified under the orthogonal model as

$$G_i = \mu + w_{i1}\theta_1 + w_{i2}\theta_2 + w_{i3}\theta_3 + w_{i4}\theta_4 \quad (i=0,1,...,4)$$

The total genetic variance V_G, contributed by allele segregation at the QTL, can be partitioned into four independent components of variance. Each variance component is contributed by its own corresponding genetic parameter as

$$\sigma_t^2 = \frac{\left(\sum_{i=0}^{4} f_i G_i w_{it}\right)^2}{\left(\sum_{i=0}^{4} f_i w_{it}^2\right)} \qquad \text{Eqn 4}$$

where t = 1,2,…,4 corresponds to the four orthogonal scales defined for the monogenic, digenic, trigenic and quadrigenic genetic effects, and i (= 0,1,…,4) indicates the number of A alleles in the QTL genotype. The significance of the estimated genetic effects can be tested using the one- or two-tailed t-test, with the standard error given by $\sqrt{\sigma'^2_t}/\sqrt{n}$ (degree of freedom equals to n-4), where σ'^2_t is the estimated variance for the t^{th} contrast and can be calculated by $\hat{\sigma}_e^2 \cdot \sum_{i=0}^{4} w_{it}^2 / f_i$, where $\hat{\sigma}_e^2$ is the estimated residual variance and n is the sample size. In this work, we characterize and illustrate the model (1) in two specific populations, an S_2 population (below) and a randomly mating population (Supporting Information Method S2), though the model is generic for populations with any given genetic structure. It should be noted that the variance components here refer to genetic variances contributed by monogenic, digenic, trigenic or quadrigenic effects in the model, rather than the variances of the contrasts.

One locus model for an S_2 population

In the second generation segregating population, denoted by S_2, created from crossing two parental autotetraploid lines with genotypes *AAAA* and *aaaa*, the frequencies of the offspring genotypes can be expressed in terms of α, the coefficient of double reduction at the QTL, as, $f_0 = (1+2\alpha)^2/36$, $f_1 = 2(1-\alpha)(1+2\alpha)/9$, $f_2 = [3-4\alpha(1-\alpha)]/6$, $f_3 = 2(1-\alpha)(1+2\alpha)/9$ and $f_4 = (1+2\alpha)^2/36$. The corresponding orthogonal contrast scales are summarized in Table 2. The genotypic values G_A = $(G_4\ G_3\ G_2\ G_1\ G_0)^T$ can be presented in a matrix form given by

$$G_A = S_A E_A = \begin{bmatrix} 1 & 2 & (5-2\alpha)/3 & 2(1-\alpha)/3 & (1-\alpha)\left(4\alpha^2-4\alpha+3\right)/\left(12(2+\alpha)\right) \\ 1 & 1 & (1-4\alpha)/6 & -(1+2\alpha)/6 & -(1+2\alpha)\left(4\alpha^2-4\alpha+3\right)/\left(24(2+\alpha)\right) \\ 1 & 0 & -(1+2\alpha)/3 & 0 & (1-\alpha)(2\alpha+1)^2/\left(12(2+\alpha)\right) \\ 1 & -1 & (1-4\alpha)/6 & (1+2\alpha)/6 & -(1+2\alpha)\left(4\alpha^2-4\alpha+3\right)/\left(24(2+\alpha)\right) \\ 1 & -2 & (5-2\alpha)/3 & -2(1-\alpha)/3 & (1-\alpha)\left(4\alpha^2-4\alpha+3\right)/\left(12(2+\alpha)\right) \end{bmatrix} \begin{bmatrix} \mu \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{bmatrix}$$

The genetic effects of the QTL genotypes can be calculated from $E_A = S_A^{-1} G_A$ where

$$S_A^{-1} = \begin{bmatrix} (1+2\alpha)^2/36 & 2(1+2\alpha)(1-\alpha)/9 & (4\alpha^2-4\alpha+3)/6 & 2(1+2\alpha)(1-\alpha)/9 & (1+2\alpha)^2/36 \\ (1+2\alpha)/12 & (1-\alpha)/3 & 0 & -(1-\alpha)/3 & -(1+2\alpha)/12 \\ \dfrac{(1+2\alpha)(5-2\alpha)}{12(2+\alpha)} & \dfrac{(\alpha-1)(4\alpha-1)}{3(2+\alpha)} & -\dfrac{(4\alpha^2-4\alpha+3)}{2(2+\alpha)} & \dfrac{(\alpha-1)(4\alpha-1)}{3(2+\alpha)} & \dfrac{(1+2\alpha)(5-2\alpha)}{12(2+\alpha)} \\ 1/2 & -1 & 0 & 1 & -1/2 \\ 1 & -4 & 6 & -4 & 1 \end{bmatrix}$$

General two locus model

The one locus method described above is extended to two biallelic loci, A and B, in an autotetraploid population with a specified genetic structure. There will be twenty-five possible genotypes at the two loci (without accounting for linkage phase). A general form for the two-locus tetraploid genotype may be given as $A_i a_{(4-i)} B_j b_{(4-j)}$ with $i = 0,1,...,4$ for the number of *A* alleles and $j = 0,1,...,4$ for the number of *B* alleles in the genotype. The genotypic value and genotype frequency are denoted by G_{ij} and f_{ij}. The marginal frequencies of the genotypes at locus A and locus B are denoted by $f_{i.}$ and $f_{.i}$ ($i = 0,1,...,4$). Without loss of generality, locus A is assumed to be closer to the centromere than locus B and the coefficients of double reduction at the two loci are denoted by α and β, respectively. A linear model for the genotypic value is comprised of genetic effects at each of the two loci and epistatic effects between the genes at the two loci, and is fully characterized by a total of twenty-five parameters in the form of a regression model of allelic effects analogous to equation (1), as follows:

$$\begin{aligned} G_{ij} = \mu &+ x_1\theta_1 + x_2\theta_2 + x_3\theta_3 + x_4\theta_4 + y_1\zeta_1 + y_2\zeta_2 + y_3\zeta_3 + y_4\zeta_4 + z_{\theta_1\zeta_1} I_{\theta_1\zeta_1} + z_{\theta_1\zeta_2} I_{\theta_1\zeta_2} + \\ & z_{\theta_1\zeta_3} I_{\theta_1\zeta_3} + z_{\theta_1\zeta_4} I_{\theta_1\zeta_4} + z_{\theta_2\zeta_1} I_{\theta_2\zeta_1} + z_{\theta_2\zeta_2} I_{\theta_2\zeta_2} + z_{\theta_2\zeta_3} I_{\theta_2\zeta_3} + z_{\theta_2\zeta_4} I_{\theta_2\zeta_4} + z_{\theta_3\zeta_1} I_{\theta_3\zeta_1} + \\ & z_{\theta_3\zeta_2} I_{\theta_3\zeta_2} + z_{\theta_3\zeta_3} I_{\theta_3\zeta_3} + z_{\theta_3\zeta_4} I_{\theta_3\zeta_4} + z_{\theta_4\zeta_1} I_{\theta_4\zeta_1} + z_{\theta_4\zeta_2} I_{\theta_4\zeta_2} + z_{\theta_4\zeta_3} I_{\theta_4\zeta_3} + z_{\theta_4\zeta_4} I_{\theta_4\zeta_4} \end{aligned} \qquad \text{Eqn 5}$$

where μ is the population mean, θ_i (or ζ_i) (i = 1, …, 4) are accordingly monogenic, digenic, trigenic and quadrigenic genetic effects at locus A (or locus B), and x_i (or y_i) (i=1, …, 4) are the design variables for the corresponding genetic effects at locus A (or locus B), as summarised in Table 3. $I_{\theta_i\zeta_j}$ are the epistasis parameters between the effects θ_i and ζ_j (i=1,…,4; j=1,…,4). For example, $I_{\theta_1\zeta_1}$ is the monogenic x monogenic effect of loci A and B. The corresponding design

variables are given by $z_{\theta_i\zeta_j}$. A full definition of all 25 genetic effect parameters is given in Supporting Information Table S2.

