Genetic Algorithm Based Approach to Optimize Phenotypical Traits of Virtual Rice

Weilong Ding [1], Lifeng Xu [1], Yang Wei [1], Fuli Wu [1], Defeng Zhu [2], Yuping Zhang [2], Nelson Max[3]

[1] College of Computer Science & Technology, Zhejiang University of Technology, Hangzhou 310026, China

[2] State Key Laboratory of Rice Biology, China National Rice Research Institute, Hangzhou 310006, China

[3] Department of Computer Science, University of California, Davis, 95616, USA

E-mail: wlding@zjut.edu.cn

Abstract: How to select and combine good traits of rice to get high-production individuals is one of the key points in developing crop ideotype cultivation technologies. Existing cultivation methods for producing ideal plants, such as field trials and crop modeling, have some limits. In this paper, we propose a method based on a genetic algorithm (GA) and a functional-structural plant model (FSPM) to optimize plant types of virtual rice by dynamically adjusting phenotypical traits. In this algorithm, phenotypical traits such as leaf angles, plant heights, the maximum number of tiller, and the angle of tiller are considered as input parameters of our virtual rice model. We evaluate the photosynthetic output as a function of these parameters, and optimized it using a GA. This method has been implemented on GroIMP using the modeling language XL (eXtended L-System) and RGG (Relational Growth Grammar). A double haploid population of rice is adopted as test material in a case study. Our experimental results show that our method can not only optimize the parameters of rice plant type and increase the amount of light absorption, but can also significantly increase crop yield.

Key Words: functional-structural model; genetic algorithm; plant type; optimal design

0. Introduction

The qualities of a rice plant's type and canopy structure play key roles in increasing its production and improving its quality. Therefore, more and more scientists in this area focus on the problem of optimizing phenotypical traits of rice in order to obtain ideotype individuals with high or super-high production.

Two main strategies are used to investigate ideotype breeding (Qi et al. 2010; Andrivon et al. 2013): 1) the experiment based approach (Peng et al. 2008; Srbislav 1994); and 2) the plant model based approach. For the first approach, the breeding experts implement numerous test-cross-screenings on crop plants in the field and then carefully observe, analyze, and summarize their phenotypical traits. Finally, ideotypes can be obtained after long term breeding and observation. Although this method can find an optimal plant type close to the ideotype we want, it is time-consuming and difficult to verify whether the cultivated plant types achieves the optimal shapes or not (Su et al. 2008). In addition, this approach takes little consideration of environmental and other random factors which have influence on the growth of plants. The other method of finding a crop ideotype, the plant models-based approach (Yin et al. 2003; Cilas et al. 2006; Markus et al. 2007; Letort et al. 2008; Suriharn et al. 2011; Milo and Last 2012; Semenov et al. 2014; Van Oijen et al. 2016) can more accurately describe the relationship between the growth and development of plants and environmental factors. Using this method to study plant growth behavior not only compensates for the shortage of field experiments, but also saves time and resources. However, the method does not contain the morphological information of plants, so it cannot be used to study the relationship between plants and the environment by combining it with

a three-dimensional morphology method. Therefore, it is necessary to explore a new method for cultivating the ideotype crop.

Considering the above problems and previous research, we propose a method based on a genetic algorithm (GA) and a functional-structural plant model (FSPM). This model has the ability to optimize virtual rice by dynamically adjusting its phenotypical traits. A plant functional-structural model effectively combines the structures of plants with their physiological functions, and can quantitatively simulate the growth and development of plants and make the whole process visible (Lu et al. 2014; Room et al. 1996; Sarlikioti et al. 2011a and 2011b). In this model, through the calculation of photosynthetic yield in a crop canopy, the assimilation in different organs and their final productions, we can obtain the phenotypical traits of a rice plant which can maximize the amount of light radiation intercepted and the grain yield.

As an effective optimization method, the genetic algorithm (Qi et al. 2010; Quilot-Turion et al. 2012; Drewry et al. 2014) has been applied in some complicated optimization problems of different plant types and achieved good results. Our virtual rice model is built according to the principle of a plant functional-structural model. In this model a genetic algorithm is employed to study how to optimize the parameters of rice plant type, such as leaf inclination angle, plant height, maximum tiller number, and tiller angle. The new model provides a method for developing a crop ideotype.

1. Functional-structural Model of Rice

A rice plant is made up of roots, stem, leaves and panicles, as shown in Fig.1. The jointed stem of the rice consists of a series of nodes and internodes. The composition of a rice panicle includes panicle axis, primary and secondary branch, and spikelet (before maturity)/grain (after maturity). Rice plants have three growth stages: 1) vegetative (from germination to panicle initiation); 2) reproductive (from panicle initiation to flowering); and 3) ripening (from flowering to mature grain). For more information about the rice plant, please refer to the literatures (Sharma et al. 1999) and (Karen et al. 2002).
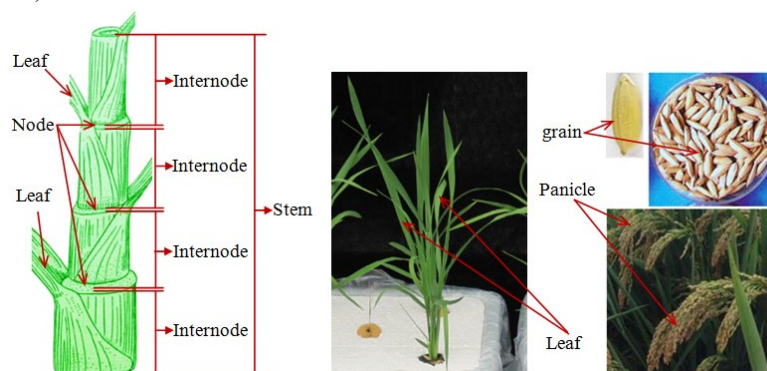

Fig. 1 the organs of rice plant

The Functional-structural model described in this article is constructed using the eXtended L-System modeling language (XL) (Kniemeyer 2004) on an open-source platform, GroIMP (Kniemeyer et al. 2008). XL is developed based on the Java programming language, and further integrated with the parallel rewriting rules of the L-system (Kniemeyer et al. 2004). GroIMP is an interactive platform for virtual plant development, including model construction, visualization, interactive modules, etc. Using grammars of the XL language and the Java programming language, a model of virtual plant can be constructed based on various rules and algorithms.

