University of Reading

Evaluation of recurrent selection as a method to achieve rapid re-adaptation of faba bean to a niche agro-climate

PhD Thesis

School of Agriculture Policy and Development

Vasiliki Tagkouli

September 2020

Declaration of original authorship

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Vasiliki Tagkouli

September 2020

Στο Θάνο..

Acknowledgements

Over the last four years, my doctoral studies have been one of the most incredible experiences of my life. Despite the challenges it was a joyful and fruitful journey.

Firstly, I am deeply grateful to my supervisors Donal O'Sullivan and Sheila Alves for giving me insightful comments, suggestions, support and guidance throughout this process. Their knowledge and experience made this work possible and I am grateful for the time and effort that they have given to me.

I received generous support and hence I want to thank the technician teams within the School of Agriculture. Both teams at CEL (Controlled Environment Laboratory) and Sonning farm provided excellent experimental conditions, plant care and help during the long experimental seasons. Special thanks to Caroline Hadley and Richard Casebow for their kind support and help on my last time-tight field experiment.

I owe thanks to everyone in my research group and office colleagues for creating an excellent working environment and for continuously inspiring and supporting me. I especially thank Dr Deepti Angra who has always been supporting me scientifically, informationally and emotionally. I also appreciate the feedback and suggestions offered on several occasions by the academic staff of the school.

Further to the academic support, I would like to thank my best friend Lucie Meyer for inspiring me and always providing endless support, joyful discussions and moments of laughter that helped me get through my PhD research.

I would like to thank my parents and brother for providing emotional and moral support since the beginning, no matter the distance, or the time we are apart.

I owe a very important thank you to the most important person in my life, my husband and greater supporter Thanos, for always standing by my dreams and believing in me more than anyone, and for that I will be everlasting grateful.

Abstract

Faba bean (*Vicia faba* L.) is a nitrogen-fixing, high protein grain legume crop with high yield potential in temperate maritime climates. Across Europe as a whole, the small size of the faba bean seed market supports relatively small numbers of commercial breeding programmes.

The overarching aim of this thesis was to design and validate a deliberately small-scale, low cost faba bean breeding programme suited to the purpose of obtaining faba bean varieties tailored to the uniquely mild, high-rainfall Irish climate. A Modified Recurrent Selection Scheme (MRSS) was established in two sites (Reading, UK and Oakpark, Ireland) based on principles of extreme simplicity – after the creation of the founding pool (219 F_1 hybrid combinations), outcrossing was left entirely to bee-assisted pollination and selection was based on one simple-to-measure trait – individual plant seed yield. In each selective generation the top c. 4% highest yielding plants was advanced to winter bulking. In the fourth year of the project, co-ordinated replicated yield plot trials were conducted at both sites to compare yield performance of the foundation bulk, the penultimate and latest selections from both sites together with the founding spring varieties and a newer additional spring variety 'Lynx' as checks. Selection on individual plant seed yield in Reading had in effect achieved a total gain in yield of 16.5% compared to the founding bulk, while the equivalent parallel selection at Oakpark was 9.6% higher yielding than the founders. The increase in yield was accompanied by a substantial increase in seed weight.

Stochastic simulations were carried out using the *R* package *AlphaSim* in which the main breeding parameters – Population Size (PS) and Selection Intensity (SI) – were varied in factorial combinations with the main unknown parameter (effective outcrossing rate 'OR') to find the optimal and most cost-effective combinations. The highest

simulated SI invariably produced higher initial rates of gain (e.g. <5 years), though in lower PS, the rapid elimination of variation through sharp SI limited the long-term gain. High OR diminished gain in the short term but gave substantially higher long-term gain in all scenarios and weakened the effect of increasing SI as well as delaying the crossover point where low SI overtook high SI in cumulative genetic gain. Since SI can be varied for free but cost grows almost linearly with PS, the most cost-effective scenarios involved low PS with high SI with the caveat that over the medium to long term, either bringing in new genetic material to replenish diversity or attempts to lift OR would be required to sustain high rates of gain.

Finally, genomic responses to selection were investigated in the Reading selections by genome-wide genotyping of the founders, all selected individuals and a sample of unselected individuals from each selective generation. The total number of alleles present diminished in each selection cycle, as simulations predicted, by c.5% per cycle. In contrast, after an initial drop from the founding generation, heterozygosity trended upwards in all subsequent selections. Comparisons of selected and unselected cohorts in each generation showed a strong tendency for highly heterozygous individuals to be selected, suggesting that hybrid vigour contributes strongly to individual seed yield performance raising the concern that heterosis would therefore mask true additive variation. In order to look for signals of directional selection at particular loci, characteristic of underlying additive variation for the trait under selection, an F_{ST} scan was conducted, resulting in the detection of numerous loci that appeared to have undergone convincing directional selection.

Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Description
AFLP	Amplified fragment length polymorphism
BNF	Biological nitrogen fixation
BSL	Breeding scheme language
CAP	Common agricultural policy
CMS	Cytoplasmic male sterility
CTAB	Cetyl trimethylammonium bromide (cationic detergent)
FST	Fixation index
F_{ST}	Index of genetic differentiation
FTE	Full-time equivalent
GAI	Green area index
GG	Genetic gain
GH	Glasshouse
GHG	Greenhouse gases
GS	Genomic selection
GV	Genic variance
GWAS	Genome wide association study
Ho	Observed heterozygosity
HSW	Hundred seed weight
HSW	Hundred seed weight
IPSY	Individual plant seed yield
IRL	Ireland
ISSR	Inter-simple sequence repeats
LD	Linkage disequilibrium
LOESS	Locally weighted least squares regression
LVC	Low vicine-convicine
MAF	Minor allele frequency
MLMM	Multiple locus mixed linear model
MRS	Modified recurrent selection
MRSS	Modified recurrent selection scheme
OR	Outcrossing rate
PC	Protein content (%)

Abbreviation	Description (continued)
PGRO	Processors and growers research organisation
PGRO	Plant genetic resources
PP	Plant population per m^2
PS	Population size
QC	Quality control
QTL	Quantitative trait locus
RAPD	Random amplification of polymorphic DNA
RCBD	Randomised complete block design
RFLP	Restriction fragment length polymorphism
RIL	Recombinant inbred line
RL	Recommended list
RS	Recurrent selection
SD	Standard deviation
SI	Selection intensity
SN	Seed number in one m^2
SNP	Single nucleotide polymorphism
SPP	Seeds per plant
SRAP	Sequence-related amplified polymorphism
SSD	Single seed descent
SYN-X	Synthetic-X
VAT	Value added tax
VC	Vicine-convicine
VICCI	Virtual Irish Centre for Crop Improvement
YLD	Plot yield in t/ha

1. Chapter 1 – Literature Review

1.1. Introduction

1.1.1. *Vicia faba* L.- Origin and importance of the crop

Vicia faba is one of the oldest domesticated crops in the world, having been cultivated since the early Neolithic period (Cubero, 1974; Caracuta *et al.*, 2015). The first faba bean seeds were found in Tell el-Kerkh in 10,000 B.C. and were wedge-shaped, featuring the wild ancestor. Similar to faba beans, as we know them today, plant remains found at el-Kerkh were erect and had thick stems a fact that probably indicates growing in heavy soils as in the Mediterranean area (Tanno and Willcox, 2006). Thus, the Near East is considered as the centre of origin with four simultaneous routes of dispersal whereby Neolithic farmers brought the crop northwards to Europe, westwards via North Africa to Spain, eastwards via Mesopotamia to India and southwards along the river Nile to Ethiopia (Cubero, 1974). The long journey of faba beans around the globe explains its great adaptability and seed variation. Although faba beans must have been domesticated around the eastern Mediterranean, no extant wild progenitor species has yet been found with chromosome number (2n = 12) and content matching *Vicia faba* (Duc, 1997) although of course, a wild progenitor might exist in areas that have not yet been adequately surveyed such as Iraq, Iran or Afghanistan (van de Wouw, 1999). *Vicia narbonensis* L. received great attention as a potential faba bean progenitor; however, it has a different number of chromosomes (2n=14) and karyotype (van de Wouw, 1999).

Faba bean (*Vicia faba* L.) belongs to the *Leguminosae* or *Fabaceae* family within the *Fabeae* tribe and is also known by a variety of common names such as broad/ fava/ bell/ field/ tic or horse bean. The species *Vicia faba* was initially divided into subspecies of '*faba'* and '*paucijuga'* with paucijuga representing a primitive small seeded form of *Vicia faba* was considered similar to the wild type and to be closely related to a

hypothetical extinct progenitor (Cubero and Suso, 1981). Later, the '*faba*' subspecies was subdivided into botanical types based on the size of its seeds: *V. faba subsp. faba var major* designating types with seeds of more than 1g weight, *V. faba subsp. faba var minor* that represents small seeded types with weight less than 0.5g and *V. faba subsp. faba var equina* which are varieties with intermediate seed size (Maxted, 1995).

Faba bean (*Vicia faba* L.) plays an important role in human and animal nutrition. Since the crop is rich in protein, fibre and nutrients (Duc, 1997; Multari, Stewart and Russell, 2015; Warsame, O'Sullivan and Tosi, 2018), in most countries the seeds are valued mainly as a protein-rich human food and the straw can be used to feed ruminant livestock. Faba beans are widely cultivated in 65 countries around the world, with the largest producers being China, Ethiopia, Australia, United Kingdom and Morocco (FAO, 2020). The major reason for its cultivation as a food crop is its high protein content because it can replace more expensive animal protein in the diet. Apart from its basic use as dry beans, soaked and boiled in soups and stews, faba beans can also be used in several other forms, such as roasted and salted beans as a snack, its green fresh pods as a delicacy in the Mediterranean region and its green seeds can be consumed raw, canned or frozen.

More generally, all grain legumes are rich in protein content compared to some other staple foods (minimum 19% protein content). As a consequence, grain legumes provide one third of protein used by humans (Smýkal *et al.,* 2015) and the *Fabaceae* family to which they belong represents the second most economically important family after *Poaceae* family that contains the cereal staples of wheat, maize and rice. In surveys of grain legume protein content, faba beans at around 30% protein are ranked only below soybean (36% protein) (Duodu and Apea-Bah, 2017). However, amongst grain legumes with a comparable or higher protein content, faba beans are distinguished by higher yield potential (Cernay *et al*., 2015) and their ability to fix nitrogen through a symbiotic

interaction with *Rhizobium leguminosarum* biovar *fabae* in root nodules at an especially high rate (Watson *et al*., 2015).

Apart from protein and beneficial levels of fibres, resistant starches and micronutrients such as iron and zinc, faba beans can also contain anti-nutritional compounds, such as tannins and vicine and convicine, with the latter responsible for a potentially lethal threat to humans. Tannins reduce the digestibility of protein (Crépon *et al*, 2010), while vicine and convicine cause oxidation of glutathione in human red blood cells which can lead, in genetically predisposed individuals, to a severe form of haemolytic anaemia (Khazaei *et al*, 2018).

In addition to its credentials as a key pulse in human and animal nutrition, faba bean has been characterised as an environmentally beneficial species, as it is a legume that can meet all its nitrogen (N) needs through biological nitrogen fixation (BNF) (Herridge, Peoples and Boddey, 2008; Liu *et al.* 2019) as well as leaving behind a residue that can partly satisfy the nitrogen requirement of subsequent crops in the rotation (López-Bellido *et al*., 2003). Several fast-growing soil bacteria have been identified as faba-bean-nodulating microsymbionts but most commonly the crop fixes atmospheric nitrogen in symbiosis with *Rhizobium leguminosarum biovar viciae* (Allen and Allen, 1981). Other species like *R. laguerrereae* (Saïdi *et al*., 2014), *R. fabae* (Tian *et al*., 2009), *R. etli* and *Agrobacterium tumefaciens* are also able to establish nitrogen fixation symbiosis with faba beans (Youseif *et al*., 2014).

Faba beans can be cultivated from both autumn and spring sowing, depending on their seasonal adaptation, and are thus referred to as winter or spring types. Flowering time is advanced by exposure to cold (vernalization) in winter beans. Winter faba beans are more photophilous compared to spring varieties and although vernalization advances flowering when it happens during the embryo development, prolonged exposure to low

temperatures in later developmental stages may delay flowering (Evans, 1959). Moreover, winter beans benefit from synchrony of their life cycle with the wettest parts of the year and escape late-season drought as they mature earlier than spring varieties. Spring beans have a shorter period of vegetative growth than winter types. Unlike winter beans, spring beans develop rapidly with a single-pod bearing stem and have a higher optimal sowing density, whereas the long and slow growth of winter beans means they are selected to be sown at lower density and produce several pod-bearing branches.

The geographical range of faba bean production covers a range of climatic zones from temperate to semi-arid, with correspondingly diverse cultivars and crop-management techniques (López-Bellido *et al*., 2005). Multi-site trials, of spring-sown faba bean types, over a large geographic scale encompassing three mega-environments - Continental, Oceanic and Mediterranean - have demonstrated that quite different faba bean germplasm types and different sowing seasons are needed to yield optimally in each of these environments (Flores et al., 2013). Different environments also affect the performance of winter type faba beans and in experiments testing 15 faba bean cultivars in contrasting climatic sites almost all of the winter cultivars perform better over oceanic and continental environments than at the Mediterranean environment (Flores et al., 2012). In the Mediterranean basin, some varieties can be sown at the end of summer to be harvested at the end of autumn (Cubero, 2017). Within Europe, the best yields for faba bean are observed in the Oceanic zone (United Kingdom and Ireland included) where both winter and spring cultivation is possible.

Extremely cold weather affects the crop and thus in cooler agroclimatic zones sowing is postponed to the end of winter or beginning of spring to avoid prolonged, hard frost (Sallam, Martsch and Moursi, 2015). Excessively wet weather conditions can raise fungal disease pressure and affect the standing ability of the crop and cause lodging

problems. In faba bean, as in all crops, a variety of factors can impact crop productivity, notably abiotic stress (winds, extreme temperatures, drought etc) and biotic stress imposed by living organisms (bacteria, viruses, parasites, fungi, weeds etc). Mild, cool conditions are preferable for its development. Whole faba bean plant yield is negatively affected by heat stress during floral development, which suggests that climate warming will limit the performance of the crop (Bishop, Potts and Jones, 2016). Pollinator activity can partly mitigate pollen damage due to heat stress by bringing viable pollen from less stressed plants or flowers to those whose pollen is no longer fertile, causing in the process a shift from self- to cross-pollination (Bishop *et al.*, 2017). This latter observation highlights the relevance of faba bean reproductive biology to faba bean breeding and production and so will be covered in more detail in the next section.

1.1.2. Reproductive biology of *Vicia faba* L.

Understanding of the plant reproduction system is crucial for the design of breeding schemes. Faba bean is a partially allogamous and entomophilous (insect-pollinated) outcrossing species. Several biological features underpin its dependence on pollinating insects, including its floral morphology (Figure 1.1), which is suited to bee visitation with distinct petal spots and vein markings that are used by insects as nectar guides, cone-shaped epidermal petal cells that provide 'grip' to pollinators, production of nectar both at the base of the floral corolla (Bailes, Pattrick and Glover, 2018) and in extra-floral nectaries (Davis, Peterson and Shuel, 1988) on the stipules at the base of each leaf node that furnishes an energetic reward to visitors that successfully visit the flower, and a flower closure mechanism that restricts the nectar reward to insects large enough to exceed the opening force required to push the standard and wing petals apart. Pollinators such as bumblebees (*Bombus spp*), honeybees (*Apis mellifera*) and solitary bees can

pollinate faba bean flowers (Stoddard and Bond, 1987), with bumblebees being the most effective pollinating species, at least in the UK (Garratt et al., 2014).

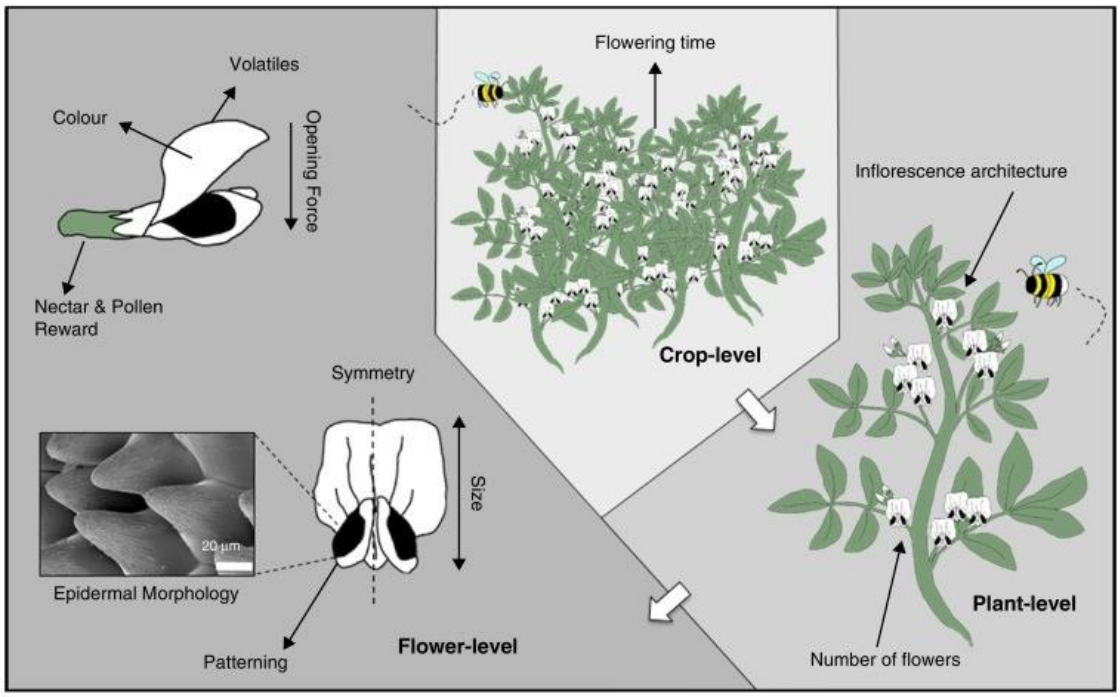


Figure 1.1. Faba bean floral traits that affect pollinator visitation and are therefore beneficial for crop yield (source: Bailes *et al*., 2015).

The rate of outcrossing is therefore related to bee activity with inbreeding occurring either when pollinators transfer self-pollen from flowers of the same plant or by selfing of unvisited flowers. In field conditions, cross-pollination varies from 5-50%, but most commonly reports estimate 20-50% (Fyfe and Bailey, 1951; Rowlands, 1958; Hanna and Lawes, 1967; Poulsen, 1973; Sjödin, 2009). The importance of cross-pollination on faba beans has been recognised for years. In principle, pollinators improve yield, however, the outcrossing rate varies and depends both on genotype (since the level of autofertility varies greatly) and the environment (Bishop, Garratt and Breeze, 2020).

The outcrossing habit and pollinator dependence have several important consequences for breeding. On one hand, the natural tendency to avoid self-pollination

without the physical stimulus of visitation by a pollinating insect (also known as 'tripping') means that heterogeneous open-pollinated varieties where pollinator visitation and therefore outcrossing rates are high can benefit from a degree of hybrid vigour. On the other hand, since all natural populations and most varieties of faba bean do consist of open-pollinated outcrossing populations, the associated heterogeneity and hybridity can make it difficult to identify true-breeding characters such as resistance to a specific disease without first bringing each accession to be screened through a number of generations of inbreeding. All generation and maintenance of genetic stocks, whether to inbreed to homozygosity or to create new hybrids, must be carried out in an environment which excludes pollinating insects. In practice, this often means in a screenhouse or pollinator exclusion cage in summer or in a screened glasshouse when growing out of season. This contrasts with inbreeding crops such as wheat or barley which can be readily inbred or outcrossed at scale in open outdoor or unscreened glasshouse conditions and therefore its partially allogamous habit imposes a large cost overhead on faba bean genetics and breeding.

Hybridization allows genetic recombination and as a result better exploitation of the genetic potential of the crop. Controlled crosses need to be protected from pollinating insects by being conducted in a pollinator-free environment. For experimental purposes, if specific crosses are of interest to the breeder, manual emasculation and crossing is used. Although there are several techniques for hand crossing, all of them work in a similar way. The breeder chooses the plant that will carry the cross (female plant) and emasculates the bud before anthers mature. Fresh pollen is then transferred from the anthers of the plant that has been chosen as pollen donor (male parent) to the stigma of the female emasculated bud. The position of the flower on the stem and the time of day of crossing affect the success of the crossing.

1.2. Breeding

1.2.1. Selection: a method of conventional plant breeding

Plant breeding aims to produce plants with a combination of desirable characteristics, including crop adaptation to a target environment, higher yield, quality/nutrition traits and disease resistance.

Although plant breeding research started long ago, most breeding research has focused on the mega-crops wheat, maize and more recently soybean while little work has been done on faba beans, most of which started after world crisis in the soybean market in 1971 (Duc, 1997).

Plant breeding entails a wide variety of methods which can be loosely described as **conventional**, such as selection, hybridization, polyploidy, induced mutation or **modern**, including biotechnological (*in vitro* cultivation of plant cells), genetic engineering (transfer of precise stretches of DNA within and between species) and marker-assisted selection and genomics-assisted breeding.

Selection, which is the oldest of the conventional breeding methods, entails the evaluation of the desired trait or traits so that comparatively superior individuals can be advanced to the next generation according to the breeding objectives. Although conceptually simple, efficient selection can benefit from both genetic insights into the causal relationships between genetic variation and phenotype and mathematical models that offer the possibility to monitor and predict outcomes. Since Mendel introduced the laws of inheritance, it is established that there are hereditary factors (genes, hereditary material) that control plant characteristics and at least in a handful of model organisms, such as thale cress (*Arabidopsis thaliana*) as a model Brassica or the barrel medic (*Medicago truncatula*) as a model legume, near complete and well-annotated genome assemblies as well as directories of both natural and induced genetic variants have been

assembled, all of which provide a growing understanding of the causal relationship between genetic and phenotypic variation. In the absence of a comprehensive understanding of what every gene and gene variant do in our species of interest, there is also a mathematical expression for heritability which can prove useful to the breeder:

$$P= G + E +GE$$

where:

- P = phenotype or trait
- G = phenotypic value determined by the genotype or genes
- E = phenotypic value determined by the environment
- GE = the interaction of genotype and environment

which explains the total variation observed for a phenotypic trait e.g. plant height is equal to the environment-independent genetic height potential of the plant + the general environmental influence on height + the specific way in which the particular genotype interacts with the environment. The G term above represents the overall phenotype expressed by the whole genetic complement of a genotype. When it comes to transferring the G component of P from a parent to its progeny, it becomes necessary to consider the genetic complexity of the trait i.e. the number of genes controlling it - this can range from monogenic (for qualitative traits) to polygenic (for quantitative traits) inheritance (Acquaah, 2017). The most important trait of all - yield - is highly polygenic.

Whilst the formula for calculation of heritability can inform the breeder as to how feasible it will be to breed for a particular trait, it says nothing about how to go about that breeding process. A fundamental equation of plant breeding which provides a degree of insight into how quickly or efficiently a given degree of directional change in phenotype

(genetic gain) can happen is the breeder's equation. The breeder's equation calculates the response to selection in one generation as follows:

$$\Delta G = ih^2\sigma \text{ or } R = ih^2\sigma = \bar{x}o - \bar{x}p$$

where:

- ΔG = genetic gain
- R = the gain in one generation of selection (response to selection)
- i = intensity of selection
- h^2 = narrow sense heritability (h^2=VP/VA)
    - VA = the total phenotypic variance
    - VP = the variance due to additive genetic effects
- σ = phenotypic standard deviation of the parental population
- $\bar{x}o$ = mean phenotype of the offspring of selected plants
- $\bar{x}p$= mean phenotype of the whole parental generation

There are three factors that can change the response to selection:

a. The phenotypic variation of the population

b. The heritability of the trait of interest

c. The threshold of selection that is set by the breeder and based on which plants are selected to progress to the next season.

The higher the heritability, the smaller the number of selected genotypes needed to achieve a set level of genetic gain, and vice versa. When there is no environmental variance (VA=VP) then the heritability is unity and the progress of selection perfect (Rajcan *et al*., 2011).

1.3. Methods of faba bean genetic improvement

Crop improvement refers to any process that enhances the performance or quality characteristics of a crop plant species. However, there is a major distinction to be drawn between those improvement activities generally referred to as trait genetics or pre-breeding that focus on improvement of one specific trait at a time and breeding per se, where all relevant characteristics are taken into consideration in selecting candidate varieties.

1.3.1. Pre-breeding or single trait improvement

The objective of pre-breeding is to find and manipulate heritable but as yet unexploited trait variation to meet a need or solve a specific production problem or vulnerability. The generic approach is to survey diverse germplasm for extended or novel variation in the trait of interest, to develop segregating populations from novel trait donor material, followed by an analysis of genetic architecture of the trait, possibly culminating in discovery of either tightly-linked markers, which can be used as proxy selection targets when breeding for the trait, or ideally the causative mutation, allowing for use of a so-called perfect marker for the trait. Several examples of pre-breeding research on traits which are considered important in faba bean breeding are discussed below.

Anti-nutritional compounds (vicine and convicine) heads up the list of traits for nutritional improvement in faba beans. Mendelian inheritance of a single locus controlling accumulation of vicine and convicine and potential to genetically map the gene was demonstrated by (Gutierrez *et al*., 2006). Commercial varieties and a number of other accessions were examined and compared and Low Vicine-Convicine (LVC) lines were clearly distinguished (Khamassi *et al*., 2013). A segregating population from a cross between high and low VC lines from that study was later genotyped and a

Quantitative Trait Locus (QTL) for vicine-convicine was localised on chromosome I (Khazaei *et al*., 2015). This locus was further fine-mapped and a causative mutation in a gene controlling a key step in the biosynthesis of vicine and convicine identified (Björnsdotter *et al*., 2020). The impact progress in understanding the molecular basis for vicine/convicine expression will have for favism sufferers is reviewed in Khazaei *et al*. (2019).

Major abiotic factors such as frost tolerance have also been the object of some breeding interest. Japanese lines were found that successfully developed flowers under snow cover (Fukuta and Yukawa, 1998), while one French landrace (Côte d'Or) and one Chinese inbred line (BPL4628) were found to be tolerant to frost (Stoddard *et al*., 2006). A modified mass selection scheme was used to identify new potential sources of winter-hardiness by bulking populations for three cycles. That research generated comparable winter-hardiness material between Washington locations and northern European populations, with different percentage gain among the survivors (Landry *et al*., 2017). Furthermore, experiments have shown that faba bean seedlings growing under a cold temperature treatment (12/5^{o}C, day/night) are more freezing tolerant than seedlings that grow under a warm (17/12^{o}C) pre-acclimation environment (Landry and Hu, 2018).

Amongst all legumes, faba beans are the most drought sensitive. Lack of water during flowering and pod setting stage results in reduction of shoot mass, root mass and subsequently the yield (Torres *et al*., 2010). Rhizotron experiments with different water regimes indicated association between root architecture and shoot measurements (Belachew *et al*., 2019). Genotypes adapted to water deficit conditions were screened for differences in stomatal conductance, transpiration efficiency, leaf temperature and carbon isotope discrimination (Khan *et al*., 2007).

Numerous fungal pathogens infect the crop, leading to serious disease outbreaks, yield losses and if not addressed, in the extreme, to a steady reduction of the cultivated area in many countries (Torres *et al*., 2006). *Ascochyta fabae* Speg., causing ascochyta blight, *Uromyces viciae-fabae* and *Botrytis fabae* are among the most important fungi affecting the faba beans. The weedy root parasitic plant *Orobanche crenata* or crenata broomrape, is also threatening faba bean performance.

Ascochyta blight causes necrotic lesions on leaves, stems, pods and seeds. It can cause great yield losses depending on the environmental conditions and the type of the cultivar. Its global spread is thought to have been exacerbated by the international exchange of plant germplasm as it is mainly a seedborne pathogen (Kaiser, 1997). Genetic resistance in ascochyta tested by infecting plants in the field, and resistant genotypes to ascochyta blight attributed to a single dominant gene were produced (Kohpina *et al*., 2000). The most resistant line for ascochyta is INRA 29H and was created at INRA-Rennes, (Station d' Amelioration des Plantes, Le Rheu, France) by Drs P. Berthelem and J. Le Guen (Maurin and Tivoli, 1992). Ascochyta blight has been studied extensively in studies screening several faba bean accessions for resistance to *Ascochyta fabae* by rating the disease severity based on lesion size (Sillero *et al*., 2001) and others describing the pathogen and introducing the importance of the choice of the appropriate isolate and techniques for field, growth chamber and detached leaf evaluation methods (Tivoli *et al*., 2006). Multi-location experiments in Europe testing 484 *V. faba* accessions for *A.fabae* resistance revealed strong genotype × environment interactions resulting in instability of resistant phenotype expression across environments (Rubiales *et al*., 2012). Management of *Ascochyta fabae* under Mediterranean conditions showed that the genotypes and sowing date affects the disease severity (early sowing higher disease progress than in late sowing) (Ahmed *et al*., 2016). One of the latest reports that

used the line 29H was from Atienza *et al*. (2016) validating mapped QTLs for ascochyta located on chromosome II (*Af*2) with other studies and identifying additional sources of resistance on chromosome III (*Af*3).

Another constraint for faba bean world cultivation is chocolate spot, caused by the fungus *Botrytis fabae*. Studies in Egypt showed 25-100% yield loss due to chocolate spot depending on the severity of the epidemic conditions (El-Komy, 2014). In both monocropping and intercropping faba bean systems, nitrogen application can increase the appearance of chocolate spot, as its application reduces the temperature and increases the relative humidity of the canopy, creating favourable conditions for the fungi (Guo *et al*., 2020). Disease symptoms and severity can be scored using a visual evaluation infection type scale (Beyene, Derera and Sibiya, 2018). Genetic resistance to faba bean chocolate spot was first reported by Hanounik and Hawtin (1982), where several accessions (NEB 938, NEB 519) were found to be highly resistant (less than 1% infection rate), while cultivars Giza 1 and Giza 3 also showed a moderate resistance. Hanounik (1988) found that numerous ILB accessions (ILB 2282, 3025, 3026) appeared highly resistant to chocolate spot after field detached leaf tests. Accessions ILB 438 and 938 and BPL 710, 1179 and 1196, found by ICARDA to be consistently partially resistant across multiple environments over many years, were also used as sources of resistance in breeding programs (Duc, 1997). Through similar screening, more and more sources of chocolate spot resistance were reported (Bond et al., 1993; Tivoli et al., 2006).

Obligate biotrophic parasites such as rust are devastating plant pathogens (Brown and Hovmøller, 2002). In Europe, rust usually develops late in the season and hastens the maturation of the crop; though in Canada yield losses up to 50% on faba beans due to rust have been reported (Bernier and Conner 1982). The first report of molecular tagging of a gene controlling race specific resistance to faba bean rust was Uvf-1 associated with

molecular markers by Bulked Segregant Analysis (Avila *et al*., 2003). Rust resistance is expected to be controlled by different genes, as several sources of resistance have been reported (Sillero, Moreno and Rubiales, 2000). Breeding studies for rust resistance on faba beans following detached leaf assays were conducted, and genotypes with intermediate response were identified (Herath, Stoddard and Marshal, 2001). Segregating populations from crosses between faba bean cultivars "Sakha 3" and "Misr 1" were screened for rust and chocolate spot resistance, resulting in demonstration of moderate heritability of resistance of diseases, low genetic gain and involvement of epistatic interactions for rust and high genetic gain for chocolate spot (El-Rodeny *et al*., 2020).

Broomrape (*Orobanche crenata)* is a root parasitic herbaceous plant which inflects devastating losses on faba beans (El-Ghareib *et al*., 2019) from 50-80% (Gressel *et al*., 2004). Since attempts to reduce the parasite, as late sowing, crop rotation, use of herbicides, nitrogenous fertilization and soil solarization (Rubiales *et al*., 2006; Parker, 2009; Stoddard *et al*., 2010) have not been totally successful (Rubiales *et al*., 2009), genetically resistant lines is the most appropriate and low cost means of controlling the parasite (Pérez-de-Luque *et al*., 2007). RAPD, ISSR and SRAP markers were used to map Orobanche tolerance segregating in diallel crosses between six Egyptian faba bean genotypes, resulting in the identification of QTLs for spike number and height per plant (Abd El-Fatah and Nassef, 2020).

1.3.2. Plant breeding and breeding schemes for faba beans

As stated above, many abiotic and biotic stresses affect the performance of crops. Crop yields and adaptation to stress can be increased via plant breeding (Moose and Mumm, 2008). Improved cultivars can be developed generating novel genotypes from the hybridisation of carefully chosen parental/founder germplasm, followed by selection and

stabilization of best performing genotypes or populations. These basic steps form the basis of plant breeding.

There are too many variations of the basic methods for founder germplasm selection, crossing methods, selection methods and strategies for development of stable and uniform genotypes or populations to be discussed fully here, so this section will focus on those methods which are applicable in faba bean. Faba bean methods of improvement can be categorized according to the nature of the end product: 1. Outcrossing open-pollinated populations, 2. Synthetic varieties, 3. Inbred lines 4. Hybrids and 5. Autogamous lines. Current commercial varieties are developed either as synthetic varieties (Gallais, 1992), open-pollinated populations from mass selection or through pedigree breeding as inbred lines, while hybrids and autogamous lines are mainly theoretical or experimental options. These five categories will be defined and discussed in turn.

Open-pollinated populations derive by self or cross-pollination that happens through random mating in the field by wind or pollinator activity. All landraces of an allogamous crop like faba bean are by definition open-pollinated populations that have evolved over time in adaptation to the environment and cultivation practices and preferences of farmer selectors. However, more formal approaches to population improvement – namely recurrent selection and mass selection – exist and will be discussed later.

Synthetic cultivars are developed by selecting and recombining in all possible combinations several inbred lines and bulking together equal amount of seeds per cross. Commercial faba bean breeding and production programmes are mainly based on synthetic cultivars (usually comprising 5 to 8 different components) and improved populations. Synthetics are expected to perform higher and produce more stable yields

than improved populations as they are taking advantage of heterosis and heterogeneity (Link *et al*., 1997; Torres *et al*., 2006). In a breeding scheme, there is always a question of how much inbreeding should be utilized. As far as synthetics are concerned, well-evaluated inbred lines hold the answer.