In a similar but algebraically more tedious way to the one locus model, we derived the orthogonal contrast scales for the two-locus tetrasomic model under two different settings regarding the mutual dependency of genotypes at the two loci: linkage equilibrium (Supporting Information Method S3) and linkage disequilibrium (Supporting Information Method S4). We would like make it clear that given a feasible sample size, it is impractical to estimate all of the parameters in the above two-locus model (Eqn 5). To tackle this practical limitation, we suggest use of a reduced model, in which the focus is on the interaction parameters of interest, as shown in Supporting Information Method S5.

Detection of major gene segregation in an outbred autotetraploid population

The segregation analysis models the trait phenotype data distribution as a mixed distribution in which each component distribution corresponds to a particular genotype of the major QTL. We illustrate a trait phenotype based segregation analysis by modeling the trait phenotype data using the following likelihood function of *m* mixed normal distributions

$$L\left(G,\sigma^2\left|Y,G_{P_1},G_{P_2},\alpha\right.\right)=\prod_{i=1}^{n}\sum_{j=0}^{m-1}f_j\left(G_{P_1},G_{P_2},\alpha\right)g_j\left(y_i;G_j,\sigma^2\right) \qquad \text{Eqn 6a}$$

where *m* represents the number of segregating QTL genotypes with the genotypic value vector, $G=(G_0\,\mathrm{K}\,G_{m-1})$, σ^2 is the residual variance, G_{P_1} and G_{P_2} denote the two parental QTL genotypes, $Y=\{y_1,y_2,\mathrm{K},y_n\}$ represents the offspring trait data, α denotes the coefficient of double reduction, $f_j\left(G_{P_1},G_{P_2},\alpha\right)$ (j=0,…,m-1) indicates the frequency of the QTL genotype Q_jq_{4-j} and $g_j\left(y_i;G_j,\sigma^2\right)$ is the probability density function of a normal distribution with mean G_j and variance σ^2.

To estimate the genetic effect parameters, we first need to calculate the mean for each QTL genotype from the offspring population, which is equivalent to estimating the means for a finite mixture of component distributions. For any given parental QTL genotypes and the coefficient of double reduction at the putative major QTL, the parameters can be estimated from standard normal mixture model analysis using the EM algorithm (Dempster, 1977). In the present context, offspring QTL genotypes were unknown but can be inferred either from the individual's genotype

information at marker loci nearby to the QTL, as we developed previously (Luo *et al.*, 2000; Luo *et al.*, 2004), or from their parental genotypes at the QTL. A modified version of equation (6a) incorporating parental marker information, given by M_{P_1}, M_{P_2}, and offspring marker information given by O_i, is given as follows

$$L\left(G,\sigma^2\left|Y,G_{P_1},G_{P_2},\alpha\right.\right)=\prod_{i=1}^{n}\sum_{j=0}^{m-1}f_{ij}\left(G_{P_1},G_{P_2},\alpha,r,M_{P_1},M_{P_2},O_i\right)g_j\left(y_i;G_j,\sigma^2\right) \qquad \text{Eqn 6b}$$

where f_{ij} is the QTL genotype frequency for the *i*th individual and the *j*th QTL genotype, calculated according to equations (3) and (4) in the multi-locus linkage analysis we developed previously (Leach *et al.*, 2010).

Assuming biallelic segregation at a putative QTL, there are a total of twelve possible autotetraploid parental genotype configurations, listed as (1) $aaaa\times Aaaa$, (2) $aaaa\times AAaa$, (3) $aaaa\times AAAa$, (4) $Aaaa\times AAaa$, (5) $Aaaa\times AAAa$, (6) $Aaaa\times AAAA$, (7) $AAaa\times AAAa$, (8) $AAaa\times AAAA$, (9) $AAAa\times AAAA$, (10) $Aaaa\times Aaaa$, (11) $AAaa\times AAaa$, and (12) $AAAa\times AAAa$. We conducted a scan of the likelihood function (Eqn 6a) over all twelve parental genotype configurations and over all different levels of double reduction from its minimum value of 0.00 to the maximum of 0.25, at every increment of 0.005. Given a parental genotype configuration $\left(G_{P_1},G_{P_2}\right)$, the frequency of QTL genotype Q_jq_{4-j} denoted $f_j\left(G_{P_1},G_{P_2},\alpha\right)$ (*j*=0, …, *m*-1), can be calculated as a function of the coefficient of double reduction α.

The EM algorithm is initialised with starting values for the QTL genotypic values by using k-means cluster analysis. The sample variance is used to initialise σ^2. It then involves iterating the E-step that calculates the conditional probability of the i^{th} individual having the QTL genotype Q_jq_{4-j}, i.e.

$$\omega_{ij}=\frac{f_j\left(G_{P_1},G_{P_2},\alpha\right)g_j\left(y_i;G_j,\sigma^2\right)}{\sum_{k=0}^{m-1}f_k\left(G_{P_1},G_{P_2},\alpha\right)g_k\left(y_i;G_k,\sigma^2\right)} \qquad \text{Eqn 7}$$

and the M-step that calculates the maximum likelihood estimates (MLEs) of the model parameters given the conditional probabilities from the above E step from the following formula

$$\hat{G}_j=\sum_{i=1}^{n}\omega_{ij}y_i\,/\sum_{i=1}^{n}\omega_{ij} \qquad \text{Eqn 8}$$

$$\hat{\sigma}^2 = \sum_{i=1}^{n}\sum_{j=0}^{m-1} \omega_{ij}(y_i - \hat{G}_j)^2 \Big/ n \qquad \text{Eqn 9}$$

The E-step and M-step are repeated iteratively until convergence.

We calculated the log-likelihood ratio statistic (LRT)

$$LRT = 2\left[L\left(\hat{G}^*, \hat{\sigma}^{*2} \middle| Y, G_{P_1}, G_{P_2}, \alpha\right) - L\left(\bar{G}, \bar{\sigma}^2 \middle| Y, G_{P_1}, G_{P_2}, \alpha\right)\right] \qquad \text{Eqn 10}$$

as a statistical test for significance of major QTL segregation in the population under study. In Eqn 10, $\hat{G}^*$ and $\hat{\sigma}^{*2}$ are the MLEs of the genotypic means and residual variance, while $\bar{G}$ and $\bar{\sigma}^2$ are the mean and variance of the trait calculated from all individuals. Each model was compared with the null model assuming no major gene to be segregating in the population, by applying the likelihood-ratio test (LRT). The LRT statistic in the present context asymptotically follows a chi-square distribution with *m*-1 degrees of freedom, with *m* equal to the number of QTL genotypes segregating as defined above.

Estimation of overlap between normal densities

We proposed an average overlapping coefficient (*aOVL*) to define a disparity index for quantifying the average difference between any two component normal distributions. The overlap coefficient between two normal distributions has been defined (Inman & Bradley, 1989) as

$$OVL = 2\Phi\left(-\frac{|\mu_1 - \mu_2|}{\sigma}\right) \qquad \text{Eqn 11}$$

where Φ denotes the cumulative distribution function of the standard normal distribution, μ_1 and μ_2 are the means of the two component normal distributions, and σ^2 is the variance for the component normal distribution. In the tetraploid case, there are *k* (*k* = 2,…, 5) components in the mixture normal distribution and the corresponding *aOVL* could be calculated by

$$aOVL = \sum_{i=1}^{k}\sum_{j=i+1}^{k} 2\Phi\left(-\frac{|G_i - G_j|}{\sigma}\right) \Big/ \binom{2}{k} \qquad \text{Eqn 12}$$

aOVL takes a value between 0 and 1, with larger values indicating that the component normal distributions are less well separated.

Simulated autotetraploid populations

Simulated populations were created by developing programs to mimic the gametogenesis of an autotetraploid individual with a given genotype and random union of gametes to generate a zygote. Segregation and recombination of alleles at the loci of interest were simulated under tetrasomic inheritance, as explained in detail elsewhere (Luo *et al.*, 2000). Given a simulated genotype for any offspring individual, the phenotype of the individual was determined as the sum of the genotypic value calculated from the corresponding simulated model, developed here as shown in equations (1) or (5), or developed by Killick (1971) (see Supporting Information Note S1), and a variable randomly sampled from a normal distribution, $N(0,\sigma^2)$. The residual variance was calculated according to the prior phenotypic variance of the trait in question and the heritability of the simulated QTL.

