SUNFLO: A joint phenotyping and modelling approach to analyse and predict the differences in yield potential of sunflower genotypes

Jérémie Lecoeur[1], Richard Poiré-Lassus[1], Angélique Christophe[2], Lydie Guilioni[1]
[1]Montpellier SupAgro, UMR 759 LEPSE, 2 place Viala, F-34060 Montpellier, France,
E-mail: jlecoeur@supagro.inra.fr
[2]INRA, UMR 759 LEPSE, 2 place Viala, F-34060 Montpellier, France

ABSTRACT

This work was focussed on improving the description of organogenesis, morphogenesis and metabolism in a biophysical plant model. A greenhouse experiment was carried out to identify and to estimate the phenotypic traits involved in plant productivity variability of 26 genotypes. The ability of the biophysical model to discriminate the genotypes was tested on previous results of a field survey focussed on evaluating their genetic progress since 1960. Plants were phenotyped on 4 areas: phenology, architecture, photosynthesis and biomass allocation. 12 traits or genotypic parameters were finally chosen to account for the phenotypic variability. A biophysical model was especially built to integrate the genotypic parameters and to evaluate their respective contribution to the variability of yield potential. A large phenotypic variability was found for each term of the energetic approach of above-ground biomass production. The biophysical model was able to account for 80 to 90% of observed variability in yield potential. This model was an interesting tool for analyzing the phenotypic variability of complex plant characteristics such as light interception efficiency. This model showed that several ways are possible to reach high yields. Unlike a classical statistic analysis, this approach allowed to highlight some efficient parameter combinations used by the most productive genotypes. The next steps will be to evaluate the genetic determinisms of the genotypic parameters and to test the reliability of the phenotyping approach.

Key words: biophysical model – *Helianthus annuus* – phenotypic expression – phenotypic characterization – sunflower – yield potential.

INTRODUCTION

Current knowledge in biology does not presently allow to link "whole plant" and "molecular" approaches. As a result, complex plant characteristics such as crop yield cannot be grasped by using "molecular" knowledge in "bottom up" approaches. Consequently, journals are crammed with attempts to identify genes that might explain the build up of plant phenotypes in responses to environment. However, strong difficulties have been encountered in attempting to quantitatively relate the information at gene level to its expression in complex phenotypic traits at plant level (e.g. Sinclair et al., 2004). This actual gap between the identification of an allelic combination at genome level and the corresponding phenotype at plant level greatly limits the potential benefits of the bottom-up approaches in improving our understanding of the genotype-environment interactions and the phenotypic plasticity (e.g. Sinclair and Purcell, 2005). At the same time, the biophysical approaches progressed in understanding and in formalizing the interactions between physical environment and plant responses and regulations (e.g. Jones, 1992).

A possible approach to reducing the gap between the molecular and plant levels is the use of models representing the plant as a biophysical system decomposed as a set of functions determining the phenotype built up in response to environment (e.g. Jeuffroy et al., 2006). To get a coherent system, two types of equations have to be combined, the energy and mass balance equations and the biological regulation equations in response to environment. From one degree of breaking down the plant functioning in elementary processes, the parameters of the equations used to describe these elementary processes may be compared to genotypic characteristics (Yin et al., 2004). Then, it is possible to use the quantitative genetic methods, especially heritability calculations and QTL determinations, to evaluate the genetic determinism and the variability of the studied process. Depending on the way the plant response is taken into account, the use of this genetic information in a set of equations describing the plant functions may allow to account for the plant phenotypic plasticity. This approach has been explored for complex traits such as the expansion rate of a single leaf (Reymond et al., 2003). These examples are far from crop yield, in terms of complexity and time and space levels. The more suitable plant representations to tackle the yield variability would be the crop models (Sinclair and Seligman, 1996). Recent studies have attempted to integrate biochemical and physiological information in crop model to improve the heuristic performance of these models in the analysis of phenotypic plasticity (e.g. Hammer et al., 2004).