Inbred lines are a homozygous breeding material that through repeating selfing over numerous generations maintains a nearly identical genotype due to long inbreeding. Although genetic diversity is important, genotypic adaptation is equally attractive. Inbred lines lack of genotypic diversity and for a given environment (and/or stress factor) the best adapted inbred genotype (pure line) performs better than any mixture of this genotype with others (Ghaouti, Vogt-Kaute and Link, 2008). Selection of selfed individuals although simplifies the breeding procedure is difficult to be achieved in faba beans as offspring tend to express higher outcrossing rate compared to their selected parents with low outcrossing rate (Suso and Maalouf, 2010).

Hybrids are generated by deliberate control of pollination (cross-pollination of specific varieties). Unlike open-pollinated populations, hybrids are not used in northern Europe for commercial production yet. Hybrid varieties can contribute to yield improvement and the improvement of its components. Higher and more stable yield can be obtained by hybrid combinations, where effective allelic and inter-allelic gene interaction happens (Bishnoi *et al*., 2012). Depending on the trait and the genetic profile of the parents, heterotic effects can be either positive or negative (Lal *et al*., 2019). Hybrids and synthetics are based on heterosis, which results in progeny with greater characteristics compared to the parental lines. The future of synthetics and hybrids and their potential have also been given by Bond (1982), Picard *et al*. and Bond (Picard *et al*., 1982; Bond, 1989). Despite some efforts to develop a hybrid breeding system (Link *et al.* 1997), it has been proved difficult to obtain stable cytoplasmic male sterility (CMS)

(Bishnoi and Panchta, 2012) and therefore true hybrid breeding methods cannot currently be implemented.

In the past, Lawes (1980) proposed that *Vicia faba* should and could be developed as an autogamous species. **Autogamous lines** are self-pollinating plants where each flower pollinates the ovules with its own gametes. Although autogamous lines speed up the breeding operations and can increase the response to selection, they do not take advantage of heterosis. Rare cleistogamous (closed flowered) mutants have been found and this character would permit efficient production of autogamous lines. (Poulsen and Martin, 1977). However, beyond isolation of mutant lines with closed flowers, this concept does not appear to have been put into practice.

The most used breeding methods that can be practically followed in faba bean - synthetic population development, recurrent selection and mass selection, all methods that can be implemented in cross-pollinated species - are presented in Figure 1.2 below.

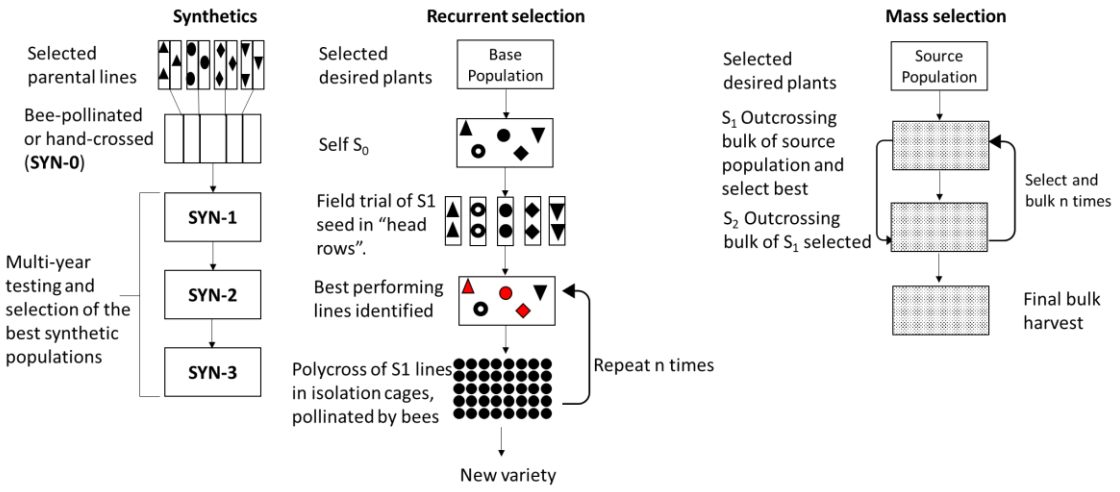


Figure 1.2. Schematic representation of the mechanics of some of the most used breeding schemes in plants. The notation S_x will be used hereafter to refer to the xth generation of selfing. SYN-X similarly refers to the xth generation of propagation as an outcrossing population following constitution of the synthetic components in SYN-0. In the recurrent selection scheme, distinct genotypes are represented by different shaped symbols.

Generally, **synthetic breeding** methods have been developed for allogamous species, where large-scale hybridization is difficult or impossible to be achieved. Parental lines are usually selected by pedigree breeding methods and synthetic varieties are created by inter-mating of those desirable inbred parental lines (SYN-0), followed by random mating of the SYN-0 to create the SYN-1, intermating of SYN-1 resulting in the creation of SYN-2 and so on (Márquez- Sánchez and Sahagún-Castellanos, 2002). After SYN-0, the individual lineage is lost and the plants are treated as a population. A certain number of synthetics are developed and their performance is evaluated as they go through the generations gradually eliminating all but the best performing synthetics. Synthetics take advantage of heterosis, but in this scheme it is difficult to define the optimal number of parental lines so as to be high enough to generate significant heterosis but low enough to allow selection among all possible synthetic genotypes coming from the parental lines (Gallais, 1992).

Recurrent selection is a widely used method for population improvement and the implementation represented in Figure 1.2. In recurrent selection, a number of plants holding exceptional characteristics (the Base population) is bulk-selfed (S_0). Progeny of the S_0 (S_1) undergo field trial in 'head rows' and the best lines are being identified and polycrossed in isolation exclusion cages pollinated by bees. This procedure can be repeated n times until the trait of interest meets the breeding scheme expectations. This process allows the creation of novel genotypes in each cycle through genetic recombination which increases the genetic diversity of the population. At the end of this process the new variety can be released.

Mass selection is the simplest method where the only intervention is in selection of individuals to advance to the next generation. At the extreme of low intervention, no phenotypes are measured and a random, representative sample of harvested seed is

advanced, relying in this case on natural selection by the environment to reduce frequency of deleterious alleles over many generations; however, to obtain higher rates of gain or to focus gain on traits that are not selected by the environment, seed of only individuals chosen based on their apparent superior phenotype may be advanced. In the latter case, the harvested seeds of the best performing individuals in each generation are bulked and the cycle repeated *n* times. Mass selection repeated *n* times is recurrent selection in its most basic form. The advantage of this scheme is that can effectively improve characteristics with high heritability, however, it is not effective for poorly heritable traits as the selected individuals might be superior due to environmental influence and not due to genetic gain.

The question of which of the above methods is the most effective has been addressed by several authors, According to Bond (1987), who made a comparison between pedigree breeding and recurrent selection scheme, pedigree breeding cannot be achieved under open-pollinated conditions even when there is reduced pollinator activity, as neighbouring plants will have the opportunity for uncontrolled cross pollination and accurate pedigree records cannot be kept. The pedigree method is considered "dangerous" for faba beans, when the trait under improvement is yield, as the chosen best performing individuals is likely to come from F_1 out-crossed plants, if selfing is not strictly controlled (Hawtin, 1982).

Ibrahim (2015) compared empirical outcomes of the application of three different breeding schemes (pedigree, single seed descent and mass selection) using the same F_2 populations made from two crosses as starting material. F_6 selections made using each of the above methods were compared regarding their yield performance. Among the three methods, single seed descent gave higher values of seed yield, number of pods and number of seeds per plant. It also achieved to maintain phenotypic and genotypic

variation within the population and a higher genetic gain and heritability values. However, this was a report of a limited source variation as the team used Giza 843 (Egyptian, hundred seed weight 'HSW' 60- 65 g) and International Lines of Broad Beans (ILB) ILB 450 and ILB 312 (from ICARDA with HSW of 80 and 90g, respectively).

Gharzeddin *et al* (2019) compared self- and open- pollinated breeding methods, following pedigree selection for the first category, and recurrent selection and synthetics for the later. Open pollinated lines showed increased yield performance and more specifically synthetics revealed greater yield compared to lines coming from recurrent selection while, lines from recurrent selection method had greater yield than lines from pedigree selection.

1.4. Molecular tools for faba bean breeding

Vicia faba L. is diploid and has 12 chromosomes (2n=12) and a 13Gb genome size. The physical chromosomes are similar in size (7-9 μm length) apart from chromosome one (18 μm length). Although the generous size of the chromosomes facilitated cytogenetic studies, the progress of molecular breeding is slow compared to other crops, due to the complexity of the genome.

The first molecular marker-based studies in faba bean involving generation of linkage maps using restriction fragment length polymorphism (RFLP) and random amplification of polymorphic DNA (RAPD) were produced by Van de Ven *et al*. (1991), on 1993 and 1995 by Torres *et al*. and Satovic *et al.* in 1996.

AFLP-based diversity studies were used to study the genetic diversity of a large set of faba bean inbred lines coming from Asian, African and European sources and showed that among those 79 lines only the Asian lines clustered a distinct group based on Jaccard's similarity coefficient (Zeid, Schön, and Link 2003). ALFP markers were

also used to study the diversity of faba bean germplasm from China in comparison to faba bean germplasm from outside China, and principal component analysis showed that Chinese germplasm was clustering separately from the outside of China originated lines (Zong *et al.* 2009). Later, RAPDs and AFLPs were used to detect QTLs explaining frost tolerance and to conduct comparative mapping across backcross families, respectively (Arbaoui *et al*., 2008; Ali, 2015).

A significant advance was made when sequence-based markers began to be used, allowing collinearity between pulses, as faba beans, and sequenced legume genomes, such as that of *M. truncatula* genome to be described (Ellwood *et al*., 2008). Intron targeted amplified polymorphic (ITAP) markers used in *M. truncatula*, lupins and soybean used to generate genetic map of faba beans, while 235 expressed sequence tag-SSRs (EST-SSRs) enable the construction of six linkage groups containing 552 loci (El-Rodeny *et al*., 2014). Also, in 2014, a SNP-based linkage map of faba bean, consisting of 12 linkage groups, was published with 551 single nucleotide polymorphism markers (SNPs) and 71 SSRs (Kaur *et al*., 2014).

The latest maps are mainly or solely based on SNP markers all assigned to six linkage groups corresponding to the six physical chromosomes of faba bean (Satovic *et al.*, 2013; Webb *et al*., 2016). These detailed syntenically anchored maps have made it possible to examine gene content and pursue candidate gene hypotheses when traits have been mapped to genetically defined intervals. For example, Webb et al, 2016 report a WD40 transcription factor underlying the zero tannin/white flower phenotype in faba bean, which emerged as a candidate gene due to the existence of an orthologous gene in called TTG1 with a similar biological function in the colinear Medicago interval.

1.5. Methods for detection of signatures of selection

The species evolution studies go back to Darwin and Wallace (Darwin, 1870) who firstly reported the principle of natural selection. Natural selection is the process where fitness-enhancing traits, associated with a species improvement (adaptation to environmental factors, tolerance or survival of stresses or ability to reproduce), tend to establish and become more frequent in the population over time.

Recurrent selection scheme theory exploits the benefit of cross-pollination in outcrossing populations which enhances their phenotypic performance. Molecular markers, and specifically SNP markers are a promising tool which allows quick, effective and low-cost genomic studies. Investigating the microevolution and adaptation of an organism that has been exposed to myriad environments and selection pressures, it is important to apply functional studies to characterize putative selected alleles and QTLs.

There are several statistical methods and statistical interpretations that are used to study the genomic evolution and detect the response to selection, or in other words alleles that repeatedly arise suggesting targeted evolution. Those detecting methods can be categorised at macroevolutionary and microevolutionary levels (Vitti, Grossman and Sabeti, 2013). At macroevolutionary level, the methods that are used are mainly compare homologous traits or sequences among species, families or taxonomic classes (Hurst, 2002). Commonly used methods in this category are the McDonald-Kreitman test (MKT), the Hudson-Kreitman-Aguade test (HKA) and phenotypic method tests that compare traits among related species (McDonald and Kreitman, 1991; Hudson, Kreitman and Aguade, 1987; Romero, Ruvinsky and Gilad, 2012). At microevolutionary level, methods are more directed at detecting positive selection that change allele frequencies reaching fixation at the extreme, noting that the more loci which have been fixed as a consequence of selection, the less the genetic diversity on the population (Smith and

Haigh, 1974). Commonly used methods at microevolutionary level are the Tajima's D test and linkage disequilibrium (LD) based tests (Tajima, 1989; Sabeti *et al*., 2002). A key concept flowing from the existence of directional selection, is differentiation, a measure of how genetically different two populations are from each other. The most widely used metric for differentiation is the Wright's fixation index (Wright, 1949; Holsinger and Weir, 2009).

In evolutionary biology, characterising the genetic structure of a population is very important. In 1949, Sewall Wright introduced "F-statistics" to describe the population structure (Wright, 1949; Weir, 2012). F statistic values can be calculated by population genetic data and include parameters like F_{IT}, F_{IS} and F_{ST}.

The Wright equation of F-statistics is:

$$(1-F_{IT}) = (1-F_{IS})\ (1-F_{ST}), \text{ where:}$$

- $F_{IT}= 1 – (H_I/H_T)$
- $F_{IS}=1 – (H_I/H_S)$
- $F_{ST=}1 – (H_S/H_T)$

and:

- H_T = total gene diversity or expected heterozygosity in the total population as estimated from the pooled allele frequencies
- H_I = intrapopulation gene diversity or average observed heterozygosity in a group of populations
- H_S = average expected heterozygosity estimated from each subpopulation
- F_{IS} = the deficiency or excess of average heterozygotes in each population
- F_{ST} = the degree of gene differentiation among populations in terms of allele frequencies (also known as fixation index)
- F_{IT} = the deficiency or excess of average heterozygotes in a group of populations

The F_{ST} takes values in the range of 0 to 1,

where,

0: no genetic divergence

1: fixation for alternate alleles in different subpopulations (CropGeneBank, 2020)

The F_{ST} index is used to make inferences about selection at individual loci. The majority of polymorphisms have either neutral or deleterious effect on the allele frequencies in populations of finite size (Kimura, 1968; Ohta, 1973) and as a result, the effective population size, the outcrossing rate and mutation rate determine the rate of differentiation within species (Biswas and Akey, 2006). Identifying targets of positive selection assists the understanding of the role of selection in the evolutionary process. Functionally important regions of the genome can be delimited through the study of positive selection using F_{ST} permitting association of the phenotypic diversity of the trait under selection with the genetic variation. F_{ST}s can be used to functionally annotate the genome and have been applied extensively in human and plant studies (Biswas and Akey, 2006). For artificial selection, the method of F_{ST}s for interpopulation comparisons is used for an approach known as "selective sweep mapping" which helps to identify genetic regions associated with the phenotype under selection (Schlötterer, 2003). The rationale of this approach is that artificial selection for the phenotype of interest should generate a detectable signature of altered allele frequencies centred around causative loci that carry variation affecting the phenotype under selection. An early example of the application of this approach was provided by allozyme data from *Drosophila simulans,* where outlier F_{ST} targets of directional selection were identified (Vitalis, Dawson and Boursot, 2001).

F_{ST} statistics and F_{ST} outlier-value method are used by plant breeders to retrospectively look at genome-wide signatures of selection over several breeding generations. Local haplotypes which tended towards fixation, after over 60 years of

breeding durum wheat for grain yield, quality traits, protein and phenotypic characteristics, were identified using the F_{ST} method for scanning for signatures of selection (N'Diaye *et al*., 2018). Fst was also used to estimate differentiation between wheat landraces and modern cultivars at a set of genes known to influence yield, quality or adaptation to environment thus pointing to those loci which had been under historic positive selection (or not) versus those with further potential to become selection targets (Zhao *et al*., 2019). Selective sweep analysis, applied to study the origins of domesticated narrow-leafed lupins, used F_{STs} to detect domestication loci, although no priori candidate loci showed strong selective sweep (Mousavi-Derazmahalleh *et al*., 2018). Ascochyta blight resistance in chickpeas was also assessed under the F_{ST} framework, resulting in the identification of candidate genes involved in the resistance via F_{ST} genome-scan (Li *et al.*, 2017).

1.6. Project Outline

1.6.1. Faba bean production in Ireland

Ireland imports over 3 million tonnes of protein for animal feed annually, most of it coming from soya bean and maize. Faba beans (*Vicia faba* L.) constitute a rich alternative, and potentially home-grown, source of protein and are currently used mostly in coarse rations. It is also a crop well-suited to the Irish climate. Despite all the stress factors described above, the Irish climate offers most of the time temperate weather conditions ideal for the crop. However, occasionally, winter sowing leads to excessive vegetative growth and spring sowing can be compromised by the hard texture of the soil (Teagasc, 2020).

The faba bean crop faces both agronomic and breeding challenges, as there are reports from Irish farmers and agronomists regarding gaps in knowledge, such as the

sowing dates, seed rates, disease/weed/insect control, and nutrient responses, while, there are no current breeding programmes for Spring beans carrying out selection in the peculiarly mild and wet Irish conditions, where high yield potential is sometimes compromised by high disease pressure and faba bean has so far remained a niche crop, grown on 11,000 ha in recent years supported by a protein crop scheme (Teagasc, 2020). Farmers and markets need reassurance on the stability and performance and thus the economic profit they will receive from the crop to consider it as home-grown replacement for soybean. With various legume-friendly greening measurements being introduced as part of post-2014 CAP reforms, and the expected surge in feed demand brought about by the expansion of the dairy herd following abolition of milk quotas in 2015 an increase in bean acreage is both predicted and desired.

Taking into consideration the strategic importance of the crop, the unreliable yields that led to low levels of faba bean cultivation over the years, the absence of a breeding programme targeting the Irish climate and the paucity of published data on performance of different breeding schemes applicable to faba bean, the present project was set up to address the overarching objective of designing a breeding programme tailored to the reproductive biology of faba bean and amenable to implementation on a small scale for niche/emerging markets.

Two decisions were taken at the outset of the project as to the nature and scope of the breeding programme, both driven by necessity. First, since the primary focus was on rapidly achieving adaptation in a selective environment, and since the only dedicated human resource available to run the programme was one PhD candidate, that the only feasible type of breeding scheme was a form of recurrent selection with a single trait under selection – seed yield per plant. Mass selection was ruled out as pure mass selection retains all genotypes that survive (slowing progress towards fixation) and allows seed

number per plant to dominate early rounds of selection at the expense of seed size, whereas recurrent selection allowed equal selection of superior genotypes whether they achieved yield by virtue of high seed size or high seed number or a combination. Secondly, it was decided to rely on insect-mediated outcrossing to maintain high levels of recombination across generations relieving the need for labour-intensive manual crossing. Given that the broad nature of the breeding scheme was committed to at the outset, three specific objectives to investigate optimisation of such a breeding scheme were defined, each addressed by a separate Chapter in this thesis, in pursuit of the overall goal:

1. To establish a new Modified Recurrent Selection Scheme (MRSS) and to obtain empirical data on genetic gain in early generations of selection at two distinct selection sites (Chapter 2).
2. To conduct a theoretical simulation study of recurrent selection to guide choice of appropriate parameters that would maximise genetic gain (Chapter 3).
3. To use high density genotyping to monitor outcrossing rates and quantify changes in genetic diversity over time as well as to dissect genomic responses to selection (at one selection site) (Chapter 4).

1.6.2. Hypotheses

Behind the three objectives lie three corresponding hypotheses.

1. Following implementation of the modified recurrent selection in practise, we expect to observe a gain in yield of the final selected population over the founding population. Furthermore, the final selected population (after three rounds of selection), given that approximately half the genomes represented in the founding population were elite current varieties, is expected to exceed yield of those same elite founding varieties by virtue of having selected novel combinations of alleles from the elite founders conditioning high yield potential/adaptation with novel yield-enhancing alleles from the non-elite/diverse founders.
2. To the extent that the simulation correctly models trait architecture and heritability, it should be possible to identify the best combination to population size and selection intensity to maximise genetic gain over a defined period of time.
3. Since major features of the selective environment (e.g. soil, climate pattern, biotic stresses) will favour alleles at specific loci controlling additive variation in responses to these edaphic and biotic factors that enhance yield potential of individuals, selective sweeps on such loci will be detectable as localised changes in allele frequency. It is expected that by measuring the allele frequencies at the beginning and at the end of the programme, responses to selection will be seen in the form of selective sweeps or even fixation on specific loci.

1.7. References

Abd El-Fatah, Bahaa E. S., and D. M. T. Nassef. 2020. Inheritance of Faba Bean Resistance to Broomrape, Genetic Diversity and QTL Mapping Analysis. Molecular Biology Reports. **47** (1): 11–32. https://doi.org/10.1007/s11033-019-05101-1.

Acquaah, G. 2017. Plant Breeding, Principles. Encyclopedia of Applied Plant Sciences. **2**: 236-242. https://doi.org/10.1016/b978-0-12-394807-6.00196-9.

Ahmed, S., M.M. Abang, and F. Maalouf. 2016. Integrated Management of Ascochyta Blight (*Didymella Fabae*) on Faba Bean under Mediterranean Conditions. Crop Protection. **81**: 65-69. https://doi.org/10.1016/j.cropro.2015.12.013.

Ali, M. 2015. Genetic Linkage Maps and Homology Study of Backcross Families of German Faba Bean (*Vicia faba* L.). Journal of Crop Improvement. **56** (3): 1036-1048. https://doi.org/10.1080/15427528.2015.1053012.

Allen, O. N. and E. K. Allen. 1981. The Leguminosae. A Source Book of Characteristics Uses and Nodulation. University of Wisconsin Press, Madison, U.S.A. https://doi.org/10.1007/978-1-349-06142-6_1.

Arbaoui, M., W. Link, Z. Satovic, and A.-M. Torres. 2008. Quantitative Trait Loci of Frost Tolerance and Physiologically Related Trait in Faba Bean (*Vicia Faba* L.). Euphytica. **164**:93–104. https://doi.org/10.1007/s10681-008-9654-0.

Atienza, S., C. Palomino, N. Gutiérrez, C. Alfaro, D. Rubiales, A. M. Torres, and C. M. Ávila. 2016. QTLs for Ascochyta Blight Resistance in Faba Bean (*Vicia Faba* L.): Validation in Field and Controlled Conditions. Crop and Pasture Science. **67**(2): 216-224. https://doi.org/10.1071/cp15227.

Avila, C., J. C. Sillero, D. Rubiales. M. T. Moreno, and A. M. Torres. 2003. Identification of RAPD Markers Linked to the Uvf-1 Gene Conferring Hypersensitive

Resistance against Rust (*Uromyces Viciae-Fabae*) in *Vicia Faba* L. Theoretical and Applied Genetics. **107** (2): 353–58. https://doi.org/10.1007/s00122-003-1254-8

Bailes, E, J., J, G. Pattrick, and Be, J. Glover. 2018. An Analysis of the Energetic Reward Offered by Field Bean (*Vicia Faba*) Flowers: Nectar, Pollen, and Operative Force. Ecology and Evolution **8** (6): 3161–71. https://doi.org/10.1002/ece3.3851

Bailes, E, J., J, Ollerton, J.G. Pattrick, and B, J. Glover. 2015. How Can an Understanding of Plant–pollinator Interactions Contribute to Global Food Security? Current Opinion in Plant Biology. **26**: 72-79. https://doi.org/10.1016/j.pbi.2015.06.002.

Belachew, K. Y., K.A. Nagel, H. Poorter, and F, L. Stoddard. 2019. Association of Shoot and Root Responses to Water Deficit in Young Faba Bean (*Vicia Faba* L.) Plants. Frontiers in Plant Science. **10**: 1063. https://doi.org/10.3389/fpls.2019.01063.

Bernier, C. C., and R. L. Conner. 1982. Breeding for Resistance to Faba Bean Rust. In Faba Bean Improvement. 251-257. https://doi.org/10.1007/978-94-009-7499-9_26.

Beyene, A. T., J. Derera, and J. Sibiya. 2018. Genetic Variability of Faba Bean Genotypes for Chocolate Spot (*Botrytis Fabae*) Resistance and Yield. Euphytica. **214**:132. https://doi.org/10.1007/s10681-018-2210-7.

Bishnoi, S. and Panchta, J.S. 2012. Advances on heterosis and hybrid breeding in faba bean (*Vicia faba* L.). Forage Research. **38**. 65-73.

Bishop, J., H. E. Jones, Dl M. O'Sullivan, and S. G. Potts. 2017. Elevated Temperature Drives a Shift from Selfing to Outcrossing in the Insect-Pollinated Legume,

Faba Bean (*Vicia Faba*). Journal of Experimental Botany. **68** (8): 2055–63. https://doi.org/10.1093/jxb/erw430

Bishop, J., M. P. D. Garratt, and T. D. Breeze. 2020. Yield Benefits of Additional Pollination to Faba Bean Vary with Cultivar, Scale, Yield Parameter and Experimental Method. Scientific Reports. **10** (1): 2102. https://doi.org/10.1038/s41598-020-58518-1

Bishop, J., S. G. Potts, and H. E. Jones. 2016. Susceptibility of Faba Bean (*Vicia faba* L.) to Heat Stress During Floral Development and Anthesis. Journal of Agronomy and Crop Science. **22**(6): 508-517. https://doi.org/10.1111/jac.12172.

Biswas, S., and J. M. Akey. 2006. Genomic Insights into Positive Selection. Trends in Genetics. **22**(8):437-446. https://doi.org/10.1016/j.tig.2006.06.005.

Björnsdotter, E., M. Nadzieja, W. Chang, L. Escobar-Herrera, D. Mancinotti, D. Angra, H. Khazaei, H. Crocoll, C. Vandenberg, A. Stoddard, F.L. et al. 2020. VC1 Catalyzes a Key Step in the Biosynthesis of Vicine from GTP in Faba Bean.. *bioRxiv.* https://www.biorxiv.org/content/10.1101/2020.02.26.966523v1

Bond, D. A. 1989. A Short Review of Research on Male Sterility and Prospects for F1 Hybrid Varieties in Field Beans (*Vicia Faba* L.). Euphytica. **41**: 87-90. https://doi.org/10.1007/bf00022416.

Bond, D. A. 1982. The Development and Performance of Synthetic Varieties of *Vicia Faba* L. Faba Bean Improvement. **6**: 41-51. https://doi.org/10.1007/978-94-009-7499-9_4.

Bond, D. A. 1987. Recent Developments in Breeding Field Beans (*Vicia Faba* L.). Plant Breeding. **99** (1): 1-26. https://doi.org/10.1111/j.1439-0523.1987.tb01144.x.

Bond, D. A., G. J. Jellis, G. G. Rowland, J. Le Guen, L. D. Robertson, S. A. Khalil, and L. Li-Juan. 1993. Present Status and Future Strategy in Breeding Faba Beans

(*Vicia Faba* L.) for Resistance to Biotic and Abiotic Stresses. Euphytica. **73**: 151–166. https://doi.org/10.1007/bf00027191.

Brown, J. K. M., and M. S. Hovmøller. 2002. Aerial Dispersal of Pathogens on the Global and Continental Scales and Its Impact on Plant Disease. Science **297** (5581): 537–41. https://doi.org/10.1126/science.1072678

Caracuta, V., O. Barzilai, H. Khalaily, I. Milevski, Y. Paz, J. Vardi, L. Regev, and E.Boaretto. 2015. The Onset of Faba Bean Farming in the Southern Levant. Scientific Reports 5 (October): 14370. https://doi.org/10.1038/srep14370

Cernay, C., T. Ben-Ari, E. Pelzer, J.-M. Meynard, and D. Makowski. 2015. Estimating Variability in Grain Legume Yields across Europe and the Americas. Scientific Reports 5 (June): 11171. https://doi.org/10.1038/srep11171

Crépon, K., P. Marget, C.e Peyronnet, B. Carrouée, P. Arese, and G. Duc. 2010. Nutritional Value of Faba Bean (*Vicia Faba* L.) Seeds for Feed and Food. Field Crops Research. **3**: 329-339. https://doi.org/10.1016/j.fcr.2009.09.016.

CropGeneBank.sgrp.cgiar.org, 2020. [online] Available at: https://cropgenebank.sgrp.cgiar.org/images/file/learning_space/molecular_markers/volume2/04_Measures.pdf

Cubero, J. I. I. 2017. Leguminosas hortícolas: guisantes, judías y habas hortícolas in Cultivos Hortícolas al Aire Libre, eds J.V. Maroto and C. Baxauli (Almería: Cajamar Caja Rural), 703–741

Cubero, J. I. 1974. On the Evolution of *Vicia Faba* L. Theoretical and Applied Genetics. **45**: 47–51. https://doi.org/10.1007/bf00283475.

Cubero, J. I., and M.-J. Suso. 1981. Primitive and Modern Forms of *Vicia Faba*. Die Kulturpflanze. **29**(1): 137-145. https://doi.org/10.1007/bf02014744.

Darwin, C. 1870. On The Origin Of Species. New York: D. Appleton and Company

Davis, A. R., R. L. Peterson, and R. W. Shuel. 1988. Vasculature and Ultrastructure of the Floral and Stipular Nectaries of *Vicia Faba* (Leguminosae). Canadian Journal of Botany. **66**(7): 1435-1448. https://doi.org/10.1139/b88-198.

Duc, G. 1997. Faba Bean (*Vicia Faba* L.). Field Crops Research. **53**(1): 99-109. https://doi.org/10.1016/s0378-4290(97)00025-7.

Duodu, Kwaku G., and F. B. Apea-Bah. 2017. African Legumes: Nutritional and Health-Promoting Attributes. Gluten-Free Ancient Grains. Cereals, Pseudocereals, and Legumes: Sustainable, Nutritious, and Health-Promoting Foods for the 21st Century, Woodhead Publishing, Elsevier Ltd. 223-269 https://doi.org/10.1016/b978-0-08-100866-9.00009-1.

El-Ghareib, A., A. Azab, E. Mesbah, and N. Etiwa. 2019. Influence of Sowing Dates and Broomrape Control Methods on Yield and Yield Components of Some Faba Bean Cultivars. Journal of Plant Production. **10**(6): 427-434. https://doi.org/10.21608/JPP.2019.48288.

El-Komy, M. H. 2014. Comparative Analysis of Defense Responses in Chocolate Spot-Resistant and -Susceptible Faba Bean (*Vicia Faba*) Cultivars Following Infection by the Necrotrophic Fungus *Botrytis Fabae*. The Plant Pathology Journal. **30**(4); 355-366. https://doi.org/10.5423/ppj.oa.06.2014.0050.

Ellwood, S. R., H. T. T. Phan, M. Jordan, J. Hane, A. M. Torres, C. M. Avila, S. Cruz-Izquierdo, and R. P. Oliver. 2008. Construction of a Comparative Genetic Map in Faba Bean (*Vicia Faba* L.); Conservation of Genome Structure with *Lens Culinaris*. BMC Genomics 9 (August): 380. https://doi.org/10.1186/1471-2164-9-380

El-Rodeny, W., Eldemery, S., Solimanf, A. and Abdellatif, K. 2020. Investigation of Chocolate Spot and Rust Resistance in Egyptian Faba Bean Population Using Morphological Traits and Molecular Markers. Agrociencia, **54**(2): 15-30.

El-Rodeny, W., M. Kimura, H. Hirakawa, A. Sabah, K. Shirasawa, S. Sato, S. Tabata, et al. 2014. Development of EST-SSR Markers and Construction of a Linkage Map in Faba Bean (*Vicia Faba*). Breeding Science **64** (3): 252–263. https://doi.org/10.1270/jsbbs.64.252.

Evans, L. T. 1959. Environmental Control of Flowering in *Vicia Faba* L. Annals of Botany. **23**(4): 521-546. https://doi.org/10.1093/oxfordjournals.aob.a083675.

FAO. 2020. FAOSTAT. [online] Available at: http://www.fao.org/faostat/en/#data/QC

Flores, F., M. Hybl, J. C. Knudsen, P. Marget, F. Muel, S. Nadal, L. Narits, et al. 2013. Adaptation of Spring Faba Bean Types across European Climates. Field Crops Research. **145**: 1-9. https://doi.org/10.1016/j.fcr.2013.01.022.

Flores, F., S. Nadal, I. Solis, J. Winkler, O. Sass, F. L. Stoddard, W. Link, B. Raffiot, F. Muel, and D. Rubiales. 2012. Faba Bean Adaptation to Autumn Sowing under European Climates. Agronomy for Sustainable Development. **32**: 727-734. https://doi.org/10.1007/s13593-012-0082-0.

Fukuta, N., and T. Yukawa. 1998. Varietal Difference in Snow Tolerance and Growth Characteristics of Broad Bean (*Vicia Faba* L.). Japanese Journal of Crop Science. **67**(4): 505-509. https://doi.org/10.1626/jcs.67.505.

Fyfe, J.s L., and Norman T. J. Bailey. 1951. Plant Breeding Studies in Leguminous Forage Crops I. Natural Cross-Breeding in Winter Beans. The Journal of Agricultural Science. **41**(4): 371-378. https://doi.org/10.1017/s0021859600049558.

Gallais, A. 1992. Pourquoi Faire Des Variétés Synthétiques? Agronomie. **12**: 601-609. https://doi.org/10.1051/agro:19920803.

Garratt, M. P. D., D. J. Coston, C. L. Truslove, M. G. Lappage, C. Polce, R. Dean, J. C. Biesmeijer, and S. G. Potts. 2014. The Identity of Crop Pollinators Helps Target Conservation for Improved Ecosystem Services. Biological Conservation. **169**: 128-135. https://doi.org/10.1016/j.biocon.2013.11.001.

Ghaouti, L. and V.t-K., Werner and Link, Wolfgang. 2008. Development of locally-adapted faba bean cultivars for organic conditions in Germany through a participatory breeding approach: Participatory breeding of faba bean for organic conditions. Euphytica. **162**: 257-268. https://doi.org/10.1007/s10681-007-9603-3.