Segregating autotetraploid potato population

A first generation segregating population (S_1) of autotetraploid potato (*Solanum tuberosum*) was created by crossing two parental cultivars, with the American cultivar Atlantic as the maternal parent and the Chinese cultivar Longshu-3 as the paternal parent. A second generation (S_2) segregating population consisting of 304 full-sib individuals (S_2) was derived by crossing two individuals (5-12 and 1-20) from the S_1 population. The S_2 population was planted together with their parental lines in three different field trials in 2015, each with five replicates per individual, by propagating the individuals asexually using tubers. A series of morphological and agronomic traits were scored, including plant height and flowering time.

Data availability

Programs and data for statistical analyses presented in the paper are freely available from https://github.com/LJLeach/QuantModelTetra and the link is included on our group website at www.statisticalgenetics.info.

Results

Detecting major genes in outbred autotetraploid populations

To illustrate the use of our new method, we considered the detection of major effect QTL in a segregating population, one of the most popular quantitative genetic analyses using trait phenotype data. If a major gene makes a sufficiently large contribution to the phenotypic variation of a quantitative trait relative to the background genetic and environmental variation, then the phenotypic distribution will be multimodal (Falconer, 1989). In an outbred autotetraploid S_2 population, the distribution may be bimodal, trimodal, quadrimodal or quinquemodal under a biallelic model, depending on the parental genotype configuration and the occurrence of double reduction. Simulated examples of each are shown in Fig. 1a-d.

We simulated a quantitative trait for 300, 500 or 1,000 S_2 individuals generated by crossing parental genotypes $AAaa \times AAaa$ and with a range of heritability values for the major gene, from 10% to 35%. The monogenic, digenic, trigenic and quadrigenic effects for this gene and the population mean were all set equal to 1. The coefficient of double reduction at this QTL was equal to 0.1. Accordingly, the genetic variance of the major gene, V_G, was calculated as 1.132, according to equation (4). The residual variances of the trait were chosen based on the simulated heritability and genetic variance of the major gene.

In practice, the parental genotype configuration is usually unknown in outbred autotetraploid populations, therefore trait segregation analysis was carried out across all twelve possible parental genotype configurations, (for example $AAAa \times AAAa$), across the range of possible values (0-0.25) for the rate of double reduction (Luo *et al.*, 2006a). The value corresponding to the highest likelihood value was taken as the MLE of the double reduction parameter. For a given parental genotype configuration and coefficient of double reduction, the expected offspring genotype frequency distribution could be calculated directly in terms of α, and then used to calculate the general orthogonal scales w_{ij} for the genetic effects of genotype i for the jth contrast (j = 1,2,...,4). We implemented the EM algorithm as described in Equations (7)-(9) to calculate the MLEs of the genotypic values for each QTL genotype, $\hat{G}_j$. The genetic effect parameters for the major QTL, E_A, could then be calculated according to equation (3), and the significance of major QTL segregation was tested as shown in equation (10).

Segregation analysis of phenotypic data in autotetraploids has a very poor statistical power for major gene detection when heritability is low (10%) (Table 4), which reflects the high degree of

overlap between the component normal distributions, as indicated by a high average overlapping coefficient (aOVL>0.5). When heritability is doubled to 20%, then the statistical power reaches an acceptable level of 79%, though only when there is large population size of at least 1,000 here. Only when there is a large degree of separation between component distributions, for example $h^2 = 0.30$ and $aOVL = 0.3575$, does major gene detection have adequate power with more realistic population sizes ($n \geq 300$). When trait heritability is low, we caution that the genetic variance of major QTL is significantly overestimated by using trait segregation analysis in autotetraploids. Segregation analysis was able to correctly infer the parental QTL genotypes only in less than 10% of simulations.

Comparison with Killick's model

We used a numerical example to explore the statistical properties of our model compared with an additive-dominance model, such as Killick's model. We simulated two biallelic loci, Q_A and Q_B, with alleles segregating in linkage equilibrium at the two loci in an S_2 population created from crossing two homozygous autotetraploid parents. The population mean, all genetic effects of the two loci and epistatic effects were simulated to be equal to 1. It should be noted that the simulation was designed without incorporating an environmental variable, with a purpose to minimize the influence of random sampling variation in the comparison of the methods. Shown in Supporting Information Table S3 are the genotypic values calculated either from a two locus orthogonal contrast based model defined according to equation (5) or from Killick's model with two loci. In Scenario 1, double reduction was absent at both loci, and genotypic values were generated either under our model (Scenario 1^{O}**,** Supporting Information Table S3a) or under Killick's model (Scenario 1^{K}**,** Supporting Information Table S3b); in Scenario 2, the coefficient of double reduction was equal to 0.05 for locus Q_A and 0.10 for Q_B, and the data was simulated under our model (Scenario 2^{O}).

As expected from the orthogonal property of our model, estimates of monogenic, digenic, trigenic and quadrigenic effects under both scenarios are independent of the estimation of epistatic effects (Table 5). All genetic effects can be consistently estimated under both single-locus (reduced) models and two-locus (full) models, when data are simulated under either our model or Killick's model. For example, under Scenario 1^{O}, all estimated genetic effects take the same value of 1 regardless of whether the model is fitted for one locus or for both loci, including epistatic parameters. In contrast, estimates of additive and dominance effects are markedly biased from their true values, particularly when epistatic effects are fitted in Killick's model, suggesting the model is vulnerable to the inclusion of various effects in the genetic models. For example, in Scenario 1^{K}, additive and dominance effects are estimated to be equal to 1.94 in either reduced (single locus)

model, but all effects across both loci are correctly estimated to be equal to 1.00 when the full model is used. A further limitation of Killick's model arises because the genetic effect parameters are only defined on the basis of genotypic values and not using the genotypic frequencies (though genotypic frequencies are considered when estimating the population mean). This explains why the estimates of genetic effects from Killick's model (and other additive-dominance models in the current literature) involving both loci remained the same under Scenarios 1^o and 2^o. In contrast, our model confers a statistically appropriate and feasible way to estimate the various genetic effects in populations with various genotypic frequency distributions.

Parameter estimation under the orthogonal model

We carried out a simulation study to test for reliability of the theoretical models presented above and to explore the statistical properties of the methods developed for estimating the model parameters under the orthogonal model, including the impact of the double reduction parameter on the estimation of genetic effect parameters. We considered a biallelic quantitative trait locus, Q_A, segregating in an S_1 population created from parental genotypes *AAaa* and *AAaa*. Segregation at the simulated QTL contributed 20% of the phenotypic variance of the trait. All genetic effects at the QTL, $E_A = (\mu \ \theta_1 \ \theta_2 \ \theta_3 \ \theta_4)^T$, were set equal to 1, and the residual variance was determined accordingly. Given the parental genotypes, offspring genotypes were generated under two levels of double reduction, $\alpha = 0$ or 0.15, the former corresponding to bivalent pairing of homologous chromosomes and the latter to quadrivalent pairing in the autotetraploid meiosis.

Our previous simulation studies showed that trait segregation analysis based on phenotype data alone had limited power to detect segregation of major genes for traits with low heritability (see Table 4). It is well established that use of genetic markers is effective in recovering genotype information at QTL, leading to an increase in statistical power for detecting the QTL. We therefore simulated a chromosome with a single QTL closest to the centromere and an additional 10 genetic marker loci equally spaced at 10 cM intervals downstream from the QTL (Table 6). We implemented a modified version of the trait segregation analysis (equation 6b), which incorporates the parental and offspring marker information. This is expected to make the distribution of offspring QTL genotypes more informative, with an expected gain in statistical efficiency of parameter estimation from the mixture model.

Based on the simulated parental QTL genotype configuration (*AAaa* and *AAaa*), the EM algorithm was used as described in equations (7) to (9) to give the MLEs for the QTL genotypic values. The genetic effect parameters were then estimated using our orthogonal model based on a range of

assumed values for the coefficient of double reduction. Table 7 shows the means and standard errors of the estimated genetic effects based on 500 repeated simulations. The genetic effect parameters and their variance components were predicted adequately as long as double reduction was properly taken into account, while the corresponding estimates were comparatively poor when double reduction was ignored. It is clear from the heritability estimates (h^2) that an overestimated double reduction parameter may lead to overestimation of the genetic variance, and thus overestimation of trait heritability.

