1

Virtual Plant Breeding

Sven B. Andersen
University of Copenhagen, Dept. Agronomy and Ecology
Denmark

1. Introduction

Breeding better cultivars has become a highly efficient way to improve plant production for yield, quality and reduced input. Most breeding activities have been focussed into fewer and larger programs with less people engaged. Educational environments for the area have been correspondingly reduced in size, which makes it harder to provide inspiring environments to generate high level education, for the fewer people needed in industry. Still plant breeders, scientists as well as society have ample interest in widespread public understanding of the use of new as well as old technologies for improvement of our cultivated plants. This is not least to avoid future communication problems with the general public like experienced with genetically modified plants during recent years.

Many university environments dealing with plant breeding education face two main problems. The first problem is the strong tendency to privatise practical plant breeding, which moves the breeding operations away from educational institutions into private more closed environments. Without live breeding programs associated with the campus, meaningful training becomes a possibility only for the few. Advanced students dedicated to the area may enter private activities as part of their advanced education, but this is not an option for the majority. The second problem is the decreasing number of plant breeders needed in the increasingly more efficient breeding programs. With fewer jobs on the market the general interest among students is reduced and the entire educational environment diminishes. For a majority of students in biology the interest in plant breeding exists, but only on a shallow level. They may well see the point in understanding major principles of applied genetics, also to be used in basic plant science, but actual training in plant breeding theory and operations is far beyond both their time and dedication. To reach a wider public within education in the future, plant breeding may need new ways to communicate its methods and principles.

Simulated learning systems are now widely used to train special skills in complex environments. Clinical training of nurses and health care workers are time-consuming and resource-demanding, but can be well supplied by simulations in special virtual learning environments. It is possible through experimental studies to assess the effect on such simulated training of skills within health care (Tsai et al., 2008). Several more general software packages for simulation based training of complex skills are available on the internet. Several British law schools collaborate on a simulated virtual learning approach to train skills in legal transactions with the software SIMPLE (SIMPLE 2011). ISLE Interactive

Simulated Learning Environments gives as a service the construction of user defined training systems for complex and expensive equipment or environments (ISLE 2011). The software SIMWRITER from NexLearn can be customized to users' needs for simulation to train complex social skills (Nexlearn 2011). Virtual Property Manager is a simulation software developed to teach students within residential property management, developed to generate more interest among students for the area (Carswell & James 2007). The principles of efficient training with virtual learning environments have been outlined with the theory of *problem spaces* (Stefanutti & Albert 2003).

The software Supergene relates to the problem in plant breeding education that most operations related to plant breeding outside laboratories are very time-consuming and resource-demanding. Students have to devote an increasing part of their time to studies of rapidly expanding molecular and physiological issues. This leaves less time to come to understand the actual applications on plant, population and ecological levels. The program is an attempt to establish simulation based illustration of phenomena and operations in applied genetics and breeding based on very few assumptions. The program is written in Delphi version 5 and works on Windows platforms. The entire package can be downloaded from www.supergene.dk. The present interphase points towards plants, but the basic simulation process may be used for other eukaryotes with some modification.

2. Material and methods

2.1 Basics of the simulation process

The software simulates virtual genetic alleles each holding one byte of information, which means that one gene locus may form a maximum of 256 different mutations or alleles. Gene loci each holding one allele are arranged linearly into virtual chromosomes. One complete set of chromosomes for a species is simulated as a gamete, which has the ability to fuse with another gamete from the same species during a fertilization process to form a diploid zygote. The user defines number of chromosomes, length of individual chromosomes as well as total number of gene loci during creation of the species.

Diploid zygotes formed by fertilization between two gametes from the same species form the basis of individual plant simulation. Such zygotes have the ability to form gametes through meiosis. During meiosis homologous chromosomes are first copied and subsequently paired into bivalents and recombined through formation of randomly positioned chiasmata. Number of chiasmata is drawn from a Poisson distribution with an expectation to generate a recombination frequency of 1 % between gene loci. Chiasmata are positioned randomly on chromosome pairs according to a uniform distribution with assumption of no interference between double crossovers. The approximately one centimorgan between each gene locus implemented also sets the limit of resolution during simulated linkage analysis. The meiosis process includes mutagenesis of each parental gamete before formation of chiasmata and recombination. The number of mutations in a gamete is drawn from an approximate Poisson distribution adjusted to obtain expected number of mutations in each gamete desired by the user. Mutations are located randomly onto loci of chromosomes according to a uniform distribution, and mutation of one allele means that its byte content is changed to carry a new random byte.

In addition to nuclear chromosome sets each zygote also contains one gamete simulating a chloroplast genome and one gamete simulating a mitochondrial genome. During virtual meiosis these cytoplasmic genomes are copied into the new gametes after mutagenesis. The user can control the frequency of paternal or maternal inheritance of cytoplasmic organelles during the fertilization process. If transferred from both the male and the female gamete into the new zygote the mitochondrial genomes are recombined and randomly reduced to one copy. Chloroplast genomes, if inherited from both the male and the female gamete, are randomly reduced to one copy without recombination.