Gharzeddin, K., F. Maalouf, B. Khoury, L. Abou Khater, S. Christmann, and N. Ali Jamal El Dine. 2019. Efficiency of Different Breeding Strategies in Improving the Faba Bean Productivity for Sustainable Agriculture. Euphytica. **215**:203. https://doi.org/10.1007/s10681-019-2521-3.

Gressel, J., A. Hanafi, G. Head, W. Marasas, A. Babatunde Obilana, J. Ochanda, T.Souissi, and G. Tzotzos. 2004. Major Heretofore Intractable Biotic Constraints to African Food Security That May Be Amenable to Novel Biotechnological Solutions. Crop Protection. **23**(8): 661-689. https://doi.org/10.1016/j.cropro.2003.11.014.

Guo, Z., Y. Dong, K. Dong, J. Zhu, and L. Ma. 2020. Effects of Nitrogen Management and Intercropping on Faba Bean Chocolate Spot Disease Development. Crop Protection. **127**. https://doi.org/10.1016/j.cropro.2019.104972.

Gutierrez, N., C. M. Avila, G. Duc, P. Marget, M. J. Suso, M. T. Moreno, and A. M. Torres. 2006. CAPs Markers to Assist Selection for Low Vicine and Convicine

Contents in Faba Bean (*Vicia Faba* L.). Theoretical and Applied Genetics. **114** (1): 59–66. https://doi.org/10.1007/s00122-006-0410-3

Hanna, A. S., and D. A. Lawes. 1967. Studies on Pollination and Fertilization in the Field Bean (*Vicia Faba* L.). Annals of Applied Biology. **59**(2):289 – 295. https://doi.org/10.1111/j.1744-7348.1967.tb04438.x.

Hanounik, S. B. 1988. New Sources of Resistance in *Vicia Faba* to Chocolate Spot Caused by *Botrytis Fabae*. Plant Disease. 72:696-698. https://doi.org/10.1094/pd-72-0696.

Hanounik, S. B., and G. C. Hawtin. 1982. Screening for Resistance to Chocolate Spot Caused by *Botrytis Fabae*. Faba Bean Improvement. **6**: 243-250. https://doi.org/10.1007/978-94-009-7499-9_25.

Hawtin, G. C. 1982. The Genetic Improvement of Faba Bean. Faba Bean Improvement. **6**: 15-32. https://doi.org/10.1007/978-94-009-7499-9_2.

Herath, I. H. M. H. B., F.L. Stoddard C and Marshall, D. R. 2001. Evaluating faba beans for rust resistance using detached leaves. Euphytica, **117**: 47-57. https://doi.org/10.1023/A:1004071916288

Herridge, D. F., M. B. Peoples, and R. M. Boddey. 2008. Global Inputs of Biological Nitrogen Fixation in Agricultural Systems. Plant and Soil. **311**: 1-18. https://doi.org/10.1007/s11104-008-9668-3.

Holsinger, Kent E., and Bruce S. Weir. 2009. Genetics in Geographically Structured Populations: Defining, Estimating and Interpreting FST. Nature Reviews Genetics. **10**: 639-650. https://doi.org/10.1038/nrg2611.

Hudson RR, Kreitman M, Aguade M. 1987. A test of neutral molecular evolution based on nucleotide data. Genetics **116**(1):153–59

Hurst LD. 2002. The Ka/Ks ratio: diagnosing the form of sequence evolution. Trends Genet, **18**(9):486. https://doi.org/10.1016/S0168-9525(02)02722-1

Ibrahim, H. M. 2015. Effectiveness of Breeding Methods for Production of Superior Genotypes and Maintenance of Genetic Variance in Faba Bean (*Vicia Faba* L.). American Journal of Life Sciences. **3**(1): 11-16. https://doi.org/10.11648/j.ajls.20150301.13.

Kaiser, W. J. 1997. Inter- and Intranational Spread of Ascochyta Pathogens of Chickpea, Faba Bean, and Lentil. Canadian Journal of Plant Pathology. **19**(2): 215-224. https://doi.org/10.1080/07060669709500556.

Kaur, S., R. B. E. Kimber, N. O. I. Cogan, M. Materne, J. W. Forster, and J. G. Paull. 2014. SNP Discovery and High-Density Genetic Mapping in Faba Bean (*Vicia Faba* L.) Permits Identification of QTLs for Ascochyta Blight Resistance. Plant Science. **217-218**: 47-55. https://doi.org/10.1016/j.plantsci.2013.11.014.

Khamassi, K., F. B. Jeddi, D. Hobbs, J. Irigoyen, F. Stoddard, D. M. O'Sullivan, and H.Jones. 2013. A Baseline Study of Vicine–convicine Levels in Faba Bean (*Vicia Faba* L.) Germplasm. Plant Genetic Resources. **11**(3): 250-257. https://doi.org/10.1017/s1479262113000105.

Khan, H. R., W. Link, T. J. Hocking, and F. L. Stoddard. 2007. Evaluation of Physiological Traits for Improving Drought Tolerance in Faba Bean (*Vicia Faba* L.). Plant and Soil. **292**: 205-217. https://doi.org/10.1007/s11104-007-9217-5.

Khazaei, H., D. M. O'Sullivan, H. Jones, N.Pitts, M. J. Sillanpää, P. Pärssinen, O. Manninen, and F. L. Stoddard. 2015. Flanking SNP Markers for Vicine–convicine Concentration in Faba Bean (*Vicia Faba* L.). Molecular Breeding. 35-38. https://doi.org/10.1007/s11032-015-0214-8.

Khazaei, H., R. W. Purves, J. Hughes, W. Link, D. M. O'Sullivan, Alan H. Schulman, E. Björnsdotter, et al. 2019. Eliminating Vicine and Convicine, the Main Anti-Nutritional Factors Restricting Faba Bean Usage. Trends in Food Science and Technology. **91**: 549-556. https://doi.org/10.1016/j.tifs.2019.07.051.

Khazaei, H., W. Link, K. Street, and F. L. Stoddard. 2018. ILB 938, a Valuable Faba Bean (*Vicia Faba* L.) Accession. Plant Genetic Resources: Characterization and Utilization. **16**(5): 478-482. https://doi.org/10.1017/s1479262118000205.

Kimura, M. 1968. Evolutionary Rate at the Molecular Level. Nature. **317**: 624-626. https://doi.org/10.1038/217624a0.

Kohpina, S., R. Knight, and F. L. Stoddard. 2000. Genetics of resistance to ascochyta blight in two populations of faba bean. Euphytica. **112**: 101-107. https://doi.org/10.1023/a:1003853126862.

Lal, K., C. B. Yadav, S. Nath, and D. K. Dwivedi. 2019. Heterotic Responses for Yield and Its Components in Faba Bean (*Vicia Faba* L.). International Journal of Current Microbiology and Applied Sciences. **8**(6): 662-677. https://doi.org/10.20546/ijcmas.2019.806.077.

Landry, E. J., C. J. Coyne, R. J. McGee, and J. Hu. 2017. A Modified Mass Selection Scheme for Creating Winter-Hardy Faba Bean (*Vicia Faba* L.) Lines with a Broad Genetic Base. Euphytica. **213**: 72. https://doi.org/10.1007/s10681-017-1843-2.

Landry, E.J., and J. Hu. 2018. Increasing pre-acclimation temperature reduces the freezing tolerance of winter-type faba bean (*Vicia faba* L.) J. Agron. Crop Sci. **205**:12. https://doi.org/10.1111/jac.12289

Lawes, DA. 1980 Recent developments in understanding, improvement and use of *Vicia faba* In: R.J. Summerfield and A.H. Bunting (eds.), Advances in Legumes

Science. Proceedings of the International Legume Conference, Kew, 31 July-4 August 1978, Royal Botanic Garden, Kew, the Missouri Botanical Garden, and the University of Reading, UK, **108**: 625-636.

Li, Y., P. Ruperao, J. Batley, D. Edwards, J. Davidson, K. Hobson, and T. Sutton. 2017. Genome Analysis Identified Novel Candidate Genes for Ascochyta Blight Resistance in Chickpea Using Whole Genome Re-Sequencing Data. Frontiers in Plant Science **8**: 359. https://doi.org/10.3389/fpls.2017.00359

Link, W., W. Ederer, R. K. Gumber, and A. E. Melchinger. 1997. Detection and Characterization of Two New CMS Systems in Faba Bean (*Vicia Faba*) *. Plant Breeding. **116**(2): 158-162. https://doi.org/10.1111/j.1439-0523.1997.tb02171.x.

Liu, Liting, J. D. Knight, R. L. Lemke, and R. E. Farrell. 2019. A Side-by-Side Comparison of Biological Nitrogen Fixation and Yield of Four Legume Crops. Plant and Soil. **442**: 169-182. https://doi.org/10.1007/s11104-019-04167-x.

López-Bellido, F. J., L. López-Bellido, and R. J. López-Bellido. 2005. Competition, Growth and Yield of Faba Bean (*Vicia Faba* L.). European Journal of Agronomy. **23**(4): 359-378. https://doi.org/10.1016/j.eja.2005.02.002.

López-Bellido, R. J., L. López-Bellido, F. J. López-Bellido, and J E. Castillo. 2003. Faba Bean (*Vicia Faba* L.) Response to Tillage and Soil Residual Nitrogen in a Continuous Rotation with Wheat (*Triticum Aestivum* L.) under Rainfed Mediterranean Conditions. Agronomy Journal. **95**(5): 1253-1261. https://doi.org/10.2134/agronj2003.1253.

Márquez-Sánchez, F. and Sahagún-Castellanos, Jaime. 2002 Synthetic varieties or composite varieties? Maydica. **47**: 103-105.

Maurin, N., and B. Tivoli. 1992. Variation in the Resistance of *Vicia Faba* to *Ascochyta Fabae* in Relation to Disease Development in Field Trials. Plant Pathology. **41**(6): 737-744. https://doi.org/10.1111/j.1365-3059.1992.tb02557.x.

Maxted, N. 1995. An Ecogeographical Study of *Vicia* Subgenus *Vicia*. Systematic and ecographic studies on crop genepool. No 8. International Plant Genetic Resources Institute (IPGRI), Rome

McDonald, J. H., and M. Kreitman. 1991. Adaptive Protein Evolution at the Adh Locus in Drosophila. Nature **351** (6328): 652–54. https://doi.org/10.1038/351652a0

Moose, Stephen P., and R. H. Mumm. 2008. Molecular Plant Breeding as the Foundation for 21st Century Crop Improvement. Plant Physiology **147** (3): 969–77. https://doi.org/10.1104/pp.108.118232

Mousavi-Derazmahalleh, M., B. Nevado, P. E. Bayer, D. A. Filatov, J. K. Hane, D. Edwards, W. Erskine, and M. N. Nelson. 2018. The Western Mediterranean Region Provided the Founder Population of Domesticated Narrow-Leafed Lupin. Theoretical and Applied Genetics. **131** (12): 2543–54. https://doi.org/10.1007/s00122-018-3171-x

Multari, S., D. Stewart, and W. R. Russell. 2015. Potential of Fava Bean as Future Protein Supply to Partially Replace Meat Intake in the Human Diet. Comprehensive Reviews in Food Science and Food Safety. **14**(5): 511-522. https://doi.org/10.1111/1541-4337.12146.

N'Diaye, A., J. K. Haile, K. T. Nilsen, S. Walkowiak, Y. Ruan, A. K. S., F. R. Clarke, J. M. Clarke, and C. J. Pozniak. 2018. Haplotype Loci Under Selection in Canadian Durum Wheat Germplasm Over 60 Years of Breeding: Association with Grain Yield, Quality Traits, Protein Loss, and Plant Height. Front. Plant Sci. **871**:1589. https://doi.org/10.3389/fpls.2018.01589

Ohta, T. 1973. Slightly Deleterious Mutant Substitutions in Evolution. Nature **246** (5428): 96–98. https://doi.org/10.1038/246096a0

Parker, C. 2009. Observations on the Current Status of Orobanche and Striga Problems Worldwide. Pest Management Science **65** (5): 453–59. https://doi.org/10.1002/ps.1713

Pérez-de-Luque, A., M. D. Lozano, M. T. Moreno, P. S. Testillano, and D. Rubiales. 2007. Resistance to Broomrape (*Orobanche Crenata*) in Faba Bean (*Vicia Faba*): Cell Wall Changes Associated with Prehaustorial Defensive Mechanisms. Annals of Applied Biology. **151**(1): 89-98. https://doi.org/10.1111/j.1744-7348.2007.00164.x.

Picard, J., P. Berthelem, G. Duc, and J. Le Guen. 1982. Male Sterility in *Vicia Faba*. Faba Bean Improvement. 53-69. https://doi.org/10.1007/978-94-009-7499-9_5.

Poulsen, M. H., and A. Martin. 1977. A Reproductive Tetraploid *Vicia Faba* L. Hereditas. **87**:123–126. https://doi.org/10.1111/j.1601-5223.1977.tb01253.x

Poulsen, M. H. 1973. The Frequency and Foraging Behaviour of Honeybees and Bumble Bees on Field Beans in Denmark. Journal of Apicultural Research. **12**(2): 75-80. https://doi.org/10.1080/00218839.1973.11099733.

Rajcan, I., J. G. Boersma, and E. J. Shaw. 2011. Plant Genetic Techniques. Comprehensive Biotechnology. **4**: 133-147. https://doi.org/10.1016/b978-0-08-088504-9.00252-x.

Romero, I. G., I. Ruvinsky, and Y. Gilad. 2012. Comparative Studies of Gene Expression and the Evolution of Gene Regulation. Nature Reviews Genetics. **13**: 505-516. https://doi.org/10.1038/nrg3229.

Rowlands, D. G. 1958. The Nature of the Breeding System in the Field Bean (*V. Faba* L.) and Its Relationship to Breeding for Yield. Heredity. **12**: 113-126. https://doi.org/10.1038/hdy.1958.9.

Rubiales, D., A. Pérez-de-Luque, M. Fernández-Aparico, J. C. Sillero, B. Román, M. Kharrat, S. Khalil, D. M. Joel, and C. Riches. 2006. Screening Techniques and Sources of Resistance against Parasitic Weeds in Grain Legumes. Euphytica. **147**: 187-199. https://doi.org/10.1007/s10681-006-7399-1.

Rubiales, D., C. M. Avila, J. C. Sillero, M. Hybl, L. Narits, O. Sass, and F. Flores. 2012. Identification and Multi-Environment Validation of Resistance to Ascochyta Fabae in Faba Bean (*Vicia Faba*). Field Crops Research. **126**(14): 165-170. https://doi.org/10.1016/j.fcr.2011.10.012.

Rubiales, D., M. Fernández-Aparicio, K. Wegmann, and D. M. Joel. 2009. Revisiting Strategies for Reducing the Seedbank of *Orobanche* and *Phelipanches* pp. Weed Research. **49**(s1): 23-33. https://doi.org/10.1111/j.1365-3180.2009.00742.x.

Sabeti, P. C., D. E. Reich, J. M. Higgins, H. Z. P. Levine, D. J. Richter, S. F. Schaffner, S. B. Gabriel, et al. 2002. Detecting Recent Positive Selection in the Human Genome from Haplotype Structure. Nature **419** (6909): 832–37. https://doi.org/10.1038/nature01140

Saïdi, S., M.-H. Ramírez-Bahena, N. Santillana, D. Zúñiga, E. Álvarez-Martínez, A. Peix, R. Mhamdi, and E.Velázquez. 2014. *Rhizobium Laguerreae Sp. Nov*. Nodulates *Vicia Faba* on Several Continents. International Journal of Systematic and Evolutionary Microbiology **64** (Pt 1): 242–47. https://doi.org/10.1099/ijs.0.052191-0

Sallam, A., R. Martsch, and Yasser S. Moursi. 2015. Genetic Variation in Morpho-Physiological Traits Associated with Frost Tolerance in Faba Bean (*Vicia Faba* L.). Euphytica. **205**: 395-408. https://doi.org/10.1007/s10681-015-1395-2.

Satovic, Z., A. M. Torres, and J. I. Cubero. 1996. Genetic Mapping of New Morphological, Isozyme and RAPD Markers in *Vicia Faba* L. Using Trisomics. Theoretical and Applied Genetics. **93**: 1130-1138. https://doi.org/10.1007/bf00230136.

Satovic, Z., C. M. Avila, S. Cruz-Izquierdo, R. Díaz-Ruíz, G. M. García-Ruíz, C. Palomino, N. Gutiérrez, et al. 2013. A Reference Consensus Genetic Map for Molecular Markers and Economically Important Traits in Faba Bean (*Vicia Faba* L.). BMC Genomics 14 (December): 932. https://doi.org/10.1186/1471-2164-14-932

Schlötterer, C. 2003. Hitchhiking Mapping – Functional Genomics from the Population Genetics Perspective. Trends in Genetics. **19**(1): 32-38. https://doi.org/10.1016/s0168-9525(02)00012-4.

Sillero, J. C., C. M. Avila, M. T. Moreno, and D. Rubiales. 2001. Identification of Resistance to Ascochyta Fabae in *Vicia Faba* Germplasm. Plant Breeding. **120**(6): 529-531. https://doi.org/10.1046/j.1439-0523.2001.00653.x.

Sillero, J. C., M. T. Moreno, and D. Rubiales. 2000. Characterization of New Sources of Resistance to Uromyces Viciae-Fabae in a Germplasm Collection of *Vicia Faba*. Plant Pathology. **49**(3): 389-395. https://doi.org/10.1046/j.1365-3059.2000.00459.x.

Sjödin, Jan. 2009. Induced Morphological Variation in *Vicia Faba* L. Hereditas. **67**(2): 155-179. https://doi.org/10.1111/j.1601-5223.1971.tb02371.x.

Smith, J. M., and J. Haigh. 1974. The Hitch-Hiking Effect of a Favourable Gene. Genetical Research **89** (5-6): 391–403. https://doi.org/10.1017/S0016672300014634

Smýkal, P., C. J. Coyne, M. J. Ambrose, N. Maxted, H. Schaefer, M. W. Blair, J. Berger, et al. 2015. Legume Crops Phylogeny and Genetic Diversity for Science and Breeding. Critical Reviews in Plant Sciences **34** (1-3): 43–104. https://doi.org/10.1080/07352689.2014.897904

Stoddard, F. L., A. H. Nicholas, D. Rubiales, J. Thomas, and A. M. Villegas-Fernández. 2010. Integrated Pest Management in Faba Bean. Field Crops Research. **115**(3): 308-318. https://doi.org/10.1016/j.fcr.2009.07.002.

Stoddard, F. L., and D. A. Bond. 1987. The Pollination Requirements of the Faba Bean. Bee World. **68**(3): 144-152. https://doi.org/10.1080/0005772x.1987.11098923.

Stoddard, F. L., C. Balko, W. Erskine, H. R. Khan, W. Link, and A. Sarker. 2006. Screening Techniques and Sources of Resistance to Abiotic Stresses in Cool-Season Food Legumes. Euphytica. **147**: 167-186. https://doi.org/10.1007/s10681-006-4723-8.

Suso, M. J., and F. Maalouf. 2010. Direct and Correlated Responses to Upward and Downward Selection for Outcrossing in *Vicia Faba.* Field Crops Research. **116**(1-2): 116-126. https://doi.org/10.1016/j.fcr.2009.12.001.

Tajima, F. 1989. Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. Genetics **123** (3): 585–95.

Tanno, K., and G. Willcox. 2006. The Origins of Cultivation of *Cicer Arietinum* L. and *Vicia Faba* L.: Early Finds from Tell El-Kerkh, North-West Syria, Late 10th

Millennium B.p. Vegetation History and Archaeobotany. **15**: 197-204). https://doi.org/10.1007/s00334-005-0027-5.

Teagasc. Field Beans - Teagasc | Agriculture and Food Development Authority. https://www.teagasc.ie/crops/crops/research/research-programme/cropquest/field-beans/.

Tian, C., C. Tao Gu, L.Wu, W.Chen, W. Chen, En Tao Wang, J. G. Gu, and T. X. Han. 2009. *Rhizobium Fabae* Sp. Nov., a Bacterium That Nodulates *Vicia Faba*. International Journal of Systematic and Evolutionary Microbiology. **59**(5): 2871-2875. https://doi.org/10.1099/00207713-59-5-1258.

Tivoli, B., A. Baranger, C. M. Avila, S. Banniza, M. Barbetti, W. Chen, J. Davidson, et al. 2006. Screening Techniques and Sources of Resistance to Foliar Diseases Caused by Major Necrotrophic Fungi in Grain Legumes. Euphytica. **147**: 223-253. https://doi.org/10.1007/s10681-006-3131-4.

Torres A.M., N.F.Weeden and A.Martin. 1993. Linkage among isozyme, RFLP and RAPD markers in *Vicia faba*. Theor. Appl. Genet. 85: 937–945. https://doi.org/10.1007/BF00215032

Torres, A. M., B. Román, C. M. Avila, Z. Satovic, D. Rubiales, J. C. Sillero, J. I. Cubero, and M. T. Moreno. 2006. Faba Bean Breeding for Resistance against Biotic Stresses: Towards Application of Marker Technology. Euphytica. **147**: 67-80. https://doi.org/10.1007/s10681-006-4057-6.

Torres, A. M., C. M. Avila, N. Gutierrez, C. Palomino, M. T. Moreno, and J. I. Cubero. 2010. Marker-Assisted Selection in Faba Bean (*Vicia Faba* L.). Field Crops Research. **115**(3): 243-252. https://doi.org/10.1016/j.fcr.2008.12.002.

Torres, A. M., Z. Satovic, J. Canovas, S. Cobos, and J. I. Cubero. 1995. Genetics and Mapping of New Isozyme Loci in *Vicia Faba* L Using Trisomics. Theoretical and Applied Genetics. **91** (5): 783–89. https://doi.org/10.1007/BF00220960

van de Ven, W.T.G., Waugh, R., Duncan, N., Ramsay, G., Dow, N. and Powell, W. 1991 Development of a genetic linkage map in *Vicia faba* using molecular and biochemical techniques. Aspects of Applied Biology, **27**:49-53.

van de Wouw, M., Enneking, D., Maxted, N. and Robertson, La. 1999. Genetic resources of mediterranean *Vicia* species. In: Maxted N, Bennett SJ, eds. Plant genetic resources of legumes in the Mediterranean. Dordrecht: Kluwer, 132–157

Vitalis, R., K. Dawson, and P. Boursot. 2001. Interpretation of Variation across Marker Loci as Evidence of Selection. Genetics **158** (4): 1811–1823. PMC1461744

Vitti, J.J., S. R. Grossman, and P. C. Sabeti. 2013. Detecting Natural Selection in Genomic Data. Annual Review of Genetics. **47**: 97-120. https://doi.org/10.1146/annurev-genet-111212-133526.

Warsame, A. O., D. M. O'Sullivan, and P. Tosi. 2018. Seed Storage Proteins of Faba Bean (Vicia Faba L): Current Status and Prospects for Genetic Improvement. Journal of Agricultural and Food Chemistry. **66**(48): 12617-12626. https://doi.org/10.1021/acs.jafc.8b04992.

Watson, C., Stoddard, F., Carlsson, G., SteenJensen, E. and Murphy-Bokern, D., 2015. European Research on Legume Based Cropping Systems. [online] Donausoja.org.

Webb, A., A. Cottage, T.s Wood, K. Khamassi, D. Hobbs, K. Gostkiewicz, M. White, et al. 2016. A SNP-Based Consensus Genetic Map for Synteny-Based Trait Targeting in Faba Bean (*Vicia Faba* L.). Plant Biotechnology Journal **14** (1): 177–85. https://doi.org/10.1111/pbi.12371

Weir, B. S. 2012. Estimating F-Statistics: A Historical View. Philosophy of Science **79** (5): 637–43. https://www.journals.uchicago.edu/doi/10.1086/667904

Wright, S. 1949. The Genetical Structure of Populations. Annals of Eugenics. **15**(1): 323-354. https://doi.org/10.1111/j.1469-1809.1949.tb02451.x.

Youseif, S. H., F. H. Abd El-Megeed, Amr Ageez, E. C. Cocking, and S. A. Saleh. 2014. Phylogenetic Multilocus Sequence Analysis of Native Rhizobia Nodulating Faba Bean (*Vicia Faba* L.) in Egypt. Systematic and Applied Microbiology **37** (8): 560–69. https://doi.org/10.1016/j.syapm.2014.10.001

Zeid, M., C.Schön, and W. Link. 2003. Genetic Diversity in Recent Elite Faba Bean Lines Using AFLP Markers. Theoretical and Applied Genetics. **107** (7): 1304–14. https://doi.org/10.1007/s00122-003-1350-9

Zhao, J., Z. Wang, H. Liu, J. Zhao, T. Li, J. Hou, X. Zhang, and C. Hao. 2019. Global Status of 47 Major Wheat Loci Controlling Yield, Quality, Adaptation and Stress Resistance Selected over the Last Century. BMC Plant Biology **19** (1): 5. https://doi.org/10.1186/s12870-018-1612-y

Zong, X, X. Liu, J. Guan, S. Wang, Q. Liu, J. G. Paull, and R.t Redden. 2009. Molecular Variation among Chinese and Global Winter Faba Bean Germplasm. Theoretical and Applied Genetics. **118** (5): 971–78. https://doi.org/10.1007/s00122-008-0954-5

2. Chapter 2- Modified Recurrent Selection Scheme (MRSS): A promising method for small-scale breeding of faba beans

2.1. Abstract

Maintaining seed yield under a niche agro-climate requires use of a targeted breeding scheme. The aim of this chapter was to create and test a "*sur mesure*" breeding programme, for the peculiar Irish weather conditions, promising to improve seed yield. In a four-year single-operator project, a diverse faba bean germplasm of 219 F_1 hybrids, consisted of the high-yielding spring varieties 'Fanfare' and 'Vertigo', sources of seed quality, disease resistance and some diverse inbred lines, created the initial Foundation population. This Foundation population entered a Modified Recurrent Selection Scheme (MRSS) consisting of a sequence of outcrossing generations at two experimental sites (site one: UoR in the UK and site two: Teagasc in Carlow, Ireland)- of two generations per year (winter GH and spring cage) at site one and one generation per year (spring cage) at site two, respectively. The selection generations were confined to prevent ingress of pollen from external sources and captive bumblebees were ensuring high outcrossing strictly among the individuals of each generation. The top 3.7% yielding individual plants were selected at the end of each spring cage cycle and the chosen individuals were advanced to the next cycle. In the fourth year, coordinated yield trials were established in both sites testing the in-between selected populations along with high yielding market varieties ('Fanfare', 'Vertigo' and 'Lynx'). In these field trials, the outcrossing populations that were tested at UoR showed yield progress of 13.9% (after two selection cycles) and 15.2% (after three selection cycles) over the founding population. Furthermore, the latest UoR selection was the highest yielding entry in the UoR evaluation trial and competitive with the market varieties in the Irish trial.

2.2. Introduction

Despite the low level of faba bean production in Ireland in recent decades, Ireland's temperate maritime climate (as for other arable crops) permits a wide range of sowing dates, a long growing season and high yield potential. In view of renewed interest in sustainable protein cropping, Ireland's Food and Agriculture Development Authority (Teagasc) has conducted agronomic trials aimed at supporting increased production of faba bean with appropriate advice on variety performance and agronomy. One of the most remarkable findings from the first years of this programme was that the leading so-called 'spring' variety of the time 'Fuego' sown in late autumn can out-yield all 'true' winter varieties (Figure 2.1.), suggesting that the seasonal ideotypes developed by continental and UK breeders may not apply in the Irish context. The second notable observation from these trials was that the main deficiency of 'Fuego' when it was out-performed by winter varieties from autumn sowing was its comparative susceptibility to fungal disease in the mild, wet Irish winter conditions (Figure 2.1.). These observations led to the conclusion that the unique climatic conditions prevalent in Ireland required a new, bespoke ideotype combining characteristics previously found only separately in winter and spring genepools.

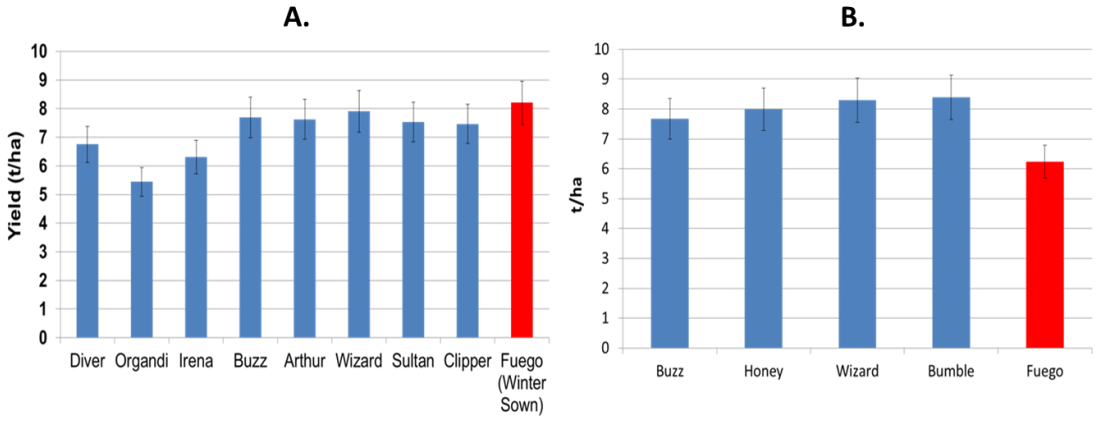


Figure 2.1. Yield performance of commercial faba bean varieties in **A.** 2013-14 (sown on 12th of November) and **B**. 2014-15 (sown on 29th of October) late autumn-sown trials, characterised by low (**A**) and high (**B**) disease pressure respectively [John Carroll (Teagasc), pers. comm.].

Current advice given by Teagasc to Irish bean growers reflects this confusing and unsatisfactory situation, stating:

"*While spring sowing type varieties dominate [the Irish field bean seed market]; and can be sown as early as November, winter varieties for October/November only sowing are also available.*"

This study was designed to develop and test a faba bean breeding scheme capable of selecting a new Irish-adapted ideotype permitting optimal yield in late autumn-sown Irish growing conditions to meet the demand for high-yielding varieties that can be sown from November to January where heavy soils make spring establishment difficult or impossible. Three main elements were taken into consideration in designing the new breeding scheme: the specificities of faba bean reproductive habits and breeding technologies available, technical and financial resources available, and available germplasm suited to the breeding objectives and these will be introduced in turn.

2.2.1. Breeding options for faba bean

As discussed in Chapter 1.3.2, the allogamous nature of faba bean makes traditional forms of pure line/pedigree breeding unwieldly and expensive due to the need to prevent uncontrolled cross-pollination. Pure line breeding does not exploit heterosis, which can be very strong in faba bean (Bond *et al.*, 1964; Zeid *et al.*, 2004). We can also ignore doubled haploidy and hybrid breeding as these technologies have not been successfully developed for faba bean. Genotyping technologies that might permit genomic selection (GS) had just become available at the outset of this project (Webb et al, 2016), but it was considered that GS is a method that can enhance the efficiency of an established selection scheme whereas the need being addressed was to set a breeding scheme up from scratch rather than to enhance an established scheme with genomic prediction of phenotype.

In reality, most commercial breeding is of synthetic varieties, which involves separate steps of line breeding followed by creation of intermating populations of a fixed number of lines and selection of those intermated 'synthetic' populations showing the best agronomic performance, so this was considered as a potential option. Recurrent selection is proposed as one of the most efficient and dynamic procedures for partially allogamous species under open pollinating conditions (Hallauer and Darrah, 1985). In faba beans, although in the past there were several reports highlighting the benefits of recurrent selections and its results in breeding (McVetty and Nugent-Rigby, 1984; Rowland, 1987; Bond *et al.*, 1993), the method seems to have fallen out of favour- in academic circles at least-, as no recent reports were found in northern Europe, and particularly in Ireland and UK, of recurrent selection in faba beans. Nonetheless, for its simplicity, recurrent selection was also given consideration as an option.

Finally, mass selection, being the simplest of all breeding scenarios, having been practised since the dawn of agriculture by simply re-sowing what survived the previous

season, was considered, although it has the drawbacks of relying completely on natural processes of pollinator-mediated outcrossing at natural rates to recombine genotypes, and purely on natural selection to drive gains in yield.

2.2.2. Technical and financial resources available

Considering the small size of the current field bean seed market in Ireland, commercial breeding is restricted to a process of screening varieties developed for other nearby markets, principally the UK. As with many under-utilized crops, the most likely source of investment to allow a minor or under-utilized crop to breakthrough into levels of profitability and uptake that would encourage private investment is the public purse (Stamp *et al*, 2012). Concerning faba bean in Ireland, such an investment was made in the form of a 5-year project funded by the Department of Agriculture, Food and the Marine under its Research Stimulus programme entitled the Virtual Irish Centre for Crop Improvement (VICCI). In recognition of policy objectives to increase agrobiodiversity and reduce the current protein production deficit, and the aforementioned issues surrounding the suitability of imported varieties for the Irish climate, a 4-year PhD project tasked with exploring the possibilities for a small-scale breeding programme was included in the VICCI programme. The dimensions of this investment – one full-time researcher with a four-year reporting time horizon, and a commensurate budget for research expenses were important in narrowing down the options considered in designing the new scheme. In effect, simplicity of operation was key, since all crossing, propagation, field nurseries phenotyping and seed handling had to be carried out by a single individual. Thus, a synthetic breeding scheme, which relies substantially on manual crossing in the pedigree phase and requiring upwards of 7 years to complete a

single complete cycle, was ruled out. This left recurrent selection, mass selection or variations thereof as the main options.

2.2.3. Selection of germplasm to meet breeding objectives

The foundation of any breeding program focused on production targets must be elite common knowledge varieties proven through extensive replicated trials to be competitive in terms of their yield in the target environment. Elite adapted material will have flowering times and durations that allow the growth cycle to make optimum use of the local seasonal distribution of rainfall, light and warmth as well as having adequate tolerance to prevalent biotic and abiotic stresses. The UK and Ireland both run Recommended List trials – in the UK, these are organised by the non-statutory levy body the Processors and Growers Research Organisation (PGRO) and in Ireland by the Irish Department of Agriculture, Food and the Marine (DAFM), whose purpose is to provide yield and agronomic performance data on the best current faba bean varieties. Agronomic performance data includes data on tolerance to the most prevalent pests and diseases. Every year, the recommended list (RL) of bean varieties is updated to add superior new varieties and remove those which have been outclassed.