Case study of flowering time and plant height in autotetraploid potato

To demonstrate the application of the methods we developed for modeling and analyzing real experimental data collected from autotetraploid species, we analyzed the phenotype data of two quantitative traits, flowering time and plant height, scored on 304 offspring from a cross between two varieties of cultivated autotetraploid potato (*Solanum tuberosum*). We found significant variation in the trait phenotype scores across three separate field trials, showing the major effect of the environment on both traits (ANOVA, $p < 0.001$, Supporting Information Tables S4, S5). We also observed a significant genotype by environment (G x E) interaction for plant height (ANOVA, $p < 0.001$, Supporting Information Table S5).

We observed highly significant statistical evidence for segregation of major QTL genes for both flowering time and plant height. Figure S2 shows the LOD score profiles for different configurations of parental genotypes of QTL for both traits, which enable identification of the most likely parental QTL genotype(s). For flowering time, there was significant evidence for a major QTL ($p < 0.0001$) under three alternative configurations, namely [5] ($Aaaa \times AAAa$), [10] ($Aaaa \times Aaaa$) and [11] ($AAaa \times AAaa$), each with similar LOD score profiles and maximum LOD scores (Fig. S2a). Under each alternative model, the estimated mixture normal distribution is composed of five component normal distributions. In the absence of any marker information, model [11] was chosen as the most likely parental genotype configuration (Table 8, Fig. 1e), because the average parental trait scores (illustrated by arrows in Fig. 1e) were most likely to have come from the third component normal distribution corresponding to genotype *AAaa*. Meanwhile, model [5] was deemed unlikely because the parental phenotype scores were unlikely to have come from component genotypes *Aaaa* and *AAAa*, as shown in Fig. S3a. However, model [10] was only slightly less likely than model [11], as shown in Fig. S3b, illustrating the difficulty in clearly distinguishing parental genotype configuration through trait segregation analysis in autotetraploids.

For plant height, there was significant evidence for a major QTL (p = 0.003) under genotype configurations [3] $aaaa \times AAAa$, [9] ($AAAa \times AAAA$) and [11] ($AAaa \times AAaa$) (Fig. S2b). Model [9] was chosen as the most likely parental genotype configuration (Table 8, Fig. 1f), because the average parental trait scores were more likely to have come from the component normal distributions for genotypes *AAAa* and *AAAA* (Fig. 1f) than from the component distributions for genotypes *aaaa* and *AAAa* (Fig. S3c) or *AAaa* and *AAaa* (Fig. S3d). Preliminary prediction of parental genotype configuration in this way can be valuable for various downstream analyses, including linkage and linkage disequilibrium based QTL analyses.

We estimated the narrow heritability of flowering time and plant height using the linear mixed model, as described in Supporting Information Method S6, based on the most likely value of the coefficient of double reduction inferred from the trait segregation analysis (Table 8). High values of narrow heritability were estimated for both flowering time (79%) and plant height (73%), and the assumed values for the double reduction rate had very little effect (Supporting Information Table S6). For both traits, the major QTL effect explained a significant proportion of the total phenotypic variance (29-39%), and up to one half of the total genetic variance (40-50%, Table 8).

All of the estimated genetic effects at the major QTL for both traits were highly significant (p < 0.0001). These estimates reveal valuable information on the genetic architecture of the trait, which may be useful for identifying useful QTL for selection purposes. For flowering time, a monogenic effect of 5.92 indicates that the average effect of replacing allele *a* with allele *A* at the major QTL in any autotetraploid genotype will be to delay flowering by around 6 days. Significance of the trigenic and quadrigenic genetic effects means that the interaction between the increasing alleles (*A*) and decreasing alleles (*a*) depends on the identity of the third and fourth alleles in an autotetraploid genotype. However, the corresponding genetic variance components at trigenic and quadrigenic levels contributed very little to the total genetic variance of the major QTL (0.07% and 0.32%), suggesting that higher level allelic interactions exert limited effects on the genotypic values for this trait. For plant height, we observed a monogenic effect of 12.46, meaning that on average, replacing allele *a* with allele *A* at the major effect QTL will increase height by around 12cm, while a significant digenic effect of -6.84 suggests the importance of two-way allelic interactions at the major QTL for plant height, though such interactions make a relatively small contribution (2.48%) to the total genetic variance.

Discussion

In recent decades, considerable progress has been made in the genetic analysis of quantitative traits in diploid plant, animal and human species. For example, new technologies and statistical methods have enabled genome-wide mapping of genetic variation and led to the detection of individual genomic regions (Quantitative Trait Loci), or more rarely individual quantitative trait nucleotides, that directly or indirectly influence trait phenotypic variation. However, progress has been limited by many factors (Hill, 2012), including difficulties in disentangling pleiotropic and epistatic effects of genes, and the complicated inheritance systems in polyploid species.

A crucial foundation for all quantitative genetic analyses in species of any ploidy level, including QTL analysis and evaluation of quantitative genetic parameters, is a quantitative genetic model that links together the genetic effects of genes with the quantitative trait phenotype. Historically, the field of quantitative genetics has focused on diploid species, and as such, the quantitative genetic model, theory and methods for various quantitative genetic analyses have been well established and routinely practiced (Mather & Jinks, 1971; Falconer, 1989; Lynch & Walsh, 1998). Meanwhile, progress in the genetic analysis of polyploid species, particularly autotetraploids, has lagged far behind. Built upon Kempthorne's model, Hackett *et al.,* (2001) proposed a quantitative genetic model for use in interval mapping of QTL in autotetraploids, though this model was not based on a strict tetrasomic inheritance model, and does not possess the orthogonal property between different model parameters and their estimates.

In this article, we have contributed a quantitative genetic model for autotetraploid species based on the orthogonal contrast scales model developed for diploids (Cockerham, 1954). The model relates the phenotype score of an autotetraploid individual for a quantitative trait to the alleles at the loci that contribute to trait variation, in terms of monogenic, digenic, trigenic, and quadrigenic effects at individual loci, and epistatic effects between loci. The orthogonal property of the model ensures that the genetic effects can be independently estimated for one or two loci, assuming only pair-wise epistatic effects. This property is very useful for obtaining reliable estimates of genetic effects and genetic variance components, even when the number of QTL involved is unknown, which is usually the case (Zeng *et al.*, 2005). The model could be extended to three or more loci, assuming epistasis only occurs between pairs of loci. It has been recognised that use of an orthogonal model for QTL mapping would be advantageous in various ways (Kao & Zeng, 2002; Zeng *et al.*, 2005). For example, different QTL genetic effects may be tested and estimated separately. Parameter estimates are also independent of which, if any, epistatic effects are fitted in the model, which will simplify genetic interpretation of the underlying genetic architecture.

Our model decreases the number of parameters used to describe the genetic effects of QTL from 255 (Kempthorne, 1955; Kempthorne, 1957) down to 24 for a two-locus analysis, making it statistically tractable for real data analysis. We have shown its suitability for use in populations with various genetic structures, including segregating populations (e.g. S_2) and natural populations of unrelated individuals, either in linkage equilibrium or linkage disequilibrium. While there is evidence for multiple allele segregation at individual QTL genes in diploid populations (Barton & Keightley, 2002), each allele may only increase or decrease the trait phenotype, therefore the biallelic setting is appropriate for quantitative genetic analysis in autotetraploids.

We have shown that our model can accurately estimate the genetic effects in a segregating population under either one locus (reduced model) or two locus (full model) settings. Our model takes proper account of the complex features of autotetrasomic inheritance, including double reduction, unlike other quantitative genetic models developed for autotetraploids (Killick, 1971). We showed that the double reduction parameter has a significant impact on the genetic parameter estimation, and thus advise that this parameter should be taken into account in any quantitative genetic analysis in autotetraploid species to avoid bias in the estimation of genetic effects. We have provided statistical tests for the significance of double reduction and methods for its accurate estimation using molecular marker data in our previous work (Luo *et al.*, 2004; Luo *et al.*, 2006a). Since double reduction is a location dependent parameter, the marker data can provide an approximation of the double reduction parameter near to the QTL.

