

EFFICIENT INTROGRESSION OF VALUABLE NEW ALLELES FROM LUPIN GENE BANKS INTO ELITE LUPIN BREEDING PROGRAMS

Wallace Cowling

School of Plant Biology, Faculty of Natural and Agricultural Sciences, The University of Western Australia, 35 Stirling Highway, Crawley WA 6009, Australia

*Corresponding author: wcowling@cyllene.uwa.edu.au

ABSTRACT

Lupin gene banks have ample supplies of valuable new alleles (VNAs), defined as minor alleles that have the potential to improve valuable quantitative traits in domesticated lupin species. Population genetics theory indicates that VNAs are readily lost during the backcross process, well before their value is realised in crop plants. In elite breeding programs they can also be lost by genetic drift due to their low allele frequency in the population – as a result of small effective population size in most breeding programs. This paper provides a model to move VNAs from wild to domesticated lupins and to capture their value in an elite breeding program. The model is based on the binomial probability distribution for segregating alleles in backcross populations, with at least 95% chance of retaining each VNA in fully domesticated offspring derived from the BC₂ generation. The model allows for simultaneous selection for 5 or 6 major domestication genes. The VNA is fixed by selfing BC₂-derived progeny – this is crucial to improving the chances of retaining the VNA in subsequent crossing to elite parents. The model will generate a continuous flow of VNAs for assessment in elite breeding material, and consequently will help to release the value of lupin genetic resources currently tied up inside international lupin gene banks.

INTRODUCTION

The world is relatively well endowed in lupin genetic resources with 17 genetic resource centres holding a minimum of 200 accessions, and more than 25,000 accessions in gene banks worldwide (Cowling *et al.* 1998a). Valuable new alleles (VNAs) for crop lupins are hypothesised to be present in these lupin gene banks, but they are largely inaccessible to lupin breeding programs as a result of formidable genetic barriers as will be described in this paper.

Crop-type lupins share a problem with many other crop species – the domestication process has dramatically reduced the size of the crop-type populations. Cultivated sweet narrow-leafed lupin (*Lupinus angustifolius* L.) has a very narrow genetic background as a result of domestication from a handful of genotypes during the 20th century (Cowling *et al.*

1998b, Cowling 1999, Gladstones 1982). For most crops, one or more genetic ‘bottlenecks’ have occurred during domestication as a result of the founder effect (Ladizinsky, 1985). The founder effect limits the potential for genetic improvement of most crop species. The corollary of the founder effect is that wild relatives harbour valuable genetic variability with high breeding potential (Ladizinsky, 1985). These VNAs are defined as minor alleles that have the potential to improve valuable quantitative traits in domesticated species.

Only five or six domestication genes separate cultivated *L. angustifolius* from the wild genepool (Cowling *et al.* 1998b, Gladstones, 1970, see Materials and Methods), which is similar to the situation in many other crop plants (Ladizinsky, 1985). Nevertheless, this presents a major obstacle to introgression of VNAs from wild to domesticated *L. angustifolius*. Only one out of 4096 domestication genotypes segregating in the F₂ of a cross between wild and domesticated *L. angustifolius* are homozygous for crop-type alleles at all 6 domestication loci. These rare plants can be selected over subsequent generations of selfing, but are likely to be agronomically inferior to the elite breeding lines since 50% of the alleles on average come from the wild parent.

The challenge of introgressing genetic diversity from wild relatives, as the genetic gap between improved and ‘unimproved’ gene pools increases, has limited the use of exotic germplasm in elite breeding programs. The performance of breeding populations is almost always negatively affected when exotic germplasm is introduced (Rasmusson and Phillips, 1997). The exception is the routine backcrossing and transfer of major genes (such as disease resistance). Some breeders have concluded that minor alleles cannot be transferred across the domestication bridge: “What is precluded is the transfer of useful genetic diversity for important quantitative traits” (Rasmusson and Phillips 1997).