Successive cycles of meiosis and fertilization form the basic simulation of sexual propagation of plants in populations. Chromosome doubled haploid plants are generated through meiosis with subsequent chromosome doubling instead of fertilization. Cloning uses simple copying of a zygote without meiosis and fertilization, but with mutation. Cybridization of protoplasts to transfer cytoplasmic male sterility (*cms*) or other cytoplasm inherited factors transfers only the chromosomal genomes of the recipient. Organellar genomes are combined from both protoplast parents following the rules above for biparental inheritance. In addition to the mutation and recombination, gametes of a plant may be transformed, which means that one or more transgenic constructs (inserts) are substituted into the chromosomes instead of the ordinary gene loci. Such transgenic alleles carry one random byte value termed the "expression value" to simulate different expression levels of different inserts.

2.2 Major features of the user interface

Input and output for the software has been arranged in a set of material windows presenting plants, populations and groups of populations. Operation windows make it possible to perform basic breeding operations like crossing, selfing, doubled haploid generation, genetic transformation of plants and cybridization with protoplasts. Also it is possible to perform polycross, topcross and testcross to estimate combining abilities of plants and populations and use them for evaluation and selection based on offspring values. Some basic analyses like analysis of MxN crosses and diallel sets of crosses have been implemented in special windows. For markers it is possible to perform simple analysis of association between traits and markers in populations. It is possible to analyze biparental mapping populations and download all polymorphic marker scores together with trait values for subsequent quantitative trait loci (QTL) analysis. Markers found linked to genes of interest can subsequently be transferred to a special marker list in the program and used for marker assisted selection (MAS).

Other windows make it possible to generate and edit traits associated with virtual genes in the genome of the species simulated. Such definitions of a trait are methods for calculation of values of the trait for each plant, based on the alleles the plant inherited from its father and mother. The trait values for each plant or population can subsequently be used for selection of individuals or populations with desirable alleles.

For simulation of molecular markers the basic byte values, are read out directly as marker alleles (Fig. 2.). These allele values may be thought of as the migration distance on a gel. The marker allele value zero is considered a null allele that does not generate a detectable signal. Heterozygous co-dominant marker loci show both allele values, while homozygous marker loci show only one allelic value. Null alleles are not shown as a separate value when

NR	NOR.Yield	CA.Yield	NOR.Carbohydrat	CA.Carbohydrate
1	1022.5590	1568.2407	0.6869	0.6676
2	1716.3158	1845.6825	0.7118	0.6917
3	1672.9583	1974.7544	0.7106	0.6745
4	2176.4598	1790.4046	0.8117	0.6891
5	1544.1726	1571.1863	0.6585	0.6882
6	2449.9753	1633.6382	0.6497	0.6774
7	2079.8218	1835.6891	0.6731	0.6540

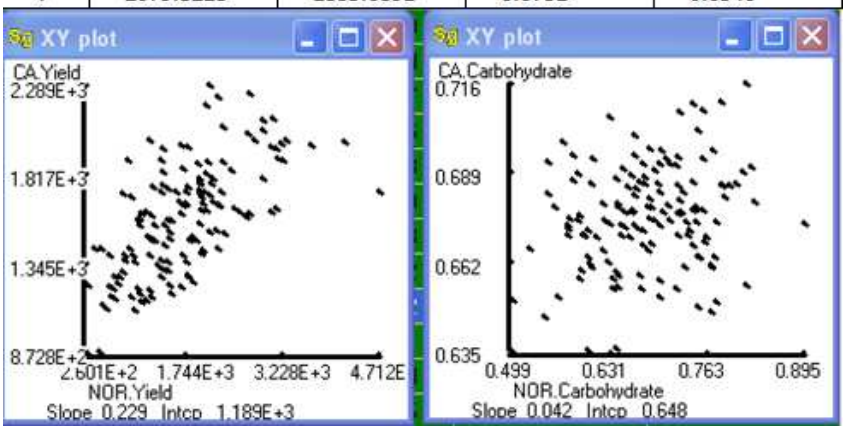


Fig. 1. Ryegrass plants after polycross evaluation showing yield of parental plants (NOR.Yield, kg/ha) and yield of their corresponding polycross offspring (CA.Yield). The offspring versus parent regression shows high heritability for yield, while carbohydrate content (per cent total carbohydrate of dry matter) show low heritability.

heterozygous, but appear as the value 0 when homozygous. The user can control the frequency of markers simulated as dominant or co-dominant. Highly co-dominant markers are like Simple Sequence Repeats (SSR), while more dominant systems behave like Amplified Fragment Length Polymorphisms (AFLP).

F2	NR	AthBIO2b	AthGPA1	nga1145
Plant1	1	134	110 81	152
Plant2	2	198 134	81	152
Plant3	3	198 134	110 81	152
Plant4	4	198 134	110 81	0
Plant5	5	134	110	0
Plant6	6	198	110 81	152
Plant7	7	198 134	110 81	0

Fig. 2. Virtual *Arabidopsis thaliana* plants segregating for two co-dominant markers (AthBIO2b and AthGPA1), each showing two different allelic values for heterozygous and only one allelic value for homozygous genotypes. The marker nga1145 segregates dominantly with a null allele, which shows the value 0 only if homozygous.

2.2 Simulation of traits

The software enables user definition of traits calculated for each plant based on the allelic values inherited in its genes. Values of SIMPLE traits are calculated as a function of the allelic values in one or more genetic loci from either chromosomal or cytoplasmic genomes. Values for TRANSGENIC traits are calculated like for SIMPLE traits, but only based on expression values of transgenic inserts in the chromosomes. COMBINED traits are user defined functions of one or more SIMPLE, TRANSGENIC or COMBINED traits, which enable construction of very complex and correlated traits.