We carried out trait segregation analysis to illustrate the practical application of our quantitative genetic model for the analysis of trait phenotype data in autotetraploids. Trait segregation analysis has been an important topic in the history of quantitative genetics in diploids (Falconer, 1989). It serves as an important intermediate step prior to collection of genomic marker data, allowing major genes affecting quantitative traits to be detected prior to designing further genomic analyses, such as QTL analysis, and also enabling more efficient selection in breeding programs of agronomic traits in autotetraploid crops or animals (Falconer, 1989).

Segregation analysis involves the estimation of normal mixtures, which is well known to be an ill-posed problem, particularly when the disparity between the component distributions is small (Xiao *et al.*, 2007; Lourens *et al.*, 2013). Methods for segregation analysis may therefore suffer from low statistical power in diploids, and even more so in autotetraploids (Xiao *et al.*, 2007). We extended the concept of the overlapping coefficient (Inman & Bradley, 1989) to quantify the disparity

between multiple component normal distributions for segregation analysis in autotetraploids. Our results echo previous work showing that maximum likelihood estimation may perform poorly when component distributions are poorly separated, and substantial bias may be observed when OVL exceeds 0.45 (Lourens *et al.*, 2013). Additionally, when the disparity is low, within population variance will be underestimated; in the present context, the proportion of the trait phenotypic variation explained by the QTL, *i.e.* its heritability, may therefore be markedly overestimated. This could be an inherent weakness of the statistical method implemented for segregation analysis and care must thus be taken when interpreting the results of phenotype-based analysis in autotetraploids. We observed the statistical power to detect QTL segregation is low when heritability is low (≤ 20%) and the OVL is greater than 0.4, while the power became adequate for detecting QTL segregation when heritability was at least 30%, in populations with a modest size of at least 300.

We have demonstrated the utility of our quantitative genetic model for autotetraploids by analyzing real data on flowering time and plant height in a segregating population of potato. We estimated high values of narrow heritability for both flowering time (79%) and plant height (73%), consistent with various other potato populations (Khan *et al.*, 2013; Ozturk & Yildrim, 2014), and also crops such as barley, which have shown high heritability (>90%) for flowering time (Maurer *et al.*, 2015). Trait segregation analysis showed evidence for segregation of a major gene affecting flowering time in this population. Work in *Arabidopsis*, rice and tomato, led to the identification of FLOWERING LOCUS T (FT) as the mobile signal "florigen" that plays a central role in the floral transition, travelling from the leaves to the shoot apical meristem to promote flowering (Turk *et al.*, 2008). More recently, several functional homologues of the key *Arabidopsis* flowering time genes have been identified in potato, including *StSP3D* as the mobile signal "florigen", and *StSP6A* as the mobile signal "tuberigen" responsible for the stolon to tuber transition (Navarro *et al.*, 2011). A CONSTANS (CO) homologue, *StCO*, has also been discovered with a role in repression of tuberisation (Gonzalez-Schain *et al.*, 2012), as well as homologues of other key flowering time genes, including *StCDF1* (Kloosterman *et al.*, 2013). The major regulators of flowering time in potato are therefore conserved with *Arabidopsis*, but have also been recruited to control the developmental switch involved in storage organ formation.

The quantitative genetic model we presented here lays the foundation for quantitative genetic analysis in autotetraploid species. In addition to the classical phenotype based analysis of quantitative traits, such as segregation analysis, estimation of breeding values, and genetic variance components, the model fulfills an essential requirement for DNA molecular marker assisted QTL analysis under both linkage and linkage disequilibrium based settings. The model will therefore

facilitate future studies of the genetic architecture and evolution of quantitative traits in important crop species such as potato (D'Hoop *et al.*, 2008; Massa *et al.*, 2015), and coffee (del Pilar Moncada *et al.*, 2016), forest legumes such as alfalfa (Yu *et al.*, 2016), and ornamental species such as rose (Gitonga *et al.*, 2016).

Acknowledgements The study was supported by research grants from BBSRC (BB/N008952/1) in the United Kingdom, the National Nature Science Foundation of China (grant numbers 81172006 and 91231114) and the Leverhulme Trust.

Author contribution Z.L. conceived of and designed the study. Z.L. designed the theoretical model and statistical methods. J.C. implemented the statistical methods. Z.L., F.Z. and L.W. created the potato segregating population, implemented field trials and collected the phenotypic data. J.C. analysed the data with inputs from Z.L. L.L., Z.L., J.C. and L.L. wrote the paper.

References

Adams KL, Wendel JF. 2005. Novel patterns of gene expression in polyploid plants. *Trends in Genet*ics **21:** 539-543.

Barton NH, Keightley PD. 2002. Understanding quantitative genetic variation. *Nature Reviews Genetics* **3:** 11–21.

Cockerham CC. 1954. An extension of the concept of partitioning hereditary variance for analysis of covariance among relatives when epistasis is present. *Genetics* **39:** 859-882.

Comai L, Tyagi A, Winter K, Holmes-Davis R, Reynolds SH, Stevens Y, Byers B. 2000. Phenotypic instability and rapid gene silencing in newly formed *Arabidopsis* allotetraploids. *The Plant Cell* **12:** 1551-1568.

Danzmann RG, Gharbi K. 2001. Gene mapping in fishes: a means to an end. *Genetica* **111:** 3-23.

Del Pilar Moncada M, Tovar E, Montoya JC, Gonzalez A, Spindel J, McCouch S. 2016. A genetic linkage map of coffee *Coffea Arabica* L. and QTL for yield, plant height, and bean size. *Tree Genetics and Genomes* **12:**5.

Dempster AP. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B* **39:** 1-22.

D'hoop BB, Joao Paulo M, Mank R, van Eck HJ, van Eeuwijk FA. 2008. Association mapping of quality traits in potato *Solanum tuberosum* L. *Euphytica* **161:** 47-60.

Falconer DS, Mackay TFC. 1989. *Introduction to Quantitative Genetics.* Pearson Prentice Hall.

Fisher RA. 1947. The theory of linkage in polysomic inheritance. *Philosophical Transactions of the Royal Society B* **23:** 55-87.

Gitonga VW, Stolker R, Koning-Boucoiran CFS, Aelaei M, Visser RGF, Maliepaard C, Krens FA. 2016. Inheritance and QTL analysis of the determinants of flower color in tetraploid cut roses. *Molecular Breeding* **36:** 143.

Gonzalez-Schain ND, Diaz-Mendoza M, Zurczak M, Suarez-Lopez P. 2012. Potato CONSTANS is involved in photoperiodic tuberization in a graft-transmissible manner *The Plant Journal* **70:** 678-690.

Hackett CA, Bradshaw JE, McNicol JW. 2001. Interval mapping of quantitative trait loci in autotetraploid species. *Genetics* **159:** 1819-1832.

Haldane JBS. 1930. Theoretical genetics of autopolyploids. *Journal of Genetics* **22:** 359-372.

Hill WG. 2012. Quantitative genetics in the genome era *Current Genomics* **13:** 196-206.

Inman HF, Bradley EL. 1989. The overlapping coefficient as a measure of agreement between probability distributions and point estimation of the overlap of two normal densities. *Communications in Statistics- Theory and Methods* **18:** 3851-3874.

Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS et al. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* **473:** 97-100.

Kao CH, Zeng ZB. 2002. Modeling epistasis of quantitative trait loci using Cockerham's model. *Genetics* **160:** 1243-1261.

Kempthorne O. 1955. The correlation between relatives in a simple autotetraploid population. *Genetics* **40:** 168-174.

Kempthorne O. 1957. *An Introduction to Genetic Statistics*. New York: John Wiley & Sons.

Li CC. 1957. The genetic variance of autotetraploids with two alleles. *Genetics* **42:** 583-592.

Khan MF, Tabassum N, Latif A, Khaliq A, Malik M. 2013. Morphological characterization of potato Solanum tuberosum L. germplasm under rainfed environment. *African Journal of Biotechnology* **12:** 3214-3223.

Killick RJ 1971 The biometrical genetics of autotetraploids: I. Generations derived from a cross between two pure lines. *Heredity* **27:** 331-346.

Kloosterman B, Abelanda JA, del Mar Carretero Gomez M, oortwijn M, de Boer JM, Kowitwanich K, Horvath BM, van Eck HJ, Smaczniak C, Prat S et al. 2013. Naturally occurring allele diversity allows potato cultivation in northern latitudes. *Nature* **495:** 246-252.