At the outset of this project, in 2016, nine varieties were listed in the UK Spring RL (PGRO, 2016) and three in the Irish RL (DAFM, 2017). Pale hilum 'Vertigo' and 'Fanfare' were the highest yielding varieties in the UK at 103 and 102% of control yield respectively, while there was a new entrant in its first year of provisional recommendation, 'Lynx', which matched 'Fanfare's' 102% of control yield but with better levels of resistance to downy mildew. 'Boxer', 'Fury', 'Fuego', 'Pyramid' and 'Babylon' were the less well performing pale hilum varieties (93-98% of control yield) with 'Maris Bead', an old, black hilum tick (small-seeded) variety completing the UK

RL. In Ireland in 2017 (DAFM, 2017), ‘Fanfare’ was the highest yielding variety listed at 104% of control yields, which together with ‘Boxer’ (102%) ‘Fuego’ (97%) were the only varieties listed. ‘Fanfare’ was the most resistant to downy mildew in Ireland, but none of the three Irish-listed varieties were resistant to chocolate spot or rust.

By 2019 in the UK (PGRO, 2019), ‘Boxer’, ‘Fury’ and ‘Babylon’ had been removed from the RL and the pale hilum ‘Mallory’, ‘Cartouche’, ‘Victus’ and ‘Tiffany’ were introduced. The nine varieties were composing the UK RL list on 2019 in descending yield performance order were the pale hilum ‘Lynx’, ‘Victus’, ‘Fanfare’, ‘Vertigo’, ‘Mallory’, ‘Tiffany’, ‘LG Cartouche’, ‘Fuego’ (ranging 99-104% of control), and finally the less well yielding (82% of control) ‘Maris Bead’. ‘Lynx’ is reported to have the best standing ability to harvest in both years (2016 and 2019). In Ireland, by 2019 (DAFM, 2019), ‘Fuego’ was removed from the RL and ‘Lynx’ was added, having the highest reported yield (106% control). ‘Boxer’ and ‘Fanfare’ continued being in the RL, both yielding 101% of controls.

Genetic variation is another essential ingredient of any selection scheme. Genetic resources capturing variation in morphology, seed size, seed composition, yield, adaptation to abiotic stresses and pests should all be considered for inclusion in the founding genepool of a breeding programme. Genetic resources held at the University of Reading included the materials listed in Table 2.1 and as these included seed size, seed quality and disease resistance as well as collectively encompassing a wide diversity of ecogeographic origins, the pragmatic decision was taken to base the selection of non-elite germplasm on working collections near to hand.

However, more generally when setting up a completely new breeding programme, consideration should be given to collections of plant genetic resources (PGR) conserved in agricultural genebanks. Genes impacting adaptation to biotic and abiotic stress factors

may be found in these collections and are always attractive to breeders. As the quantity of germplasm is likely to be far greater than actually required, pre-screening that PGR for the purpose in hand is the first step for a successful pre-breeding program. As climate change is a current and future challenge, maintaining the biodiversity of PGR is essential for both agriculture and environment. The focused identification of germplasm strategy (FIGS) is a method of exploring germplasm diversity for climate-adaptive traits and for one such target - drought tolerance- the FIGS approach has been used successfully to draw a manageable selection of materials for detailed screening from a huge gene bank catalogue (Khazaei *et al.*, 2013). ICARDA holds the largest collection of faba bean genetic resources globally (Maalouf, 2011).

2.2.4. Definition of the MRSS

As discussed in Chapter 1, recurrent selection could be an ideal scheme however, given the tight timeframe of the PhD that would be limited to only spring cage outcrossing cycles. Taking advantage of the winter outcrossing in the glasshouse, as an outcrossing and not selection step, what was proposed, finally, is a scheme which has elements of both mass selection and recurrent selection, which will be referred to from hereon in as Modified Recurrent Selection Scheme (MRSS). The MRSS proposed here is based on the following conjectures that run against conventional wisdom:

1 - **outcrossing should be maximised, not minimised**. Outbreeding causes both recombination – needed to create novel genotypes on which selection can act, and heterosis – which boosts yield, but is not heritable (except by clonal propagation) and at the same time masks locus-specific genetic merit. However, in the absence of doubled haploid technology to accelerate inbreeding, the time and effort required to obtain sufficiently inbred material to obtain reliable phenotypes that reflect heritable genetic

merit may be outweighed by the ability to select continuously without the need to follow the selective steps with any specific further measures to exploit heterosis.

2 – **single plant seed yield, whilst individually error-prone and unreproducible, is nonetheless an acceptable proxy for whole population yield** when applied repeatedly to the whole population. Again, taking into account the low seed multiplication rates obtained in faba bean and low/variable rates of selfing if not manually tripped, the effort to make faba bean 'conform' to the demand that phenotypic selection is based only on homozygous plant populations may outweigh the benefits.

The combination of these two conjectures inspired the MRSS design illustrated in Figure 2.2.

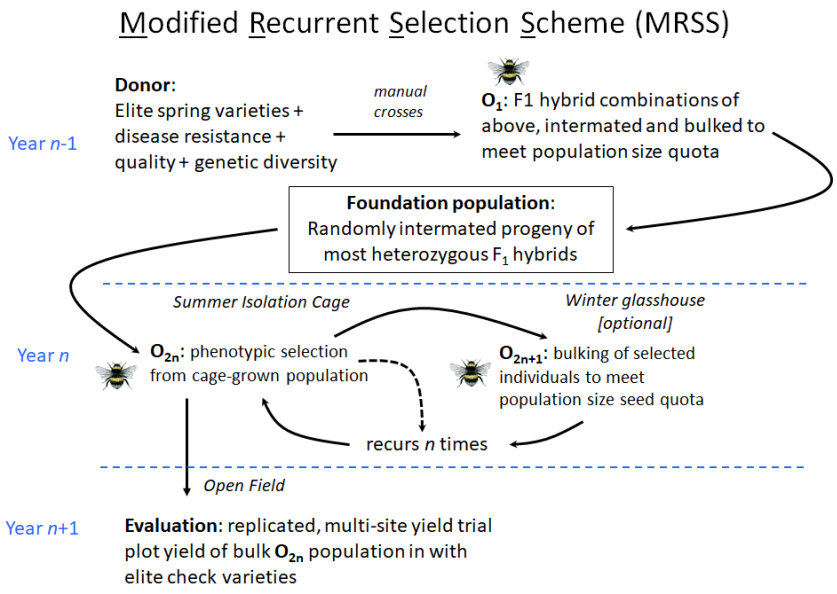


Figure 2.2. Schematic representation of the generic Modified Recurrent Selection Scheme. The scheme has only outcrossing steps. Generation names are given in bold and are joined by arrows from the beginning to the end of the scheme to indicate the sequence of steps involved. A bumblebee symbol is included to indicate steps where captive colonies are confined with the population to drive high rates of outcrossing.

Briefly, following creation of a founding population, which will be composed of the progeny of randomly intermated F_1 hybrid individuals created from a collection of

donor individuals which capture all the traits to be incorporated in the desired ideotype. Where breeding for spring cropping, each round of selection involves two generations per year – a summer (April to September) pollinator exclusion cage where the entire population is grown in isolation with captive bumblebee colonies to drive outcrossing and where phenotypic selection is conducted on the basis of individual plant seed yield.

An equal number of seed from each of the selected individuals[1] is grown on in the subsequent winter glasshouse (October – March) generation whose purpose is mainly to provide a chance for further outcrossing amongst selected genotypes while simultaneously generating enough seed to have a large enough seed bulk for the subsequent spring cycle but non-field based screens could be incorporated here (e.g. pathotest, marker-based selection). When selecting for long season (autumn-sown) conditions, as is the case for the Irish-targeted ideotype, the winter bulking step must be omitted. After this process has been repeated for *n* generations, yield of the bulk population from the last selective generation, remnant bulks of previous selection generations and all relevant checks can be evaluated side-by-side in replicated field trials. The operation is simple, low cost and can be readily scaled to the extent of winter glasshouse available. Selective gains are made in each growing season and immediately recombined.

This Chapter will describe the establishment of recombined foundation population generated by intermating a set of faba bean lines that collectively possess all the required characteristics. The details of how the generic MRSS described above was implemented to produce two parallel RS populations, one targeting a spring-sown UK-adapted type cycling through two generations per year in Reading, and the other targeting an autumn-

[1] to ensure that all selected genotypes contribute equally to the next generation, rather than allowing genotypes whose seed yield derives from a large number of small seeds from predominating over genotypes whose seed yield derives from a small number of large seeds.

sown ideotype going through one generation per year in Oakpark (Ireland) will then be presented. Three primary hypotheses in relation to this two-pronged scheme were tested:

1. RS would produce an initially steep genetic gain in yield compared with the performance of the founding population.
2. If pursued for a sufficient number of cycles, the performance of the selected populations should exceed the performance of the best progenitor.
3. Since sowing dates and key climatic variables were quite different between the two selection sites, selections made in the Reading environment should out-yield Oakpark selections from the same founding population when tested at Reading and vice versa.

Since the breeding strategy was based on a closed, but freely intermating, population, check varieties could not be included within the selection environment (cage) in order not to dilute the pollen population with 'historic' genotypes and therefore it was not possible to compare aggregate or individual plant yields across generations or between sites. Genetic progress for yield and site-wise comparisons were therefore assessed retrospectively in the final year of the project in replicated open field plot trials comparing yield and general agronomic characteristics of final selections, remnant seed from intermediate selections and the foundation population, as well as elite progenitor and current recommended varieties.

2.3. Materials and Methods

2.3.1. Genetic resources for the MRSS

Prior to the start of this project, a set of faba bean lines (Table 2.1.), coming from a variety of sources and collectively possessing the characteristics required to establish a founding genepool, from which better adaptation to late autumn sowing under Irish maritime

climate along with many more desirable traits can be selected, was collected and grown in the glasshouse. As Table 2.1 shows, lines included the zero tannin and low vicine traits associated with better animal feed quality (Crépon et al. 2010), neither of which are currently available in the Irish RL varieties, sources of disease resistance, as well as some less relevant traits like crimson flower or yellow testa and exotic origins that signify potential generic sources of unique genetic diversity that might drive high levels of heterosis.

A large number of intercrosses were made amongst the members of this founding genepool, mainly to ensure that elite spring varieties 'Fanfare' and 'Vertigo' were paired in as many combinations as possible with sources of quality, disease resistance and exotic material. F_1 hybrid seed from *x* unique crossing combinations were planted to establish the Founder population which was grown in outcrossing conditions and is referred to in the MRSS as the first outcrossing (O_1) generation. The initial set of outcrosses was not a full diallel but rather a targeted selection of pairings designed to favour elite spring material over diverse, poorly adapted and mostly inbred material (Figure 2.3.).

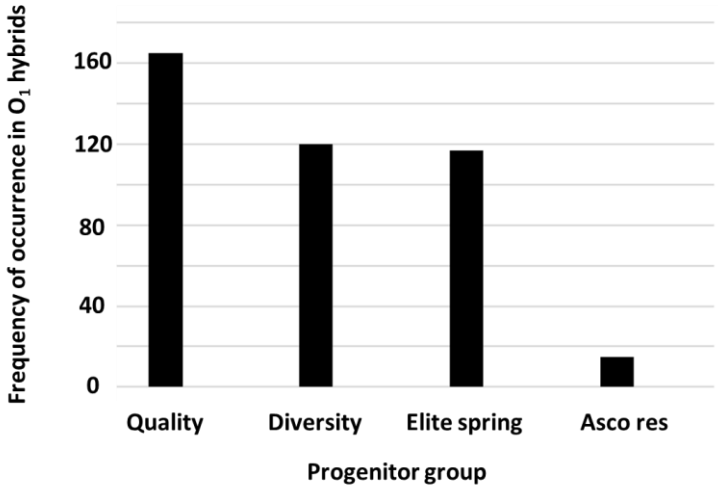


Figure 2.3. Contribution of each main progenitor group to the O_1 foundation population was calculated by assigning each parent of the 219 O_1 hybrid individuals to one of 5 categories.

Table 2.1. List of lines/varieties intercrossed to generate the MRSS Foundation Population showing origin of the line/variety and any notable traits or characteristics. Institutional abbreviations used in this table: University of Helsinki (UoH), Georg August University Gottingen (GAUG), Agricultural Research Center (ARC) in Egypt.

Source ID	Accession Code	Donor	Germplasm Status	Unique Trait	Flower Colour
Fanfare	-	LS Plant Breeding NPZ-Lembke (DE)	cultivar	RL	Normal
Vertigo	-	LS Plant Breeding NPZ-Lembke (DE)	cultivar	RL, low tannin	Normal
Hedin	NV639	GAUG	inbred line from cultivar	highly autofertile, highly inbred	Normal
Albus	NV643	NIAB	inbred line from cultivar	zero-tannin *(zt1)*	White
Kasztelan	NV644	NIAB	inbred line from Polish cultivar	quality	White
BPL10	NV648-1	ICARDA	inbred line	quality	Normal
INRA 29H	NV657	INRA	inbred line	ascochyta resistance	Normal
Mélodie	NV735	UoH	inbred line from cultivar	LVC	Normal
Betty	NV866	Betty (AFP 33/229)	inbred line from cultivar	LVC	Normal
Disco/2	NV866	NIAB	inbred line from cultivar	zero-tannin (*zt2*), LVC	White
F5 from NV644x NV153	NV873-13 (F4)	NIAB	recombinant inbred line	dwarf	White
HEL170	RV322	UoH	Chinese inbred line	early-flowering	Normal
Casata midwinter	RV503	UK broad bean grower selection	inbred line from cultivar	diversity	Crimson
Crimson Flowered-3	RV504	UoR germplasm collection	inbred line from heirloom cultivar	diversity	Crimson
Cuscan Super Yellow-1	RV506	UoR germplasm collection	partially inbred line from landrace	diversity, yellow testa	Normal
Iantos-3	RV507	UoR germplasm collection	partially inbred line from landrace	diversity, yellow testa	Normal
Mustard Yellow	RV508	UoR germplasm collection	partially inbred line from landrace	diversity	Normal
Sakha	RV509	ARC, Egypt	inbred line from cultivar	diversity	Normal
Nubaria	RV510	ARC, Egypt	inbred line from Spanish cultivar	tall cultivar, low tannins	Normal
Misr3	RV511	ARC, Egypt	inbred line from cultivar	quality	Normal
Giza716	RV512	ARC, Egypt	inbred line from cultivar	quality	Normal

2.3.2. Implementation of the MRSS

In total, ten experiments took place (Figure 2.4.). At the University of Reading (site number one) two generations per year, one overwinter glasshouse and one spring-summer cage experiment, were established annually. At Teagasc, Carlow (site number two), two parallel cage experiments were established each year. The locations of the experiments in both sites are given in Figure 2.5.

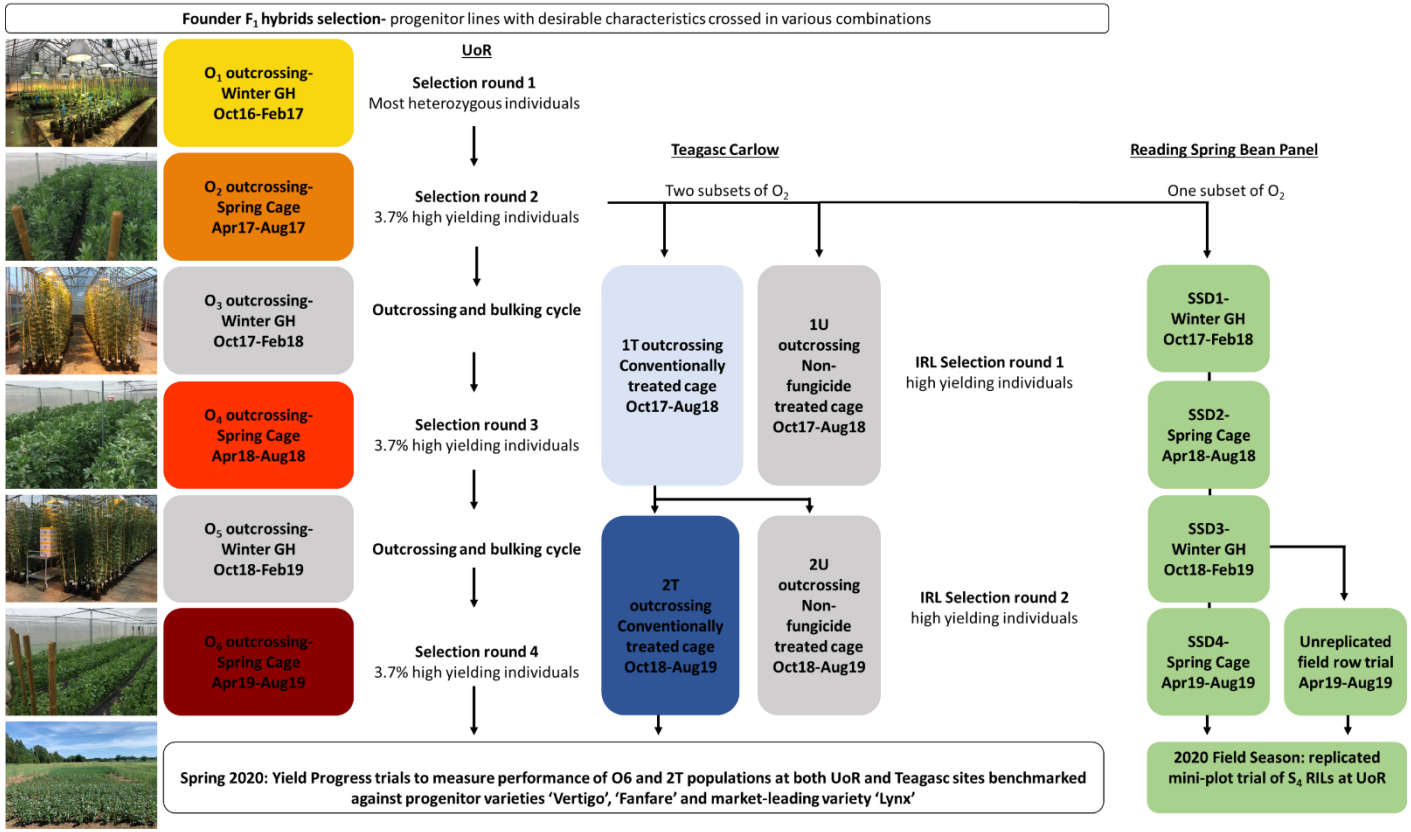


Figure 2.4. Implementation of the MRSS scheme showing how the foundation population was created and chronology of selections carried out in parallel at two sites (UoR and Teasgasc, Carlow). As part of a separate PhD project (Ahmed Warsame), a subset of the O_2 population has been put through three rounds of SSD with the aim of creating a panel of RILs – termed the Reading Spring Bean Panel - that represent the genetic diversity segregating in the MRSS. The relevance of this parallel project is explored in Chapter 5.

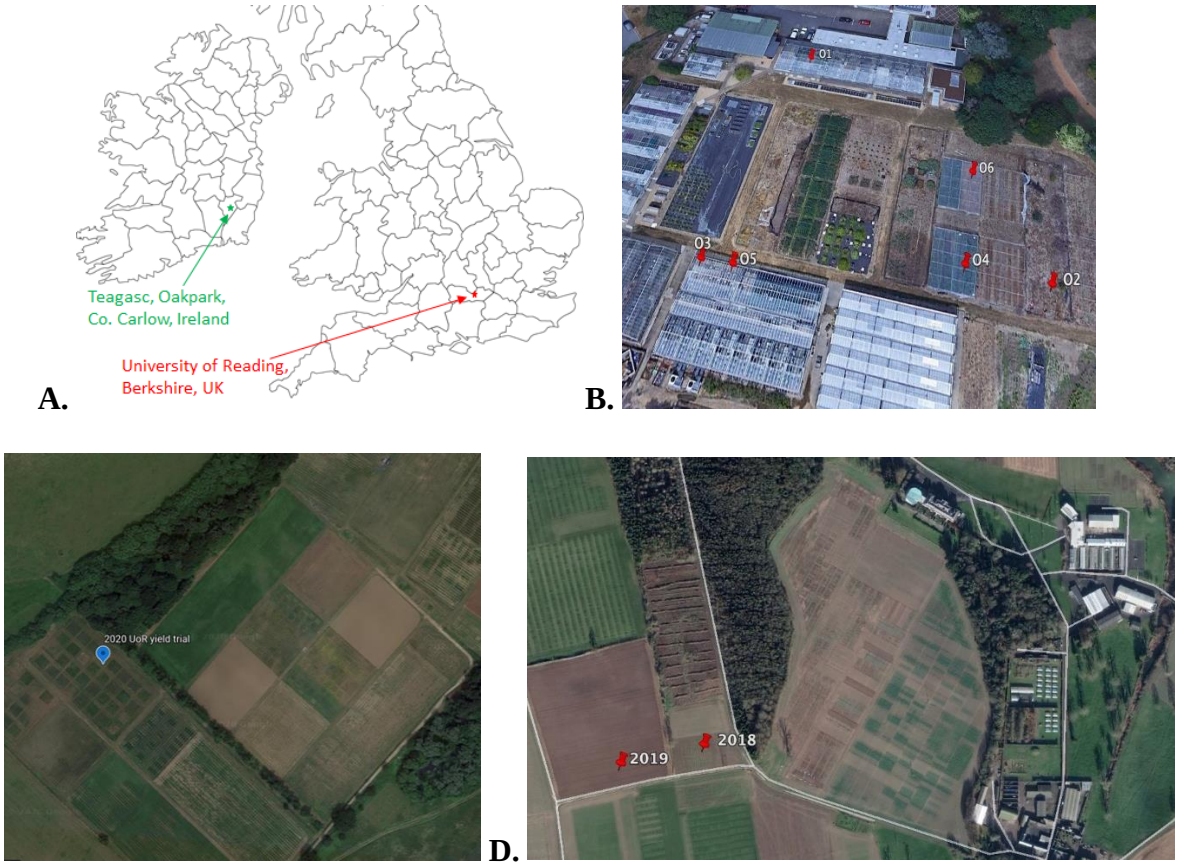


Figure 2.5. Locations of the experiments. **A.** Map of UK and Ireland showing geographical locations of the sites. **B.** and **C.** Google Earth images of the UoR Crop and Environment Laboratory and Sonning Farm sites showing locations of individual experiments. O_1, O_3 and O_5 are locating the glasshouse experiments and O_2, O_4, O_6 show the cage experiments. **D.** Google Earth image showing locations of the 2018 (1T, 1U) and 2019 (2T, 2U) experiments in Teagasc, Carlow.

2.3.2.1. Conduct of selection cycles and selection criteria

In mid-October 2016, a collection of 219 F_1 hybrids coming from crosses between the listed lines in Table 2.1 was assembled to form the O_1 Foundation Population. The O_1 population was grown over-winter in the glasshouse with outcrossing promoted by both bee-mediated pollination and manual crossing, where needed to ensure cross-pollination of asynchronous lines. In order to confirm hybridity and assess overall diversity, DNA was extracted from each individual plant from this recombining generation with a modified CTAB method (Fulton, Chunwongse, and Tanksley, 1995), carefully quantitated and genotyped using 40 highly informative (Polymorphic Information Content PIC>0.4), well-distributed SNP markers (Figure 2.6.). Since more F_1s were put into the Founding Population O_1 than were strictly needed to fulfil the seed requirement for the first selective generation O_2, this SNP profile was also used as a criterion to select the most heterozygous individuals whose progenies would constitute the O_2 population.

Individuals composing the summer selective generations (O_2, O_4 and O_6) were randomly distributed across four blocks within the isolation cage. All cages in Reading were irrigated and treated according to best local management practices (Appendix B). Selection in the O_2, O_4, O_6, 1T and 2T generations were conducted using the same method. Plants were individually harvested at maturity, dried, threshed and counted. The seed weight and seed number were recorded against each individual plant ID and a selection threshold applied to the list ranked by Individual Plant Seed Yield (IPSY) identify the top 3.7% individuals whose progeny would make up the next generation. Two subsets of the bulked seed from the O_2 generation were created to establish field-based selection in Ireland (Teagasc, Oak Park, Co. Carlow) over the 2018 and 2019 seasons. Two pollinator exclusion cages (with same size and layout with the ones at the UoR, but without supplementary irrigation) were constructed for selection at Oakpark. Each year, one cage was treated (T) as per standard commercial faba

bean cultivation practice while fungicide treatments were withheld from the untreated (U) cage in order to exercise selection under enhanced disease pressure. However, as no significant disease was present in either 2018 or 2019, the idea of conducting parallel section under high disease pressure was abandoned and no data from 1U or 2U will be presented here.

A 3.7% selection intensity (SI) was decided for the first selection round O_2 as being the most stringent selection that would still give enough seed to supply O_3, 1T and 1U and this SI was strictly adhered to in all further cage generations. Equal numbers of seed of each selected genotype were always advanced to the next generation. No selection was practiced in the overwinter O_3 and O_5 bulking generations, but captive bumblebees were present throughout flowering and an equal number of progenies from each plant was pooled to form the bulk population for the subsequent selective generation. Further details of glasshouse management are presented in Appendix A.

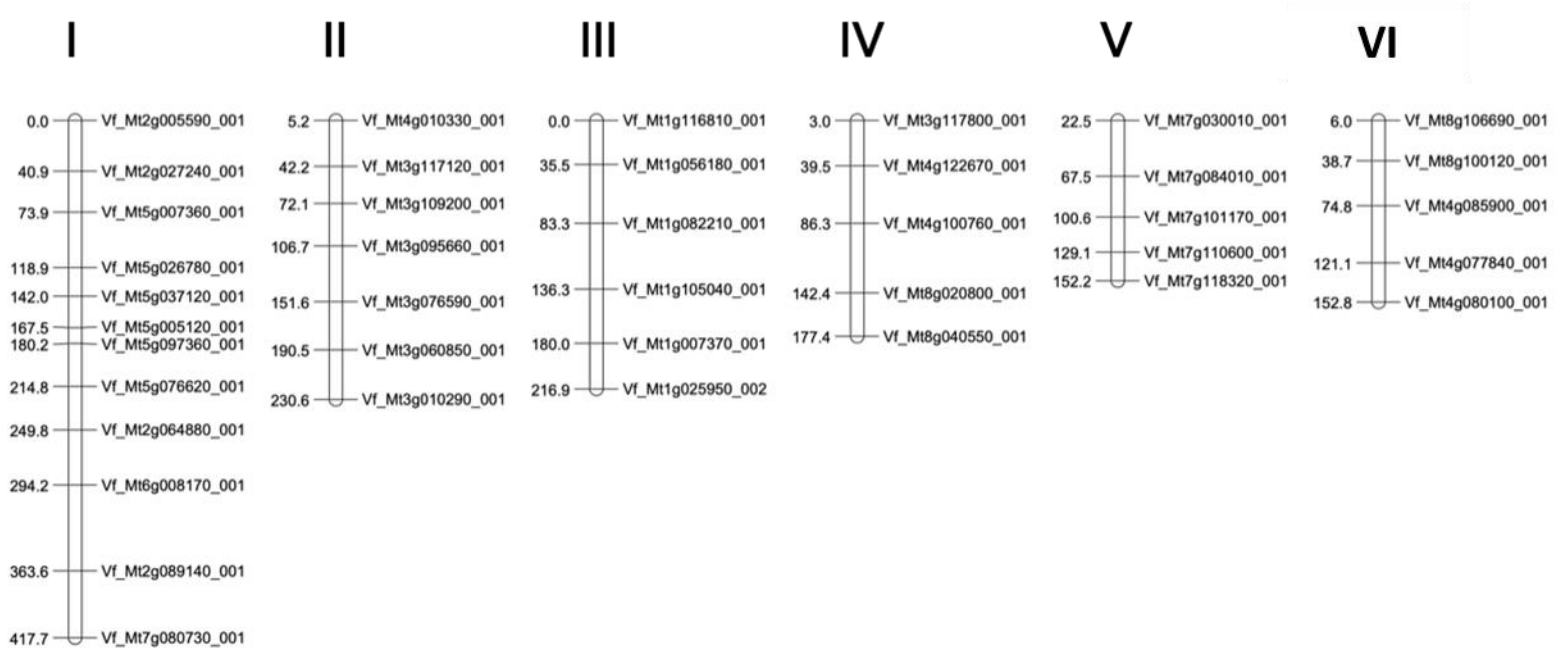


Figure 2.6. Distribution of the 40 SNP markers used to assess diversity and heterozygosity in the O_1 generation across the six *Vicia faba* chromosomes.

2.3.3. Yield Progress field trial

In spring 2020, two co-ordinated field experiments, one at the University of Reading's arable trial ground at Sonning Farm and one at the Teagasc Oakpark Crops Research Centre in Carlow, were conducted to assess the real gain through the breeding scheme.

The MRSS selection generations had not included any check varieties as the inclusion of populations of check varieties would contaminate the genepool through pollen transfer from checks to breeding population and thus defeat the purpose of the selective strategy. These evaluation trials were therefore the only opportunity to judge the progress or not of seed yield through the selection cycles and the adaptation of the two final selections (O_6 and 2T) in the respective environments in which they had been selected.

2.3.3.1. Experimental design of 2020 Yield Progress trial

Both trials at UoR and at Teagasc, Carlow followed a Randomised Complete Block Design (RCBD). Randomisation within blocks was done using the OPTEX Procedure that searches for optimal experimental designs in SAS/QC (Atkinson, Donev, and Tobias 2007).

Entries for the UoR trial consisted of remnant seed of a subset of the Foundation population O_1, two UoR selections (O_4 and O_6), the latest Teagasc selection (2T) and 'Fanfare', 'Vertigo' and 'Lynx' as check varieties (supplied by LSPB, Cambridge, UK). Market varieties 'Fanfare', 'Vertigo' and 'Lynx' were replicated twice in each block while the selection genotypes O_1, O_4 and O_6 were single entries in each block. The trial was drilled on 7th of April 2020.

At Teagasc the trial consisted of the O_6 selection, 1T, 2T, 'Fanfare' and 'Lynx' (supplied by Seedtech, Waterford, IRL), 'Vertigo' 'Fanfare' and 'Lynx' (supplied by

LSPB, Cambridge, UK) and the winter variety 'Wizard'. Figure 2.7. A and B present the layout of the two field experiments. The trial was drilled on 20th of March 2020. Five entries (O_6, 2T, 'Fanfare' UK, 'Vertigo' and 'Lynx' UK) were common in the two sites.

Where the same variety had been obtained from different sources, the entries in these trials are henceforth referred to with the suffix _UK or _IRL depending on whether they were sourced from the UK agent (LSPB) or the Irish agent (Seedtech). This could be of importance as preservation of optimum heterosis during seed multiplication of synthetic varieties depends on outcrossing driven by natural pollination services in each seed crop, which could vary from location to location (Wright, 1977). Counted packets of cleaned and germination-tested seed of all entries were drilled at a sowing density of 40 seeds/m^2.

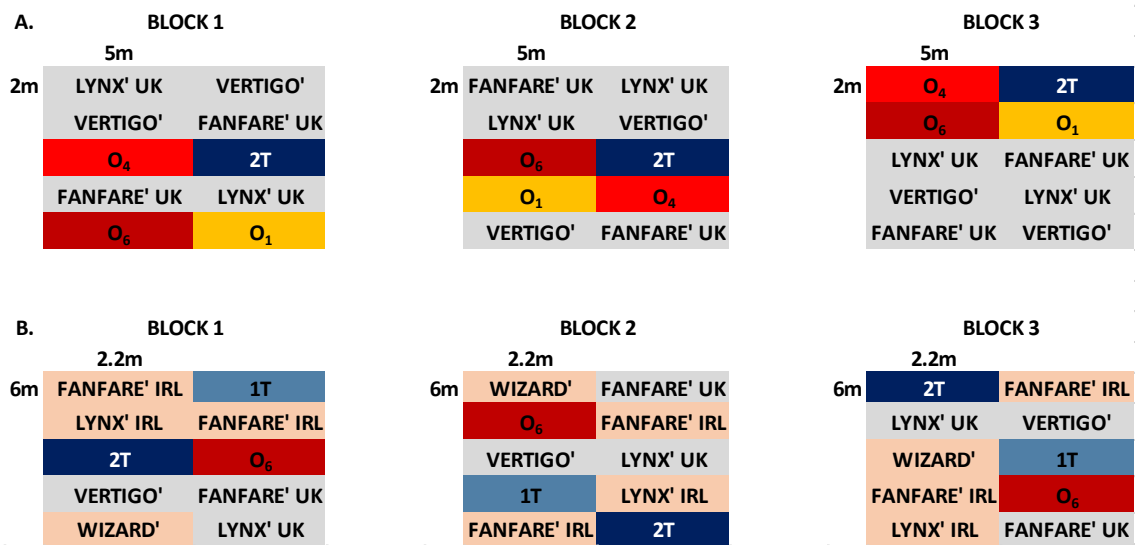


Figure 2.7. RCBD design of Yield Progress field trials. **A.** Experimental layout at UoR Sonning University Farm. In each block each population is represented once and each check variety twice. The plots have size of 2m x 5m. **B.** Experimental layout at Teagasc, Carlow. In each block each population and most checks are represented once and the check variety 'Fanfare IRL' twice. The Plot dimensions were size of 2.2m x 6m.

2.3.3.2. Plant population and Green Area Index

At UoR, the total number of fully germinated plants in each plot (the plant population) were counted, while at Teagasc the germinated plants in $2m^2$ of each plot were recorded to give a comparable estimate of plant population.