From personal experience, I was not able to achieve a useful result from crossing domesticated and wild *L. angustifolius* over 17 years of active lupin breeding. Inevitably, after selfing for 5-6 years, the few surviving fully domesticated progeny were inferior agronomically to the parent elite lines. I had reached a breeding ‘dead-lock’ – having invested a large amount of time and

resources, I could not justify further crossing into the elite population, so any VNAs in the surviving progeny were therefore lost. Most breeders have not overcome this problem, despite the fact that it severely limits the long-term effectiveness of plant breeding.

It is likely that VNAs are present in wild germplasm, and it should be possible to reveal their effect in desirable agronomic backgrounds. Advanced backcross analysis (Tanksley and Nelson, 1996) has been used to prove that quantitative trait loci (QTLs) from wild rice (*Oryza rufipogon* Griff.) are associated with positive contributions to yield and quality in cultivated rice (*O. sativa* L.) (McCouch *et al.* 2007, Moncada *et al.* 2001, Thomson *et al.* 2003). This confirms that wild relatives contain valuable minor alleles that may contribute to improvement in economically important quantitative traits, such as grain yield, in elite breeding programs.

This paper describes an efficient and low-cost method to access potentially valuable genetic diversity in the wild lupin gene banks. The recent development of a genetic map in *L. angustifolius* (Nelson *et al.* 2006) will aid the incorporation of advanced backcross QTL analysis (Tanksley and Nelson, 1996) or association mapping (Brescghello and Sorrells, 2006) into this process.

MATERIALS AND METHODS

There are six major domestication genes in *L. angustifolius* (Cowling 1999, Cowling *et al.* 1998b, Gladstones 1970). *Leucospermus* (*Leuc*) controls pigment production in seeds, cotyledons and flowers and the recessive form results in white flowers, green cotyledons and relatively white seeds in the domesticated crop. *Iucundis* (*Iuc*) controls alkaloid production and the recessive form results in 'sweet' seeds in the domesticated crop. *Mollis* (*Moll*) controls water permeability of seed and the recessive form allows immediate germination upon sowing – this maternal trait is expressed in the seed coat of progeny seeds. *Tardus* (*Ta*) and *Lentus* (*Le*) control pod shattering and both are needed in the recessive form to prevent pod shattering at harvest. Early flowering (loss of vernalisation requirement) is controlled by *Ku*, and is the only dominant domestication allele – it is important for adaptation to short cropping seasons.

The introgression model involves crossing the wild lupin (W) to three adapted varieties (D1, D2, D3) such that $F_1 = D1 \times W$, $BC_1 = F_1 \times D2$, and $BC_2 = BC_1 \times D3$ (Fig. 1). An allele *A'* from the wild parent is followed through the process of backcrossing to the BC_2 population based on binomial probability (Fig. 1). The same adapted variety could be used ($D1 = D2 = D3$) which may make the VNAs more 'visible' as a result of greater uniformity in BC_2 -derived lines, but this reduces the opportunity for recombination in the elite population, and it reduces the effective population size of the BC_2 due to greater inbreeding. The fate of domestication genes from the domesticated parent(s)

must be considered in parallel (Fig. 2) to the fate of *A'* alleles from the wild parent (Fig. 1).

The binomial distribution for segregating progeny indicates that there is a probability of $P = 0.984$ that at least 1 out of 6 BC_1 progeny will contain the *A'* allele in the heterozygous form, so at least 6 BC_1 progeny should be used in crossing to produce the BC_2 . The BC_2 field trial in year 2 should contain at least 6 rows from each of 6 BC_2 families (minimum 36 rows) with at least 48 viable seeds in each row (it is necessary to grow 48 BC_2 plants in each row in order to select for domestication genes as described in Fig. 2). There is a probability of $P = 0.968$ (0.984 in $BC_1 \times 0.984$ in BC_2) that at least one of the 36 BC_2 rows is segregating for *A'* (Fig. 1). In this row, at least 7/48 plants are homozygous for this allele ($P = 0.973$). Homozygous plants *A'A'* may show a visible improvement to stress or disease pressure and be selected for harvest – in any case, a minimum number of plants should be selected in the BC_2 and selfed for further testing. BC_2 -derived individuals contain on average 87.5% elite alleles, and are expected to be acceptable for agronomic characteristics. Since all *A'* alleles are subjected to the same probability of retention during the backcross process, it may be concluded ($P > 0.95$) that the entire genome of the W parent will be dispersed across the 36 $BC_2S_{0:1}$ family rows with 48 plants in each row (total 1728 plants).