The rice FSPM integrates physiological processes, morphogenesis rules, and related environmental parameters together in one model. It is capable of simulating the dynamics of the production and consumption of biomass based on several models, i.e. the photosynthesis model with parameters of leaves and intercepted light calculated based on topological and geometric information, the carbon partitioning model with prediction of source-sink relations, and the growth model with organ formation and morphogenesis processes (Xu et al. 2009).

The structural part of the rice FSPM describes the process of the plant morphological construction and development of the stems, grains, and leaf blades. A cylinder is used to simulate one internode, and multiple connecting internodes made a stem or a spikelet. A grain is constructed using a curved surface and texturing. Furthermore, an instance variable j is designed within each leaf object, indicating the amount of intercepted light, which is a key factor in influencing the assimilation rate and the amount of photosynthetic yield. Detailed descriptions of the structural model can be found in articles of Xu et al. (2009 and 2011).

There are five main modules contained in the rice FSPM: the light environment module, the radiation model, the photosynthesis model, the carbon partitioning model, and growth functions for organs.

1) Light environment module

In the light environment module, direct and diffuse light radiation components are established to simulate the distribution of light in the sky. The former is solar radiation directly reaching the Earth's surface in the form of a parallel solar beam, while the later is solar radiation being scattered after passing through the atmosphere. The sky is simulated as a set of light sources (Evers et al. 2010; Buck-Sorlin et al. 2011) which represent light with different light intensity from different positions. Rays emitted by all light sources are uniformly distributed parallel rays, and uniformly put into the scene. Stocking density is controlled by the number of rays, and the respectively formulae of the direct radiation E_{dir} (W/m^2) and diffuse radiation E_{dif} (W/m^2):

$$E_{dir} = 0.85 \times E_0 \times (r_o/r)^2 \times \sin\beta \qquad (1)$$
$$E_{dif} = 0.15 \times E_0 \times (r_o/r)^2 \times \sin\beta \qquad (2)$$

where E_0 is the solar constant 1368W/m^2, r_o is the mean Earth-Sun distance, r is the real Earth-Sun distance, and β is an elevation angle of the light source.

2) Radiation model

The amount of Photosynthetic Active Radiation (PAR), intercepted by the canopy of rice is calculated using a radiation model in the platform of GroIMP (Hemmerling et al., 2008). This model was based on a Monte-Carlo ray-tracer, which simulates the certain number of rays emitted by all light sources as well as the number of times a ray is followed on reflection. This Monte-Carlo simulation is done on an isolated rice plant. An instance of the radiation model is invoked and carried out at each step, and Shaders (Kniemeyer et al., 2008) are used to map onto the geometrical objects representing plant organs in our rice FSPM. A combination of 100,000 rays and 10 reflections per ray are used as the two parameters to instantiate the light object within the scene in this case. The path of a ray in the scene and the likelihood with which it will be absorbed, reflected or transmitted depend on the geometry and distribution of objects in the scene as well as on their optical properties (Buck-Sorlin et al., 2011), which can be calculated by a method defined in the platform of GroIMP, radiation.compute(). The amount of PAR (*par*) absorbed by a leaf (*lf*) can be calculated as (shown in XL):

par=*radiation*.getAbsorbedPower(*lf*).integrate()*GRPAR;

where the method getAbsorbedPower() of radiation returns the spectrum of the absorbed PAR of a leaf, method integrate() is used to accumulate absorbed PAR of each patch of a leaf, and GRPAR is a coefficient to convert spectrum to Photosynthetic Photon Flux Density.

3) Photosynthesis module

The photosynthesis model was established based on a model named LEAFC3, which was proposed by Nikolov et al (1995). It can be used to estimate steady short term carbon dioxide, water vapor and heat flux in the leaves of C3-plants. In this study, the photosynthesis model is updated by changing the species-specific parameters and the environmental parameters accordingly (Xu et al. 2009). With the prediction of the physiological and biochemical responses, stomatal conductance, leaf energy balance, and the other relative physiological processes, photosynthetic rate (the current average leaf photosynthetic rate, mol/m^{2}s) can be estimated. And in each growth day the accumulation of photosynthetic production of $P(t)$ (unit: g) from a leaf with the area of a (unit: m^2) can be further obtained.

4）Assimilation Distribution Module

A public assimilate pool (Xu et al. 2009) is used to store the products of local assimilation of each single leaf, which is calculated as above based on the light distribution and instantaneous assimilation rate. The photosynthetic products $P(t)$ are added to the common assimilate pool (CAP), so that the common pool assimilates biomass production $CAP(t)$ (unit: g) on the t-th day:

$$CAP(t)=CAP(t\text{-}1)+P(t) – C_{mr}(t) \qquad (3)$$

where $C_{mr}(t)$ represents the assimilates consumed by respiration. Maintenance respiration is computed as a fixed proportion (0.014 g glucose g^{-1} dm) of structural biomass (Goudriaan and van Laar, 1994). The amount of obtained assimilates in the growth step t for biomass increment ($G_o(t)$, unit: g) of the specific organ O on the t day, can be calculated as,

$$G_o(t) = \min(S_o(t), (S_o(t)/ S_{total}(t))\times CAP(t)) \qquad (4)$$

where $S_o(t)$ is the sink strength of organ O, whose value is variable during its life and determined by a flexible mathematical function based on the beta function (Yin et al. 2003) and its parameters are estimated by inverse methods (Mathieu et al. 2007).The sink strengths of the same kind organs are equal at the same stage of growth. $S_{total}(t)$ is total sink strength of all the organs. At growth day t, the biomass ($D_o(t)$, g) of the organ O can be calculated as