At UoR, additional measures of canopy growth were taken through the season using a mobile phenotyping cart (the Phenocart) to capture changes in Green Area Index (GAI) over time (Figure 2.8.). Two Red-Green-Blue (RGB) images per plot were captured weekly throughout the growing season using the Phenocart. The images were taken using Canon EOS 6D camera at a fixed height on the Phenocart overhead camera arm. Custom software (PhenoHarvest) were used to count green pixels in the raw images and return the proportion of green area in the two sub-plot images which were averaged to obtain the plot GAI (Figure 2.9.).

Figure 2.8. Phenocart consists of two mountain bikes welded to a steel frame that holds the camera and gps sensor.

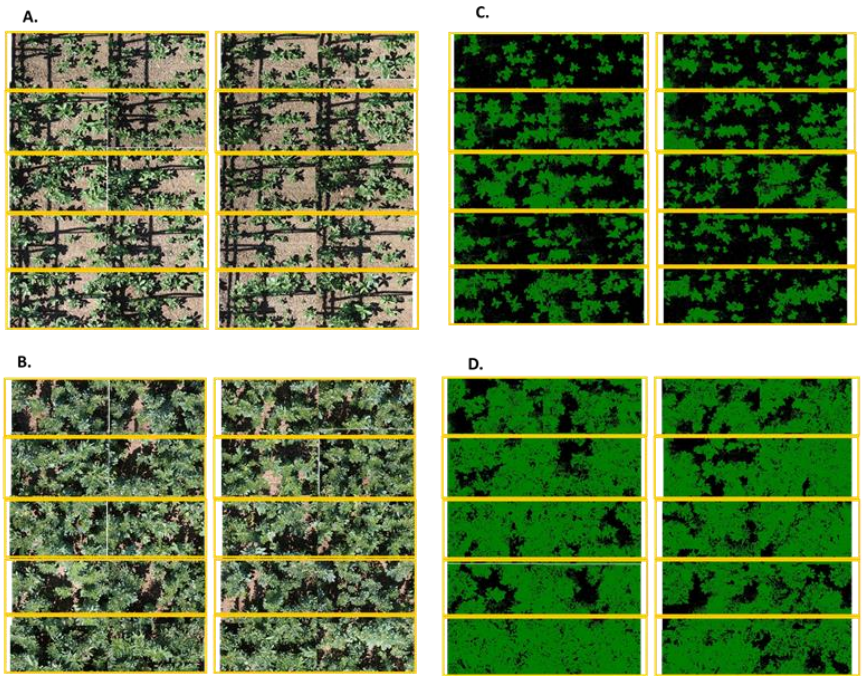


Figure 2.9. Illustration of time series image analysis using PhenoHarvest software. Panels **A.** and **B.** are photomontages of the eight plots forming Block 1 at the Sonning site assembled from two subplot jpeg images per plot. The subplot fields of view are non- overlapping but have been presented with gaps removed and yellow lines separating plots. The images were taken on two dates (A. On 19/5/2020 and B. 24/6/2020 respectively) while images **C** and **D** are the equivalent processed images used to extracting the percentage of green pixels used as a proxy for GAI.

Best Linear Unbiased Predictions (BLUPs) were used to account for variation due to spatial effects in the field. BLUPs were calculated using the *Lme4* package in R (Bates *et al*., 2015) and the correlations among the measured phenotypes were estimated using the package *corrplot* in R (Wei and Simko, 2017).

2.3.3.3. Yield, Hundred Seed Weight and Seed Number per m^2

In both evaluation trials, whole plots were harvested with a plot combine and plot yield (YLD) expressed in tonnes per hectare was measured. At UoR, the number of seeds was counted in a random weighed sample of 200g per each plot to calculate Hundred Seed Weight (HSW) as well as to estimate the total number of seeds per plot (total plot weight in grams*number of seeds per 200g/200) and seed number per m^2 (SN = seeds per plot/harvested plot area). At Teagasc seed number count was not based on a whole plot data but was calculated on a representative subplot sample of one m^2. BLUPs for YLD, HSW and SN were calculated using the *Lme4* package in R.

ANOVA, implemented in R was used to calculate means, standard errors and to detect statistically significant differences in seed yield (g) among the genotypes in both evaluation experiments individually and as a multilocation analysis. The R package *emmeans* was used to calculate estimated marginal means (EMMs) for the genotypes (Searle *et al*, 1980).

2.4. Results

2.4.1. Phenotypic distribution and selection thresholds during recurrent selection

2.4.1.1. Selection in O_1 generation on heterozygosity

From the 219 individuals of the O_1 Foundation population, 93 plants (O_1 selection) with the most heterozygosity (by setting a threshold of >25% heterozygosity i.e. more than 10 heterozygous loci out of 40) were selected (Figure 2.10.) and exactly 19 progeny seed per selected genotype pooled to create the O_2 outcrossing population. It should be noted that although just over half the founding plants that formed the O_1 generation – the O_1 unselected group - did not contribute seed to the O_2 bulk, having been grown intermingled with the O_1sel group in a confined glasshouse compartment with a captive bumblebee

colony and contributed to the pollen pool, the O_1 unselected individuals will have fathered a proportion of seed set by the O_1sel, so it is unlikely that this initial enrichment for heterozygosity caused any allele loss.

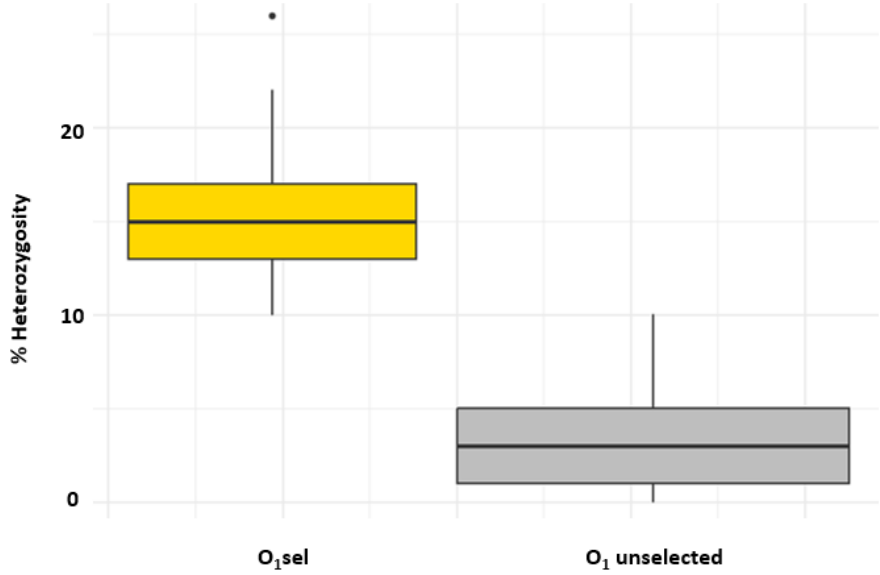


Figure 2.10. Distribution of heterozygosity amongst selected and unselected parts of the O_1 Foundation population based on genotyping using 40 well-distributed and informative SNP markers.

2.4.1.2. Selection from O_2 onwards on Individual Plant Seed Yield in caged field environments

In total, 4,646 faba bean plants were individually harvested, threshed and weighed from five isolation cages over 3 seasons: O_2 (2017), O_4 (UoR - 2018), O_6 (UoR - 2019), 1T (Teagasc – 2018) and 2T (Teagasc – 2019). In each case, a constant selection intensity was applied corresponding to the 3.7% of plants with highest Individual Plant Seed Yield (IPSY). As can be seen from the IPSY distributions of the O_2 – O_4 – O_6 UoR recurrent selection series boxed in red in Figure 2.11, there is not a constant gain in this parameter from generation to generation.

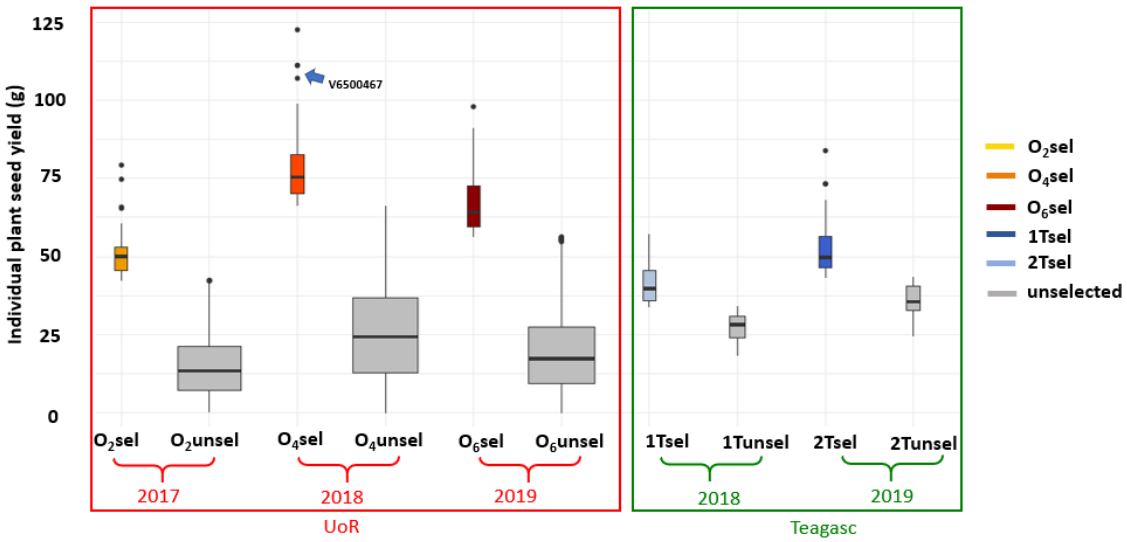


Figure 2.11. Distribution of individual plant seed yield (g) amongst selected and unselected portions of the O_2, O_4, O_6, 1T and 2T generations. Boxplot widths are proportional to the square root of the sample sizes in each group. It should be noted that in Teagasc trials, only the 10% heaviest individuals were threshed and weighed, so the full distribution of weights in the unselected population is not available for 1T or 2T. The individual plant (V6400467) with the highest seed number (and third highest IPSY) across the whole dataset is highlighted by the blue arrow (see section 2.4.3.).

This is not unexpected as yield potential of the same genotype can vary hugely due to shifts in weather patterns from year to year. Although IPSYs were only generated for the better performing plants in the 1T – 2T Teagasc recurrent selection series, the fact that yield potential was higher in Teagasc in 2019 compared with 2018 (this was the case also for other bean trials conducted by Teagasc; 2017-2018= 2.8 t/ha and 2018-2019= 5.5 t/ha,), whereas in UoR, there was an opposing trend, further emphasises the fact that any underlying genetic gain that may accrue is overlaid with a potentially far greater environmental response. The assumption underlying the MRSS is that whether selection takes place in a high, medium or low yield potential season, those genotypes that rank

the highest will be always enriched for yield-enhancing alleles and depleted of genes that constrain yield potential compared to those below the selection threshold. As mentioned in section 2.3.3, it was not possible to include checks consisting of a stable, unchanging control variety within the selection cages solely for the purposes of quantifying year-to-year variation in yield potential as such an attempt would not change the selection approach but would certainly slow genetic progress in randomly intermating environment by displacing pollen of "improved" lines with that of a "static" genotype.

2.4.1.3 Responses of seed number and seed weight to environment and selection (UoR)

In order to further examine factors driving IPSY during selection, distributions of yield components – Seeds Per Plant (SPP) and Hundred Seed Weight (HSW) - were plotted for UoR, where all plants in each generation had been individually harvested (Figure 2.12.). Interestingly, while seed number per plant changes mirrored IPSY, being substantially higher in the intermediate O_4 selection cycle compared to either O_2 or O_6, mean HSW appeared to increase steadily in each successive generation ($O_2 < O_4 < O_6$). This suggested that the overall yield potential in a given year was primarily a function of number of seeds set, whereas seed weight was more likely to be responding to directional selection.

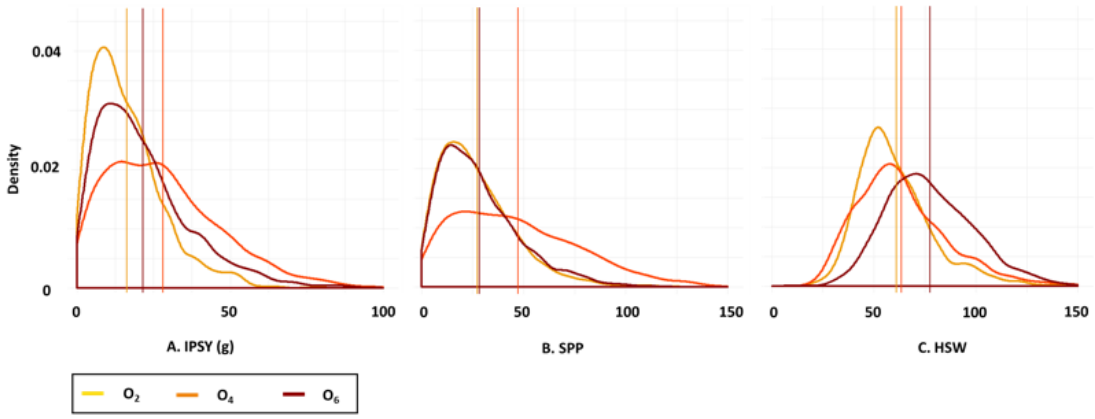


Figure 2.12. Distribution of **A.** IPSY (g), **B.** number of seeds per plant (SPP) and **C.** HSW by generation for UoR. The vertical lines represent the mean of each generation.

Table 2.2 shows that though a similar number of plants were harvested in each selective generation, the 2018 O_4 generation was outlying in terms of having an exceptionally large cage yield and mean SPP. Weather data for the three growing seasons were accessed to examine the possibility that a more favourable growing season in 2018 could explain the exceptional performance of the O_4 cage.

Table 2.2. Key population statistics in UoR summer cage selective generations.

	O2	**O4**	**O6**
No of harvested plants	1,454	1,581	1,611
Total Cage Yield (kg)	23.8	44.5	34.7
Mean SPP	27.4	47.43	28.4
Max SPP	194	201	124
Mean IPSY (g)/plant	16.4	28.11	21.54
Max IPSY (g)/plant	79.3	122.51	97.93
HSW	61.2	63.5	77.64

2.4.2. Influence of weather conditions on yield during the selection cycles (UoR)

Average temperature per month and hours of sunshine are important environmental factors especially during flowering and pod set for faba beans as the developmental rate is a function of thermal degree days and the photosynthetic capacity is highly dependent on the total solar radiation interception. Since adequate irrigation was provided in all three experiments, rainfall was not considered as an explanatory factor as it would be in a rainfed crop. Table 2.3. shows the pertinent local weather patterns during the 2017-2019 growing seasons.

Table 2.3. Average daily sunshine hours (h) and daily average temperature (^{o}C) throughout the 2017-2019 growing seasons (source: Dr Roger Bruce, Reading University Atmospheric Observatory "RUAO").

Year	Month	Average Temperature (^{o}C)	Average Sunshine (h)
2017	April	9.3	6.9
	May	13.8	5.5
	June	17.3	7.7
	July	18.1	5.7
	August	16.4	6.2
2018	April	10.9	4.1
	May	14.1	8.6
	June	17.1	8.4
	July	20.7	9.7
	August	18.1	6.5
2019	April	9.6	6.2
	May	12	6.4
	June	15.3	6.3
	July	18.7	7.9
	August	18.1	7.3

For spring sown faba beans, May, June and July are the months of flowering and pod set, thus the weather conditions during these months are important for the yield

performance of the crop. In 2018, the May to July environmental conditions were more favourable compared to either 2017 or 2019. The average temperature was higher throughout the whole season, but especially so at the pod set, and as irrigation was provided, the warmer 2018 conditions would not have been accompanied by soil moisture deficit-induced heat stress. The exceptionally high seed yields exhibited by the population in 2018 in Reading illustrate the strong influence of environment on yield and demonstrate that comparisons at the population or individual genotype level cannot be made unless the genotypes/populations being compared have been grown side-by-side.

2.4.3 Anatomy of a high-yield potential plant

As noted in the previous section, high yield potential in 2018 was accompanied by similarly high SPP. To illustrate what high seed set looks like in the extreme, the structure of the individual plant from with the highest number of individual seeds in any generation within the MRSS scheme is shown in Figure 2.13. The exceptional number of seeds borne by this plant was due to an exceptionally high conversion of floral organs into successful, well-filled pods, with18 podding nodes on the main stem and 13 podding nodes on a secondary basal branch, for a total of 51 pods (1.6 per node) and a mean of 3.94 seeds per pod. The mean HSW (53.2g) for this exceptionally prolific individual was 16% below the population mean (63.5g) but with over four times the mean SPP, its IPSY was the third highest overall.

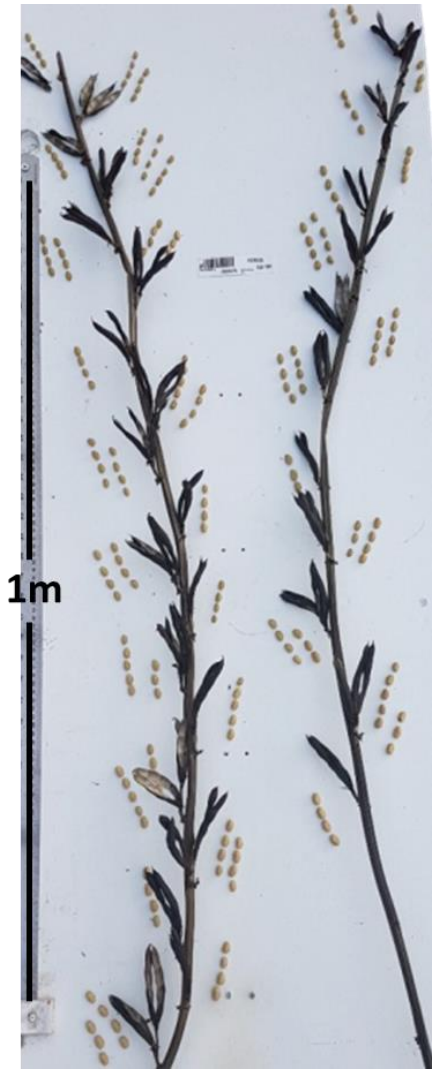


Figure 2.13. High yielding individual (V6500467) of the O_4 generation showing a single plant seed yield of 107g from 201 seeds (HSW – 53.2g) distributed over 51 pods. Seed from each pod has been removed and placed in "pod-rows" at the node of origin to illustrate seed and pod size, number and distribution. This single plant has two branches which have been detached for the purposes of the photo.

2.4.4. Measurement of Yield progress

As noted above in section 2.3.3., it is not clear from year to year during the selection process whether there is any underlying progress in yield, nor indeed whether the apparent directional shift in yield (HSW (g)/plant) (Figure 2.12.) is due to differences in

environment or genetic gain. However, this question was addressed in the fourth year of the project by conducting an open field yield plot trial in which combine yield of the foundation population and intermediate generations as well as the final selections was measured and benchmarked against performance of current elite check varieties.

2.4.4.1 Overview of Yield Progress trials

Co-ordinated yield plot trials were conducted at the University of Reading's Sonning Farm (site abbreviation – UoR) and at Teagasc Oakpark in Ireland (site abbreviation – IRL), with five common entries corresponding to three check varieties and the latest recurrent selection bulks from the respective sites. Entries where limited seed was available were grown at one site only. Common measurements at both sites included combine yield (YLD), and potential explanatory variables of Plant Population per m^2 (PP), Hundred Seed Weight (HSW), Seed Number per m^2 (SN). At UoR only, Green Area Index (GAI) was measured weekly to track canopy development and in IRL only, Protein Content (PC) of harvested seed was determined. BLUPs for each of the above traits were calculated separately for each trial and for the common entries in a combined multi-environment analysis and the BLUPs and *p*-values for ANOVA tests of significance of Genotype (as well as Site and Genotype x Site interaction where appropriate) are set out in Table 2.4.

The trial was designed primarily with the objective of detecting any significant differences in yield between entries (Genotypes), and the significance of the 'Genotype' factor for YLD means that at least in UoR, there were clear winners and losers. PP, early GAI and HSW were also significant for 'Genotype' at UoR, indicating that YLD could have been explained in part by how well the plots established and how quickly the canopy closed as well as by HSW. In IRL, 'Genotype' was not a significant factor for YLD,

suggesting that there was less clear separation of genotypes at that site, but PP, PC and HSW were all significant, so there were at least some yield components and explanatory variables that differed significantly between entries.

In the multi-environment analysis, ‘Site’ was a highly significant factor for all measured traits, confirming that the differences in soil, weather and season length between UoR and IRL did have a significant impact on trait values. Finally, regarding the multi-environment analysis, ‘Genotype’ was very significant for HSW and PP; the ‘Genotype x Site’ interaction was highly significant for SN and PP, and although n ‘Genotype’ was not significant for YLD itself, the ‘Genotype x Site’ interaction for YLD was significant at the 5% level. This can be interpreted as showing that there was a degree of genotype-dependent site adaptation in some important yield components, although this did not ultimately give rise to significant separation of genotypes for YLD. The results will be discussed in detail in the following sections starting with establishment, then discussing canopy development and finally yield components.

Table 2.4. Summary of 2020 yield trial phenotypes showing ANOVA *P*-values for each factor modelled and BLUPs per genotype for each measured phenotype at each site.

Trait	Units	Site	ANOVA Factors				BLUPs										
			Genotypes	Blocks	Site	Genotypes x Site	Entries in both sites					Entries in UoR		Entries in IRL			
			P-value	*P*-value	*P*-value	*P*-value	O6	2T	'Lynx' (UK)	'Fanfare' (UK)	'Vertigo'	O1	O4	1T	'Fanfare' (IRL)	'Lynx' (IRL)	'Wizard'
GAI_15MAY		UoR	<0.001***	0.015*	NA	NA	0.34	0.24	0.14	0.20	0.20	0.21	0.31	NA	NA	NA	NA
GAI_19MAY		UoR	<0.001***	<0.001***	NA	NA	0.43	0.35	0.19	0.27	0.31	0.30	0.42	NA	NA	NA	NA
GAI_26MAY		UoR	0.027*	<0.001***	NA	NA	0.37	0.34	0.27	0.32	0.36	0.35	0.37	NA	NA	NA	NA
GAI_9JUNE		UoR	0.242	<0.001***	NA	NA	0.50	0.48	0.44	0.48	0.50	0.53	0.48	NA	NA	NA	NA
GAI_15JUNE		UoR	0.215	<0.001***	NA	NA	0.56	0.53	0.49	0.53	0.55	0.56	0.55	NA	NA	NA	NA
GAI_24JUNE		UoR	0.002**	0.001**	NA	NA	0.83	0.80	0.68	0.74	0.76	0.79	0.78	NA	NA	NA	NA
GAI_30JUNE		UoR	0.027*	<0.001***	NA	NA	0.82	0.78	0.71	0.77	0.80	0.79	0.83	NA	NA	NA	NA
GAI_7JULY		UoR	0.236	<0.001***	NA	NA	0.74	0.69	0.67	0.72	0.73	0.73	0.75	NA	NA	NA	NA
GAI_13JULY		UoR	0.540	0.017*	NA	NA	0.81	0.76	0.78	0.81	0.79	0.77	0.78	NA	NA	NA	NA
GAI_22JULY		UoR	0.562	0.064	NA	NA	0.66	0.61	0.59	0.66	0.65	0.64	0.71	NA	NA	NA	NA
GAI_28JULY		UoR	0.593	0.012*	NA	NA	0.39	0.35	0.41	0.40	0.36	0.42	0.43	NA	NA	NA	NA
YLD	**(t/ha)**	UoR	0.036*	0.038*	NA	NA	2.97	2.70	2.37	2.53	2.44	2.61	2.86	NA	NA	NA	NA
HSW	**(g)**	UoR	<0.001***	0.944	NA	NA	59.8	55.7	48.4	53.4	55.3	55.5	63.3	NA	NA	NA	NA
SN	**m^{-2}**	UoR	0.665	0.048*	NA	NA	548	508	495	486	448	460	465	NA	NA	NA	NA
PP	**m^{-2}**	UoR	0.036*	0.007**	NA	NA	34.0	35.8	26.7	29.7	28.9	30.9	34.5	NA	NA	NA	NA
YLD	**(t/ha)**	IRL	0.112	0.179	NA	NA	5.02	4.80	5.40	5.50	4.80	NA	NA	5.40	3.80	5.50	4.70
HSW	**(g)**	IRL	<0.001***	0.5	NA	NA	59.6	51.6	44.4	44.7	47.8	NA	NA	47.7	48.7	45.4	56.5
SN[1]	**m^{-2}**	IRL	0.048	0.244	NA	NA	722	749	743	942	805	NA	NA	808	880	870	743
PP	**m^{-2}**	IRL	0.023*	0.742	NA	NA	34.8	36.3	27.9	42.6	34.3	NA	NA	32.8	34.4	32.1	28.5
PC	**%**	IRL	0.002**	0.532	NA	NA	28.00	27.90	29.30	26.97	26.83	NA	NA	28.70	27.83	28.10	26.93
YLD	**(t/ha)**	UoR + IRL	0.061	0,259	<0.001***	0.04*	4.00	3.78	3.40	3.52	3.28	NA	NA	NA	NA	NA	NA
HSW	**(g)**	UoR + IRL	<0.001***	0.988	0.002**	0.28	59.7	53.6	46.9	50.5	52.8	NA	NA	NA	NA	NA	NA
SN	**m^{-2}**	UoR + IRL	0.160	0.496	<0.001***	0.006**	615	618	572	643	576	NA	NA	NA	NA	NA	NA
PP	**m^{-2}**	UoR + IRL	<0.001***	0.0147*	<0.001***	<0.001***	35.8	36.0	26.0	35.2	30.6	NA	NA	NA	NA	NA	NA

NA: Not applicable

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

PC: Protein content

[1]based on one m^2 subplot

2.4.4.2. Germination and emergence

All plots at both sites were drilled with the same sowing density of 40 seeds/m^2 and with seed which had high germination rates (>97%) targeting a final plant population of 35 plants/m^2. Actual Plant Population (PP) in IRL was closer to the targeted population density than in UoR. Whereas in IRL, all the recurrent selections and three check varieties were in the 30-35 plants/m^2 range and estimated marginal means analysis showed no statistically significant differences (Figure 2.14 C), at UoR, there was an unfortunate divergence between the recurrent selections and the three check varieties, with O_1, O_4, O_6 and 2T all falling in the 31-38 range, with 'Fanfare' UK, 'Lynx' UK and 'Vertigo' all falling below target in the 27-29 range, such that of the three checks at UoR, only 'Fanfare' UK was considered to have statistically comparable PP to the recurrent selections. In the multi-environment analysis, 'Fanfare' UK, O_6 and 2T outcrossing populations all fell within the target range of 35-40 plants/m^2 with the estimated marginal means showing that 'Lynx' UK is the only statistically significant outlier (Figure 2.14. A.).

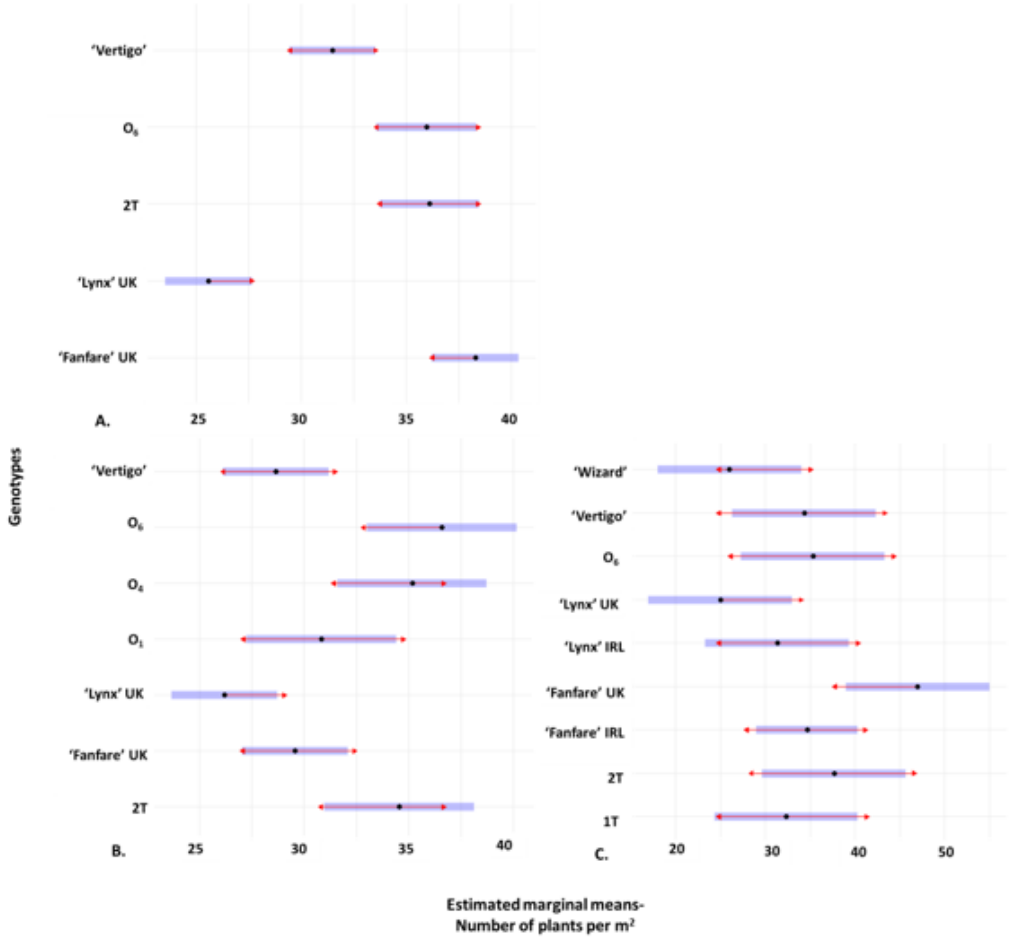


Figure 2.14. Estimated marginal means for the Plant Population (PP) per m^2 analysed for **A.** Combined analysis for UoR and IRL trial. **B.** UoR trial. **C.** IRL trial. The black dot is the mean, the blue bars are the confidence intervals for the estimated marginal means and the red arrows explain the comparisons among them. Overlapping arrows among the entries mean that there is no statistically significant difference based on the Tukey test for a=0.05.

The unexpectedly high PP of 'Fanfare' (UK) (47 plants/m^2) in IRL was also visually confirmed by the drone image taken on the 15th of June (Figure 2.15.).

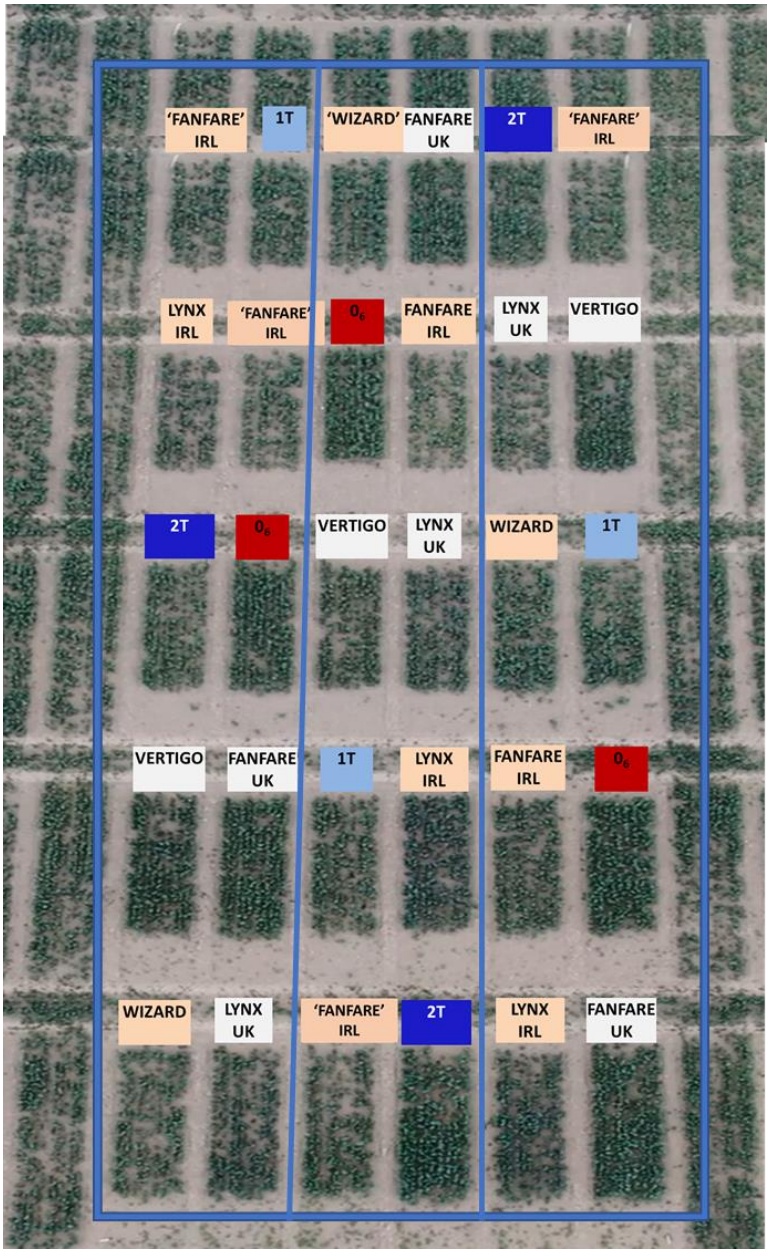


Figure 2.15. Drone image taken on 15/06/2020 over the IRL field trial at Teagasc (Oakpark) illustrates differences in establishment among the entries.

These differences in establishment could be explained in two ways. First, even though counted packets of seeds were used (thus eliminating error in sowing rate), the seed drilling equipment can introduce small variations in the sowing density; for example, this is the only possible explanation for how 'Fanfare' UK reached a higher PP

in IRL than was theoretically possible given the number of seeds provided. Second, where establishment conditions are challenging, real differences in the quality and vigour of the seedlots could make some more likely to establish poorly than others. Differences in seed quality can be environmental or genetic, with the latter of greater interest in the context of this thesis. For example, it would be expected that significant differences in seed weight drive faster and more even emergence due to the inherently greater nutrient reserves in the cotyledons of larger seeds (Ellis, 1992). The fact it has been confirmed in faba bean that seed weight-related seedling vigour in turn is influenced by the degree of heterozygosity may also be relevant, and the evidence regarding the degree of hybridity/heterozygosity of the recurrent selections tested here are explored in Chapter 4.