The domestication alleles will be selected in a certain order during this backcrossing process (Fig. 2). Following the same logic as above, it is possible to select most of the major domestication alleles before sowing in the $BC_2S_{0:1}$ field row trial.

The time taken to reach the $BC_2S_{0:1}$ field row trial will be approximately 1.6 years from planting the initial parents (Fig. 2). The domestication genes that are segregating in this trial include the non-shattering gene *ta* and early flowering gene *Ku*. The $BC_2S_{0:1}$ family rows will segregate 1:1 for non-shattering:shattering pods (*ta ta:Ta ta*). In segregating rows, at least 7 of 48 plants ($P = 0.973$) will be homozygous *ta ta*. The $BC_2S_{0:1}$ family rows will also segregate 1:1 for *Ku Ku:ku ku*. Segregation for late flowering will occur in some rows and at least 7 of 48 plants ($P = 0.973$) will be homozygous *ku ku* or *Ku Ku* in these rows.

It follows that at least 1 of 48 plants ($P = 0.955$) in a segregating row will be both *A'A'* and *ta ta* (i.e. homozygous for the VNA as well as the allele preventing pod dehiscence). Many rows will be uniformly homozygous for *ta ta*; therefore the number of individuals homozygous for *A'A'* and *ta ta* will be higher in these rows. Imposing a 95% probability limit for selection of domestication genes during the backcross process is a conservative approach, designed to improve the chances of fixing most VNAs in BC_2 -derived plants. If the success rate for selecting domestication alleles is higher than in the model, there will be a corresponding increase in the number of homozygous plants for *A'A'* and all domestication genes

RESULTS AND DISCUSSION

The introgression of wild alleles into elite germplasm is a specialised backcrossing/selfing procedure which combines selection for domestication alleles in parallel to fixing VNAs from the wild parent during two backcross generations to elite parents. There is a high probability ($P = 0.955$) that at least 1 plant will be homozygous for all the domestication alleles and the potentially valuable minor wild allele A' in every segregating $BC_2S_{0:1}$ row. Each selected plant from $BC_2S_{0:1}$ rows will contain on average 87.5% elite alleles.

This model provides a 95% probability of sampling all the alleles from the wild parent in a background of 87.5% adapted germplasm among 36 $BC_2S_{0:1}$ rows with 48 plants in each row (total population 1728 plants). Those plants that look 'interesting' should be selfed, harvested and retained for further testing. These plants will provide an extremely valuable source of novel genetic diversity for selection and crossing to begin the next phase of breeding. Selected $BC_2S_{1:2}$ lines have a good chance of being homozygous $A'A'$ which improves heritability of the VNAs in the next round of crossing, when the aim is to increase the proportion of elite background increases to at least 93.75%. At this point, the positive effects of VNAs should be clearly seen in elite breeding lines.

It is important that the effective population size (N_e) and selection coefficient (s) of elite breeding programs are kept in balance to avoid the effect of genetic drift. If s exceeds $\frac{1}{2}N_e$, selection will dominate drift (Falconer and Mackay, 1996) and favourable A' alleles will be selected – assuming that selection is effective in BC_2 -derived lines. Modelling in animal populations suggests that a genetic bottleneck of 20 individuals ($N_e = 20$) is sufficient to potentially eliminate low-frequency alleles ($p < 0.1$) at selectively-neutral loci, compared with alleles at frequencies above 0.1, due to genetic drift (Luikart *et al.* 1998). In other words, low frequency alleles are most sensitive to the impact of genetic drift in small crop breeding populations. It is important to increase the frequency of favourable A' alleles above 0.1, and the effective population size above 20, in elite crop breeding programs in order to enhance the introgression of VNAs into the elite population.