The objective of this study was to evaluate the ability of a phenotyping approach combined with a dedicated simple biophysical model to account for the genotypic variability of yield potential. Our assumptions was that the genotypic variability of seed yield could be accounted for by using a set of robust equations, well-tested in crop modelling studies, coupled with a few parameters taking into account the observed phenotypic variability of the studied genotypes. This approach included three parts (i) the development of a biophysical model taking into account the specificities of the sunflower biology, (ii) the estimation of genotypic parameters from measurements on a limited number of isolated plants grown in greenhouse and (iii) an independent set of data obtained in a field experiment for the model evaluation. A panel of 26 genotypes was studied mixing historical commercial hybrids (Vear et al., 2006), experimental hybrids and introgression lines between *Helianthus annuus* and *Helianthus mollis*. This panel was interesting for two reasons. First, the 26 genotypes displayed a wide range of phenotypic differences. Previous observations reported differences related to phenology, light interception, biomass production and allocation (Debaeke et al., 2004). Secondly, there were very large differences in seed yield between genotypes. The seed yield of the most productive genotype is five time higher than the introgession lines. Even among the commercial hybrids, differences higher than 40% were observed (Vear et al., 2003). This trait variability and seed yield scale are relevant for evaluating a modelling approach. The genotypic parameters were chosen according to their ability to integrate the specificities of the sunflower biology in the terms of the Monteith generic approach of the biomass production (Monteith, 1977). Beyond the objective to model the seed yield phenotypic variability from genotypic characteristics, this second aim was to evaluate which plant traits highly contribute to the seed yield variability.

MATERIALS AND METHODS

Model development: The model SUNFLO estimates the above-ground biomass production of a sunflower crop from incident radiation and mean air temperature. It works in daily time steps and describes the plant phenology, the plant leaf expansion, the biomass production and allocation. It takes into account the behaviour of various genotypes by the mode of some parameters which are genotype dependent.

The plant phenology is driven by the thermal time. Cumulative thermal time was calculated as the sum of the daily mean air temperature from emergence using a base temperature of 4.8°C common to all genotypes. Four key stages, expressed as thermal dates with genotypic values, were used to delimited periods of plant cycle with changes in plant physiology: the floral bud appearance (E1), the beginning of flowering (F1), the beginning of grain filling (M0) and the physiological maturity (M3) (Table 1).

Assuming the canopy is a homogeneous absorber, the daily radiation interception efficiency (RIE) was estimated from Beer's law using daily LAI and an extinction coefficient (k) determined for each genotype (Table 1). LAI was calculated from the plant density and the plant leaf area able to intercept photosynthetically active radiation. This latter was estimated as the difference between total leaf area and senescent leaf area. Because in sunflower the distribution of the leaf area along the stem showed a bell-shape (Dosio et al., 2003), plant leaf area was calculated from leaf number (N) with a logistic equation with 3 genotypic parameters, A1, A2 and A3, respectively, the maximal plant leaf area, the rank and the area of the largest leaf of the plant (Table 1). The number of leaves increases linearly with cumulative thermal time from emergence to the beginning of flowering. Then the leaf number decreases linearly from the beginning of seed filling to plant maturity as nitrogen moves from leaves to seeds during the monocarpic leaf senescence (Sinclair and deWit, 1975).

The radiation use efficiency (RUE) represents the ability of the crop to convert the intercepted energy into biomass. RUE is known to change during the plant growth cycle (e.g. Lecoeur and Ney, 2003 on pea). A single general pattern of change in RUE over crop development was used for all genotypes. RUE was equal to a minimum up to 300°Cd, then it increased linearly to reach a maximum level at the beginning of flowering. RUE remained constant until the beginning of seed filling, then it declined exponentially to zero upon the plant death. This general pattern was modulated through a genotypic parameter taking into account the different photosynthetic capacities of genotypes relative to Melody. A depressive function of non optimal temperature was applied to RUE, calculated from daily mean air temperature. The above-ground biomass production was calculated from Monteith's formula (1977) linking dry matter production to incoming photosynthetically active radiation through two radiation efficiencies.

The allocation of biomass to seeds was estimated by using two allocation coefficients. The first coefficient determined the fraction of total biomass allocated to the capitulum (HIcap). It changes with thermal time and was modelled as a single logistic function. However, its maximum value reached at physiological maturity was genotype-dependent. A second coefficient corresponded to the fraction of

capitulum biomass allocated to the seeds (HIseed). It was also genotype-dependent. Finally, the seed yield was calculated by multiplying the final biomass and the two allocation coefficients.

Estimation of the genotypic parameter: The estimation of the genotypic parameters was carried out in a greenhouse experiment (Montpellier, southern France). Plants were grown in pots of 7.5 l filled with mixtures of loamy soil and organic compost. There were 10 pots per genotype and they were arranged in order to mimic an agronomic culture density of 6 plants per m². The soil was continuously maintained at water retention capacity by irrigations at least once a day with a modified one-tenth Hoagland solution corrected with minor nutrients. Air temperature and radiative conditions were managed in order to obtain thermal conditions and photoperiod similar to those classically observed in field conditions.