2.4.4.3. Canopy Development (UoR)

At UoR, GAI as a measure of canopy cover, was recorded starting on 15/5/2020 based on overhead RGB images taken at regular intervals until all plots reached senescence and BLUPs reported in Table 2.4.

Figure 2.16. shows the evolution of green canopy cover over time for each genotype. The GAI curves are consistent with the ANOVA tests which indicated that 'Genotype' was most significant at the first two timepoints (15/5 and 19/5) and on 24/6, On these three dates, it is particularly evident that O_4, O_6 and 2T have developed more rapidly and reached higher, earlier peak GAI compared to the rest, while 'Lynx' UK trails all other entries, stretching the variance in the other direction. This is entirely consistent with the differences in emergence shown in the previous section, but the almost 2-fold gap in GAI value between recurrent selections and check varieties is only partly explained by the 10-20% differences in PP with the majority of the difference being attributed

following qualitative observation of the plots to the generally larger size of the recurrent selection plants.

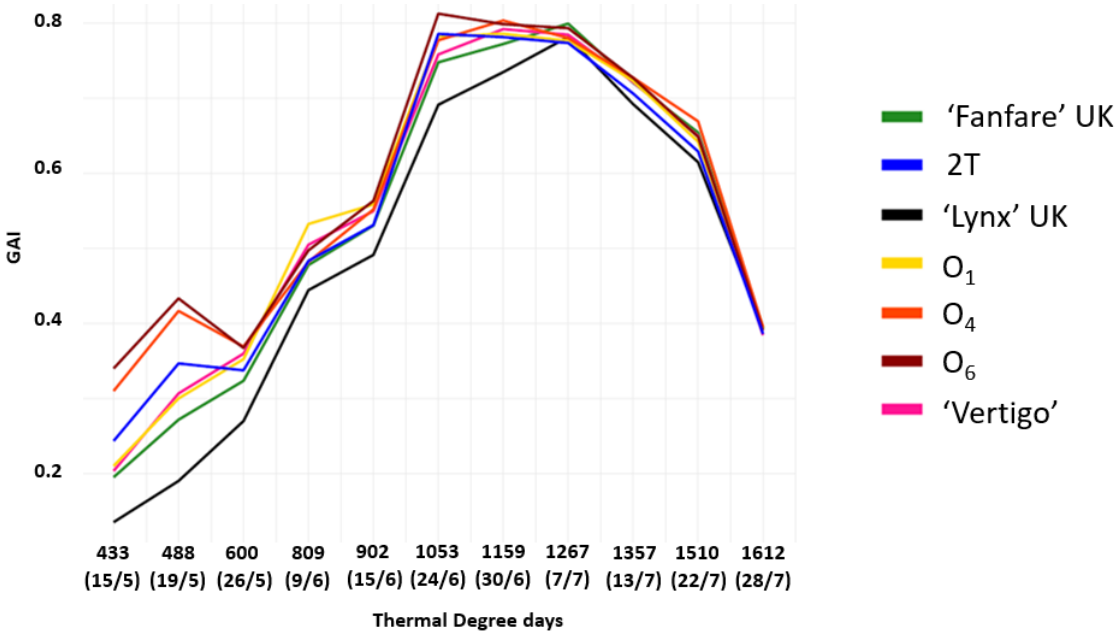


Figure 2.16. Changes in GAI. X-axis is in thermal degree days calculated for the dates at which the pictures were captured starting from the 7th of April (crop establishment) and y-axis the % of green area (BLUPs). Each line corresponds to a different entry.

The other notable feature of the GAI curves in Figure 2.16 is the downward trajectory of the O_4, O_6 and 2T plots between the 19th and 26th May. As confirmed by rainfall data shown Figure 2.17., there was an extended period of drought during May which towards the end of the droughted spell, visibly differentially stressed the more advanced plots composed of larger plants with higher water demand, resulting in genotype-dependent leaf rolling and apparent loss of GAI.

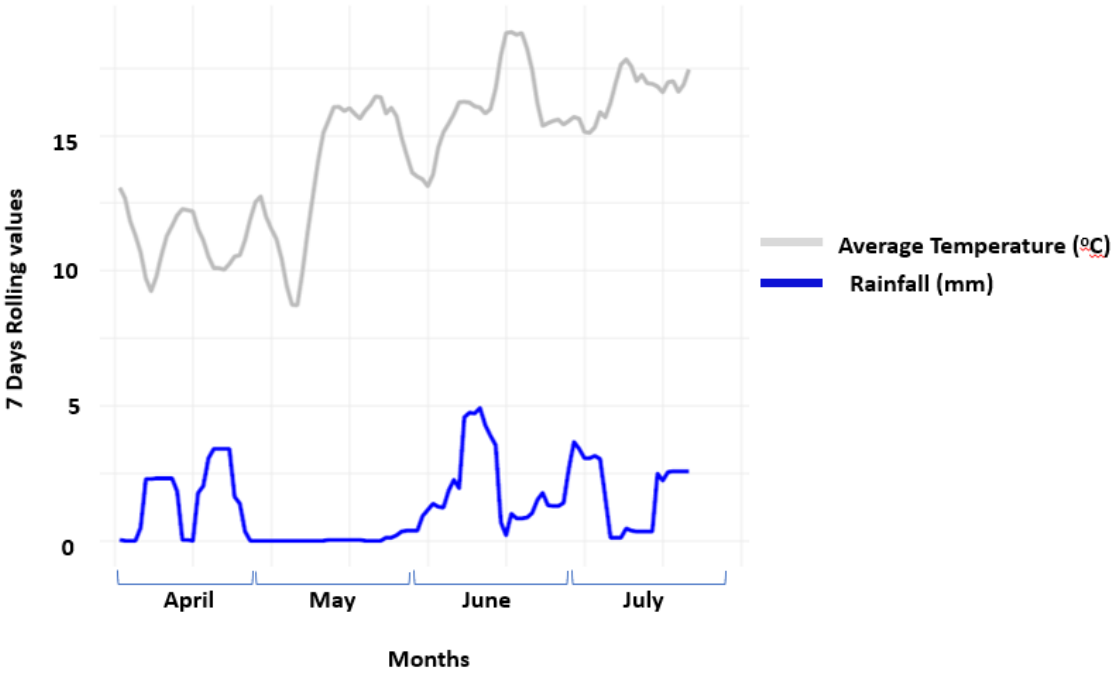


Figure 2.17. Average temperature (ºC) and rainfall (mm) from the establishment of the crop (7th of April) until the end of the growing season (source: Caroline Hadley, weather station at Reading University Crop Research Unit "CRU").

2.4.4.4. Yield response to selection

Finally, genotype rankings and relationships between plot yield (YLD) and its components of Hundred Seed Weight (HSW) and Seed Number per m^2 (SN) were examined first at the individual trial locations and then across sites.

At UoR, overall yields were low, ranging from 2.5 to 3.1 t/ha. For context, control yields in the provisional Spring Bean 2020 RL trial results posted in PGRO website 13th September 2020) was 2.84 t/ha, well below the 5-year control yields for the 2020 RL (5.2 t/ha), indicating that 2020 has been a difficult year for spring beans due to late establishment followed by early (May) drought, so it should be borne in mind that this first year of evaluation of yield progress may not have permitted the full yield potential of any entry in the trial to be evaluated. Nonetheless, there were significant differences

in yield with O_6, the latest recurrent selection from the UoR site having the highest yield (3.2 t/ha), significantly above check varieties 'Lynx' UK and 'Vertigo' (Figure 2.18).

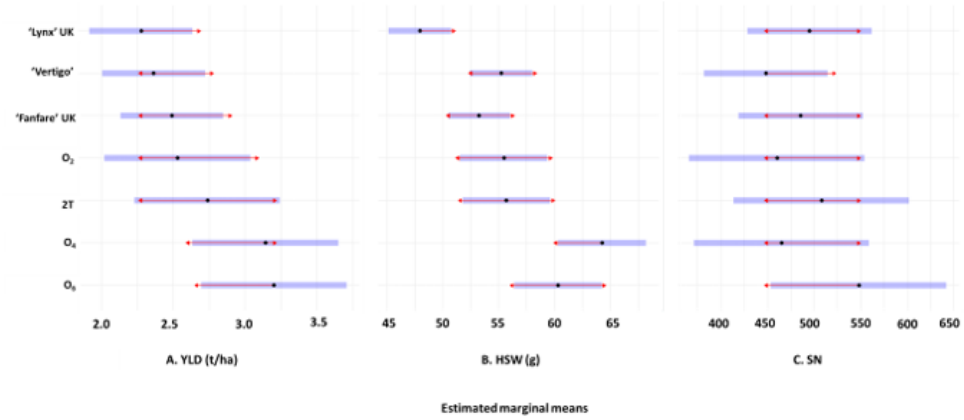


Figure 2.18. Estimated marginal means for **A.** YLD (t/ha), **B.** HSW (g) and **C.** SN calculated overall for UoR. The black dot is the mean, the blue bars are the confidence intervals for the estimated marginal means and the red arrows explain the comparisons among them. Overlapping arrows among the entries mean that there is no statistically significant difference based on the Tukey test for a=0.05.

This latter result is important as 'Vertigo', together with 'Fanfare' were the only elite spring UK founders in the MRSS, demonstrating that in just three rounds of selection and six generations from the original elite x diverse crosses, the population YLD means being achieved are competitive with the elite progenitor material. Furthermore, the three MRSS bulks O_1 (foundation), O_4 (intermediate) and O_6 ("final") show progressively increasing yields as would be expected under high selection intensity for the measured trait. Of course, the observed YLD increments during these few generations of selection are not statistically significant, but they are at least a sign of underlying genetic gain and confirmation that IPSY may be used as a proxy for YLD.

Of the yield components, only HSW shows significant differences between entries, suggesting HSW as the better candidate to explain YLD differences. In fact, as

the plots in Figure 2.19 show, YLD is strongly driven by HSW in UoR (p=0.0068) whereas the suggestive correlation with SN is non-significant (p=0.29). This result appears to settle the question raised in section 2.4.1.3 in favour of HSW being the yield component most responsive to selection.

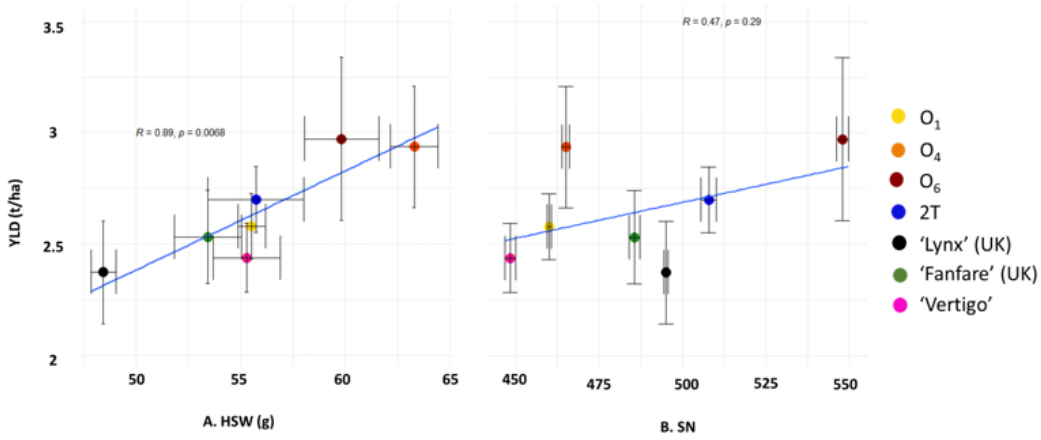


Figure 2.19. Correlation of BLUPs of YLD (t/ha) to HSW (g) (**A**) and SN (**B**) in the UoR trial. X- axis shows HSW (g) in **A**. and SN (g) **B.** Y-axis shows the YLD (t/ha). Horizontal error bars (SE) correspond to X-axis and vertical to Y-axis.

In IRL, overall yields were far higher than in UoR, ranging from 3.8 – 5.5 t/ha. Estimated marginal means analysis (Figure 2.20) confirms that although YLD does not significantly vary between entries, HSW and SN do. Unlike at UoR, in IRL, the latest IRL recurrent selection bulk (2T) did not outyield the theoretically less-adapted 1T, suggesting that whatever "progress" had been made between 1T and 2T did not advantage 2T in this single yield trial. Nonetheless, 2T yielded comparably to the O_6, which had gone through three more rounds of outcrossing and one more round of selection (albeit at UoR) and was ranked in the middle of the check varieties, so the same competitiveness with elite progenitors claimed earlier for O_6 applies here regarding performance of IRL selections in IRL.

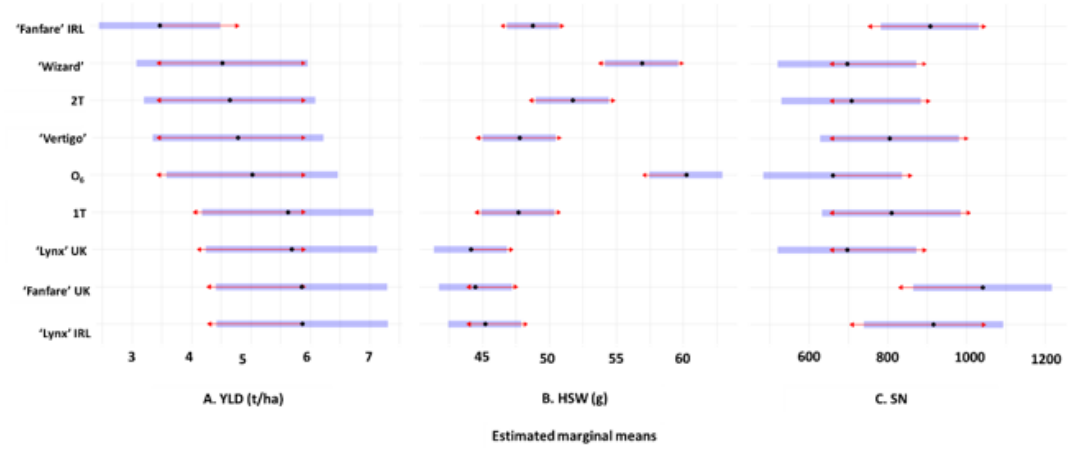


Figure 2.20. Estimated marginal means for **A.** YLD (t/ha), **B.** HSW (g) and **C.** SN calculated overall for IRL (Teagasc. Oakpark). The black dot is the mean, the blue bars are the confidence intervals for the estimated marginal means and the red arrows explain the comparisons among them. Overlapping arrows among the entries mean that there is no statistically significant difference based on the Tukey test for a=0.05.

Different weather conditions between UoR and IRL could possibly explain the difference in the overall yield performance between the two sites. As previously shown on Figure 2.17. at UoR experimental site a severe drought right after the sowing of the crop and temperature extremes during the pod filling stage, which probably affected the establishment and performance of all entries, hence the overall less YLD in UoR compared to IRL. In contrast, mild and wet weather conditions in Ireland (Figure 2.21) between the sowing date (20th of March) and germination of the genotypes were ideal conditions for the establishment of the experiment.

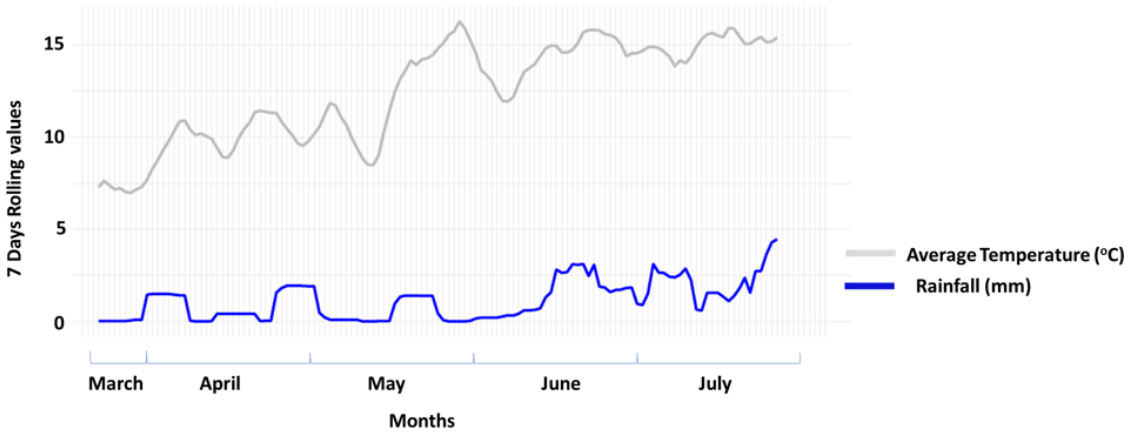


Figure 2.21. Illustration of the average temperature (°C) and rainfall (mm) from the establishment of the crop (20th of March) until the end of the growing season.

As in UoR, at Teagasc there was a negative correlation between YLD and HSW (p=0.31) and no relationship between YLD and SN (Figure 2.22).

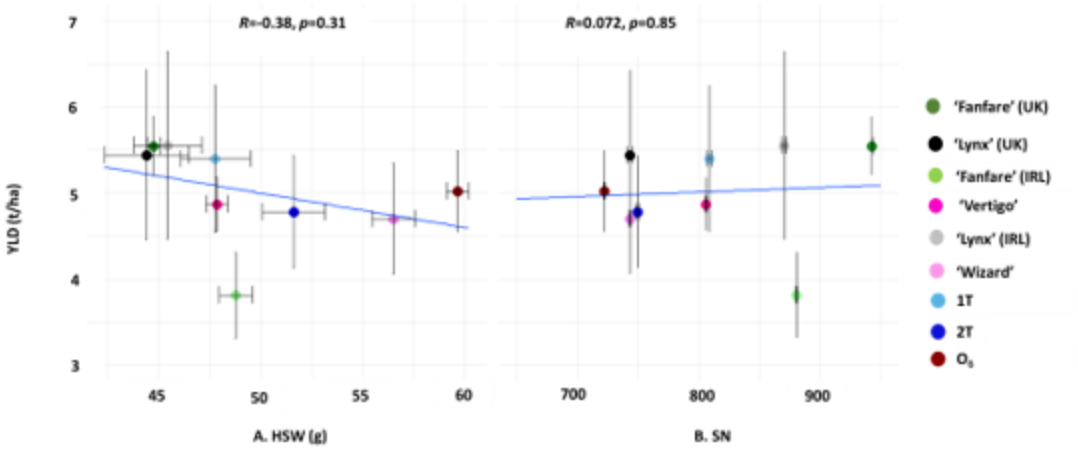


Figure 2.22. Correlation of BLUPs of YLD (t/ha) to HSW (g) (**A**) and SN (**B**) at Teagasc, Oakpark. X- axis shows HSW (g) in **A**. and SN (g) **B.** Y-axis shows the YLD (t/ha). Horizontal error bars correspond to the X-axis variable Standard Error (SE) and vertical to YLD SE. SN is a derived value by YLD and HSW.

Protein Content (PC) was not a focus of this investigation; yet it is precisely for its protein content that faba bean is valued. PC was measured in IRL and values found in the range of 26.8-29.3%, typical for spring beans in IRL. Importantly, PC was found to be uncorrelated with YLD (Table 2.4.), which bodes well for faba bean breeding as it indicates that even at overall YLD of 5.5 t/ha, it is possible to maintain PC of 28.7%, and there is no sign of any negative correlation between PC and YLD as has been reported for some other crops where N supply is limiting at the higher end of yield potential (Evans and Bhatt 1977; Panthee et al. 2005).

And overall, there were no significant differences among genotypes on YLD and SN while there were significant differences for the HSW (Figure 2.23). Across site BLUPs could give a glimpse of which of the common entries was most broadly adapted in this twin site yield trial and in this respect, both UoR and IRL recurrent selection bulks O_6 and 2T were ranked above any of the three common check varieties (Figure 2.24).

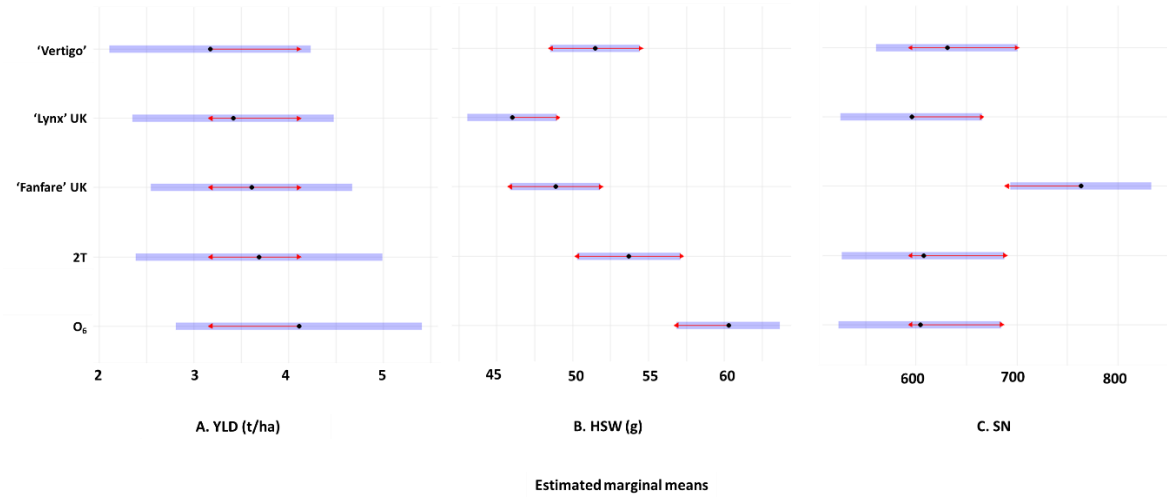


Figure 2.23. Estimated marginal means for **A.** YLD (t/ha), **B.** HSW (g) and **C.** SN calculated overall for the two Locations. The black dot is the mean, the blue bars are the confidence intervals for the estimated marginal means and the red arrows explain the comparisons among them. Overlapping arrows among the entries mean that there is no statistically significant difference based on the Tukey test for a=0.05.

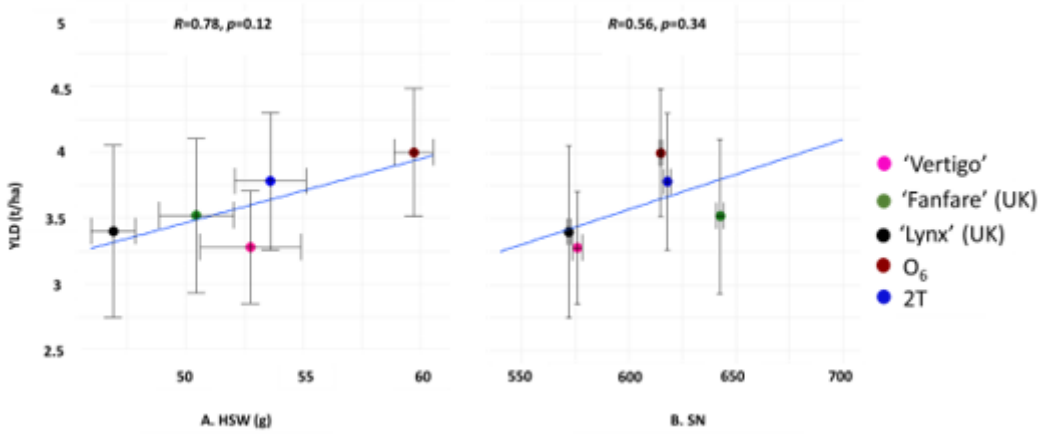


Figure 2.24. Correlation of BLUPs of YLD (t/ha) to HSW (g) (**A**) and SN (**B**) calculated overall for the two Locations. X- axis shows HSW (g) in A. and SN (g) B. Y-axis shows the YLD (t/ha). Horizontal error bars correspond to the X-axis variable SE and vertical to YLD SE.

Pearson's Correlation Coefficient (PCC) was calculated among the recorded phenotypes GAI, YLD, HSW and PP at UoR and for YLD, HSW, PP and PC at IRL. The overall correlation table (Figure 2.25) of all the quantitative phenotypes captured during the season helps provide an overview of how yield was developed and consolidated, what characteristics of genotypes were driving yield and the extent to which the genotypes responded similarly across sites.

This overview correlation matrix supports many of the findings already commented on; namely, the impact in UoR of genotype-dependent emergence on PP, persisting into effects on early GAI and the strong correlation between YLD and HSW in both sites. However, the opportunity to look at across-site correlations (even though based only on 5 common entries) revealed a surprising correlation between IRL_YLD and UoR GAI 'GAI_28JULY'. Further inspection of this unsuspected correlation revealed that the most successful yielding variety in Ireland 'Lynx' (UK) (among the ones being common in both sites), together with 'Fanfare' UK, which also did comparatively well in IRL, had high persisting green area when most of the other entries have declined (Figure 2.26.). This raises the intriguing possibility that extended canopy duration/late senescence could be specifically advantageous in the Irish climate, a question that would be relatively straightforward to test by integrating measurements of canopy cover and greenness and a greater number of entries into future across-site trials.

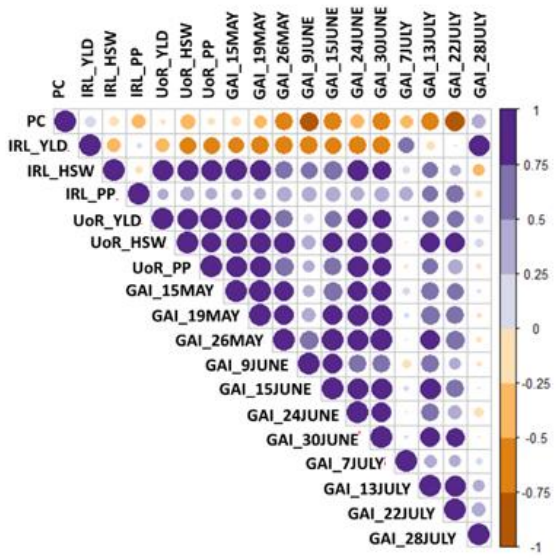


Figure 2.25. Pearson's correlation coefficient for the measured phenotypes of the evaluation trial at UoR and Teagasc. 'UoR_' and 'IRL_' are used to distinguish YLD, HSW and PP in the two sites (UoR and Teagasc in Ireland). Positive correlations are displayed in purple and negative in orange, the size and intensity of the colour show the significance.

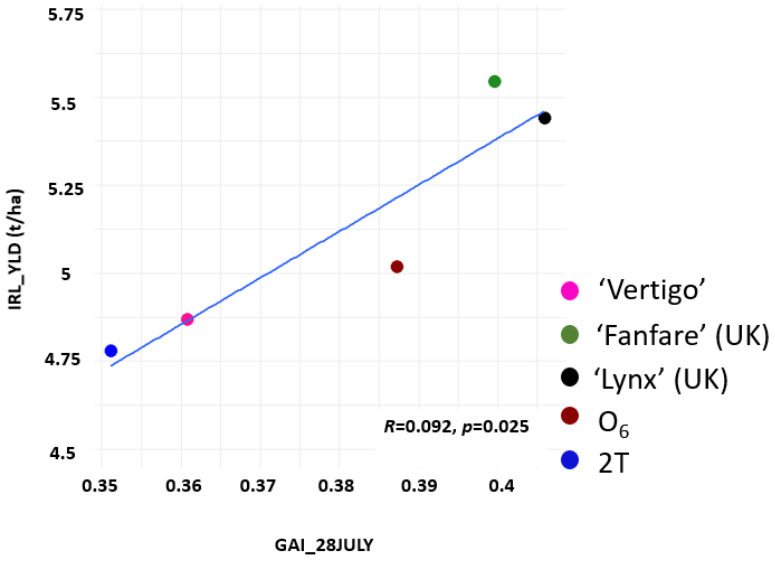


Figure 2.26. Correlation of BLUPs of YLD (t/ha) to GAI_28JULY. X- axis shows values for GAI_JULY and Y-axis shows the BLUPs for IRL_YLD (t/ha) for the five common varieties.

2.5. Conclusion

The aim of this Chapter was to conduct an empirical test of the concept that a simple, small-scale recurrent selection using captive bumblebee colonies as the sole agent of recombination and selection on IPSY as the sole selection criterion could produce competitive long-term improvement in population yield. In fact, two variants of the MRSS – one targeting a UK-adapted spring bean ideotype with two generations per year, and one targeting an Irish-adapted winter-sown bean ideotype with one generation per year - were implemented, though only the UK-based one could be studied in detail. The investigation culminated in a reciprocal transplant-type trial (for a recent example, see van Frank *et al*., 2020) where the latest generation of recurrent selections from both UoR and IRL environments were tested both in their 'native' and 'unseen' environments.

This pilot breeding exercise was modestly successful, in so far as in the 2020 yield plot trial, the products of the most intense selection (the O_6 bulk selected at UoR) performed better in trial than any of: remnant bulks of the foundation population and intermediate selections, any elite founder or current elite check variety. There was equally support in the data for the premise that spring sowing in the southeast of England and winter sowing in the southeast of Ireland represent two distinct growing environments in which different genotypes would thrive from the fact that in each site, the top-ranked entry was a bulk selected at that same site.

In terms of understanding how this yield progress and adaptation to the target environment was achieved, detailed characterization of the growth and development of entries in the UoR trial revealed that high yielding entries were characterised by high germination and rapid emergence, early vigour, early peak canopy development and high seed weight. The positive correlation of all these traits, together with the heavy emphasis on maximising outcrossing in the MRSS, suggested that heterosis achieved by

maintaining high levels of diversity and heterozygosity might have been the key factor in the performance of the UoR selections.

Concerning the performance of selections made in Ireland, we should recognise that the evaluation year came too early – after only two outcrossing rounds - and maybe was not a fair test for assessing the outcrossing populations under conditions in which the populations did not have the time to fully adapt. Heterozygosity of the O_2 generation has been genetically explored and will be presented in Chapter 4, however, as no genotype data was generated to verify the level of outcrossing achieved in the Irish selections, 1T's superior performance compared to 2T might stem from a collapse in heterozygosity of 2T.

Recurrent selection might not be the method of choice for a commercial breeder, due to the difficulty in making an outcrossing population conform to uniformity and stability standards needed to secure variety rights, but at least from the point of view of yield progress, this breeding method appears not to be too risky since we demonstrate encouraging yield progress from O_1 to O_6 in this single-person operation selecting on IPSY as a proxy for YLD.

The question of whether the empirical gains in yield observed reflect heritable genetic merit or if it is mainly as a result of hybrid vigour and heterosis will be investigated using high density genome-wide genotyping in Chapter 4.

2.6. References

Atkinson, A. and Donev, A. and T., Randall. 2007. Optimum Experimental Designs, with SAS. Oxford University Press. ISBN: 9780199296606

Bates D.s, Mächler M, Bolker B, Walker S. 2015. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, **67**(1), 1–48. https://doi.org/10.18637/jss.v067.i01.

Bond, D. A. 1987. Recent Developments in Breeding Field Beans (*Vicia Faba* L.). Plant Breeding **99** (1): 1–26. https://doi.org/10.1111/j.1439-0523.1987.tb01144.x

Bond, D. A., G. J. Jellis, G. G. Rowland, J. Le Guen, L. D. Robertson, S. A. Khalil, and L. Li-Juan. 1993. Present Status and Future Strategy in Breeding Faba Beans (*Vicia Faba* L.) for Resistance to Biotic and Abiotic Stresses. Euphytica. **73**: 151-166. https://doi.org/10.1007/bf00027191.

Bond, D. A., Drayner, J.M., Fyfe, J.L. and Toynbee-Clarke, G. 1964. A male sterile bean (V. faba L.) inherited as a Mendellian recessive character. J. Agric. Sci. Camb., **63**: 223-234.

Crépon, K., P. Marget, C. Peyronnet, B. Carrouée, P. Arese, and G. Duc. 2010. Nutritional Value of Faba Bean (*Vicia Faba* L.) Seeds for Feed and Food. Field Crops Research. **115**(3): 329-339. https://doi.org/10.1016/j.fcr.2009.09.016.

DAFM. 2017. Irish Recommended List For 2017 Sowing. Agriculture.Gov.Ie. https://www.agriculture.gov.ie/media/migration/farmingsectors/crops/seedcertification/cropvarietyevaluationcvepublications/2017/SpringBeanRecommendedList2017030117.pdf

DAFM. 2019. Irish Recommended List For 2019 Sowing. Agriculture.Gov.Ie. https://www.agriculture.gov.ie/media/migration/farmingsectors/crops/SpringBeanRecommendedList080119.pdf.

Ellis, R. H. 1992. Seed and Seedling Vigour in Relation to Crop Growth and Yield. Plant Growth Regulation. **11**: 249-255. https://doi.org/10.1007/bf00024563.

Evans, L. E., and G. M. Bhatt. 1977. Influence of Seed Size, Protein Content and Cultivar on Early Seedling Vigour in Wheat. Canadian Journal of Plant Science. **57**:929-935. https://doi.org/10.4141/cjps77-133.

Fulton, T. M., J. Chunwongse, and S. D. Tanksley. 1995. Microprep Protocol for Extraction of DNA from Tomato and Other Herbaceous Plants. Plant Molecular Biology Reporter. **13**: 207-209. https://doi.org/10.1007/bf02670897.

Hallauer, A.l R., and L. L. Darrah. 1985. Compendium of Recurrent Selection Methods and Their Application. Critical Reviews in Plant Sciences. **3**(1): 1-33. https://doi.org/10.1080/07352688509382202.