In most crop breeding programs, the effective population size is less than 20. For example, 5 soybean introductions account for 55% of the genetic background of public soybean cultivars in the USA and significant regional specialisation and isolation has occurred since breeding began (Gizlice, 1994). The effective population size of Australian canola breeding programs after 30 years of closed recurrent selection was approximately 11, based on the number of ancestral parents in pedigrees, and cumulative inbreeding resulted in the loss of 21% of the alleles present in 1970 by

random genetic drift in Australian canola varieties released in 2000 (Cowling, 2007). By comparison, 8 or 9 ancestral parents were present in pedigrees of lupin varieties released in the late 1990s in Australia (Cowling, 1999). Lupin breeders in Australia are in a similar situation to breeders of most of the world's elite crop breeding programs, and should take steps to increase the effective population size of their breeding program to avoid the loss of VNAs during the introgression process.

While new varieties may not arise from the introgression program outlined in this paper, the next round of crossing to elite parents may result in competitive progeny with high yield and quality, and with valuable improvements in some quantitative traits. Introgression is rapid – only 2.5 years from initial parents to harvest of fully domesticated $BC_2S_{1:2}$ lines – and relatively low in cost and size in terms of resources. The first 1.5 years can be managed on a glasshouse bench. The $BC_2S_{0:1}$ field trial will require approximately 36 10-m rows, each with 48 plants at 5 plants m^{-1} . The process should be within the budget of a 'normal' commercial breeding program.

The introgression process presents the breeder with a pool of novel genetic diversity in a fully domesticated background. Every selected $BC_2S_{1:2}$ line contains novel genetic diversity for further testing and crossing in the elite breeding program. If the budget is limited to one $W \times D$ cross each year, this will provide a continuous stream of new and potentially valuable germplasm into the breeding program on an annual basis. New 'D' lines will be used each year as new elite varieties are developed by the breeding program, as done in the RIPE system for barley (Kannenberg and Falk, 1995).

Tanksley and Nelson (1996) proposed a method for introgressing favourable QTLs from wild relatives, while eliminating unfavourable DNA, during backcross generations. Tanksley *et al.* (1996) demonstrated that these methods were useful to transfer mapped QTLs from wild to cultivated tomatoes to improve the phenotype of cultivated tomatoes. This introgression approach was used to move valuable minor alleles from wild to cultivated rice (McCouch *et al.* 2007, Moncada *et al.* 2001, Thomson *et al.* 2003). This provides convincing evidence that valuable minor alleles exist in wild germplasm that can improve complex traits with high economic value in domesticated germplasm.

A low-cost and efficient system of introgressing novel alleles is presented here, that is available to all breeders, whether or not they have access to molecular markers. If a system of map-based markers is available, such as developed for *L. angustifolius* by Nelson *et al.* (2006), then association mapping techniques may be used to enhance selection of VNAs from wild or non-adapted parents and improve valuable quantitative traits in elite crop plants (Bresghele and Sorrells 2006). A

regular system of introgressing VNAs into elite breeding programs, such as outlined here, will help to release the value of lupin genetic resources currently tied up inside international lupin gene banks (Cowling *et al.* 1998a).

ACKNOWLEDGEMENTS

This paper was based on a collaborative project between myself and Duane Falk (University of Guelph, Ontario, Canada) and Bevan Buirchell (Department of Agriculture and Food, Western Australia) to model the introgression of wild alleles into elite crop plants. This work was submitted by the three authors for publication as a journal article at the time of writing this conference paper, and is in the review process.