The dates of occurrence of developmental stages E1, F1, M0 and M3 were determined according to the notation proposed by the CETIOM for sunflower on six plants per genotype, twice a week. In addition, the number of visible and senesced leaves were also counted. Architectural measurements were made at the end of flowering when all the vegetative organs were fully expanded. Height, length, width and distance from stem of each blade were measured on 6 plants per genotype with a ruler (± 0.5 mm). In addition, their zenithal angles were measured with a digital protractor (± 0.5°, Pro 360, Mitutoyo, Paris, France). Then, the individual leaf area was estimated from an allometric relationship with the length and the width of the leaf blade. The light interception efficiency, and, thus, the extinction light coefficient (k), were estimated using 3D virtual scenes and a radiative balance model (Rey et al., 2008). The photosynthetic parameter (PS) was estimated from leaf photosynthetic activities measured with a portable photosynthesis system (CIRAS, PP system, UK) with a control radiation level of 1500 µmol m^{-2} s^{-1}. All the genotypic values were normalized with respect to those obtained for Melody.

Yield index and model validation: For a given genotype, the yield index (I) was defined as the ratio between the seed yield of this genotype and the average seed yield of the five oldest genotypes (see Vear et al., 2003 for more details). A yield index was calculated (Imod) for each of the 26 genotypes running SUNFLO with the mean climate conditions observed in Montpellier during the last 25 years. Imod was compared to a reference yield index (Iref). For the genotypes 1 to 20, Iref was taken from Vear et al. (2003). Iref of genotypes 21 to 26 came from experiments where some of genotypes 1 to 20 were grown in addition to the considered genotypes (Vear, pers. comm. for genotypes 21 to 23; Seryes, pers. comm. for genotypes 24 to 26).

Model simulations were tested with an independent data set collected in a field experiment carried out in Montpellier in 2002 with 5 genotypes (Albena, Heliasol, Melody, Prodisol and Vidoc) (see Rey et al., 2008 for more details on experimental design and measurements).

RESULTS

12 phenotypic traits displayed statistical differences between the 26 genotypes and were then considered as genotypic parameters (Table 1). The architectural parameters presented the highest variability with a CV value higher than 20%. The thermal date of the four key developmental stages presented a similar range of variation with CV around 10%. More surprising was the high variability in parameter k, which corresponds to the efficiency of the plant leaf area to intercept the incident radiation. This parameter is generally considered as a species characteristic and close to 0.80 for cultivated sunflower. The obtained range of k values is close to what is observed in the plant kingdom. The 10% difference in photosynthetic activity (PS) was also surprising because it was observed to be among the best commercial genotypes which are considered as optimized on this trait. In term of biomass allocation, almost no variability was observed in the proportion of biomass allocated to the capitulum (HIcap). On the other hand, the capitulum biomass allocated to the seed (HIseed) displayed a high variability with a significant increase in this value among the recent commercial genotypes.

Table 1. Minimum, maximum and mean values of the 12 genotypic parameters displaying significant differences between the studied genotypes

Genotypic parameter		Minimum	Maximum	Mean	CV
E1	(CDD)	425	690	525	0.12
F1	(CDD)	863	1253	989	0.09
M0	(CDD)	1136	1460	1253	0.07
M3	(CDD)	1578	2242	1772	0.09
Nmax	(#)	18.8	42.5	27.3	0.20
A1	(cm²)	1939	7430	5095	0.25
A2	(#)	11.0	31.5	14.9	0.25
A3	(cm²)	138	466	343	0.21
K		0.52	0.96	0.85	0.11
PS		0.92	1.02	0.97	0.03
Hicap		0.5	0.55	0.52	0.03
Hiseed		0.40	0.70	0.59	0.14

The ability of SUNFLO to account for the crop functioning was evaluated by comparing the values of a set of simulated variables to observed independent values obtained in Montpellier in 2002. The chosen variables tested the model performance on its major parts which are phenology, architecture and biomass production and allocation (Table 2). A good consistency was seen between observed and simulated values whatever the considered variables. The mean errors on phenology, architecture and total biomass were close to 10% of the observed values. The capitulum biomass displayed the highest mean error with approximately 30%.

Table 2. Comparison of observed or simulated values of the SUNFLO model.