Calculation of a trait value consists of three steps: 1) calculation of a BASIC VALUE as a sum of contributions to the trait from each gene locus, 2) addition of environmental effects and 3) scaling.

$$Basic\ value = \sum_{all\ loci} Bconstant \pm (Male + Female \pm ABS(Male - Female) * dom)/2$$

Where *Bconstant* is a user defined constant to set the general level of the trait. *Male* and *Female* are the byte values from the alleles of the locus after transformation to have an approximate normal distribution with expectation zero and variation defined for each locus by the user. *dom* is a user defined factor of dominance between alleles for each locus, which can take values between 0..1. Transformation of the uniformly distributed byte values of each allele to approximately normally distributed values has the effect that alleles with extreme contribution in positive or negative direction become rare. This means that in an ordinary population, plants with extreme genotypes of the trait become rare. The user defined variation of the allelic value distribution describes different contribution of each locus to the trait in addition to the actual allelic values inherited. The $\pm$ in the summation is a sign bit, which means that the contribution from the locus to the trait is either positive or negative. This sign bit is initially set randomly by the program, but can be controlled by the user. It has been introduced to avoid unidirectional correlation between different traits affected by the same gene loci. For a TRANSGENIC trait, the basic value is *Bconstant* plus the sum of expression values from each insert of the construct. For a COMBINED trait the basic value is the value of evaluation of a function of other traits defined by the user.

Random environmental effects are added to the basic trait value to simulate non-genetic variation. A random number is drawn from a normal distribution with expectation zero and standard deviation between zero and one. The size of this environmental standard deviation is provided by the user. The random number is subsequently multiplied with the basic trait value to generate a multiplicative environmental effect before being added to the basic trait value. This type of environmental effects simulation combined with the subsequent scaling of the trait introduces a multiplicative type of gene by environment interaction.

Scaling of traits is performed to keep the trait levels within some desired bounds. A trait like yield should be limited to non-negative values. In principle, yield should be able to go to infinity, however the illusion is lost if the trait goes much above levels observed in reality. The software uses exponential scaling functions (Table 1) to introduce such limits and the user can then use the scaling constant of the trait and the variation of the genetic loci to limit the unbounded values of the trait. A logistic function is used to scale traits with both upwards and downwards bounds. With a large scaling constant this function will also

High bound	Low bound	Trait Value
$+\infty$	$-\infty$	$Const * Bvalue$
Max	*Min*	$Min + (Max - Min) * \frac{e^{Const*Bvalue}}{1 + e^{Const*Bvalue}}$
Max	$-\infty$	$Max - e^{Const*Bvalue}$
$+\infty$	*Min*	$Min + e^{Const*Bvalue}$

Table 1. Functions used to scale traits with different boundaries. *Max* and *Min* are upwards and downwards limits of the trait. *Bvalue* is the basic trait value including added environmental effect, and *Const* is a user provided scaling constant.

effectively transform a basically quantitative trait into a two level qualitative trait. The scaling will introduce some types of epistasis and the degree of dominance of a trait is also affected by the scaling.

There are three traits with special functions generated automatically when a new species is initiated. "Malefertile" and "Femalefertile" determine the frequency of viable male and female gametes, respectively. The trait "Compatibile" determines the frequency of fertilization with pollen similar to the stigma in one or more gene loci using either a gametophytic or sporophytic model. All three traits are initiated with trait value 1.0, which means full fertility, but the user can make them dependent on genetic loci in the species.

The software provides a text editing window system named a "Tutorial" to write simple linear guides. Tutorials can be written by anyone to generate a sequence of windows, which can guide students through a set of operations on associated virtual plant material. Students can subsequently play the tutorial sequence and perform the operations at their own speed.

3. Results

3.1 Adaptation of virtual material

When initiating a new species the software will generate a single plant with maternal and paternal genome sets and random byte values in each gene allele. If multiplied for some generations with a high mutation and cross breeding frequency, this plant will give rise to a diversified base population. The base population, if big enough, will contain all possible alleles of all gene loci in equal frequency as long as no selection is performed.

If subsequently traits are defined to depend on different gene loci of the species then recurrent selection for different traits on the base population will lead to adapted subpopulations with changed allele frequencies. Such recurrent selection for many generations can be performed

easily using a special "AutoSelect" operation window. Simultaneous index selection for several traits can be performed with a "Combined" trait. During such recurrent selection, mutation rate of the species may be elevated to introduce new alleles in the existing material. This will generate highly diversified sub populations from the original base population.

During the adaptation process of populations, trait definitions can be rescaled at any time to find suitable settings.

The original base population may be considered the undomesticated or wild material of the species. If a trait like yield, controlled by hundred or more gene loci, is defined with a lower bound of zero then single plants or lines from this population will rarely show extreme yield values (Fig. 3, generation 0). This is because of normally distributed simulation effects of allelic values, which will mostly have effects close to zero and rarely more extreme contributions. Thus without selection, it is unlikely that plants or lines with many high effect alleles for the yield trait will show up. With recurrent selection, possibly combined with an elevated mutation rate, however, domesticated material with high frequency of rare alleles contributing to yield can be generated (Fig. 3). Such domestication if performed without