REFERENCES

- Breseghele, F. and M.E. Sorrells. 2006. Association analysis as a strategy for improvement of quantitative traits in plants. *Crop Science* 46: 1323-1330.
- Cowling, W.A. 1999. Pedigrees and characteristics of narrow-leaved lupin cultivars released in Australia from 1967 to 1998. *Bulletin* 4365. Agriculture Western Australia, Perth.
- Cowling, W.A. 2007. Genetic diversity in Australian canola and implications for crop breeding for changing future environments. *Field Crops Research* 104: 103-111.
- Cowling, W.A., B.J. Buirchell and M.E. Tapia. 1998a. Lupin. *Lupinus spp. Promoting the conservation and use of underutilised and neglected crops*. 23. Institute of Plant Genetics and Crop Plant Resources, Gatersleben/International Plant Genetic Resources Institute, Rome, Italy. 105 pp.
- Cowling, W.A., C. Huyghe and W. Swiecicki. 1998b. Lupin breeding. pp. 93-120. In: Gladstones, J.S., Atkins, C.A. and J. Hamblin (eds) *Lupins as Crop Plants: Biology, Production and Utilisation*. CAB International, London.
- Cowling, W.A., B.J. Buirchell and D.E. Falk. 2008. A model for introducing novel genetic diversity from wild relatives into elite crop populations. Submitted to *Australian Journal of Agricultural Research* July 2008.
- Falconer, D.S. and T.F.C. Mackay. 1996. *Introduction to Quantitative Genetics*. 4th Ed. Pearson Education Ltd, Harlow.
- Gizlice, Z., T.E. Carter Jr and J.W. Burton. 1994. Genetic base for North American public soybean cultivars released between 1947 and 1988. *Crop Science* 34: 1143-1151.
- Gladstones, J.S. 1970. Lupins as crop plants. *Field Crop Abstracts* 23: 123-148.
- Gladstones, J.S. 1982. Breeding lupins in Western Australia. *Journal of Agriculture – Western Australia (4th Series)* 23: 73-76.
- Kannenberg, L.W. and D.E. Falk. 1995. Models of activation of plant genetic resources for crop breeding programs. *Canadian Journal of Plant Science* 75: 45-53.
- Ladizinsky, G. 1985. Founder effect in crop-plant evolution. *Economic Botany* 39: 191-199.
- Luikart, G., F.W. Allendorf, J.-M. Cornuet and W.B. Sherwin. 1998. Distortion of allele frequency distributions provides a test for recent population bottlenecks. *The Journal of Heredity* 89: 238-247.
- McCouch, S.R., M. Sweeney, J. Li, H. Jiang, M. Thomson, E. Septiningsih, J. Edwards, P. Moncada, J. Xiao, A. Garris, T. Tai, C. Martinez, J. Tohme, M. Sugiono, A. McClung, L.P. Yuan and S.-N. Ahn. 2007. Through the genetic bottleneck: *O. rufipogon* as a source of trait-enhancing alleles for *O. sativa*. *Euphytica* 154: 317-339.
- Moncada, M., C. Martínez, J. Tohme, E. Guimaraes, M. Chatel, J. Borrero, H. Gauch and S. McCouch. 2001. Quantitative trait loci for yield and yield components in an *Oryza sativa* x *Oryza rufipogon* BC2F2 population evaluated in an upland environment. *Theoretical and Applied Genetics* 102: 41-52.
- Nelson, M.N., H.T.T. Phan, S.R. Ellwood, P.M. Moolhuijzen, J. Hane, A. Williams, C.E. O'Lone, J. Fosu-Nyarko, M. Scobie, M. Cakir, M.G.K. Jones, M. Bellgard, M. Książkiewicz, B. Wolko, S.J. Barker, R.P. Oliver and W.A. Cowling. 2006. The first gene-based map of *Lupinus angustifolius* L. – location of domestication genes and conserved syntony with *Medicago truncatula*. *Theoretical and Applied Genetics* 113: 225-238.
- Rasmusson, D.C. and R.L. Phillips. 1997. Plant breeding progress and genetic diversity from de novo variation and elevated epistasis. *Crop Science* 37: 303-310.
- Tanksley, S.D. and J.C. Nelson. 1996. Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. *Theoretical and Applied Genetics* 92: 191-203.
- Tanksley, S.D., S. Grandillo, T.M. Fulton, D. Zamir, Y. Eshed, V. Petiard, J. Lopez and T. Beck-Bunn. 1996. Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative *L. pimpinellifolium*. *Theoretical and Applied Genetics* 92: 213-224.
- Thomson, M.J., T.H. Tai, A.M. McClung, M.E. Hinga, K.B. Lobos, Y. Xu, C. Martinez and S.R. McCouch. 2003. Mapping quantitative trait loci for yield, yield components, and morphological traits in an advanced backcross population between *Oryza rufipogon* and the *Oryza sativa* cultivar Jefferson. *Theoretical and Applied Genetics* 107: 479-493.