Khazaei, H., K. Street, A. Bari, M.l Mackay, and F. L. Stoddard. 2013. The FIGS (focused Identification of Germplasm Strategy) Approach Identifies Traits Related to Drought Adaptation in Vicia Faba Genetic Resources. PloS One **8** (5): e63107. https://doi.org/10.1371/journal.pone.0063107

Maalouf, F. 2011. Faba bean Improvement at ICARDA: Constraints and Challenges. Grain Legumes. **56**: 13. Available at: http://webcache.googleusercontent.com/search?q=cache:e_VevDGl1IoJ:orbit.dtu.dk/files/5678198/Grain%2520Legumes%252056.pdf+andcd=8andhl=enandct=clnkandgl=uk

McVetty, P. B. E., and J. Nugent-Rigby. 1984. Natural Cross Pollination of Faba Beans (*Vicia Faba* L.) Grown in Manitoba. Canadian Journal of Plant Science. **64**(1): 43-46. https://doi.org/10.4141/cjps84-007.

Panthee, D. R., V. R. Pantalone, D. R. West, A. M. Saxton, and C. E. Sams. 2005. Quantitative Trait Loci for Seed Protein and Oil Concentration, and Seed Size

in Soybean. Crop Science. **45**(5): 2015-2022. https://doi.org/10.2135/cropsci2004.0720.

PGRO. 2016. Pulse Agronomy Guide. https://www.pgro.org/downloads/PGRO-AGRONOMY-GUIDE-2016.pdf.

PGRO. 2019. Pulse Agronomy Guide. https://www.pgro.org/pulse-agronomy-guide/.

Rowland, G. G. 1987. A Recurrent Selection Scheme for Faba Bean. Canadian Journal of Plant Science. **67**(1): 79-85. https://doi.org/10.4141/cjps87-009.

Searle S. R., F. M. Speed and G. A. Milliken. 1980. Population Marginal Means in the Linear Model: An Alternative to Least Squares Means. The American Statistician. **34**(4):216-221, https://doi.org/10.1080/00031305.1980.10483031

Stamp, P., R. Messmer, and A. Walter. 2012. Competitive Underutilized Crops Will Depend on the State Funding of Breeding Programmes: An Opinion on the Example of Europe. Plant Breeding. **131**(4): 461-464. https://doi.org/10.1111/j.1439-0523.2012.01990.x.

van Frank, G., P. Rivière, S. Pin, R. Baltassat, J.-F. Berthellot, F. Caizergues, C. Dalmasso, et al. 2020. Genetic Diversity and Stability of Performance of Wheat Population Varieties Developed by Participatory Breeding. Sustainability: Science Practice and Policy, R: A Language and Environment for Statistical Computing, **12** (1): 384. https://doi.org/10.3390/su12010384

Webb, A., A. Cottage, T. Wood, K. Khamassi, D. Hobbs, K. Gostkiewicz, M. White, et al. 2016. A SNP-Based Consensus Genetic Map for Synteny-Based Trait Targeting in Faba Bean (*Vicia Faba* L.). Plant Biotechnology Journal **14** (1): 177–85. https://doi.org/10.1111/pbi.12371

Wei T. and V. Simko. 2017. R package corrplot: Visualization of a Correlation Matrix (Version 0.84). Available from https://github.com/taiyun/corrplot.

Wright, A. J. 1977. Inbreeding in Synthetic Varieties of Field Beans (*Vicia Faba* L.). The Journal of Agricultural Science **89** (2): 495–501. https://doi.org/10.1017/S0021859600028410

Zeid M, Mitchell S, Link W, Carter M, Nawar A, Fulton T, Kresovich S .2009. Simple sequence repeats (SSRs) in faba bean: new loci from Orobanche resistant cultivar 'Giza 402'. Plant Breeding. **128**:149–155. https://doi.org/10.1111/j.1439-0523.2008.01584.x

3. Chapter 3- Use of stochastic simulations to inform MRSS design

3.1. Abstract

AlphaSim is a flexible software package built in R which enables breeders to design various breeding schemes by simulating multiple aspects of a breeding program. In this study, AlphaSim was used to evaluate the sensitivity of genetic gain under recurrent selection to several parameters in the control of the breeder and to measure possible tradeoffs between them. A Modified Recurrent Selection Scheme was simulated in which the genetic architecture of the target trait (yield) was determined by 50 QTLs/chromosome out of a total of 1000 segregating sites per chromosome. 18 scenarios involving different combinations of population size, selection intensity (SI) and outcrossing rate were tested for 30 selection cycles. Genetic gain (GG) and genic variance (GV) were compared across the tested scenarios. As expected, since new germplasm was not introduced during the simulation period, GG and GV are almost inversely related. Higher SI gives faster rates of GG during the early generations, however, possibly prematurely curtails genetic diversity needed to sustain GG in the long term. The smaller the population size, the greater the GG during the early generations but increasing population size was an inefficient means to increase long-term GG. Low rates of outcrossing favour rapid GG in the short term but limit long-term progress. The insight offered by these simulations provides a framework for rational design of small-scale breeding programmes.

3.2. Introduction

Plant breeding seeks to develop and deliver cultivars with improved traits. As previously discussed, the biological/reproductive characteristics and technological status of each crop can rule out use of certain breeding methods, narrowing down the range of the

schemes that could be deployed. However, even within a particular category of scheme, the breeder still has to make many active decisions regarding the source and quantity of diversity, the amount of hybridization, the population sizes, phenotyping methodology and selection intensity and the optimum combination of choices can be defined by the genetic architecture of the trait that needs to be improved. The fact that these parameters may all interact in non-obvious and non-linear ways suggests the need for a computational approach.

Modelling and simulating the structure of a breeding scheme by genetically describing the species and the modelled trait, in a user-friendly software able to stochastically predict genetic gain and genetic variation of an unlimited number of generations, has become feasible only very recently. Breeding Scheme Language (BSL) is an example of a simulation package built in R able to model breeding schemes helping breeders evaluate them and chose the best per condition (Yabe, Iwata, and Jannink, 2017). By defining the overall parameters regarding the structure of the study genome and the species architecture, BSL creates a founder population and further functions are applied to describe breeding activities like phenotype, genotype, genomic prediction and selection. The mating in BSL is flexible and allows random mating among individuals, inbreeding or creation of double-haploid progenies. For polygenic traits, BSL is constrained to a single uniform effect size across all loci. Most importantly BSL does not take into account genotype × environment interactions.

Similarly to BSL, AlphaSimR is a software, originally developed as a standalone package (Faux *et al*. 2016), then re-built in R (Gaynor, Gorjanc and Hickey, 2020), which gives breeders the opportunity to simulate outcomes when key parameters of a breeding scheme are varied and thus helps identify the optimum combinations of parameters to implement in practice. AlphaSimR considers additive, dominant, epistatic effects and

genotype × environment interactions. Moreover, AlphaSimR can simulate multiple traits and offers a variety of mating choices: biparental crossing, inbreeding, double haploid and open pollinating populations with varying degrees of selfing. The additional flexibility in modelling trait architecture made AlphaSimR the most suitable package to model the MRSS.

A modelling approach to breeding program design can save considerable time and money compared to a more empirical, trial-and-error approach. Such simulations have for example been used to identify the importance of choosing optimal crosses to maximise long-term gain in a recurrent genomic selection scheme (Gorjanc, Gaynor, and Hickey, 2018). In this work, a wheat breeding simulation study testing phenotypic selection, genomic selection, genomic selection at headrow stage and recurrent genomic selection was designed to run for 20 generations and compared genetic gain, genetic diversity, genomic accuracy and efficiency of converting genetic diversity into genetic gain. The results showed differences in genetic gain and genetic diversity in each year depending on the selection program, the number of parental lines and the population size and point to optimal solutions to the breeding objectives in question.

The opportunity to harness faba bean's natural outcrossing habit to drive recombination as well as the necessity for speed, simplicity and small scale (i.e. manageable by a single person) were prime considerations that led to the Modified Recurrent Selection Scheme (MRSS) described in Chapter 2. The hypothesis that such a simple breeding scheme could generate rapid genetic gain was borne out by the results of the 2020 evaluation trial, but given these promising results, several questions arise and should be considered in any further or future refinement of a similar MRSS. Making certain assumptions regarding the genetic architecture of yield, the purpose of this work was to identify the ideal population size and selection intensity that serves the purpose of

achieving the highest rate of genetic gain in the population and yet maintains sufficient genetic variation to permit continued gains over the long-term. Although the ability to modulate bee-assisted outcrossing rate is rather limited, a further aim was to investigate the consequences of varying outcrossing rate on the genetic gain/genic variance tradeoff. As no breeding scheme simulations have yet been published for faba bean, this investigation will fill an important gap in the literature and provide a basis for a more scientific approach to small-scale breeding.

3.3. Materials and methods

3.3.1. Overview of the breeding scheme

Using AlphaSimR (Gaynor, Gorjanc and Hickey, 2020) in the R environment, a simulated breeding scheme was modelled as closely as possible on the MRSS implementation described in Chapter 2. As MRSS had two experimental sites (site one: two generations per year GH-no selection and spring cage-selection and site two: one selection generation annually), the AlphaSimR code is more similar to the Irish pipeline whereby we go from selection generation to selection generation. The main features of the simulation are described in genetic terms below, while the full R code with comments is presented in Appendix C.

A set of faba bean lines possessing the required characteristics was generated and subjected to one round of intercrossing without selection to generate a recombined foundation population (Founder/ F_0 generation) of 2000 individuals, mimicking the creation of the O_1sel bulk referred to in Chapter 2. A modified recurrent selection scheme was encoded in which equal numbers of progeny of the best-performing plants from each selective generation are pooled to seed the subsequent generation. The software creates simulated progeny genotypes implementing specified levels of outcrossing and assesses

phenotypic values of individuals according to the trait QTL they have inherited and applies a specified selection threshold to determine what the "best-performing" genotypes are. The average trait value of each selective generation is used to track genetic gain over time and the changes in genic variance calculated from the simulated genotypes. Genetic gain and genic variance are the main output variables of interest. The results were averaged over 100 simulations per cycle, deriving from the same simulated founder population.

3.3.2. Input parameters of the breeding simulations

The simulated faba bean genome was designed to have 6 chromosomes (the actual number of chromosomes in faba bean) and 1000 randomly assigned non overlapping segregating sites (QTL) per chromosome for the simulated trait – corresponding roughly to the number of mapped markers per chromosome which can be tracked using the current generation of high density array-based genotyping technology (see Chapter 4). As yield, our trait of interest, has highly complex inheritance, the trait architecture in the simulations was set as polygenic. Although there is a near-complete absence of literature on the genetic architecture of yield in faba bean and quite limited data for other species, yield was assumed for the purposes of this simulation to be governed by 300 QTLs (50 per chromosome) i.e. 1 in 20 of segregating sites. The effective size of the QTL was assumed to derive from a gamma distribution with scale=1.0 and shape=0.5. Phenotypic variance was expressed by genotypic variance=0.3, environmental variance=0.1 and genotypic x environmental variance=0.3 giving narrow sense heritability (h^2) =0.3, as expected for a polygenic trait. Dominance and epistasis effects were not simulated.

Genic variance (Vg) is calculated from the observed gene frequencies assuming a perfect Hardy-Weinberg equilibrium (no dominance effect) (Bulmer, 1976), and is given by the sum of this expression across all loci:

$$Vg = 2pq * (1 + F) * a\text{^}2$$

where p is the frequency of the allele "A" and q the frequency of the allele "a", a is the additive effect and F is an inbreeding coefficient used to measure the statistically expected strength of heterozygosity in population genetics when compared to Hardy-Weinberg equilibrium and is calculated by the expression:

$$F = 1 - (Observed\ frequency\ of\ Aa)/(Expected\ frequency\ of\ Aa)$$

So, F is the probability of two alleles from any individual plant of the population and at any locus to be identical by descent. In a case of a homozygous population with equal proportions of the two genotypes the allele frequency would be 0.5 and F=1. For outcrossing populations, F is set to 0.5.

A base population of 2000 Founder plants (F_0) was used for all recurrent selection scenarios. The F_0 population along with every subsequent recurrent generation were considered to be randomly mating. Recurrent selection was performed by simulating three different selection intensities of: top 2% (selection intensity: high - sH), 4% (selection intensity: medium - sM) and 10% (selection intensity: low - sL) at three different population sizes of 2,000 (population: Small - pS), 10,000 (population: Medium - pM) and 20,000 (population: LArge - pLA) individuals. These 9 scenarios were tested with the two different outcrossing rates of 75% (outcrossing: High - oH) and 30% (outcrossing: Low - oL) generating a total of 18 scenarios in which all factorial

combinations of population size, selection intensity and outcrossing rate were explored (Table 3.1.). The scenarios were set to run for 30 generations with the aim of clarifying asymptotic endpoints to gain and genetic diversity under particular conditions, rather than because there is a realistic expectation that any recurrent selection scheme would run unmodified for so long.

Table 3.1. List of scenarios and their relative input parameters. Scenario short names following the scenario number are formatted in three parts: to show at a glance if the population size is small, medium or large (pS/pM/pLA); the selection intensity is low, medium or high (sL/sM/sH) and the outcrossing rate low or high (oL/oH). Thus, scenario #1 (1-pSsHoH) uses small population size (pS), high selection intensity (sH) and high outcrossing rate (oH).

Scenarios	Founders	Number of selected plants	Seeds per plant	Population size	Selection intensity	Outcrossing rate (%)
1-pSsHoH	2000	40	50	2000	2%	75
2-pSsMoH	2000	80	25	2000	4%	75
3-pSsLoH	2000	200	10	2000	10%	75
4-pMsHoH	2000	200	50	10000	2%	75
5-pMsMoH	2000	400	25	10000	4%	75
6-pMsLoH	2000	1000	10	10000	10%	75
7-pLAsHoH	2000	400	50	20000	2%	75
8-pLAsMoH	2000	800	25	20000	4%	75
9-pLAsLoH	2000	2000	10	20000	10%	75
10-pSsHoL	2000	40	50	2000	2%	30
11-pSsMoL	2000	80	25	2000	4%	30
12-pSsLoL	2000	200	10	2000	10%	30
13-pMsHoL	2000	200	50	10000	2%	30
14-pMsMoL	2000	400	25	10000	4%	30
15-pMsLoL	2000	1000	10	10000	10%	30
16-pLAsHoL	2000	400	50	20000	2%	30
17-pLAsMoL	2000	800	25	20000	4%	30
18-pLAsLoL	2000	2000	10	20000	10%	30

3.3.3. Cost per unit gain calculation

Each scenario gives a certain absolute value of genetic gain and genic variance at a particular timepoint. Even though maximising genetic gain seems appealing, the cost per unit gain to achieve this value should be taken into consideration in choosing the best scenario. For this study, the main direct costs per cage were calculated based on a combination of theoretical estimates and experience in running a real selection scheme. Considering the cage layouts described in Chapter 2 where we used a sowing rate of 40 seeds/m^2 and allowing for walking and working space among the plants, the total space needed to grow 2000 plants was 60m^2.

Costs for seed and fertilisers were considered zero. Pest and disease control costs were calculated according to Teagasc Crops Costs and Returns (Collins and Phelan, 2019). The variables were transformed from €/ha to £/60m^2 and the values were then scaled for 5-, 10-, 15-, 20-, 25- and 30-year timepoints. Cage costs were calculated considering the extra help of a seasonal technician for the days of sowing, harvesting and phenotyping, to help act within short good weather windows for sowing and harvesting and the right crop developmental stage for phenotyping (Table 1D- Appendix D). Postgraduate student labour costs were based on a projection of £15k per cage per year. Infrastructure, including the cage frame and the covering mesh calculated as a fixed cost of £2000 for 10 years after which replacement will take place. In total four bee colonies were calculated to be periodically introduced in each cage throughout the flowering period aiming to cover differences in flowering time. *Bombus terrestris* colonies (Koppert-Natupol) were used and the price was estimated at £60 per beehive including VAT.

3.4. Results

The output variables of cumulative genetic gain and genic variance were compared among scenarios under different population sizes, selection intensities and outcrossing rates over 30 generations. Tukey's range test for multiple pairwise comparisons run for six different timepoints, year 5, 10, 15, 20, 25, 30 and concluded that both genetic gain and genic variance differed significantly ($p<0.05$) among scenarios and depend on all three tested parameters of Population Size (PS), Selection Intensity (SI) and Outcrossing Rate (OR). ANOVA tests were conducted for homogeneity of variance and normality (Appendix E) and verified that the data are normally distributed and the variance across groups was homogeneous for all 6 timepoints.

3.4.1. Genetic gain across scenarios

In general, the trend in cumulative genetic gain across generations followed a logical pattern. Genetic gain invariably rose sharply in the earliest generations of selection as allelic diversity (and therefore breeding value) was stretched over the maximum possible range but as subsequent cycles of selection were completed, with less and less allelic diversity for recombination and selection to draw on, the rate of gain slowed until in some scenarios at least the largest effects had been fixed. Figure 3.1 shows cumulative genetic gain curves for all 18 scenarios. To avoid visual congestion, the 18 scenarios are split over 6 panels with identical axes, with each panel containing high, medium and low SI scenarios for a given population size x outcrossing rate combination. Each panel follows the same generic pattern represented in the cartoon representation of differing SI scenarios in Figure 3.2, whereby cumulative genetic gain sharper increases most rapidly during the first cycles of selection for the higher selection intensity and increases the least rapidly at lowest selection intensity.

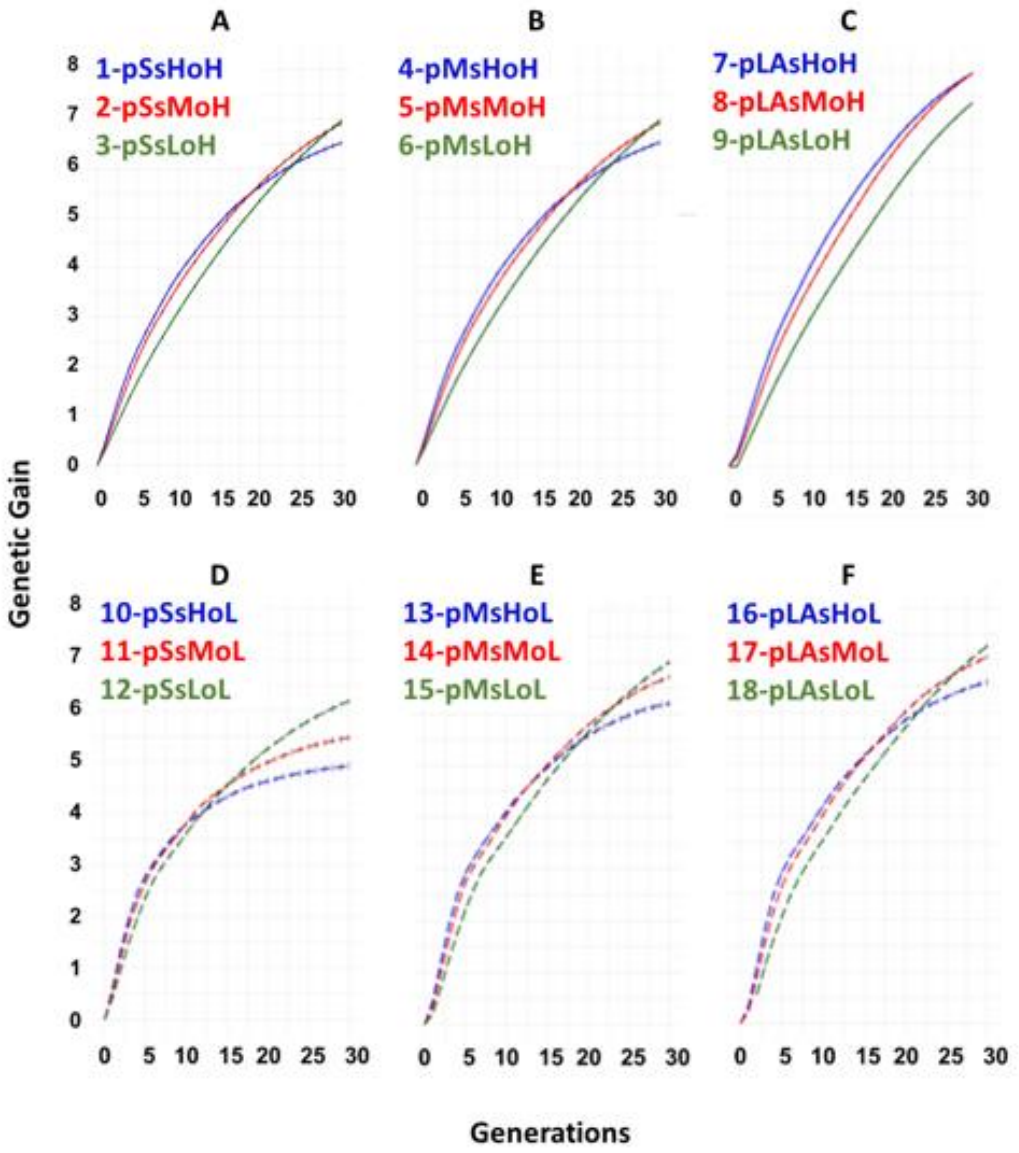


Figure 3.1. Effects of Population size, selection intensity and outcrossing rate on cumulative genetic gain. Panels group the three selection intensities of 2%, 4% and 10% and are arranged in columns by population size (Panels A, D – Small PS, B, E – Medium PS, C, F – LArge PS) and in rows by outcrossing rate (Panels A, B and C refer to the high outcrossing rate of 0.75 and panels D, E and F to the low outcrossing rate of 0.3). High, medium and low SI scenarios are shown in blue, red and green respectively. Cumulative genetic gain (y-axis) is plotted against generations (x-axis). Each datapoint is the mean of 100 simulations.

Importantly though, the high gain associated with high SI which may result in its top ranking at, say Generation 5 (G5), comes with a significant tradeoff; the faster the initial rate of gain, the more potentially beneficial alleles are eliminated before having had a chance to be selected and the more rapidly that rate of gain declines as the generations pass, such that the high SI strategy is overtaken at some point by a medium SI and eventually by the low SI strategy, which have preserved more beneficial alleles through more cycles of recombination and testing and permit higher cumulative gain in the long term (say, Generation 30 - G30). Ultimately, the rank order of SI scenarios is completely reversed between G5 and G30, which means that the best scenario depends very much on the timepoint at which gain is being measured. A similar effect can be observed for population size (PS) by comparing Figure 3.1 D, E and F, where genetic gain for the same SI at a constant low outcrossing rate reaches a higher plateau in later generations as the PS increases from S to M to LA. Again, this is because the genetic bottleneck effect is amplified by any reduction in the size of the genepool and exacerbates the premature loss of potentially beneficial alleles. The same effect of population size can also be seen in the upper row of Figure 3.1 (panels A, B and C - high outcrossing), but is harder to discern.

Finally, comparison of low vs high outcrossing rates (panels D vs A, E vs B, F vs C) shows that increased outcrossing has a similar 'straightening' effect on the genetic gain curve. When outcrossing is low, the heterozygosity of the population is low, and each given allele is dispersed over fewer individuals, making it easier to lose alleles whether by drift or due to their low breeding value.

The apparent anomaly whereby scenario 9-pLAsLoH in panel C shows almost no genetic gain after the first selection cycle illustrating a less obvious, conditional influence of founding population size on the rate of genetic gain. In a very big population size of

20,000 created by outcrossing of a founding population two orders of magnitude smaller, a permissive selection threshold (10%) takes forward more individual genotypes than the original founding population and as a result, the genetic gain is initially very weak.

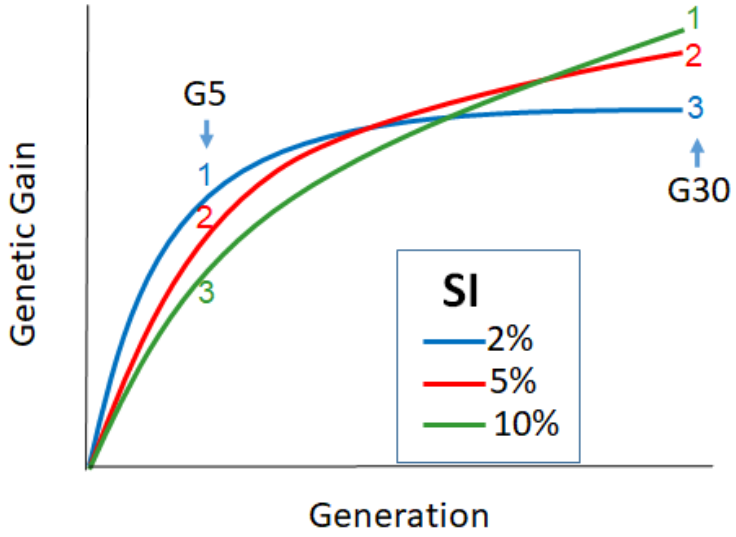


Figure 3.2. Cartoon representation of the cumulative genetic gain for scenarios differing only in SI. High, medium and low SI scenarios are shown in blue, red and green respectively. Cumulative genetic gain (y-axis) is plotted against number of generations (x-axis).

When all these interacting parameters and tradeoffs play out over 30 generations, there is considerable movement in terms of the rank order of scenarios for genetic gain, making it initially hard to discern the best combination of parameters to pick. Figure 3.3 below shows the ranking of numbered scenarios in each generation. The broad generalisation that immediately stands out is that the darker shades (which correspond to the low outcrossing scenarios) dominate the top rankings for the first 12 years/generations, but are gradually displaced from year/generation 13 onwards by high outcrossing scenarios. Individual scenarios tend to be on either an increasing or decreasing ranking trajectory. For example, scenario 10 – pSsHoL gives the highest gain

for the first three years but is the lowest ranked scenario from year/generation 17 on, whereas the top ranked scenario from year/generation 13 on – 7 – pLsHoH is initially ranked 12th but steadily climbs through the ranking.

Generations	Scenario ranking																	
	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th	12th	13th	14th	15th	16th	17th	18th
G1	10	11	1	13	12	2	16	14	3	4	5	7	17	15	8	6	9	18
G2	10	11	13	16	1	14	12	17	2	4	7	15	5	3	8	18	6	9
G3	10	13	11	16	14	17	1	12	4	2	7	15	5	18	8	3	6	9
G4	13	16	10	11	14	17	1	12	4	7	2	15	18	5	8	3	6	9
G5	13	16	10	14	11	17	12	1	4	7	15	2	18	5	8	3	6	9
G6	16	13	14	10	17	11	4	12	1	7	15	2	18	5	8	3	6	9
G7	16	13	14	17	10	11	4	7	1	12	15	2	18	5	8	3	6	9
G8	16	13	14	17	10	11	4	7	1	15	12	2	5	8	18	3	6	9
G9	16	13	14	17	7	4	10	1	11	15	12	5	2	8	18	3	6	9
G10	16	13	7	14	17	4	1	11	10	5	8	2	15	12	18	3	6	9
G11	16	7	13	17	14	4	1	11	5	8	10	2	15	12	18	3	6	9
G12	16	7	4	17	13	14	1	5	8	11	2	15	10	18	12	3	6	9
G13	7	16	4	17	14	13	5	1	8	2	11	15	18	12	10	3	6	9
G14	7	4	16	17	14	5	13	8	1	2	15	18	11	12	10	3	6	9
G15	7	4	16	17	5	8	14	13	1	2	15	18	11	12	10	3	6	9
G16	7	4	5	16	8	17	14	13	1	2	15	18	12	11	6	3	9	10
G17	7	4	8	5	17	16	14	1	13	2	15	18	12	6	11	9	3	10
G18	7	4	8	5	17	16	14	1	13	2	18	15	6	9	12	3	11	10
G19	7	4	8	5	17	16	14	2	1	18	13	15	6	9	3	12	11	10
G20	7	4	8	5	17	16	14	18	2	1	15	13	6	9	3	12	11	10
G21	7	8	4	5	17	16	14	18	2	15	1	6	9	13	3	12	11	10
G22	7	8	5	4	17	18	16	14	2	15	6	9	1	13	3	12	11	10
G23	7	8	5	4	17	18	6	16	2	9	14	15	1	3	13	12	11	10
G24	7	8	5	4	17	18	6	9	2	15	14	16	1	3	13	12	11	10
G25	7	8	5	4	17	18	6	9	15	2	14	16	3	1	13	12	11	10
G26	7	8	5	4	17	6	18	9	15	2	14	3	16	1	13	12	11	10
G27	7	8	5	4	6	9	18	17	15	2	3	14	16	1	13	12	11	10
G28	7	8	5	4	6	9	18	17	15	2	3	14	16	1	13	12	11	10
G29	7	8	5	4	6	9	18	17	15	3	2	14	16	1	13	12	11	10
G30	7	8	5	4	6	9	18	17	15	3	2	14	16	1	13	12	11	10

Colour blocking	
1	1-pSsHoH
2	2-pSsMoH
3	3-pSsLoH
4	4-pMsHoH
5	5-pMsMoH
6	6-pMsLoH
7	7-pLAsHoH
8	8-pLAsMoH
9	9-pLAsLoH
10	10-pSsHoL
11	11-pSsMoL
12	12-pSsLoL
13	13-pMsHoL
14	14-pMsMoL
15	15-pMsLoL
16	16-pLAsHoL
17	17-pLAsMoL
18	18-pLAsLoL

Figure 3.3. Descending cumulative genetic gain ranking of numbered scenarios in each generation. Darker shades correspond to the low outcrossing scenarios and lighter to the high outcrossing scenarios.

3.4.2. Genic variance among scenarios

Genic variance followed a reverse trend compared to genetic variation (Figure 3.4.). Genic variance curves went through a local maximum in early generations as the minimum population size of 2,000 individuals afforded generation of a greater diversity of novel haplotypes than were present in the founding population of 200 regardless of selection intensity, then declined rapidly at first as the bottleneck of selection eliminated the largest negative effect alleles quickly, but then less and less steeply as the remaining effects segregating become smaller. Maximum genic variance is maintained in scenarios with the largest population size and/or the selection intensity is low (10%), both of which allow more alleles to be maintained for longer in the population gene pool. Conversely, the fastest decline in genic variance is found in where low population size or high selection intensity create a sharper bottleneck effect.

The decrease in genic variance is steeper at low outcrossing rate (Figure 3.4 D, E and F) compared to panels showing high outcrossing rate scenarios (Figure 3.4 A, B and C) for the same reason advanced above to explain the steeper increase genetic gain under low outcrossing.

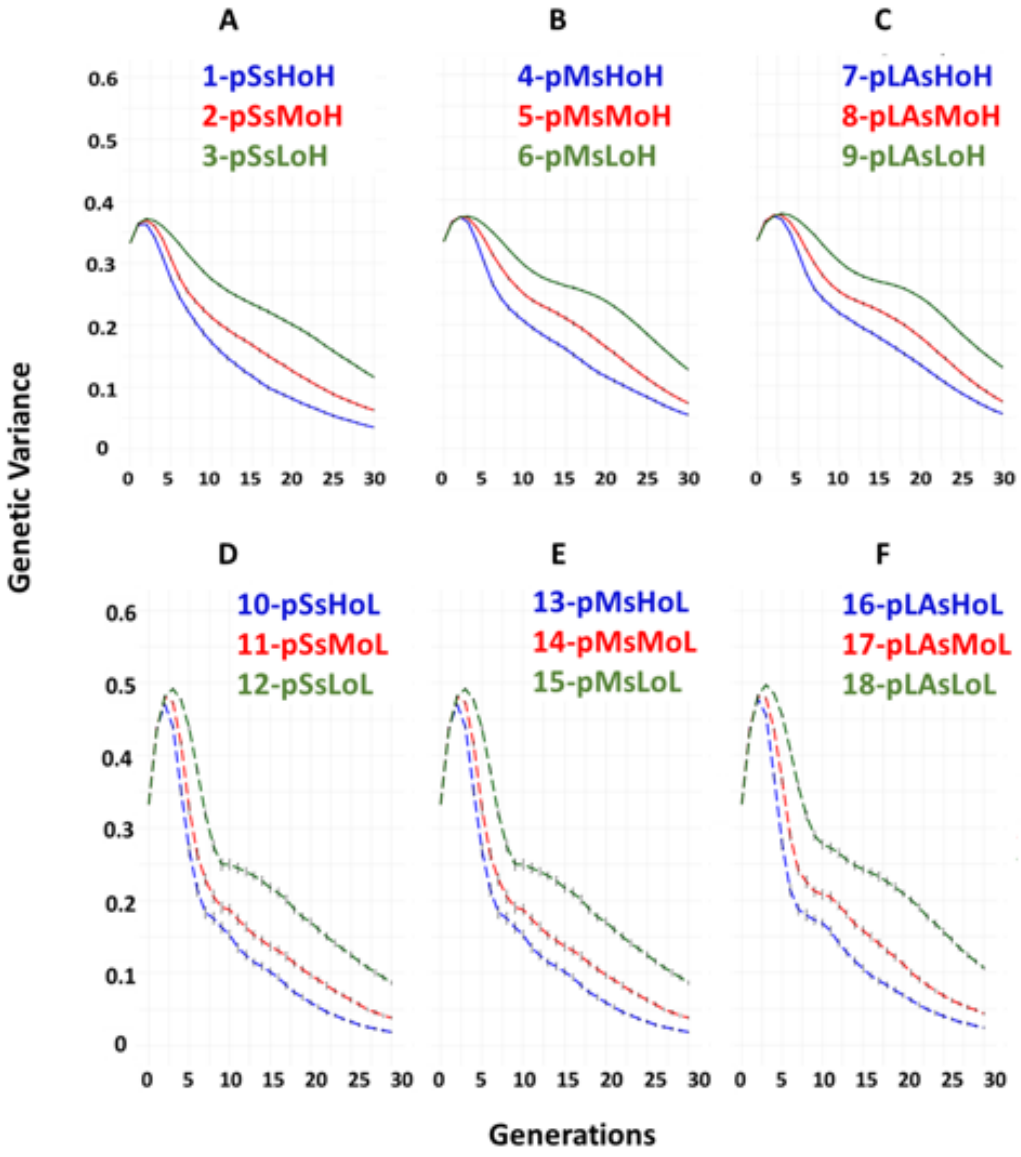


Figure 3.4. Effects of Population size, selection intensity and outcrossing rate on genetic variation. Panels group the three selection intensities of 2%, 4% and 10% and are arranged in columns by population size (Panels A, D – Small PS, B, E – Medium PS, C, F – LArge PS) and in rows by outcrossing rate (Panels A, B and C refer to the high outcrossing rate of 0.75 and panels D, E and F to the low outcrossing rate of 0.3). High, medium and low SI scenarios are shown in blue, red and green respectively. Genic variation (y-axis) is plotted against generations (x-axis). Each datapoint is the mean of 100 simulations.