$$D_o(t) = D_o(t\text{-}1) + G_o(t) \qquad (5)$$

5）Organ growth modules

The amount of organ growth in various dimensions $M_o(t)$ (growth in area for leaves, unit: m^2; growth in length for internodes, unit: m) can be calculated using the growth module, according to the relationship between the change of biomass and organ size. The formula is:

$$M_o(t)=G_o(t)\times\mu_o \qquad (6)$$

where μ_o is the conversion factor between biomass and the size of organ O, For different organs, the conversion factor μ_o values are different. But for the same organ, we assume that values of μ_o are the same so as not to increase the complexity of the model. For the blade and internode of rice, the value of μ_o respectively is 0.023 m^2/g and 0.851 m/g (Xu et al. 2011), and for the grains, it is 1, since grain size is measured as mass, so no conversion is necessary.

After obtaining organ geometry, the three-dimensional shape of rice of the current growth stage can be simulated by using visualization techniques.

2. Optimization approaches for plant types of virtual rice based on a Genetic Algorithm

Although there are many genetic, biochemical, and environmental factors which affect the growth of a plant, in this paper we consider mainly the geometric aspects of plant morphology. The reasons are: 1) the morphogenesis and architecture of a rice plant are critical factors in the yield equation (Watanabe et al. 2005), 2) the geometrical information of leaf, tiller, etc. were

always considered in developing high-yielding rice varieties through ideotype breeding (Peng et al. 2008; Yuan et al. 2015). The morphological structure of a rice plant is determined mainly by its leaf inclination angle, height, maximum tiller number and tiller angle, etc. Therefore, different plant types have different combinations of these factors and lead to various growth behaviors and final yields. Among these different combinations of plant type parameters, there will be one or more which maximize the grain yield of an individual plant. Supposing $f(x)$ denotes the mapping function of the combination of plant type factors to grain yield, the constrained optimization problem can be described as the following mathematical model:

$$\text{Maximize } f(x) \in S \quad (7)$$

where x is the vector of values for $A_1 \sim A_{n1}$, PH, TM, and TA, A_i is the angle of leaf i (unit: degrees) on the main stem, $n1$ is the number of leaves on the main stem, PH is the height of rice, TM is the maximum tiller number, TA is the tiller angle, and S is the set of yields of different plant type combinations).

Each plant type factor has its own value range. In this paper, our genetic algorithm can give the optimal combination of different input parameters which allows rice plants to generate the maximal yield.

2.1 Data encoding method and Population initialization

Symbol strings are adopted to stand for individuals in our algorithm. Therefore, each plant type factor $A_1 \sim A_{n1}$, PH, TM, and TA must be encoded as a symbol string. In this paper, the symbol string is represented as an unsigned binary integer string. Each individual is made of a chromosome complement $\{g_1, g_2, ..., g_{n_g}\}$, where g_i is the i-th plant type factor and n_g represents the total number of plant type factors which are contained in the chromosome complement. Each plant type factor g_i is formed by binary integer bits. Every information bit is 0 or 1. Therefore, g_i can be represented as:

$$g_i = \{b_i^1, b_i^2, ..., b_i^{n_b^i}\} \quad (8)$$

where b_i^j is the value of the j-th information bit; n_b^i is the string length of binary encoding of each plant type factor b_i^j, and its value is determined by both the range $[u_i, v_i]$ and the value accuracy factor s of this plant type, according to the equation:

$$2^{n_b^i} \leq (v_i - u_i) \times \frac{1}{s} \leq 2^{n_b^i + 1} \quad (9)$$

The total length of binary encoding string, L_R, is calculated by the following formula:

$$L_R = \sum_{i=1}^{n_g} n_b^i \quad (10)$$

In our algorithm, we group leaf angle, plant height, maximum tiller number and tiller angle of different leaf positions together, so an individual genotype which can represent a feasible solution is generated. Then a certain number of individuals will be randomly generated to build the initial population pool. For the binary encoding string of each individual plant type whose string length is L_R, each random information bit has equal probability of being zero or one.

2.2 Design of genetic manipulation

Genetic manipulation describes how the genetic algorithm simulates natural selection, propagation, crossover, and mutation. The genetic manipulation of this paper includes only selection, crossover and mutation. We choose individuals which have higher fitness scores from the current population and use a roulette method to generate the next generation. Elitist selection guarantees that groups to converge. The individual with the largest fitness value in a population is defined as the elite. In our algorithm the fitness value of the elite individual is compared with those of the current individuals. If the fitness value of the optimal individual is larger than the elite, the elite one is replaced by the optimal one; if not, the worst individual in the current population is replaced by the elite individual. Our crossover method can pass the original good genetic materials to the next generation, and new individuals with more complicated genetic structures will be generated. Our layered mutation method divides the mutation operation into two levels: the individual level and the information bit level. At the individual level, the mutation individual is selected by a certain individual mutation probability P_{mi}. At the information bit level, every information bit is randomly flipped by a predefined probability P_{mb}.

2.3 Design of fitness function

Good plant morphology is beneficial to the spatial configuration of plants, and the full use of available light energy and gas exchange in the canopy, which ultimately contributes to yield improvements (Ao et al. 2013). According to the study of Chen et al. (2005), the utilization ratio of light energy is increased and thus the yield of the rice is promoted. In this process, the plant morphology is a critical limiting factor for light energy utilization. The final rice yield is related to the grain number of the rice ear and quality of grain. It was assumed that for formation and maintenance of a spikelet will consume an amount D_{flower} of assimilate. If rice starts putting forth ears at the t_{grain}-th day, the maximum grain number n_{gmax} is:

$$n_{gmax} = CAP(t_{grain})/D_{flower} \qquad (11)$$

In the t-th day (d), yield $Y(t)$ is the sum of all grain. The formula is as follows:

$$Y(t) = \begin{cases} \sum_{k=1}^{n_{grain}(t)} D_{grain}(k,t), & t \le t_{grain} + 40 \\ \sum_{k=1}^{n_{gmax}} D_{grain}(k,t), & t > t_{grain} + 40 \end{cases} \qquad (12)$$

where $n_{grain}(t)$ stands for the grain number during the t-th day, $k \in [1, n_{grain}(t)]$, and D_{grain} is the assimilate quality (unit: g) that is consumed during the formation of a grain. The value of $D_{grain}(k,t)$ depend on k, t, and $CAP(t)$, and can be calculated by using formulas (4) and (5). The duration time from heading t_{grain} to the end of grain filling t_e is about 40 days (Liu et al. 2004). The fitness function $f(x)$ is

$$f(x) = Y(t_e) \qquad (13)$$

where t_e is the end of the growing season. Formula (13) can be used to evaluate each individual plant type's quality.