Leach LJ, Wang L, Kearsey MJ, Luo ZW. 2010. Multilocus tetrasomic linkage analysis using hidden Markov chain model. *Proceedings of the National Academy of Sciences* **1079:** 4270-4274.

Li J, Das K, Fu G, Tong C, Li Y, Tobias C, Wu R. 2010. EM algorithm for mapping quantitative trait loci in multivalent tetraploids. *International Journal of Plant Genomics* doi:10.1155/2010/216547.

Lourens S, Zhang Y, Long JD, Paulsen JS. 2013. Bias in estimation of a mixture of normal distributions. *Journal of Biometrics and Biostatistics* **4:** 1000179.

Luo ZW, Hackett CA, Bradshaw JE, McNicol JW, Milbourne D. 2000. Predicting parental genotypes and gene segregation for tetrasomic inheritance. *Theoretical and Applied Genetics* **100:** 1067-1073.

Luo ZW, Zhang RM, Kearsey MJ. 2004. Theoretical basis for genetic linkage analysis in autotetraploid species. *Proceedings of the National Academy of Sciences* **101:** 7040-7045.

Luo ZW, Zhang Z, Leach LJ, Zhang RM, Bradshaw JE, Kearsey MJ. 2006a. Constructing genetic linkage maps under a tetrasomic model. *Genetics* **172:** 2635-2645.

Luo ZW, Zhang Z, Zhang RM, Pandey M, Gailing O, Hattemer H, Finkeldey R. 2006b. Modeling population genetic data in autotetraploid species. *Genetics* **172:** 639-646.

Lynch W, Walsh B. 1998. *Genetics and Analysis of Quantitative Traits*. USA: Sinauer Associates Inc.

Massa AN, Manrique-Carpintero NC, Coombs JJ, Zarka DG, Boone AE, Kirk WW, Hackett CA, Bryan GJ, Douches DS. 2015. Genetic linkage mapping of economically important traits in cultivated tetraploid potato *Solanum tuberosum* L. *G3* **5:** 2357-2364.

Mather K. 1936. Segregation and linkage in autotetraploids. *Journal of Genetics* **32:** 287-314.

Mather K, Jinks JL. 1971. *Biometrical Genetics*. London: Chapman and Hall.

Maurer A, Draba V, Jiang Y, Schnaithmann F, Sharma R, Schumann E, Kilian B, Reif JC, Pillen K. 2015. Modelling the genetic architecture of flowering time control in barley through nested association mapping. *BMC Genomics* **16:** 290.

Navarro C, Abelenda JA, Cruz-Oro E, Cuellar CA, Tamaki S, Silva J, Shimamoto K, Prat S. 2011. Control of flowering and storage organ formation in potato by FLOWERING LOCUS T *Nature* **478:** 119-123.

Otto SP, Whitton J. 2000. Polyploid incidence and evolution. *Annual Reviews Genetics* **34:** 401-437.

Ozturk G, Yildirim Z. 2014. Heritability estimates of some quantitative traits in potatoes. *Turkish Journal of Field Crops* **19:** 262-267.

.

Soltis DE, Soltis PS. 1995. The dynamic nature of polyploid genomes. *Proceedings of the National Academy of Sciences* **92:** 8089-8091.

Song K, Lu P, Tang K, Osborn TC. 1995. Rapid genome change in synthetic polyploids of Brassica and its implications for polyploid evolution. *Proceedings of the National Academy of Sciences* **92:** 7719-7723.

Turck F, Fornara F, Coupland G. 2008. Regulation and identity of florigen: FLOWERING LOCUS T moves centre stage *Annual Reviews Plant Biology* **59:** 573-594.

van Geest G, Bourke PM, Voorrips RE, Marasek-Ciolakowska A, Liao Y, Post A, van Meeteren U, Visser RGF, Maliepaard C, Arens P. 2017. An ultra-dense integrated linkage map for hexaploid chrysanthemum enables multi-allelic QTL analysis. *Theoretical and Applied Genetics* **130:** 2527-2541.

Vaughan DA, Balazs E, Heslop-Harrison JS. 2007. From crop domestication to super-demonstration. *Annals of Botany* **100:** 893-901.

Wang J, Tian L. 2004. Stochastic and epigenetic changes of gene expression in *Arabidopsis* polyploids. *Genetics* **167:** 1961-1973.

Wolfe KH. 2001. Yesterday's polyploids and the mystery of diploidization. *Nature Reviews Genetics* **2:** 333-341.

Wright AJ. 1979. The use of differential coefficients in the development and interpretation of quantitative genetic models. *Heredity* **431**:1-8.

Xiao J, Wang X, Hu Z, Tang Z, Xu C. 2007. Multivariate segregation analysis for quantitative traits in line crosses. *Heredity* **98:** 427-435.

Yu L-X, Liu X, Boge W, Liu X-P. 2016. Genome-wide association study identifies loci for salt tolerance during germination in autotetraploid alfalfa *Medicago sativa* L. using genotyping by sequencing. *Frontiers in Plant Science* **7:** 956.

Zeng ZB, Wang T, Zou W. 2005. Modeling quantitative trait loci and interpretation of models. *Genetics* **169:** 1711-1725.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information tab for this article:

Fig S1 Impact of genetic effect parameters on genotypic values in the orthogonal model.

Fig S2 LOD score profiles for flowering time (a) and plant height (b) under different configurations of parental genotypes at a putative QTL.

Fig S3 Mixture normal distribution and inferred component normal distributions for flowering time and plant height under alternative parental QTL genotype configurations.

Methods S1 Notations and definition of the orthogonal model in autotetraploids.

Methods S2 One locus model for a randomly mating population.

Methods S3 Two locus model under linkage equilibrium.

Methods S4 Two locus model under linkage disequilibrium.

Methods S5 Reduced two-locus model and analysis.

Methods S6 Estimation of narrow-sense heritability in autotetraploids.

Note S1 Definition of the Killick quantitative genetic model.

Table S1 The number of digenic, trigenic and quadrigenic allelic interactions in autotetraploid genotypes.

Table S2 Definition of the 25 genetic parameters for the two locus quantitative genetic model in autotetraploids.

Table S3 Numerical examples with genotypic values generated under our orthogonal model (a) or under Killick's model (b).

Table S4 One-way ANOVA for flowering time in autotetraploid potato.

Table S5 Type II ANOVA for plant height in autotetraploid potato.

Table S6 Narrow sense heritability of two autotetraploid potato quantitative traits (flowering time and plant height) using the linear mixed model.

Figure Legends

Figure 1. Segregation analysis for quantitative traits in simulated and real outbred autotetraploid segregating populations.

The quantitative trait phenotype may show a bimodal **(a)**, trimodal **(b)**, quadrimodal **(c)** or quinquemodal **(d)** distribution in a segregating population derived from a cross between parental genotypes as indicated above each panel and with a given value of the coefficient of double reduction, α. Flowering time **(e)** in the potato segregating population showed a quinquemodel distribution with the most likely parental genotype configuration being *AAaa* x *AAaa*. Plant height **(f)** showed a trimodal distribution with the most likely parental genotype configuration being *AAAa* x *AAAA*. Average parental phenotype scores for P_1 and P_2 parental varieties are indicated using orange and green arrows respectively. Red lines indicate the mixture normal distribution and dotted blue lines indicate the inferred component normal distributions, numbered to indicate genotypes 1) *aaaa*; 2) *Aaaa*; 3) *AAaa*; 4) *AAAa*; and 5) *AAAA*.

Tables

Table 1. The general orthogonal contrast scales model for one locus.

	i	4	3	2	1	0
	Genotype	*AAAA*	*AAAa*	*AAaa*	*Aaaa*	*aaaa*
	Frequency	f_4	f_3	f_2	f_1	f_0
	Genotypic value	G_4	G_3	G_2	G_1	G_0
θ_1	W_1	w_{41}	w_{31}	w_{21}	w_{11}	w_{01}
θ_2	W_2	w_{42}	w_{32}	w_{22}	w_{12}	w_{02}
θ_3	W_3	w_{43}	w_{33}	w_{23}	w_{13}	w_{03}
θ_4	W_4	w_{44}	w_{34}	w_{24}	w_{14}	w_{04}

G_i and f_i denote the genotypic values and genotypic frequencies for the five genotypes with i copies of the A allele. θ_i (i=1,2,…,4) are the monogenic, digenic, trigenic and quadrigenic genetic effects respectively. w_{ij} (i=0,1,…,4; j=1,…,4) is the scale component of genotype i for the j^{th} contrast.