Variables	n	slope	R^2	Mean	RMSE	Bias
Number of leaves	62	1.118	0.949	16.18	2.14	-0.0214
RIE	90	1.011	0.744	0.83	0.088	0.003
Total biomass (g)	35	0.962	0.970	752.1	85.4	14.99
Capitulum biomass (g)	32	0.978	0.904	263.5	77.1	28.32

The yield was simulated for the 26 genotypes with the mean climate data observed in Montpellier during the last 25 years. The emergence date was set on the 15th April. The simulated yield ranged from 119 to 716 g m^{-2} of seeds with a mean value of 447 ± 48 g m^{-2}. The highest simulated yields were close to the values considered as being the biological potential of present genotypes (Connor and Hall, 1997). The simulated yield indices of commercial hybrids ranged from 85 to 167 for, respectively, Peredovik and Melody. The introgression lines had much lower yield indices of below 60. The comparison between the observed and simulated yield indices showed a good consistency between both values (Fig. 1). The slope of the linear regression between observed and simulated values is equal to 1 and the model accounted for more than 80% of the observed variability in the yield index. The mean quadratic error indicated that the model is able to distinguish, in terms of productivity, groups of 3 to 4 genotypes with close yield indices.
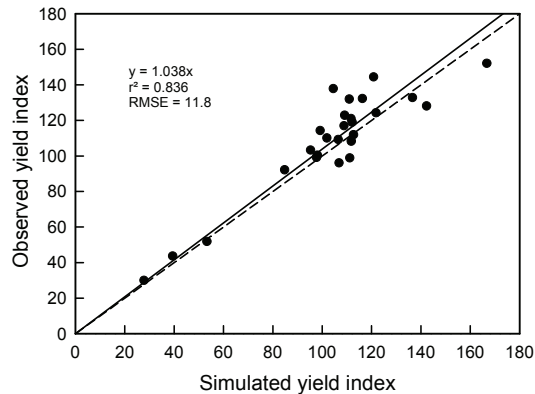
Fig. 1. Comparison of observed and simulated yield indices.

The impact of the variability of the genotypic parameters on the potential yield was estimated through a coefficient of variation (Fig. 2). To estimate this coefficient of variation, a mean value was imposed for all the genotypic parameters except one. The yield indices were then estimated by using the values observed for the 26 genotypes. This approach predicts the existing variability in *Helianthus annuus* species. As the parameter impacts are not strictly additive, the sum of the individual impacts was higher than the total observed variation in yield indices. However, this approach gave some information on the relative weight of the genotypic parameters in the yield variations.

The impact of the genotypic parameter values on the coefficient of variation of seed yield ranged from 0.5% to 14.3% for, respectively, the thermal date of E1 and the coefficient of capitulum biomass allocation to the seeds (HIseed). The other strongest individual impacts were observed for the thermal dates of F1, M0 and M3 and for the maximum plant leaf area (A1) and the position of the largest leaf (A2). When the parameters were bulked according to the major parts of the model, the ranking of the processes in term of their impact on yield variability was, first, the biomass allocation and the light interception through the plant architecture, second, the plant phenology and, far away, the photosynthesis.
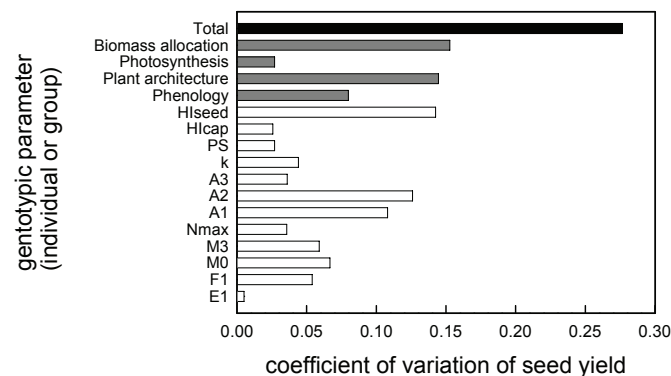

Fig. 2. Relative impact of individual or group of genotypic parameters on simulated seed yield.

DISCUSSION

The proposed approach combining phenotyping and modelling appeared to be relevant for analysing complex phenotypic traits such as seed yield. The estimation of the genotypic parameters on a few plants grown in a greenhouse gave values close to those usually observed in field conditions (Vear et al., 2003). This approach also revealed interesting traits rarely taken into account such as plant architecture, light interception efficiency and photosynthesis. The next steps would be to evaluate the robustness and the reliability of such phenotyping approaches. The more relevant traits might be the target of a more detailed analysis, especially in terms of their genetic determinism (Triboi et al., 2004). At present, this analysis would be greatly slowed down by some phenotypic measurements which are time-consuming. This suggests that simple and rapid methods in measuring phenotypic traits have to be developed. With a broader and genetically organized panel of genotypes, statistical analysis of the parameter combinations

or associations may allow to identify different ideotypes. It may also allow a subsequent analysis of the breeding strategies.