2.4 Strategies for avoiding redundant calculation

We found that there may be redundant calculation of some individual fitness scores in the genetic algorithm. This happens mainly in two situations: 1) Because of the competition operation, the elite individual of the last generation passes into the current population. In fact, its fitness has already been calculated, but it will be calculated again in the current generation. 2) Some plant

types in the current population have appeared previously in the evolutionary process, and their fitness values have been calculated, but the algorithm still calculates their fitness values again. Obviously, the above two situations increase the time cost the whole algorithm. In order to improve the efficiency, we propose here an improved version of the algorithm based on a hash table to avoid redundant calculation:

Step 1: When the algorithm starts, initialize a hash table. Plant types are regarded as keys in the hash table, and their yields are the values of the elements in the hash table.

Step 2: Before calculating individual's fitness, check if the plant type is stored in the hash table. If it is, step 3 will be executed; if not, step 4.

Step 3: This plant type's quality of yield is taken out from the hash table according to its plant type.

Step 4: Calculate the plant type's fitness, and store the plant type and its corresponding yield as a key-value pair into the hash table.

2.5 Optimization algorithm for plant types

Here we propose a genetic algorithm to obtain the optimal combination of different input parameters. The algorithm takes formula (13) as the fitness function. Before the algorithm runs, we should initialize the first population P_0 and set the parameters of the genetic operators, such as the mutation probabilities P_{mi} and P_{mb}, the crossover probability P_c, etc. In this algorithm, the individuals in a new population are produced based on genetic operators and compete with the previous elite individuals. Then the new elite individuals will be selected according to their fitness values. This process is executed repeatedly until the termination conditions of the evolution are met. After many time steps of evolution, the optimal or suboptimal solutions will be obtained. The steps are:

Step 1: Initialize the first population P_0, calculate each individual's fitness in the population.

Step 2: Define the elite to be the individual with highest fitness.

Step 3: Randomly choose two individuals among the current population and conduct an interlace operation based on the crossover probability P_c, to generate a new pair of individuals; repeat this process $n_c/2$ times to get a temporary population containing n_c individuals.

Step 4: Conduct a mutation operation on the temporary population according to the mutation probabilities (per individual P_{mi} and per bit P_{mb}).

Step 5: Look up or calculate the fitness of each individual in the temporary population.

Step 6: Update the elite individual if the temporary population contains an individual with greater fitness; otherwise replace the least fit individual by the elite individual.

Step 7: Judge whether the terminal condition is met (the step count (*iterationNum*) is equal to the maximum generation number T_{max}). If yes, go to step 8; if not, increment *iterationNum* and loop back to step 3.

Step 8: Output the current elite individual.

The complexity class of this method is $O(kn)$ where k is the iteration count and n is the number of individuals in the initial population.

3. Experiments and results

A Double Haploid (DH) population of rice was used in the implementation of the rice FSPM as a case study on the optimization of plant type. Using different combinations of the above-mentioned factors as model inputs through which the final yield can be simulated, the

specific plant type of rice with the best yield can be selected based on a large amount of sample individuals. The study was implemented on the platform GroIMP with XL and the Java programming language. The computer used in the experiments had an Intel® Core ™ 2 Duo E7300 2.66 GHz, and a 4 GB memory.