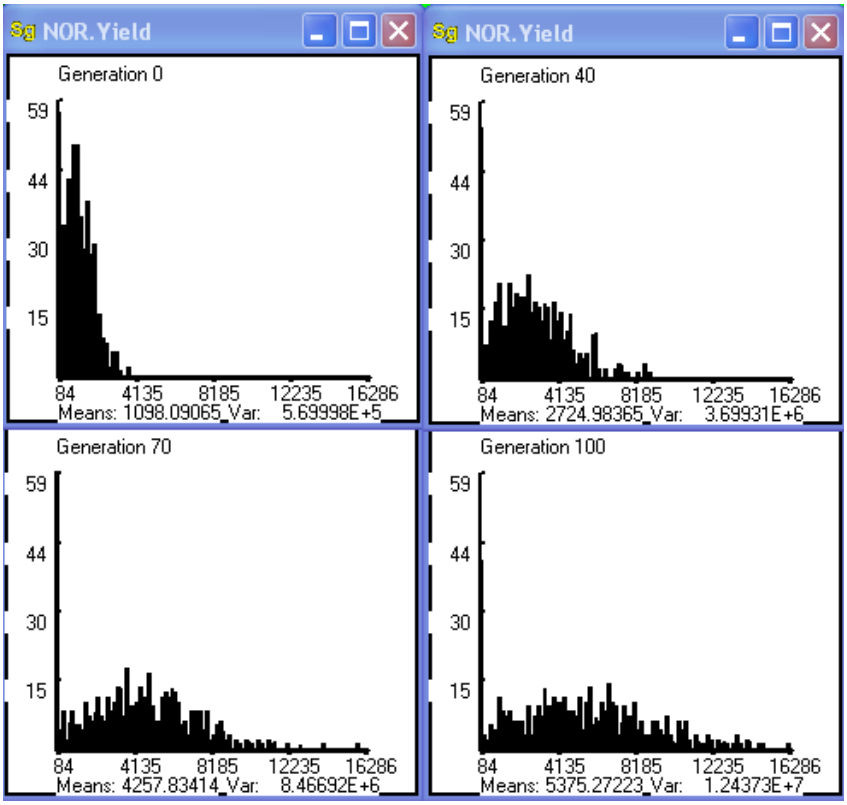


Fig. 3 Histograms illustrating adaptation of a base population through recurrent selection for a yield trait (kg seed/ha) affected by 200 different gene loci. For each generation the 20 % highest yielding plants were selected and intercrossed to obtain the next generation.

unrealistically high mutation rates, will also lead to reduced genetic diversity of the material. Of course, the domestication process may be performed for several different traits like yield, quality or disease resistance depending on different genes, either in tandem or in parallel using an index. This can generate many different domesticated or wild subpopulations. A characteristic feature of such adapted populations is that they lose their selected traits and degenerate back to the original base population if heavily mutated. This is because new alleles with extreme positive or negative effects are rare in the simulation.

Before, during or after the domestication process, genetic structure related to outbreeding tendency may be established through adjustment of the cross breeding frequency. The special trait "Compatibile" may be set to depend on one or more gene loci. Subsequent multiplication with a low degree of mutation will lead to strong selection for multiple alleles in the affecting genes and maintain low inbreeding. Also the two special traits "Malefertile" and "Femalefertile" may be set to depend on gene loci or other traits to affect male and female sterility, which will also affect genetic structure of the resulting populations. Rare mutations with strong effect on single gene traits can be selected from one or two generations of multiplication under elevated mutation frequency. If a trait named *cms* is affected by a mitochondrial gene locus then the resulting trait will be maternally inherited. Plants with rare alleles in chromosomal genes with strong effect on another trait Restorer (R) may subsequently be selected and combined with *cms* in the "Malefertile" trait, to establish a simple *cms* restoration system in the species.

3.2 Guided exercises

A number of tightly guided exercises have been developed based on the software. They have been used mainly to illustrate specific principles of breeding methods for students. Each exercise consists of a virtual plant material and an associated tutorial, which will guide students through the operations at their own speed. The virtual plant material for these exercises is of named type and species adapted to mimic as much as possible reality. The traits are selected to interest our students. Some of the exercises make reference to chapters in Sleper & Poehlman (2006). All the exercises have been used for two hours confrontation in groups of two students with approximately one teacher in the class room for each 10 students. After one week, student groups hand in a 4-5 pages report based on screen clips from the program embedded in text, explaining theoretical background, operations and results obtained. Reports are subsequently returned to the students with comments from teachers.

One exercise deals with the difference between qualitative and quantitative traits and introduces how traditionally qualitative traits, become quantitative, if the number of genes or the environmental effects are increased. It introduces heritability and simple selection theory. A second exercise gives an introduction to incompatibility and *cms* restoration systems in breeding. In addition to a basic introduction of these major fertility regulation systems this exercise also introduces the basic setup of hybrid cultivars. A third exercise introduces backcrossing with a dominant trait. A fourth exercise introduces linkage analysis between markers and a qualitative trait and performs marker assisted backcrossing of a recessive disease resistance into a high yielding cultivar. Students experience the characteristic complete absence of visible resistance in the material during the entire backcross operation, followed by reappearence of the resistance trait after final selfing of the

backcrossed material. A fifth exercise introduces polycross for estimation of combining ability and its use to construct synthetic cultivars in highly outcrossing species. In particular this exercise discusses the number of basic clones needed to avoid the risc of inbreeding depression of the synthetic cultivar during later generations of multiplication. A sixth exercise introduces the structure of hybrid cultivars based on *cms* restoration systems their maintenance and improvement of their inbred parental lines. An existing hybrid is improved through backcrossing of dominant and recessive genes for disease resistance into its parental lines and the risc of reduced hybrid performance because of linkage drag is discussed.