The simple biophysical model SUNFLO was able to account for approximately 80% of the variability in potential seed yields. This result was obtained with a highly contrasting panel of genotypes in terms of productivity. This is promising, although the resolving power of the model is still insufficient. A ten point uncertainty in yield indices is still too high to distinguish the genotypes of one same breeding generation. However, the modularity of the biophysical model is of interest for identifying the strong and weak points of a given genotype. For instance, some genotypes were very efficient in biomass allocation or in photosynthesis but none of them were optimized for all the processes described in the model. This suggests that some progress margins still exist in terms of productivity. These margins may be defined by looking for original combinations of the genotypic parameters corresponding to unknown virtual genotypes, which could be tested *in silico*. This last approach might be a useful tool for increasing the efficiency of breeding programs.

REFERENCES

Connor, D.J., and A.J. Hall. 1997. Sunflower physiology. p. 113-182. In: Sunflower Technology and Production. Agronomy Monograph 35, ASA, CSSA, SSSA , Madison, WI, USA.

Debaeke, P., A.M. Triboi, F. Vear, and J. Lecoeur. 2004. Crop physiological determinants of yield in old and modern sunflower hybrids. p 267-274. In: Proc. 16th Int. Sunflower Conf., Fargo, ND, USA.

Dosio, G.A.A., H. Rey, J. Lecoeur, N.G. Izquierdo, L.A.N. Aguirrezabal, F. Tardieu, and O. Turc. 2003. A whole-plant analysis of the dynamics of expansion of individual leaves of two sunflower hybrids. J. Exp. Bot. 54:2541-2552.

Hammer, G.L., T.R. Sinclair, S.C. Chapman, and E. van Oosterom. 2004. On systems thinking, systems biology and the in silico plants. Plant Physiol. 134:909-911.

Jeuffroy, M.H., A. Barbottin, J.W. Jones, and J. Lecoeur. 2006. Crop models with genotype parameters. p. 281-308. In: Working with crop models. INRA Editions, Paris, France.

Jones, H.G. 1992. Plants and microclimate. Cambrige University Press, Cambridge, UK.

Lecoeur, J., and B. Ney. 2003. Change with time in potentiel radiation-use efficiency in field pea. Eur. J. Agron. 19:91-105.

Rey, H., J. Dauzat, J.F. Barczi, G. Dosio, K. Chenu, and J. Lecoeur. 2008. A 3D architectural model to estimate organ light absorption in *Helianthus annuus* L.: Does heliotropism improve the light capture at plant level? Ann. Bot. (doi:10.1093/aob/mcm300).

Reymond, M., B. Muller, A. Leonardi, A. Charcosset, and F. Tardieu. 2003. Combining QTL analyis and an ecophysiological model to analyse the genetic variability of the responses of leaf growth to temperature and water deficit. Plant Physiol. 131:664-675.

Sinclair, T.R., and C.T. deWit. 1975. Photosynthate and nitrogen requirements for seed production by various crops. Science 189:565-567.

Sinclair, T.R., L.C. Purcell, and C.H. Sneller. 2004. Crop transformation and the challenge to increase yield potential. Trends Plant Sci. 9:70-75.

Sinclair, T.R., and L.C. Purcell. 2005. Is a physiological perspective relevant in a genocentric age? J. Exp. Bot. 56:2777-2782.

Sinclair, T.R., and N.G. Seligman. 1996. Crop modelling: from infancy to maturity. Agron. J. 88:698-704.

Triboi, A.M., J. Messaoud, P. Debaeke, J. Lecoeur, and F. Vear. 2004. Heredity of sunflower leaf characters useable as yield predictors. In: Proc. 16th Int. Sunflower Conf., Fargo, ND, USA.

Vear, F., H. Bony, G. Jouber, D. Tourvieille de Labrouhe, I. Pauchet and X. Pinochet. 2003. The results of 30 years of sunflower breeding for France. OCL 10:66-73.

Vear, F. 2006. Le tournesol. p. 146-160. In: C. Doré, and F. Varoquaux (eds), Histoire et amélioration de cinquante plantes cultivées. INRA Editions, Paris, France.

Yin, X., P.C. Struik, and M.J. Kropff. 2004. Role of crop physiology in predicting gene-to-phenotype relationships. Trends Plant Sci. 9:423-432.