3.1 Validation of functional-structural model of rice

In this study, we construct a morphology model and a physiological function model and establish the dynamic association mechanisms between them. The model uses one day as the growth step, and changes the model state through time to stimulate the rice growth process. In the process of establishing the model, we first use the structure model to calculate the shape data of rice, such as the number and area of each leaf, the number and size of other organs, the next growth phase organs and the topological relationship between organs, etc. Based on this data, and the rice functional model, we calculate assimilate production. Assimilates are produced by removing the consumption of respiration, then we calculate the biomass increment $G_o(t)$ of each organ, according to formula (4), based on plant morphology, environmental information and the sink strength of the various organs. According to equation (6), and the allometric relationship between organ biomass and size, we calculate the amount of change in the size of the various organs of $M_o(t)$. We calculate the production and distribution of assimilate of the next phase of the plant according to the updated plant morphology. Then we repeat the whole process. Using the functional model, we calculate the biomass. When it is not sufficient to sustain the morphology's growth, the rice stops growing. Fig.2 compares the simulated forms of the double haploid rice with the morphologies of a rice plant in different growth stages. As can be seen from Fig.2, the simulated results and the real morphologies are similar.
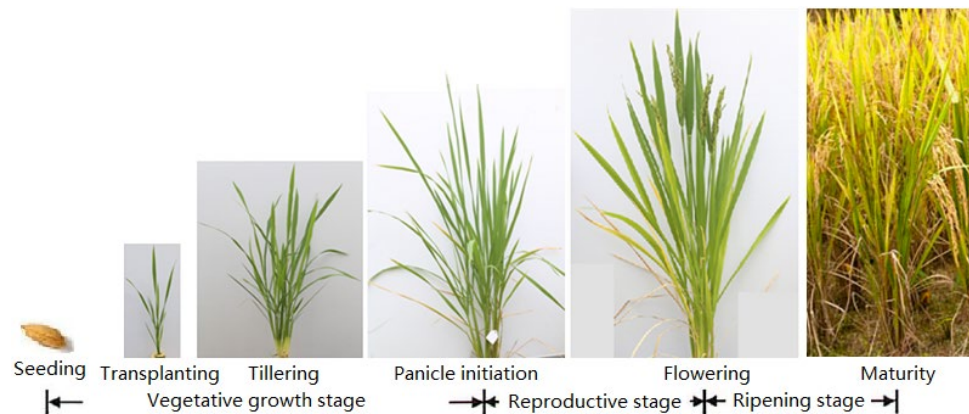
a) The morphologies of a real rice plant in different growth stages
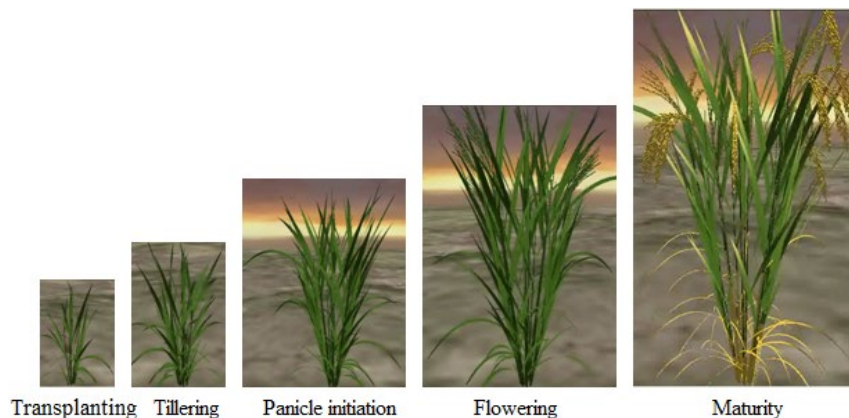
b) The simulated shapes of a rice plant in different growth stages

Fig.2 Comparison of the simulated shapes with the real morphologies of rice pant

To verify the validity of the physiological function model of rice, we study the central assimilate pool during the growth period of rice, which is shown in Fig.3. As can be seen from Fig.3, the change period of assimilates can be divided into three stages. Prophase (0-60d): leaves are in the initial-growth state. In this stage, formation and growth of organs consume all the assimilates provided from the limited leaves, and even take the available carbon from the central pool, and this is why the size of the pool decreases. During the middle and internode extension stage (60d-105d), the leaves are almost all unfolded and thus providing more assimilates than those needed by organ growth and extension, and this is reflected by the rising of the curve. In the late stage (grain filling period), the need of biomass for grain filling is taking up the dynamic source from leaf photosynthesis, and the available assimilates stored in the pool are exhausted, thus making the curve drop sharply.
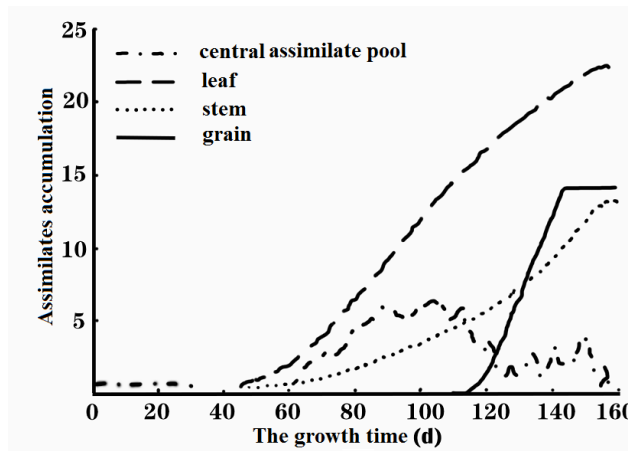

Fig.3 Dynamic change of the assimilation rate and organ's weight in the central assimilation pool

3.2 Genetic algorithm parameter setting and the optimal results of rice plant type

The GA control parameters include population size n_c, crossover probability P_c, two mutation probabilities (per individual P_{mi} and per bit P_{mb}) and the maximum generation number T_{max}. Since the simulation part of the structural-functional model is time consuming, a large population size will increase the total running time. Therefore, the population size is set to 20 in our model. Generally, the value of crossover probability P_c is set to a large number, which can speed up the convergence rate. However, if the crossover probability is too large, it may destroy excellent genetic structures (Gan et al. 2012). Considering that the elite individuals should retain some good genes, we set the crossover probability to 0.9 in order to produce a variety of genetic structures. According to the classical genetic algorithm parameter setting strategy and taking into account the efficiency of the algorithm, we set mutation probabilities (P_{mi} and P_{mb}) to 0.08 in the model. The relationship between computation time and final step count is linear. For the initial tests, we set T_{max} to 200 and repeat the program 10 times with different random number generation seeds and record the optimal fitness values of individuals of different groups each time. Fig.4 shows the functional relationship between each individual fitness value and the population diversity with increasing generation number. As shown in Fig.4a, the fitness values of individuals in the population gradually stabilized with increasing generation number. When the number of generations is from 60 to 100, the variation among fitness values becomes smaller. From Fig.4b, with the increasing of the evolution, the value of population diversity first reduces and then keeps

fluctuating within the range of 0.4 to 0.7. We set higher genetic diversities to ensure the normal function of the evolution operator, and avoid the algorithm having premature convergence to a local optimum, making it hard to find the global optimum. After 60 generations, the value of each individual fitness has only small variations, and the population diversity remains at a high level. However, with increasing evolution generation, the computation time also increases significantly. Synthesizing different factors, we set the termination step count T_{max} to 100, which meets the convergence requirements of our algorithm.