It would be quite simple also with the software to generate guided exercises for population genetics to illustrate the effect of selection and cross breeding frequency on various population genetic features like deviation from Hardy-Weinberg equilibrium. Also phenomena like genetic drift of allelic frequencies during generations and the risk of losing rare alleles may be illustrated.

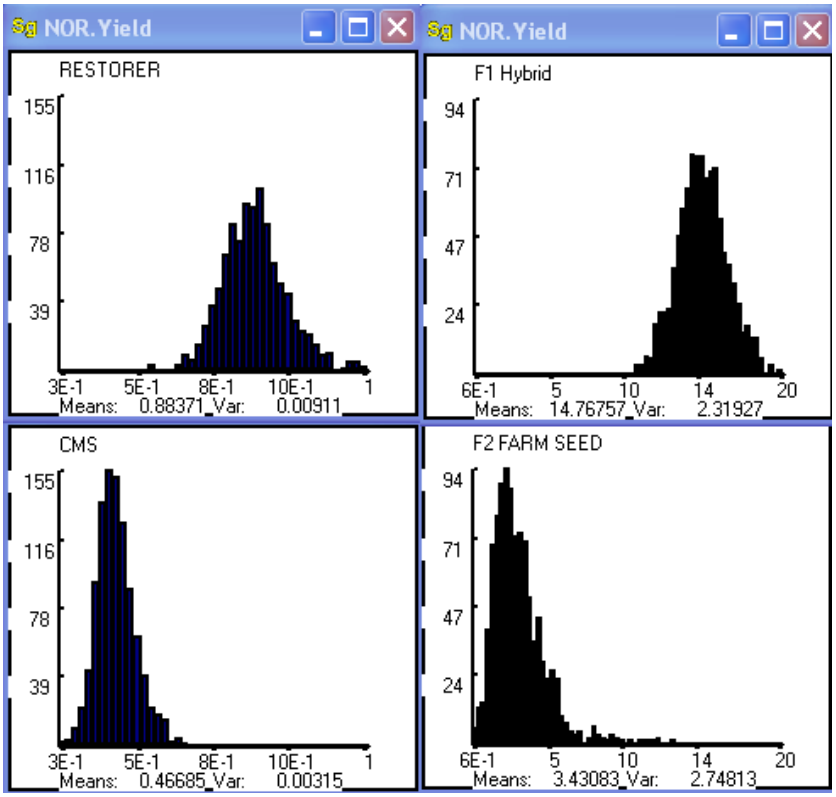


Fig. 4. Histograms showing yield levels and variation within two inbreds (Restorer and CMS), their high yielding hybrid (F1 Hybrid) and farm saved seed from the hybrid (F2 FARM SEED). Yield simulated with 200 different gene loci and approximately 500 plants per population.

A virtual *Arabidopsis thaliana* with named SSR markers from the databases has been established. It has been used in an exercise where single gene mutants for different traits are located to chromosomal areas using SSR markers. Students subsequently find markers flanking the mutation and use new markers in the interval to close map the mutation to within one centimorgan for map based cloning. In another exercise with this virtual *Arabidopsis*, students make crosses between ecotypes differing in quantitative traits. They derive mapping populations (Doubled haploids (DH), F_2 or Recombined inbred lines (RIL)) and analyse the mapping populations for all segregating markers with a marker download window. The operation will download scores of all markers for all plants of the mapping population together with trait measurement for each plant and a map of the markers. The downloaded file can be used directly with a supplied QTL mapping program SuperQTL to analyse for QTL of the trait.

3.3 Explorative learning

The simulation software has been used during several courses in molecular breeding to teach marker assisted breeding in groups of 2-3 students. A large diversified virtual material of hexaploid wheat, consisting of different groups of low yielding heterogeneous landraces and some high yielding uniform modern cultivars, was established. Students have been given this virtual genetic resource material for a two weeks working period under guidance of a teacher, but without a written tutorial. Students use various phenotypic screening and association and QTL mapping approaches to identify major and minor disease resistance genes in the landrace material followed by backcrossing into the high yielding cultivars. Some special features studied are the problem of linkage drag and associated yield loss caused by backcrossing from less adapted material. Amount of linkage drag can be studied easily using the marker system. It is possible to follow the breakdown of the linkage disequilibrium during subsequent generations to reduce the amount of introduced genetic material and markers can be used to speed up this process. Close mapping of QTL to generate diagnostic markers have been performed for subsequent selection with reduced linkage drag. The diagnostic markers have been used for pyramiding of genes for partial resistance to produce efficient disease resistance based on 2-4 different genes. Joint QTL analysis of multiple crosses have been studied to search for QTL in genetically wide materials and to study the effect of reduced size of the mapping populations. It is possible during such activities to exchange plant material between participants on different computers to generate cooperation in larger groups. The two weeks teaching periods have resulted in 25-30 pages reports composed of screen clips from the simulations, Figs and tables summarizing results, embedded in text to explain theory, ideas and results.