Table 2. Orthogonal contrast scales for one locus in a biallelic autotetraploid S_2 population.

	Genotype	*AAAA*	*AAAa*	*AAaa*	*Aaaa*	*aaaa*
	Frequency	f_4	f_3	f_2	f_1	f_0
		G_4	G_3	G_2	G_1	G_0
θ_1	W_1	2	1	0	-1	-2
θ_2	W_2	$(5-2\alpha)/3$	$(1-4\alpha)/6$	$-(1+2\alpha)/3$	$(1-4\alpha)/6$	$(5-2\alpha)/3$
θ_3	W_3	$2(1-\alpha)/3$	$-(1+2\alpha)/6$	0	$(1+2\alpha)/6$	$-2(1-\alpha)/3$
θ_4	W_4	$\frac{(1-\alpha)(4\alpha^2-4\alpha+3)}{12(2+\alpha)}$	$-\frac{(1+2\alpha)(4\alpha^2-4\alpha+3)}{24(2+\alpha)}$	$\frac{(1-\alpha)(2\alpha+1)^2}{12(2+\alpha)}$	$-\frac{(1+2\alpha)(4\alpha^2-4\alpha+3)}{24(2+\alpha)}$	$\frac{(1-\alpha)(4\alpha^2-4\alpha+3)}{12(2+\alpha)}$

$\theta_i\ (i=1,...,4)$ are the monogenic, digenic, trigenic, and quadrigenic genetic effects of locus A. f_i and $G_i\ (i=0,1,...,4)$ are the frequency and genotypic values for the i^{th} genotype $A_i a_{4-i}$, respectively. α is the coefficient of double reduction at locus A.

Table 3. The general orthogonal contrast scales model for two loci (A and B).

	Genotype	*AAAA*	*AAAa*	*AAaa*	*Aaaa*	*aaaa*		Genotype	*BBBB*	*BBBb*	*BBbb*	*Bbbb*	*bbbb*
	Frequency	$f_{4.}$	$f_{3.}$	$f_{2.}$	$f_{1.}$	$f_{0.}$		Frequency	$f_{.4}$	$f_{.3}$	$f_{.2}$	$f_{.1}$	$f_{.0}$
	G	$G_{4.}$	$G_{3.}$	$G_{2.}$	$G_{1.}$	$G_{0.}$		G	$G_{.4}$	$G_{.3}$	$G_{.2}$	$G_{.1}$	$G_{.0}$
θ_1	W_{A1}	w_{41}	w_{31}	w_{21}	w_{11}	w_{01}	ζ_1	V_{B1}	v_{41}	v_{31}	v_{21}	v_{11}	v_{01}
θ_2	W_{A2}	w_{42}	w_{32}	w_{22}	w_{12}	w_{02}	ζ_2	V_{B2}	v_{42}	v_{32}	v_{22}	v_{12}	v_{02}
θ_3	W_{A3}	w_{43}	w_{33}	w_{23}	w_{13}	w_{03}	ζ_3	V_{B3}	v_{43}	v_{33}	v_{23}	v_{13}	v_{03}
θ_4	W_{A4}	w_{44}	w_{34}	w_{24}	w_{14}	w_{04}	ζ_4	V_{B4}	v_{44}	v_{34}	v_{24}	v_{14}	v_{04}

$G_{i.}$ ($G_{.i}$) and $f_{i.}$ ($f_{.i}$) (i = 0, 1,…, 4) denote the genotypic values and genotypic frequencies for the five genotypes of locus A (locus B). θ_i (ζ_i) (i=1, 2,…, 4) are the monogenic, digenic, trigenic and quadrigenic effects for locus A and locus B, respectively. Here w_{ij} and v_{ij} (i=0,1,…,4; j=1,2,…,4), are the orthogonal contrast scales of genotype i for the j^{th} contrast, calculated separately for each locus using the general biallelic one locus model.

Table 4. Statistical power of major gene detection in outbred autotetraploid populations.

Heritability (h^2)	Sample Size (n)	*aOVL*	power (%)	Genetic Variance ($\hat{V}_{QTL}$)	
				mean	s.e.
0.10	300	0.5275	17	4.8786	0.2343
	500		18	4.1234	0.2345
	1000		23	3.4567	0.2325
0.15	300	0.4634	29	2.8886	0.1575
	500		35	2.5018	0.1477
	1000		53	2.4363	0.1306
0.20	300	0.4193	46	2.3815	0.1156
	500		54	2.1323	0.1004
	1000		79	1.6360	0.0746
0.25	300	0.3856	59	1.8227	0.0825
	500		75	1.5201	0.0703
	1000		99	1.4111	0.0504
0.30	300	0.3575	71	1.6470	0.0705
	500		93	1.4608	0.0611
	1000		99	1.2765	0.044
0.35	300	0.3329	77	1.5475	0.0626
	500		100	1.3251	0.0465
	1000		100	1.2272	0.0352

aOVL is the average overlapping coefficient between normal distributions. The empirical statistical power for major gene detection is given at significance level 5% based on 100 replicates. The simulated value of V_G was equal to 1.132.

Table 5. Estimates of genetic effects using Killick's model and the orthogonal contrast scales based model, under Scenario 1 (without double reduction, $\alpha_A = \alpha_B = 0.00$) or Scenario 2 (with double reduction, $\alpha_A = 0.05, \alpha_B = 0.10$).

Killick's model (ref. 23)

Scenario	μ	a	d_1	d_2	d_3	b	h_1	h_2	h_3	I_{ab}	I_{ah_1}	I_{ah_2}	I_{ah_3}	I_{d_1b}	$I_{d_1h_1}$	$I_{d_1h_2}$	$I_{d_1h_3}$	I_{d_2b}	$I_{d_2h_1}$	$I_{d_2h_2}$	$I_{d_2h_3}$	I_{d_3b}	$I_{d_3h_1}$	$I_{d_3h_2}$	$I_{d_3h_3}$
1^O	1.00	2.67	-2.52	-2.08	-0.85	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	1.00	-	-	-	-	2.67	-2.52	-2.08	-0.85	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	1.00	7.44	-7.03	-5.82	-2.38	7.44	-7.03	-5.82	-2.38	7.11	-6.72	-5.56	-2.28	-6.72	6.35	5.25	2.15	-5.56	5.25	4.34	1.78	-2.27	2.15	1.78	0.73
1^K	3.78	1.94	1.94	1.94	1.94																				
	3.78					1.94	1.94	1.94	1.94																
	3.78	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2^O	1.10	2.84	-2.62	-2.22	-0.91	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	1.10	-	-	-	-	2.75	-2.60	-2.15	-0.88	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	1.10	7.44	-7.03	-5.82	-2.38	7.44	-7.04	-5.82	-2.38	7.11	-6.72	-5.56	-2.28	-6.72	6.35	5.25	2.15	-5.56	5.25	4.34	1.78	-2.27	2.15	1.78	0.73

orthogonal contrast scales based model

Scenario	μ	θ_1	θ_2	θ_3	θ_4	ζ_1	ζ_2	ζ_3	ζ_4	$I_{\theta_1\zeta_1}$	$I_{\theta_1\zeta_2}$	$I_{\theta_1\zeta_3}$	$I_{\theta_1\zeta_4}$	$I_{\theta_2\zeta_1}$	$I_{\theta_2\zeta_2}$	$I_{\theta_2\zeta_3}$	$I_{\theta_2\zeta_4}$	$I_{\theta_3\zeta_1}$	$I_{\theta_3\zeta_2}$	$I_{\theta_3\zeta_3}$	$I_{\theta_3\zeta_4}$	$I_{\theta_4\zeta_1}$	$I_{\theta_4\zeta_2}$	$I_{\theta_4\zeta_3}$	$I_{\theta_4\zeta_4}$
1^O	1.00	1.00	1.00	1.00	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	1.00	-	-	-	-	1.00	1.00	1.00	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1^K	3.78	0.32	-0.81	1.94	-3.89																				
	3.78					0.32	-0.81	1.94	-3.89																
	3.78	0.32	-0.81	1.94	-3.89	0.32	-0.81	1.94	-3.89	0.03	-0.07	1.67	-0.33	-0.07	0.17	-0.42	0.83	0.17	-0.42	1.00	-2.00	-0.33	0.83	-2.00	4.00
2^O	1.10	1.08	1.08	1.07	1.07	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	1.10	-	-	-	-	1.07	1.05	1.03	1.03	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	1.10	1.08	1.08	1.07	1.07	1.07	1.05	1.03	1.03	1.05	1.04	1.02	1.02	1.04	1.03	1.01	1.01	1.03	1.02	1.00	1.00	1.03	1.02	1.00	1.00