a) Change of best individual fitness   b) Change of population diversity
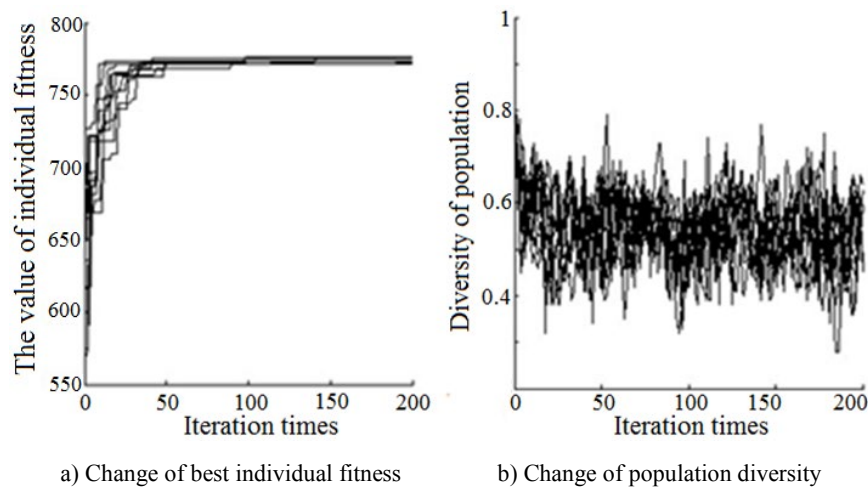
Fig.4 Changes of best individual fitness value in each generation and diversity of population

In this paper, the value of the total bit length L_R depends on the number of plant type components and the range of each plant type component. According to previous research results (Yan et al. 1999), we determine the value ranges of plant type components which need to be optimized, as shown in Table1. The values in the center column are empirical value from real rice, not the results of simulation. These empirical values are based on long-term, partly unpublished experimental studies from De-feng Zhu's laboratory. Based on them, we can set the lengths of the code $A_1 \sim A_{n1}$, PH, TM, and TA to 7, 6, 3, 5 respectively, because this is the number of bits necessary to specify an integer in the allowed range. The main stem of the simulated rice plant has 14 leaves, so $n1$ is equal to 14. After a period of running, we obtain the values of components of each optimized target plant type (as shown in Tab.1). Using the parameters of optimized plant type as the inputs of the rice function structure model, we simulate the rice plant in Fig.5. Fig.5 shows that the lower part of the canopy of the optimal plant hangs down loosely, so light can enter into the lower part much more easily. The leaves in higher part of the canopy are close to upright, so that the upper leaves can get light at each side. Moreover, it also can make more light enter into the lower part of the canopy. As a result, the heights of plants are a little bit shorter, and the numbers of grains are larger. Fig.6 shows the comparison of the amount of light absorbed by the optimal rice and the one before optimization in different growth stages. From Fig.6a we can see that the ability to absorb light of the optimized plant increases significantly. Comparing the rice individuals after optimization with those before, we can get the yields as shown in Fig.6b. From Fig.6b, we can see that after the 125th day, the plants which are optimized have larger yields than the plants which are not. From then on, the disparity grows continuously larger. In the simulation experiments of 20 runs of the rice individuals to full maturity, the average yield of individuals without optimization was 14.11g, while that of those optimized was 21.95g. The result is that the later has 55.56% more yield than the former. So combining the genetic algorithm plant type design

methods, we can estimate the best plant phenotype within the range of breeding diversity, and thus design the rice ideotype.

Tab.1 Parameter values of plant type factors before and after optimization

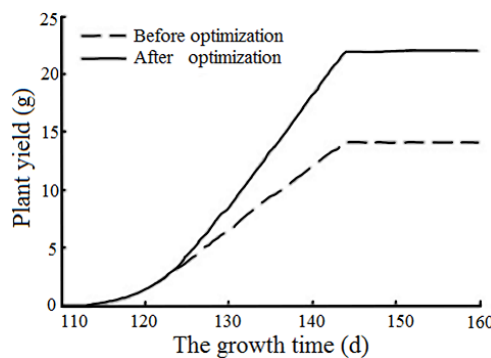| Phenotypical traits of rice	Range in real rice	The value after optimization
Height of rice plant(cm) PH	[91,136]	96
Maximum number of tiller TM	[8,15]	15
Angle of tiller TA(º)	[6,33]	22
Leaf angle $A_1{\sim}A_{14}$(º)	[16, 87]	38, 35, 33, 55, 60, 61, 66, 63, 69, 70, 72, 77, 63, 66

a)   Before optimization   b)   After optimization

Fig.5 Morphologies of reference plant
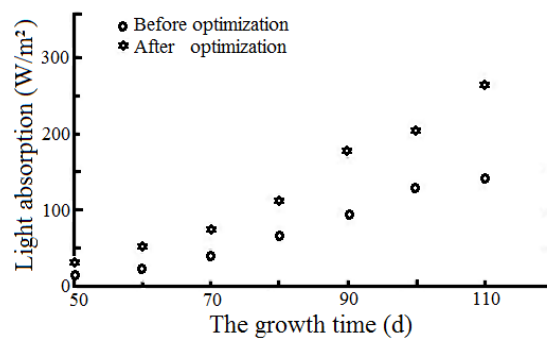
a) Light absorption change of rice canopy

b) Plant yield changes

Fig. 6 Light absorption change of rice canopy and plant yield changes of rice plant

4. Discussion

The current concentration of plant simulation studies is not on applications but rather on modeling methods. This study is a new effort towards modeling. Combining a genetic algorithm with a plant functional-structural model, and using several phenotype characters of dynamic development in the model system as genetic factors in the algorithm of virtual rice, we can get a prediction of final rice yield. This research achieves the integration of an optimization algorithm

with the model system in both temporal and spatial dimensions. It is an exploration of a new application area of plant functional-structural models.