Table 2 shows summary of results from students' joint QTL analysis of powdery mildew infection on approximately 1000 wheat DH lines from 20 different crosses. Ten different cultivars were crossed each with cultivars Grete and Jesper and about 50 DH from each F1 hybrid were used for the joint QTL analysis with approximately 3000 polymorphic markers. The table shows the positions on the chromosomes of each QTL and the LOD values obtained. In addition, additive value from each cultivar in each QTL is shown. Negative additive values indicate alleles for reduced powdery mildew infection. The two QTL with lowest LOD of 7.1 and 6.8, respectively, were deemed false positives based on permutation analysis. This status as false positives of these two QTL was also confirmed from inspection

	Chrom 1 Pos 104	Chrom 9 Pos 1436	Chrom 13 Pos 2186	Chrom 15 Pos2574	Chrom 18 Pos 3000	Chrom 20 Pos3422
Ellen	-0.1	11.8	-12.5	-13.5	6.9	1.1
Else	17.4	-3.1	3.5	9.3	6.3	0.3
Grete	-5.0	1.4	1.7	-3.3	7.0	20.0
Gunhild	10.4	0.2	-18.0	-20.2	12.6	-20.4
Jesper	7.6	-4.8	-0.7	10.9	4.5	5.6
Knud	-23.4	3.8	3.8	-9.6	-2.7	-3.5
Mads	-11.4	7.7	-0.6	-3.6	1.5	-11.4
Merete	13.7	-9.9	-2.0	-3.3	0.3	-0.3
Sigfred	-6.2	-10.9	23.5	9.3	-6.4	2.8
Soren	9.9	2.1	20.3	1.3	-18.5	4.4
Verner	-3.3	-5.0	-0.2	10.2	-1.7	-2.1
William	-9.7	6.8	-18.7	12.4	-9.8	3.6
LOD	24.0	7.1	6.8	12.0	8.4	57.0
		False	False			

Table 2. LOD values and estimated additive effects of six QTL for powdery mildew infection in 12 different cultivars.

of the trait design, which did not rely on gene loci in these chromosomal areas. Simulation of the powdery mildew trait used 20 different gene loci with 20% environmental effect, however, only four of these loci showed significant functional polymorphism in the cultivar material analyzed. The cultivar Gunhild has two desirable alleles on chromosome 15 and chromosome 20, respectively. The cultivar Knud has another desirable allele on chromosome 1. The two cultivars were subsequently crossed and markers surrounding the three QTL were used to select offspring homozygous for all three alleles. One high yielding line was found among the offspring with pyramided resistance.

4. Discussion

A large number of simulation tools related to plant breeding can be found on the internet. Many such programs target the use and analysis of molecular data from markers or data from field experiments and simulate plant performance based on a set of underlying assumptions. Other approaches use physiological or biochemical models to predict plant behaviour and connect phenotypic performance with genes (Chapman et al. 2003, Hammer et al. 2005). Softwares like QuLine (Wang & Pfeiffer, 2007) can combine many different types of genetic data now available on the internet to simulate performance of breeding material. Although not primarily made for teaching, such programs can be extremely useful for learning though mostly for people already dedicated to the area.

The simulation for clinical training in health care and the training of focussed skills in legal management are maybe somewhat different from our needs in plant breeding education. For the plant breeding area, specific training of skills may be less important. Such skills will have to be trained anyway in actual breeding programs, should the student end up with such engagement. More important in the case of plant breeding is the communication of basic theoretical ideas and phenomena to an audience with only a part of its interest and

time devoted to the area. In this situation, the most important result of such virtual plant breeding simulations is the images and associations generated in the heads of students. Stimulation of this process strongly depends on the choice of species, traits etc., which will associate positively with students intellectual background. A student with a focussed educational background in horticulture may not show much enthusiasm for a hybrid breeding program in maize, while a very similar setup named Primula or head cabbage could make a difference. In the same way, students with a highly molecular background cannot be expected to raise much interest from simulations illustrating pure traditional breeding programs. However, they may well take interest in quantitative genetics and QTL as a means to track down important genes, if wrapped in circumstances appealing to their background. Many students within plant science have interests and imaginings in genetic resources and population genetics, which can be well simulated with the present setup and used to introduce them to applied genetics. In general, however, simulations cannot substitute completely for real work with plants etc. Without hands on experience, students do not possess a background, with which the operations on the screen can generate positive associations and imaginative pictures. However, given the basic biological and ecological background, students can in a matter of hours experience complex operations and solve problems, which in real life may take several human generations and demand resources beyond imagination.

The basic setup of the present system for simulation of genetics has the advantage of being very general. The only assumptions made are the basic rules of genetics, and the byte representation of genes reduces the amount of needed computer memory. It enables more than 30 000 plants each with 3-5000 loci to be manipulated on most ordinary laptops. Still the amount of information in these genomes provides good simulation of both qualitative and quantitative traits. Like for real genetic resources the genetic diversity of populations can become exhausted if subjected to intensive inbreeding and selection. For out breeding populations, however, a quantitative trait based on 100 gene loci and a 200 plant population size will retain variation for several hundred generations of recurrent selection. Also the phenomena of heterosis and inbreeding depression are simulated very well for a trait, if its loci have been given a significant amount of dominance. As soon as ordinary populations with many heterozygous gene loci are inbred the trait shows inbreeding depression, which can be relieved through hybridization with different material. A few generations of reciprocal recurrent selection for combining ability between genetically distinct populations will efficiently build up populations showing heterosis upon hybridization. Lines derived from such reciprocally heterotic groups, generally form hybrids with good performance in the traits selected for. Such hybrids are generally quite difficult to out yield if students subsequently derive new inbreds from the original populations. Furthermore, performance of such hybrids is highly sensitive to genetic pollution from linkage drag. In general 6-7 backcrosses are needed for full recovery of hybrid performance, if one of the parentals is improved for a simple trait. In many cases, full recovery of the hybrid performance is only possible for some backcross lines, so students working in parallel, with the same starting material, can experience quite different outcomes.