914 μ is the population mean. a (b) indicate the additive effect for locus Q_A (Q_B). d_1, d_2 and d_3 (h_1, h_2 and h_3) indicate three unique dominance effects
915 for the simplex, duplex and triplex heterozygote genotypes for locus Q_A (Q_B). θ_i (or ζ_i) (i = 1,…, 4) are the monogenic, digenic, trigenic and
916 quadrigenic genetic effects at locus Q_A (or Q_B). $I_{\theta_i\zeta_j}$ denote epistasis between the effects θ_i and ζ_j (i=1,…,4; j=1,…,4).'-' indicates the parameter
917 was not estimated. The population mean, all genetic effects of the two loci and epistatic effects were simulated to be 1.0. In scenario 1, the simulation
918 data was generated under either Killick's model (1^K) or our orthogonal contrast scales based model (1^O).

Table 6. Simulation settings based on a single QTL with 10 linked marker loci.

Locus	r	Parental genotypes			Genetic parameters	
		P_1	P_2		$\mu, \theta_1, \theta_2, \theta_3, \theta_4, = 1$	
QTL	0.00	*AAaa*	*AAaa*		$\alpha = 0$	$\alpha = 0.15$
L_1	0.02	$M_1M_2M_3M_3$	$M_5M_6M_7M_8$			
L_2	0.11	$M_1M_2M_2M_4$	$M_5M_2M_2M_6$	G_4	5.458	5.215
L_3	0.19	$M_1M_1M_3M_4$	$M_3M_5M_6M_7$	G_3	1.938	1.787
L_4	0.26	$M_4M_2M_3M_4$	$M_5M_6M_6M_8$	G_2	0.708	0.622
L_5	0.32	$M_2M_1M_4M_2$	$M_4M_5M_7M_6$	G_1	0.271	0.224
L_6	0.38	$M_3M_2M_1M_1$	$M_5M_5M_6M_7$	G_0	0.125	0.089
L_7	0.42	$M_1M_3M_3M_4$	$M_4M_3M_5M_6$	σ	1.928	2.223
L_8	0.46	$M_1M_1M_3M_4$	$M_5M_2M_6M_7$			
L_9	0.50	$M_2M_2M_1M_3$	$M_1M_5M_6M_7$			
L_{10}	0.53	$M_2M_3M_4M_4$	$M_2M_5M_5M_6$			

Markers were located on the same side of the QTL, which is closest to the centromere. r denotes the recombination frequency between the QTL and marker loci. The offspring population of size n = 300 was generated under a tetrasomic inheritance model with double reduction rate set equal to 0.00 or 0.15. Heritability was assumed to be 0.2. Alleles listed in the same column had the same linkage phase.

Table 7. Means and standard errors of the parameter estimates based on 500 repeated simulations of the single QTL model.

Offspring data generated with double reduction rate $\alpha = 0$												
Simulated values				Estimated values								
				α = 0		α = 0.05		α = 0.10		α = 0.15		
μ	1.000			0.992 (0.005)		1.016 (0.005)		1.043 (0.005)		1.072 (0.005)		
θ_1	1.000	V_1	0.667	0.947 (0.006)	0.632 (0.008)	0.959 (0.006)	0.688 (0.009)	0.973 (0.006)	0.773 (0.010)	0.988 (0.006)	0.864 (0.011)	
θ_2	1.000	V_2	0.222	0.953 (0.012)	0.217 (0.005)	0.965 (0.012)	0.251 (0.006)	0.976 (0.012)	0.287 (0.007)	0.986 (0.012)	0.325 (0.007)	
θ_3	1.000	V_3	0.037	1.039 (0.029)	0.055 (0.002)	1.057 (0.028)	0.065 (0.003)	1.068 (0.028)	0.074 (0.003)	1.077 (0.028)	0.083 (0.003)	
θ_4	1.000	V_4	0.003	1.053 (0.092)	0.019 (0.001)	1.095 (0.091)	0.020 (0.003)	1.123 (0.091)	0.020 (0.001)	1.147 (0.091)	0.021 (0.001)	
σ	1.928			1.915 (0.003)		1.915 (0.003)		1.916 (0.003)		1.916 (0.003)		
h^2	0.200			0.200 (0.002)		0.217 (0.002)		0.238 (0.002)		0.258 (0.002)		
Offspring data generated with double reduction rate $\alpha = 0.15$												
Simulated values				Estimated values								
				α = 0.15		α = 0.00		α = 0.05		α = 0.20		
μ	1.000			0.999 (0.005)		0.938 (0.006)		0.950 (0.005)		1.027 (0.005)		
θ_1	1.000	V_1	0.867	0.960 (0.006)	0.814 (0.010)	0.906 (0.006)	0.561 (0.008)	0.925 (0.006)	0.641 (0.008)	0.977 (0.006)	0.908 (0.011)	
θ_2	1.000	V_2	0.311	0.969 (0.011)	0.310 (0.007)	0.903 (0.011)	0.195 (0.004)	0.936 (0.011)	0.234 (0.005)	0.981 (0.011)	0.350 (0.007)	
θ_3	1.000	V_3	0.053	1.118 (0.027)	0.086 (0.004)	1.108 (0.027)	0.059 (0.002)	1.111 (0.027)	0.068 (0.003)	1.122 (0.027)	0.095 (0.004)	
θ_4	1.000	V_4	0.004	1.312 (0.101)	0.026 (0.002)	1.435 (0.100)	0.025 (0.002)	1.343 (0.101)	0.025 (0.002)	1.313 (0.101)	0.027 (0.002)	
σ	2.222			2.212 (0.004)		2.216 (0.004)		2.213 (0.002)		2.212 (0.002)		
h^2	0.200			0.201 (0.002)		0.146 (0.001)		0.164 (0.002)		0.219 (0.002)		

μ is the population mean and θ_i (i=1,…,4) are accordingly monogenic, digenic, trigenic and quadrigenic genetic effects of the QTL. σ is the environmental error and h^2 is the heritability. V_1, V_2, V_3 and V_4 represent monogenic, digenic, trigenic and quadrigenic genetic variance components, respectively. The estimation procedure was carried out assuming a range of values for the coefficient of double reduction α. The simulated values of α are highlighted in bold.

Table 8. Inference of major QTL genes affecting flowering time and plant height in a segregating population of autotetraploid potato.

	Flowering time (days)	Plant height (cm)
G_{P_1} / G_{P_2}	*QQqq / QQqq*	*QQQq / QQQQ*
$\hat{\alpha}$	0.055	0.165
-ln(L)	939.47	1099.89
LOD score	13.83	5.70
P value	0.0000	0.003
$\hat{h}^2$	79.38	72.67
Mean	33.62	45.97
Monogenic	5.92 (25.97)***	12.46 (51.61)***
Digenic	2.83 (2.03)***	-6.84 (1.31)***
Trigenic	0.59 (0.02)***	-
Quadrigenic	-5.09 (0.09)***	-
V_{QTL}/V_{Total} (%)	39.7	29.7
$V_{QTL}/V_{genetic}$ (%)	50.1	40.8

Estimated parameters of the quantitative genetic model are given based on the most likely parental genotype configuration. Monogenic, digenic, trigenic and quadrigenic genetic effects estimated from the orthogonal contrast scales model are shown, with the genetic variance component in brackets. *** p-value < 0.0001 from the two-tailed t-test.