Lu et al. (2012) set up an ideotype model of wheat and solved it with MATLAB based on a Genetic Algorithm. But this model is a purely mathematical model and does not consider the physiological processes of wheat growth. Liu et al. (2015) presented an algorithm to automatically measure the plant type parameters based on image processing. Although the research objects are the same, the methods they adopted are different from ours. Qi et al (2010) combine an optimization algorithm with a source-sink model in a plant functional-structural model, and determine the optimal mix proportion of the weight of a corn with the total weight of its leaves and stems. This research connects the optimization algorithm with a mathematical model of the corn physiological function, and realizes the optimal production corn plant type. Quilot-Turion et al. (2012) combines Genetic Algorithms and a plant model, which can provide a reference for applying the genes in a model. However, the plant model uses a physiological function model, and the optimized target it optimizes is a parameter set of that physiological model, which is not on the same level as the phenotype of breeding practice. So the methods and results of that research cannot be easily applied to the breeding process. In contrast to Qi et al. (2010) and Quilot-Turion et al. (2012), which use parameters related to physiological functions, our research starts by connecting the physiological function to a structural model, and chooses geometric traits (plant height, length of leaf etc.) to optimize, so it provides intuitive reference information for quantitative plant design.

On the problem of how to use a functional-structural model to optimize the type of a rice plant, the better method is to simultaneously optimize the function parameters and the morphology parameters to which they are connected, to achieve a systemic prediction. For the modeling system, simultaneously optimizing the morphology parameters and the function parameters will increase the complexity of the model, because it needs to check the time varying changes with diverse functional and structural parameters. The whole process needs the support of enough experimental data. Therefore we consider the biological functions such as photosynthesis, biomass allocation and organogenesis. The photosynthetic rate has a very close connection with the rice variety and growth environment. In our model we consider the influence of different shadows and light on the photosynthesis. Biomass allocation and organogenesis are both related to the final scale of the 3-dimensional shape of each organ (such as the length of internodes and surface area of leaves). Therefore, we consider the influence of different morphology on physiologic function rather than optimizing the function parameter directly. In addition, the plant type is a very broad definition which includes not only the plant morphology, but also the layout of spacing of plants in a field and physiological aspects. The plant type is an integrative concept of the plant morphology structures, its physiological functions, etc. Currently our study is limited to morphology, and we have not yet considered these other aspects.

5. Conclusions and future work

We present an automatic design optimization method for the plant type of rice based on a functional-structural model of rice and a genetic algorithm. The experimental results demonstrate the effectiveness of our method. In the future, we will consider not only the morphology of the plant, but also take the field layout and spacing and physiological aspects into account, so that we can design a more effective high-yielding plant type. In addition, we intend to study high-efficiency and low-cost functional-structural models, fitness function design methods, and

genetic operation selection strategies in order to enhance the ability of global optimization and the convergence performance of our algorithm. We know the rice plant will be dead after the ripening stage but a new plant originates. So we will simulate the process of plant death in future. The computer system we used consists of outdated hardware and the simulation is only computed on the CPU so that the time used for simulations is long. In future we will make use of the GPU to improve the efficiency of calculation in order to save modeling time. Combined with the rice breeding process and field experimental design, the optimization method can be validated. Also with the development of the modelling technology and improvement of mechanisms of the models, a rice model combined with GA algorithms can accelerate the breeding processes.

Acknowledgment

References

Andrivon D, Giorgetti C, Baranger A, et al. 2013. Defining and designing plant architectural ideotypes to control epidemics? Eur J Plant Pathol, 135:611-617.

Ao X, Zhao M H, Zhu Q, et al. 2013. Study on plant morphological traits and production characteristics of super high-yielding soybean. Journal of Integrative Agriculture, 12(7):1173-1182

Buck-Sorlin G, deVisser P H B, Henke M, et al. 2011. Towards a functional-structural plant model of cut-rose: Simulation of light environment, light absorption, photosynthesis and interference with the plant structure. Annals of Botany，108(6), 1121-1134.

Chao Lu, Jiali Han, Fangjin Hu, Taigui Qin. 2012. Mathematical model of wheat stalk lodging-resistance during the later growth period. Mathematics in practice and theory, 42(15), 46-53. (In Chinese with English abstract)

Chen Y D, Huang, Q M, Zhang X. 2005. The breeding of rice plant-type. Shanghai Scientific & Technical Publishers. (In Chinese with English abstract)

Cilas C, Bar-Hen A, Montagnon C, et al. 2006. Definition of architectural ideotypes for good yield capacity in Coffea canephora. Annals of Botany, 97(3), 405-411.

Drewry D T, Kumar P, Long S P. 2014. Simultaneous improvement in productivity, water use, and albedo through crop structural modification. Global Change Biology, 20(6), 1955-1967.

Evers J B, Vos J, Yin X R, et al. 2010. Simulation of wheat growth and development based on organ-level photosynthesis and assimilate allocation. Experimental Botany, 61(8),2203-2216.

Gan Z B, Zhu C X, Ma Y, et al. 2012. Concurrent negotiation of relative issues based on genetic algorithm. Journal of Software, 23(11), 2987-2999. (In Chinese with English abstract)

Goudriaan J, van Laar HH. 1994. Modelling Potential Crop Growth processes. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Hemmerling, R., Kniemeyer, O., Lanwert, D., Kurth, W., Buck-Sorlin,G., 2008. The Rule-based Language XL and the Modelling Environment GroIMP Illustrated with Simulated Tree Competition. Functional Plant Biology 35, 739-750.

Karen A. K. Moldenhauer, Julia H. Gibbons. 2002. Rice Morphology and Development. In C. Wayne Smith, Robert H. Dilday (Eds). Rice: Origin, History, Technology, and Production. John Wiley & Sons, Inc., Hoboken, New Jersey. pp.103-128.

Kniemeyer O. 2004. Design and Implementation of a Graph Grammar Based Language for Functional-Structural Plant Modelling. Senftenberg, Brandenburgischen Technischen Universität.

Kniemeyer O, Hemmerling R, Buck-Sorlin G, et al. 2008. The rule-based language XL and the modeling environment GroIMP, illustrated with simulated tree competition. Functional Plant Biology, 35(9-10), 739-750.

Kniemeyer O, Buck-Sorlin G, Kurth W. 2004. A graph-grammar approach to artificial life. Artificial Life. 10(4),413- 431.