Also genetic phenomena like cytoplasmic male sterility and restoration are simulated quite well. Like for real hybrid breeding the female *cms* line can only be multiplied via pollination from a similar maintainer without restoration genes. The *cms* trait can be transferred to new

inbred maintainer types via simple backcrossing. If a rare allele is sought for restoration, then restorers in the breeding material will be rare. Otherwise the ordinary problems with restoration in the breeding material will arise, so only some inbreds can become maintainers. Transfer of the restoration trait to new inbreds, so they can be used as male lines for hybrids, must be done by backcrossing. Selection for the restoration ability during the backcrossing operation may be based on test crossing with a male sterile. Alternatively the genes affecting the trait may be mapped in a suitable mapping population and the linked or functional markers can be used for selection during backcrossing.

The program will also easily simulate transgenic engineered male sterility like the *barnase barstar* system (Denis et al. 1993) used in many *Brassica napus* breeding programs. One transgenic construct named barnase is used to generate the male sterility and another construct named barstar is used as restorer. A basta resistance trait, in addition, is linked to the barstar construct. This basta resistance trait can then be used to eliminate male fertile plants lacking the restorer after each multiplication. During the transformation process, different transgenic plants will receive different number of inserts and only some inserts show high expression of the construct. Like for real life breeding with these approaches, it is normally easier to backcross elite inserts into new breeding lines than to introduce them by *de novo* transformation.

Many of the features of self-pollinators appear naturally, if the material is suitably prepared. Populations of species with high self-pollination frequency, quickly lose genetic diversity, if selected without forced hybridization. Therefore, generation of basic subpopulations with different adaptation is most efficient with a high frequency of cross pollination. Subsequently, diversified populations, clearly different from the base population, are obtained through recurrent selection. Then cross pollination frequency is lowered for another ten generations to obtain heterogeneous populations of pure lines. From such heterogeneous land race like material then, pure lines can be extracted via single plant offspring. Subpopulations can be improved intensively for quantitative traits like yield and quality affected by many genes and give rice to cultivar like superior lines. Such cultivars will rarely be outperformed by material from the unselected base population or from populations of lines selected for other traits. They may, however, show heterosis upon hybridization, and new superior lines can be identified in the inbred offspring from such hybrids. Simple hybridization of high performing lines with non-adapted material to introduce new traits like resistance, in almost all cases destroys adaptation. Adaptation of the original line can be restored through 6-7 backcrosses with selection for the new trait. However, because of linkage drag, complete performance of the line is not always restored, and students working with the same initial material can experience different results.

Simulation of markers and genetic mapping is efficient with the present simple approach. In segregating offspring, students can count recombinants between traits and markers or between different markers. They can then calculate recombination frequencies, generate local maps surrounding genes for specific traits and decide, which markers to be used for selection or close mapping of the trait. Students during such operations will quite realistically face the questions related to lack of information from monomorphic markers, reduced information from dominant markers and the question of, which parental allele to use for selection. During marker-assisted backcrossing of e.g. a recessive trait, they will further face the chance of losing the linked trait because of recombination between marker

and the target, if the marker is not functional. During fine mapping of a qualitative trait, students will face the challenge of, how to decide on flanking markers and subsequently how to repeatedly reduce the target area based on new markers situated between the flanking ones. In dissecting quantitative traits with QTL analysis, students can realistically experience the problems with false positives, estimation of LOD thresh holds, relative effects of different QTL and problems with linkage drag, if QTL are backcrossed. Because of the simulated trait structure, positions of mapped genes or QTL can always finally be checked with the trait structure, to see whether a gene is actually simulated in the genomic area and where exactly the gene in question is situated.

A more evolution based type of teaching of applied genetics may be possible in the future with such simulation systems. For students there is much to be learned from the process of adaptation of the virtual material to simulate well, different types of breeding material or plant genetic resources. Given the necessary time, students may generate their own virtual species with trait definitions build on information from literature and databases. Underlying genes for the traits may be approximately positioned on the genome and their major effects and interactions can be modelled based on available knowledge. Different models for a genetic phenomenon may be implemented in different student groups and their performance may be compared to identify advantages and limitations of each genetic model. Relative success of major methods of breeding like hybrid programs compared with open pollinated cultivars for different types of virtual plant material could be an example. The understanding of market performance of hybrids, compared with open pollinated cultivars is complex and includes e.g. both intellectual property rights, as well as several economical issues in addition to the genetics of the plant material. The genetic dimension of this problem in itself is often too complex for many students to invest enough efforts to understand it, based on lectures and associated literature. Simulation could be an alternative way to introduce students to the genetic dimension of this discussion and then wrap the entire teaching operation into an economic environment with simple book keeping and associated seed market situations.

Also reporting from such simulated teaching may be much developed. Future students, instead of traditional written reports with screen clips, may hand in a virtual plant material and an attached Tutorial guide, which would guide the teacher or colleague students through the process and ideas of the author. To some extent the motivation of simulated learning may not so much be the actual presentation of complex phenomena to the receiver. The activation of students to create a product, like a new high yielding disease resistant cultivar, a Tutorial guide to be seen by someone else, or simply a conventional written report with colourful screen clips, may be much more motivating. Another motivation is the random number based simulation, which assures that any student product will be unique, because two simulations never produce the same result. This effectively eliminates any sentiment of communicating with a robot.

The present implementation of Supergene does not enable simulation of polyploids. Genetics of auto-polyploidy is under implementation in a planned future Java based implementation. However, general simulation of allo-polyploidy will demand a more general simulation of evolution of genomes. Also simulation of environmental effects and gene by environmental interactions are severely restricted with this simple approach. Separate environments are not simulated and therefore general gene by environment

interactions cannot be modelled. Also plant plant interactions are not simulated in the present implementation. Such plant plant interactions would be interesting for ecology. Some more complex gene by environment interactions as well as complex epistasis may be simulated using the combined trait feature of the program.

5. Conclusion

In spite of these and other limitations of the present implementation, it has been used with good results, both for illustration of various breeding phenomena and for more explorative teaching. The simulation has mainly been used for breeding operations too time or resource consuming to be performed in real life. This also means that teachers' personal contacts during the exercises have been prioritized. Only a minor part of the teacher contact has been used for questions regarding the program interface. The majority of questions and discussion with teachers generally have been related to scientific understanding of the breeding operations. In this way the software is efficient in raising questions from the students regarding why and how. Questions which would otherwise not surface based on reading a book chapter or listening to a lecture.

6. References

Carswell, A.T. & James, R.N. (2007). Virtual Property Manager: Providing a simulated learning environment in a new university program of study. *Systemics, cybernetics and Informatics*, Vol. 5, No.4, (June 2007), pp.34-40, ISSN 1690-4524

Denis, M.; Delourme, R.; Gourret, J-P.; Mariani, C. & Renard, R. (1993). Expression of Engineered nuclear male sterility in Brasscia napus. *Plant Physiology*, Vol. 101, No.4, (April 1993), pp.1295-1304, ISSN 0032-0889

ISLE (2011). http://www.biographix.com/services/pdf/ISLE.pdf

Nexlearn (2011). http://www.nexlearn.com/?q=node/6

Chapman, S.; Cooper, M.; Podlich, D. & Hammer, G. (2003). Evaluating plant breeding strategies by simulating gene action and dryland environment effects. Agronomy Journal, Vol.95, No. 1, (Jan.-Feb. 2003), pp. 99-113, ISSN 0002-1962

Hammer, G.L.; Chapman, S.; van Oosterrom, E. & Podlich, D.W. (2005). Trait physiology and crop modelling as a framework to link phenotypic complexity to underlying genetic systems: Modelling complex traits for plant improvement. *Australian Journal of Agricultural research*, Vol. 56, No.9, (2005), pp. 947-960, ISSN 0004-9409

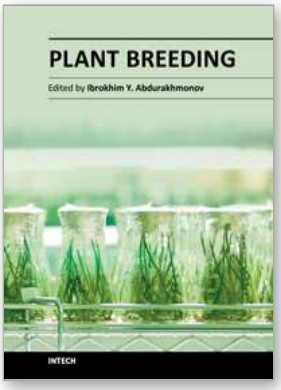
Simple (2011). http://www.ukcle.ac.uk/projects/past-projects/simpleupdate/

Sleper, D.A. & Poehlman, J.M. (2006). *Breeding Field Crops* (Fith edition), Blackwell Publishing, ISBN-13: 978-0-8138-2428-4, 2121 State Avenue, ames, Iowa, 50014

Stefanutti L. & Albert D. (2003). Skill assessment in problem solving and simulated learning environments. *Journal of Universal Computer Science*, Vol.9, No.12, (Dec. 2003), pp.1544-1468, ISSN 0948-695X

Tsai, S-L.; Chai, S-K.; Hsieh, L-F.; Lin, S.; Taur, F-M.; Sung, W-H. & Doong J-L. (2008). The use of virtual reality computer simulation in learning Port-A Cath Injection. *Advances in Health Sciences Education*, Vol.13, No.1, (March 2008), pp. 71-87, ISSN 1382-4996

Wang, J-K. & Pfeiffer, W.H. (2007). Simulation modelling in plant breeding: principles and applications. *Agricultural Sciences in China,* Vol. 6, No.8, (July 2007), pp.101-105, ISSN 1671-2927



Plant Breeding

Edited by Dr. Ibrokhim Abdurakhmonov

ISBN 978-953-307-932-5
Hard cover, 352 pages
Publisher InTech
Published online 11, January, 2012
Published in print edition January, 2012

Modern plant breeding is considered a discipline originating from the science of genetics. It is a complex subject, involving the use of many interdisciplinary modern sciences and technologies that became art, science and business. Revolutionary developments in plant genetics and genomics and coupling plant "omics" achievements with advances on computer science and informatics, as well as laboratory robotics further resulted in unprecedented developments in modern plant breeding, enriching the traditional breeding practices with precise, fast, efficient and cost-effective breeding tools and approaches. The objective of this Plant Breeding book is to present some of the recent advances of 21st century plant breeding, exemplifying novel views, approaches, research efforts, achievements, challenges and perspectives in breeding of some crop species. The book chapters have presented the latest advances and comprehensive information on selected topics that will enhance the reader's knowledge of contemporary plant breeding.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Sven B. Andersen (2012). Virtual Plant Breeding, Plant Breeding, Dr. Ibrokhim Abdurakhmonov (Ed.), ISBN: 978-953-307-932-5, InTech, Available from: http://www.intechopen.com/books/plant-breeding/virtual-plant-breeding

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821