Kniemeyer, O., Kurth, W., 2008. The Modelling Platform GroIMP and the Programming Language XL. Lecture Notes in Computer Science 5088, 570-572.

Letort V, Cournède P H, Mathieu A, et al. 2008. Parametric identification of a functional-structural tree growth model and application to beech trees (Fagus sylvatica). Functional Plant Biology, 35(10), 951-963.

Lu F, Jean-Claude M, Hervé R, et al. 2014. Comparing an empirical crop model with a functional structural plant model to account for individual variability. European Journal of Agronomy, 53, 16-27.

Lu C, Han J L, Hu F J, Qin T G. 2012. Mathematical model of wheat stalk lodging-resistance during the later growth period. Mathematics in practice and theory, 42(15), 46-53. (In Chinese with English abstract)

Liu X W, Meng Y L, Zhou Z G, Cao W X. 2004. The primary study on floret and grain development simulation in rice. Chinese Journal of Rice Science, 18(3), 249-254. (In Chinese with English abstract)

Liu C H, Chen L J, Lv C Y, Ren W T. 2015. The plant type parameters measurement algorithm of rice based on image processing technology. Journal of agricultural mechanization research. (12), 232-235. (In Chinese with English abstract)

Markus H, Shan C G, Wang P, et al. 2007. A model based ideotyping approach for wheat under different environmental conditions in North China plain. Agricultural Sciences in China, 6(12), 1426-1436.

Mathieu A, Zhang B, Heuvelink E, et al. 2007. Calibration of fruit cyclic patterns in cucumber plants as a function of source-sink ratio with the greenlab model. In Proceedings of the 5th International Workshop on Functional Structural Plant Models, J. Lovett Doust and L. Lovett Doust eds., Napier, New Zealand., No. 5, pp. 1–4.

Milo, R, Last R L. 2012. Achieving diversity in the face of constraints: lessons from metabolism. Science, 336, 1663–1667.

Nikolov T N, Massman W J, Schoettle A W. 1995. Coupling biochemical and biophysical processes at the leaf level, an equilibrium photosynthesis model for leaves of C3 plants. Ecological Modeling，80(2-3), 205-235.

Peng S B, Khush G, Virk P, et al. 2008. Progress in ideotype breeding to increase rice yield potential. Field Crops Research, 108(1), 32-38.

Qi R, Ma Y T, Hu B G, et al. 2010. Optimization of source-sink dynamics in plant growth for ideotype breeding: A case study on maize. Computers and Electronics in Agriculture, 71(1), 96-105. (In Chinese with English abstract)

Room P, Hanan J, Prusinkiewicz P. 1996. Virtual plants: new perspectives for ecologists, pathologists and agricultural scientists. Trends in Plant Science, 1(1), 33-38.

Quilot-Turion B, Ould-Sidi M M, Kadrani A. 2012. Optimization of parameters of the 'Virtual Fruit' model to design peach genotype for sustainable production systems. European Journal of Agronomy, 42 (special issue), 34-48.

Semenov MA, Stratonovitch P, Alghabari F, Gooding MJ. 2014. Adapting wheat in Europe for climate change. J Cereal Science 59:245–256

Srbislav D. 1994. Designing a wheat ideotype with increased sink capacity. Plant Breeding, 112(4), 311-317.

Su Z B, Zhan S Y, Zheng P, et al. 2008. Method for digital design of crop plant type with high photosynthetic efficiency. Transactions of the Chinese Society of Agricultural Engineering, 24(1), 203-207. (In Chinese with English abstract)

Sarlikioti V, De Visser P H B, Marcelis L F M. 2011(a). Exploring the spatial distribution of light interception and photosynthesis of canopies by means of a functional–structural plant model. Annals of Botany, 107(5), 875-883.

Sarlikioti V, de Visser P H B, Buck-Sorlin G H, et al. 2011(b). How plant architecture affects light absorption and photosynthesis in tomato: Towards an ideotype for plant architecture using a functional-structural plant model. Annals of Botany, 108(6), 1065-1073.

Sharma A R, Singh D P. 1999. Rice. In D. L. Smith, C. Hamel (Eds). Crop Yield, Physiology and Processes. Springer-Verlag, Berlin, Heidelberg, pp.109-168.

Suriharn B, Patanothai A, Boote, K J, et al. 2011. Designing a peanut ideotype for a target environment using the CSM-CROPGRO-Peanut model. Crop Science, 51, 1887–1902.

Van Oijen M, Hoglind M. 2016. Toward a Bayesian procedure for using process-based models in plant breeding, with application to ideotype design. Euphytica, 207:627-643

Watanabe T, Hanan J S, Room P M, et al. 2005. Rice Morphogenesis and Plant Architecture: Measurement, Specification and the Reconstruction of Structural Development by 3D Architectural Modelling. Annals of Botany. 95(7):1131-1143

Xu L F, Michael H, Zhu J, et al. 2009. A Rule-Based Functional-Structural Model of Rice Considering Source and Sink Functions. In Plant Growth Modeling and Applications (Baoguo Li, Marc Jaeger, Yan Guo, eds.) (PMA'09). Beijing (China), IEEE Computer Society (Los Alamitos, California), 245-252.

Xu L F, Michael H, Zhu J, et al. 2011. A functional-structural model of rice linking quantitative genetic information with morphological development and physiological processes. Annals of Botany, 107(5), 817-828.

Yan J Q, Zhu J, He C X, et al. 1999. Molecular marker-assisted dissection of genotype-environment interaction for plant type traits in rice (Oryzasativa L.). Crop science, 39(2), 538-544.

Yin X Y, Goudriaan J, ALantinga E, et al. 2003. A flexible sigmoid function of determinate growth. Annals of Botany, 91(3), 361-371.

Yuan S, Li Y, Peng S. 2015. Leaf Lateral Asymmetry in Morphological and Physiological Traits of Rice Plant. PLoS ONE 10(6): e0129832. doi:10.1371/journal.pone.0129832.