Running the scheme for 30 generations without enriching the initial diversity by introducing new germplasm results in a varying rate of decline in genic variance. Figure 3.5 shows the ranking of genic variance of the numbered scenarios in each generation. Medium and large population size maintain greater genic variance even when outcrossing rate is low (darker shaded scenarios from G1 to G7). After generation 7, lighter shaded scenarios (high outcrossing rate) dominate the top 10 ranked scenarios for genic variance although a combination of large population size (scenario-18-pLAsLoL) and low selection intensity (scenario 15-pMsLoL) make scenario 18 and 15 exceptions to this general trend.

Genetic gain and genic variance are negatively correlated overall as stated above. However, Figure 3.3. and Figure 3.5. present a common trend whereby greater genetic gain and genic variance is initially acquired by the scenarios with low outcrossing rate. However, long-term, maintaining a high outcrossing rate is the key to greater genetic gains and genic variance, and the only low outcrossing scenarios capable of above average ranking for both parameters are those with the largest population size scenarios (e.g. scenario 16-17-18- pLA) or medium population size in combination with the low SI (scenario 15-sL).

Generations	Scenario ranking																	
	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th	12th	13th	14th	15th	16th	17th	18th
G1	13	16	14	12	11	17	15	10	18	4	3	7	5	2	8	6	1	9
G2	18	15	17	14	16	12	13	11	10	8	4	5	6	3	7	9	2	1
G3	18	15	17	12	14	16	13	11	9	10	6	8	5	3	7	4	2	1
G4	18	15	12	17	14	9	16	6	8	11	3	5	13	7	2	4	1	10
G5	18	15	12	17	9	6	3	8	5	14	7	2	4	16	11	1	13	10
G6	18	15	9	6	3	8	12	5	17	7	2	4	14	1	11	13	16	10
G7	18	9	6	15	3	8	5	12	2	7	4	17	14	1	16	13	11	10
G8	9	6	18	3	8	15	5	7	2	12	4	17	1	14	16	13	11	10
G9	9	6	18	3	8	5	15	7	2	12	4	17	14	1	16	13	11	10
G10	9	6	18	3	8	15	5	7	2	12	17	4	14	1	16	13	11	10
G11	9	6	18	3	15	8	5	7	17	2	12	4	14	1	16	13	11	10
G12	9	6	18	3	15	8	5	7	2	17	12	4	14	1	16	13	11	10
G13	9	6	18	3	15	8	5	7	2	17	4	12	14	1	16	13	11	10
G14	9	6	18	3	15	8	5	7	2	12	4	17	14	1	16	13	11	10
G15	9	6	18	3	8	15	5	7	2	4	12	17	14	1	16	13	11	10
G16	9	6	18	3	8	15	5	7	2	4	12	17	14	1	16	13	11	10
G17	9	6	18	3	8	15	5	7	2	4	12	17	14	1	16	13	11	10
G18	9	6	18	3	8	15	5	7	2	12	4	17	14	1	16	13	11	10
G19	9	6	18	3	8	15	5	7	2	4	12	17	14	1	16	13	11	10
G20	9	6	18	3	8	15	5	7	2	4	12	17	14	1	16	13	11	10
G21	9	6	18	3	8	15	5	7	2	4	12	17	14	1	16	13	11	10
G22	9	6	18	3	8	15	5	7	2	4	12	17	14	1	16	13	11	10
G23	9	6	18	3	8	15	5	7	2	4	12	17	14	1	16	13	11	10
G24	6	9	18	3	15	8	5	7	2	4	12	17	14	1	16	13	11	10
G25	6	9	3	18	15	8	5	2	7	4	12	17	14	1	16	13	11	10
G26	6	9	3	18	15	8	5	2	7	4	12	17	14	1	16	13	11	10
G27	6	9	3	18	15	8	5	2	7	4	12	17	14	1	16	13	11	10
G28	9	6	3	18	15	8	5	2	7	4	12	17	14	1	16	13	11	10
G29	9	6	3	18	15	8	5	2	7	4	12	17	14	1	16	13	11	10
G30	9	6	3	18	15	8	5	2	7	4	12	17	14	1	16	13	11	10

Colour blocking	
1	1-pSsHoH
2	2-pSsMoH
3	3-pSsLoH
4	4-pMsHoH
5	5-pMsMoH
6	6-pMsLoH
7	7-pLAsHoH
8	8-pLAsMoH
9	9-pLAsLoH
10	10-pSsHoL
11	11-pSsMoL
12	12-pSsLoL
13	13-pMsHoL
14	14-pMsMoL
15	15-pMsLoL
16	16-pLAsHoL
17	17-pLAsMoL
18	18-pLAsLoL

Figure 3.5. Descending rank order of numbered scenarios for genic variance in each generation. Darker shades correspond to the low outcrossing scenarios and lighter to the high outcrossing scenarios.

3.4.3. Cost per unit gain

The preceding sections have been concerned with finding the conditions where the rate of genetic gain can be maximised and the main variable parameters within a simple recurrent selection programme that are within the control of the breeder discussed. In this final section, the economic cost of the different scenarios is estimated so that the cost-effectiveness of high gain breeding schemes can be compared.

The cumulative direct costs of running each scenario over periods ranging from 5 to 30 years were calculated as reported in Appendix D and divided by relevant cumulative genetic gain to estimate the cost per unit genetic gain, which is reported for short-term (5

year) and long-term (30-year) time horizons in Figure 3.6. The first point about these calculations is that although there are 18 different scenarios, there are only three different cost levels. This is because varying selection intensity bears no cost and outcrossing rate, though it may vary due to environmental factors is out of the control of the breeder, cannot be forced to any particular value other than zero (where no captive pollinators are supplied) so as a direct consequence of the assumptions made in the cost model, the sole driver of scenario cost is population size. Therefore, the question to be answered is whether increasing the population size is worthwhile.

Over a 5-year period, scenario 13-**pM**sHoL gives the highest genetic gain**.** Although this theoretically superior scenario is twice as cost effective as scenario 16-**pLA**sHoL (gain= 2.9), the third-ranked scenario **10-pSsHoL** (gain= 2.79) achieves 95.6% of the total possible gain at just over 20% of the cost. Over a 30-year period, scenario 7-pLAsHoH gives the highest genetic gain, but similarly to above, scenario 3-pSsLoH, though ranked 10th for gain, achieves 88% of the total potential gain of scenario 7 at just 11.4% of the cost.

The general observation that holds true for all timepoints beyond generation 3 is that the scenario giving the highest genetic gain is never the most cost-effective and the last few percent of potential gain come subject to a law of rapidly diminishing returns. This means that in the absence of strong competition to win market share in our hypothetical niche market, there are strong economic incentives to resist attempting a large scale-up of the breeding programme.

However, it should be noted that this cost saving comes at a price when optimising for short-term gain – scenario 10 is the worst scenario (18th) in terms of maintaining genic variance, so these cost-saving measures should only be taken if there is a specific intention to bring new variation into the scheme before the rate of genetic gain is

predicted to decline too far. In contrast, scenario 3, which is cost-effective over the long-term, entails no such loss of genic variance as it ranks 3rd for genic variance at year 30.

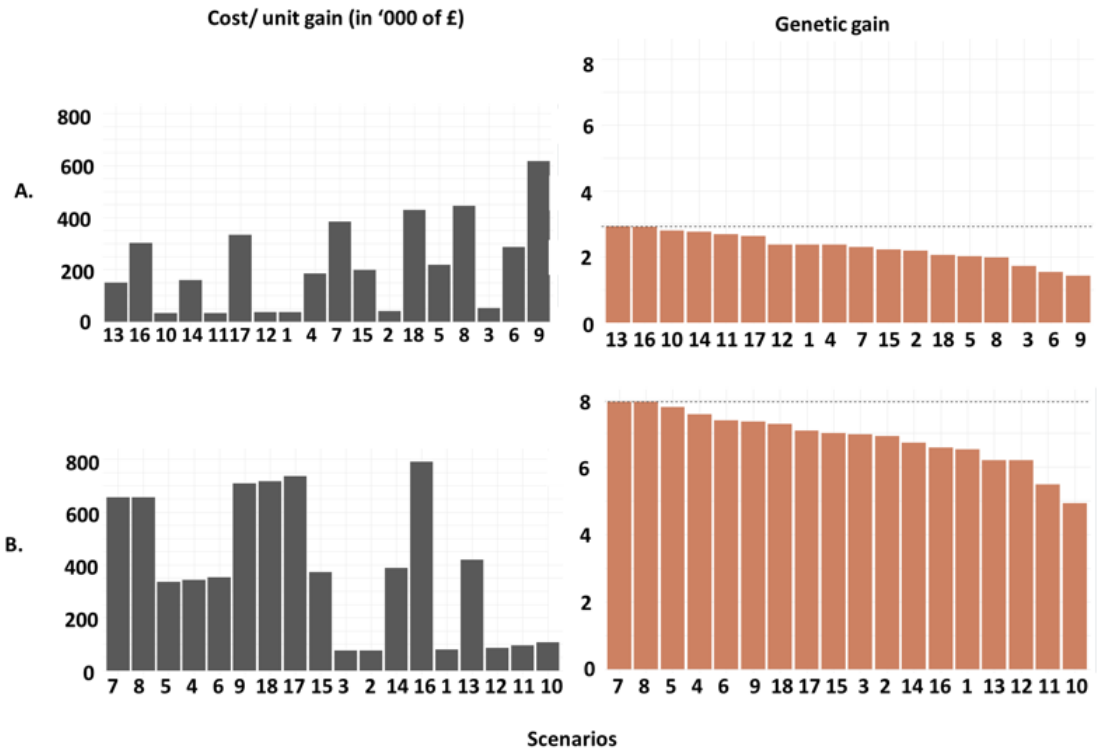


Figure 3.6. Cumulative cost in £000's per unit gain and cumulative gain between scenarios at milestones of 5 (**A.)** and 30 (**B.)** years/generations. On the x-axis, the 18 scenarios are plotted in descending order of cumulative genetic gain. A grey horizontal line is placed at the level of the highest gain scenario for each timepoint.

3.5. Conclusion

The modifiable and reproducible code developed and presented here constitutes a general framework for identifying the most powerful combinations of recurrent selection population size and selection intensity relevant to a given breeders time horizons and trait of interest. As such, it fills a decades-long gap in the faba bean breeding literature since the theoretical modelling of aspects of synthetic faba bean breeding by Wright (1977) and Gallais (1992).

In the particular case study modelled here - a recurrent selection scheme resembling the MRSS scheme implemented in practice in Chapter 2 - some important general insights were obtained. First, that within a closed genepool, there is a tradeoff between short-term speed of progress and long-term limits on magnitude of progress. Second, the degree of long vs short-term tradeoff was independently lessened by high outcrossing rates, high population sizes and low selection intensities.

Application of these general insights to the specific case in point showed that short-term speed of progress is favoured by high selection intensity in small to medium population sizes with low outcrossing rate. Furthermore, since programme cost increases linearly with population size, but genetic gain does not, the most cost-effective scenarios are always those involving small population size. Such scenarios do however lead to rapid loss of genic variance and rapidly diminishing rates of genetic gain beyond 5 years, so are not sustainable in the long term without introduction of new diversity.

Pragmatically, it is unlikely that any breeding scheme can run successfully for 30 years without introducing new germplasm. This is certainly the case where there may be (as predicted) a structural shift in climate over time or emergence of a new or substantially altered pest or pathogen that exerts a distinct new type of selection pressure. The need to bring new sources of trait variation into the scheme in response to long-term shift in biotic and abiotic pressures in conjunction with the long-term limits discussed above would suggest that future scheme design should incorporate the possibility of regularly introducing new adaptive germplasm identified through academic or pre-breeding research.

A limitation of these simulations is that they do not model hybrid vigour nor dominance or epistasis. In future simulation studies, it would be interesting to model the effects of hybrid vigour as the magnitude of hybrid vigour in faba bean is high and

well documented – mid-parent F_1 heterosis between 20 and 63 % in Bond, Fyfe, and Toynbee-Clarke (1964) and between 40 and 119% in Zeid *et al.* (2004). For complex traits like yield, additive and non-additive genetic effects can change the genetic variation and the genetic gain by influencing the response to selection (Cooper *et al*., 2009).

Another limitation of these simulations is that phenotypic predictions are based on a constant and unchanging selection environment. Understanding of Genotype by Environment interactions (G*E) is very important in plant breeding as plant performance across different environments can be highly genotype dependent (M. Cooper and Delacy, 1994). Incorporation of G*E can help predict phenotype more accurately and new tools can incorporate these interactions to prediction models (van Eeuwijk *et al*., 2019). The use of more G*E parameters in the simulations and study of the prediction accuracy would help to make simulations more robust to known variability in environmental conditions and to climate change (Heslot *et al*., 2014).

Simulations for multiple traits would also be a further step of this study. Through breeding for one trait it is possible that another trait is being purged from the population as some traits are negatively correlated (Tovignan *et al*., 2016) or simply lost due to bottleneck effect. Advances in our understanding of genetic architecture of a growing variety of traits (e.g. anti-nutritional factors) or opens up future possibilities to simulate and implement simultaneous selection on multiple traits.

3.6. References

Bond, D., Fyfe, J., and Toynbee-Clarke, G. 1966. Male sterility in field beans (*Vicia faba* L.). IV. Use of cytoplasmic male sterility in the production of F1 hybrids, and their performance in trials. The Journal of Agricultural Science. **66**(3): 369-377. doi:10.1017/S002185960006367X

Bulmer, M. G. 1976. The Effect of Selection on Genetic Variability: A Simulation Study. Genetical Research. **28**: 101-117. https://doi.org/10.1017/s0016672300016797.

Collins, C., and S. Phelan. 2019. Crops Costs and Returns 2019. Teagasc.Ie. https://www.teagasc.ie/media/website/publications/2019/Crops-Costs-and-Returns-2019.pdf.

Cooper, M., and I. H. Delacy. 1994. Relationships among Analytical Methods Used to Study Genotypic Variation and Genotype-by-Environment Interaction in Plant Breeding Multi-Environment Experiments. Theoretical and Applied Genetics. **88** (5): 561–72. https://doi.org/10.1007/BF01240919

Cooper, M., F. A. van Eeuwijk, G.e L. Hammer, D. W. Podlich, and C. Messina. 2009. Modeling QTL for Complex Traits: Detection and Context for Plant Breeding. Current Opinion in Plant Biology. **12** (2): 231–40. https://doi.org/10.1016/j.pbi.2009.01.006

Faux, A.-M., G. Gorjanc, R. C. Gaynor, M. Battagin, S. M. Edwards, D. L. Wilson, S. J. Hearne, S. Gonen, and J. M. Hickey. 2016. AlphaSim: Software for Breeding Program Simulation. The Plant Genome. **9** (3): 1-14. https://doi.org/10.3835/plantgenome2016.02.0013.

Gallais, A. 1992. Pourquoi Faire Des Variétés Synthétiques? Agronomie. **12**: 601-609. https://doi.org/10.1051/agro:19920803.

Gaynor, C., G. Gorjanc and J. M. Hickey. 2020. AlphaSimR: An R-package for Breeding Program Simulations. biorxiv. https://doi.org/10.1101/2020.08.10.245167

Gorjanc, G., R. C. Gaynor, and J. M. Hickey. 2018. Optimal Cross Selection for Long-Term Genetic Gain in Two-Part Programs with Rapid Recurrent Genomic Selection. Theoretical and Applied Genetics. **131** (9): 1953–66. https://doi.org/10.1007/s00122-018-3125-3

Heslot, N., D. Akdemir, M. E. Sorrells, and J.-L. Jannink. 2014. Integrating Environmental Covariates and Crop Modeling into the Genomic Selection Framework to Predict Genotype by Environment Interactions. Theoretical and Applied Genetics. **127** (2): 463–80. https://doi.org/10.1007/s00122-013-2231-5

Tovignan, T. K., D. Fonceka, I. Ndoye, N. Cisse, and D. Luquet. 2016. The Sowing Date and Post-Flowering Water Status Affect the Sugar and Grain Production of Photoperiodic, Sweet Sorghum through the Regulation of Sink Size and Leaf Area Dynamics. Field Crops Research. **192**: 67-77. https://doi.org/10.1016/j.fcr.2016.04.015.

Van Eeuwijk, F. A. van, D. Bustos-Korts, E. J. Millet, M. P. Boer, W. Kruijer, A. Thompson, M. Malosetti, et al. 2019. Modelling Strategies for Assessing and Increasing the Effectiveness of New Phenotyping Techniques in Plant Breeding. Plant Science: An International Journal of Experimental Plant Biology. **282**: 23–39. https://doi.org/10.1016/j.plantsci.2018.06.018

Wright, A. 1977. Inbreeding in synthetic varieties of field beans (*Vicia faba* L.). The Journal of Agricultural Science. **89**(2): 495-501. https://doi.org/10.1017/S0021859600028410

Yabe, S., H. Iwata, and J.-L. Jannink. 2017. A Simple Package to Script and Simulate Breeding Schemes: The Breeding Scheme Language. Crop Science. **57**(3): 1347-1354. https://doi.org/10.2135/cropsci2016.06.0538.

Zeid, M., C.-C. Schön, and W. Link. 2004. Hybrid Performance and AFLP- Based Genetic Similarity in Faba Bean. Euphytica. **139**: 207-216. https://doi.org/10.1007/s10681-004-3156-5.

4. Chapter 4 - Longitudinal study of diversity, heterozygosity and signatures of selection in the modified MRSS

4.1. Abstract

It has previously been noted that the modified recurrent selection scheme described in Chapter 2, allows selfing and outcrossing to go on unsupervised in parallel. The actual rate of outcrossing during selection does matter, as illustrated by simulation in Chapter 3, with high outcrossing rates increasing maximum genetic gain possible from a given genepool at cost of a slower rate of gain and slower rate of loss of genic diversity. In this chapter, we characterise the founding population and each of the selected generations at the genotypic level using genome-wide genotyping to make a retrospective empirical assessment of diversity loss and to determine what the balance between the opposing forces of outcrossing and inbreeding is in practice.

The genotyping approach used is of sufficiently high density to not only make global estimates of diversity, heterozygosity and outcrossing rates across different generations, but to scan through the genome looking for the signatures of selection. Chapter 2 described how progeny of the top-yielding individual plants from an outcrossing population were selected to form the next outcrossing generation for 3 consecutive seasons. It is hypothesised that over time, the unseen hand of selection will have selectively removed unfavourable alleles from the diverse founding population. Now, with the aid of a high-density SNP genotyping array, we can retrospectively measure changes in diversity and map locations in the genome bearing the signature of selection.

Genetic diversity describes the differences among individuals and can either decrease or increase depending on the matting pattern- inbreeding or outbreeding. The molecular data confirmed that outcrossing and inbreeding (about 60% inbreeding in each

generation) were happening at the same time. Comparison of heterozygosity rates of selected and unselected cohorts from each generation showed that selection greatly favoured highly heterozygous individuals, suggesting that blind phenotypic selection on yield was driven largely by hybrid vigour.

However, if selection was exclusively due to hybrid vigour and hence by the global level of heterozygosity, we would not expect to observe position-dependent selective sweeps. In fact, a genome-wide scan of F_{ST} showed a number of loci that were under selection, and although we can only speculate as to what the underlying traits resulting in such rapid and locus-specific shifts in allele frequency are, this does demonstrate that it is possible to simultaneously select on both underlying additive genetic variation and maintain high levels of heterozygosity and hybrid vigour.

4.2. Introduction

So far in this thesis, we documented a new scheme selecting for heterozygosity at the beginning and for yield in the following cycles, finding via empirical trials reported in Chapter 2 that it is possible to get apparently quite rapid genetic gain over the early generations of selection as predicted via simulations carried out in Chapter 3. In this chapter, we want to look genotypically at how the structure of the population has responded to the selection pressure applied.

However, even if yield has been shown empirically to respond to selection, several questions regarding the accuracy of the assumptions made in the design of the MRSS and in the simulations and the reasons for the observed gain remain unanswered. One key assumption is that by introducing captive bumblebees in every generation, we maintain a relatively high outcrossing rate, but what that outcrossing rate might be in practice is not known, even though the simulations in Chapter 3 show that variation in

this parameter has marked effects on rates of genetic gain and loss of diversity. Since we do not know the effective outcrossing rate, we also have little idea whether hybrid vigour or additive genetic effects have the greater influence on the trait under selection (individual plant seed yield) and hence the genetic basis on which high merit individuals are selected.

4.2.1. Assessing responses to selection at the genomic level

Nowadays we have very powerful tools that make it feasible to track population genetic changes in the very detailed way - at the limit, it is possible to survey genome-wide genotypes of every individual in each generation - and this opens up the opportunity for some very powerful insights.

A good example of the potential to gain gene-level insights from selection experiments is provided by the Illinois maize long-term selection experiment. In 1896 Hopkins initiated this long-term experiment to determine whether selection can change the chemical composition of corn (Hopkins, 1899) and this experiment continues until now. Forward mass selection started on the open-pollinated variety 'Burr's White' and oil and protein concentration of 163 corn ears were analysed and placed in four categories: high (IHO)and low (ILO) oil and high (IHP) and low (ILP) protein content. Forward mass selection for IHO and ILO performed for 100 generations and reverse selection initiated after the 48th generation and created four strains of Reverse high and low protein, and reverse high and low oil content (Dudley and Lambert, 2010). The evaluation trials were evaluating the latest selection along with remnant seed of previous generations. The progress of forward and reverse selection is shown in Figure 4.1.

The genetic response of this long-term selection for oil concentration of maize kernel was studied by creating a cross population between IHO and ILO 70th generation,

randomly mating for 10 generations (Laurie *et al.*, 2004). 500 selfing lines derived from the 10th generation and were genotyped with 488 markers. This study provides information on the genetic architecture of oil variation and detected QTLs with positive effects to selection. Although oil concentration is a polygenic trait, the QTL effects of the lines compared to the parental lines suggested that more than 50 QTLs could be involved in the oil variation. This study used a small number of marker and nonetheless adequate to identify QTLs, setting expectations of discovery of more associations in similar studies using a greater number of markers.

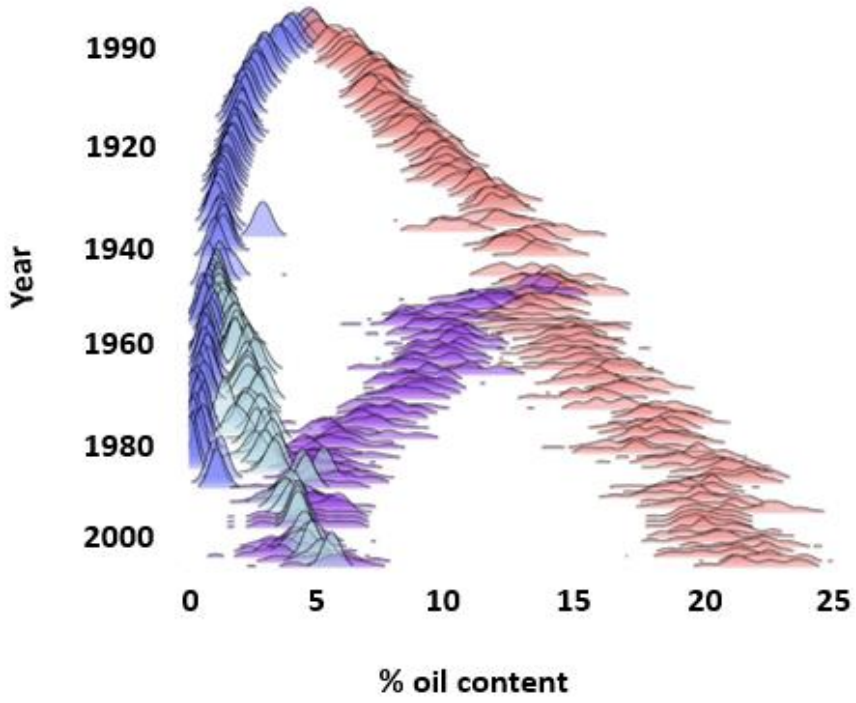


Figure 4.1. Illinois long-term selection experiment. Dark Blue phenotypic distributions for oil content of Illinois Low Oil (ILO) selections from 1896 - 2004; Dark Red: phenotypic distributions for oil content of Illinois High Oil lines over the same period. Light Blue: phenotypic distributions of reverse selections (Reverse Low Oil – RLO) practiced after 48 generations of forward selection; Purple: Reverse High Oil selections (source: Twitter Coop, 2019).

As described in Chapter 1, F statistics were developed as a means to quantify reductions in heterozygosity compared to a large, freely intermating population in Hardy-Weinberg equilibrium that can come about due to inbreeding, drift and natural or artificial selection. F-statistics can be applied at the whole genome level to make inferences about the overall degree of differentiation between populations or at the individual locus level to make inferences about the likelihood of that locus being directly (or more likely indirectly) the target of selection. Targets of positive selection were identified on 20,000 annotated genes from 56 *Medicago truncaltula* accessions (1% of sampled genes) (Paape *et al*., 2013). F_{ST} outliers (higher values- over 99 percentile) were determining loci for positive selection in sorghum research and markers with F_{ST} lower than the 99% confidence interval were associated with loci for balancing selection (Bouchet *et al*., 2012). The genome-wide fixation index has also been used for characterizing divergence between wild and domesticated types of barley (Russell *et al*., 2016; Pankin *et al*., 2018), rice (Huang *et al*., 2012) and maize (Hufford *et al*., 2012).

In chickpeas, F_{ST} have been used to identify genomic regions positively associated with yield, biomass and seed number (Lake *et al*., 2016). 20 chickpea lines with various agronomic background (adaptation, yield, seed type and phenology) were compared in six different locations in South Australia on 2013 and 2014. F_{ST}s were calculated on contrasting phenotypic groups and revealed 4 genomic regions with positive selection for seed number and five regions for yield and biomass. Genetic differentiation using F_{ST} was also used to identify signatures of selection on soybean (Kim *et al*., 2019), and common bean (Papa *et al*., 2007).

Unlike maize, where the identification of causative polymorphisms in fatty acid biosynthesis pathways underlying oil content QTL fixed under directional selection was perhaps possible due to the intensity of research on both the underlying biochemical

pathways and the excellent genomic tools available in that species, faba bean selection schemes have not been studied so intensively. However, the potential to do so is illustrated by a recent study in the Pacific North-West region of the United States, where diverse winter-hardy source material were subjected to up to four generations of mass selection (with the main selective force being over-winter survival) with free inter-mating among entries (Landry *et al.*, 2017). Despite the small number of selection cycles, a significant shift in over-winter survival and yield was observed in side-by-side evaluation of the pre- and post-selection bulks in a replicated field trial.

To date, there is no study using F-statistics as a tool to identify regions of positive selection in faba bean populations under natural or artificial selection. Since 2017, cost effective, high density genotyping has become a reality in faba bean (O' Sullivan *et al.*, 2019) and this opens up the opportunity to look at individual genome-wide genotypes of samples of the MRSS populations developed here through each generation of selection aiming to compare allele frequencies pre- and post-selection, estimate the effective rate of outcrossing and finally identify putative signatures of locus-specific selection.

4.3. Materials and Methods

The same colour scheme that was adopted in Chapter 2 to represent the advancement of the MRSS through consecutive selection cycles will be used again here. However, since in this Chapter we will need to distinguish between selected and unselected cohorts of individuals, we use shapes (as shown in Table 4.1.) to denote selection status.

Table 4.1. Symbols and colours adopted to represent the different generations and individuals (selected/unselected) in the MRSS.

Generation	Selected	Unselected
O1		
O2		
O4		
O6		

4.3.1. Establishment of a longitudinal panel of DNA samples from the MRSS

The population was structured following the modified recurrent selection scheme as described in Chapter 2. In every selection generation, DNA was extracted from each individual plant with a modified CTAB method and with a silica-based DNA extraction kit supplied by Qiagen. The number of DNA samples thus generated was too large to genotype in its entirety (O_2 – 1,454 , O_4 – 1,593 , O_6 – 1,611) so once the single plant yields for each selective generation had been determined, DNA samples corresponding to all the selected plants (O_2 – 54, O_4 – 59, O_6 – 52) and a set of random unselected plants (O_2 – 91 out of 1,454 , O_4 – 59 out of 1,593 , O_6 – 54 out of 1,611) were cherry-picked for genome-wide genotyping so that the effect of selection within each generation and across generations could be studied.

4.3.2. SNP chip and marker quality control

A *Vicia faba* SNP genotyping array (Axiom_Vfaba_v2), manufactured by Affymetrix (Thermo Fisher Scientific) and designed by Angra *et al*. (2020-unpublished) with 60,130 probesets, corresponding to 57,312 markers, was used to genotype all the selected plants

along with the random unselected sample from each selection cycle. Genotype calling was conducted using the Best Practice Workflow in the Axiom Analysis Suite software set to default parameters. Of the 57,312 markers, 60.8% (34,882) were characterised as best and recommended and 42.9% (24,602) were designated poly-high-resolution markers. 10,784 of these markers were mapped and well-distributed on the six faba bean chromosomes (Figure 4.2.). The array design and mapping represent recent unpublished work of Dr. Deepti Angra and Prof Donal O'Sullivan together with international collaborators from INRA and Aarhus University who provided SNP flanking sequences for the array design and Dr. Hamid Khazaei from the University of Saskatchewan who provided a four-way cross RIL population (Khazaei *et al.*, 2018). This work will be described in detail elsewhere in due course; this state-of-the-art genotyping tool and associated map data were made available to me prior to publication as an enabling tool for this study of signatures of selection; all DNA extraction, genotype calling and downstream analysis of MRSS population genotypes are my own work.

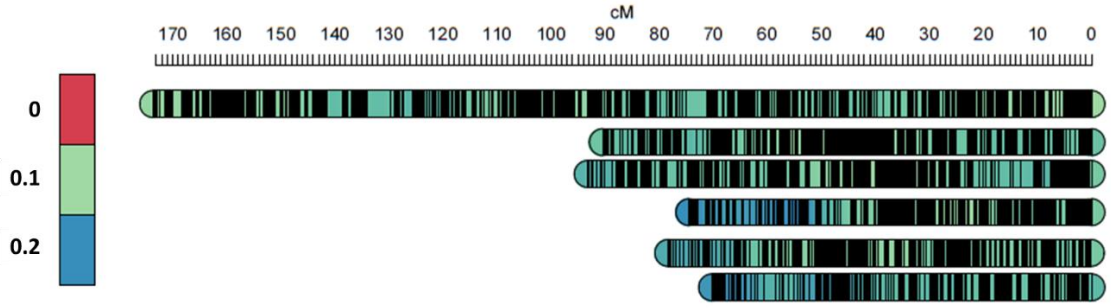


Figure 4.2. A 10,784 locus SNP-based map, showing SNP density in the six faba bean chromosomes. The colours represent the marker density in cM/locus from 0 (in red) to 0.2 (in blue). Graph created using LinkageMapView in Rstudio (Ouellette et al., 2018).

Genomic analysis was performed on the genotyping results, beginning with using packages to clean, summarize and prepare the genomic datasets for further analysis, as well as to estimate several population genetics parameters. Using *snpReady* R package

(Granato *et al*., 2018), the data and quality controlled, following which missing data were imputed. Individuals that had more than 0.5% of missing markers with >5% missing data were removed. Only the polymorphic markers are informative and were kept for further analysis steps. Markers with minor allele frequency <0.01 were removed as changes in frequencies of rare alleles cannot be reliably estimated. This QC step removed 14,382 markers by MAF=0.01 and 182 markers by call rate=0.95.

4.3.3. Calculating loss of diversity and outbreeding status through generations

Loss of diversity

All markers that are polymorphic in any generation were considered. The allele counts per generation calculated by applying the following simple formula in Microsoft (MS) Excel:

$$Allele\ number = (2 * \#\ of\ polymorphic\ loci) + (1 * \#\ of\ monomorphic\ loci)$$

Outbreeding status

As the maternal pedigree was recorded of the selected individuals, and genotyping carried out in alternate cycles, the genotype of each selected offspring could be compared with the genotype of its grand-maternal progenitor. A different category was assigned to each offspring locus according to whether they were identical in state to the grandmaternal genotype (Allele Conservation), had changed state from monomorphic to polymorphic (Allele Increase) or from polymorphic to monomorphic (Allele Loss).

The Net number of Allele Gains per offspring individual was calculated in MS Excel with the formula:

$$Net\ Allele\ Gain = \#\ gains - \#losses$$

The outcome of this function could either be a positive value, meaning allele increase, or a negative value, meaning allele loss. Broadly speaking, inbreeding would lead to allele loss (50% of polymorphic loci per generation of inbreeding), whereas outcrossing (at least with unrelated individuals) was likely to result in net allele gain and the position of an individual on the net allele gain spectrum was interpreted as a proxy measure of the likelihood that the individual in question was a product of selfing or outcrossing.

4.3.4. Calculating heterozygosity rate and inbreeding coefficient

The observed heterozygosity rate (Hoi), inbreeding coefficient (Fi) were tracked in each generation and were estimated using the *popgen* function from *snpReady*, by calculating the H_{Oi} and the F_i using the equations:

$Hoi = \frac{nHi}{m}$ and $Fi = O(Hi) - \frac{E(H)}{m - E(H)}$ where,

- nHi is the number of heterozygous genotypes (of type $A_1A_2A_1A_2$ or $A_2A_1A_2A_1$) in the individual i
- m is the number of markers
- $O(Hi)$ is the observed homozygosity for individual i
- $E(H) = Sj1 - 2pj(1 - pj)$ is the expected homozygosity across all SNPs

4.3.5. Calculating changes in allele frequency changes between the first and the last generation

Outcrossing lead to changes in the allele frequencies of the populations within the generations. Using biallelic SNPs, the allele in higher frequency (>0.5) in a locus is named reference allele and the allele with the lowest frequency (<0.5) minor allele. A

simple excel formula was used to calculate allele frequencies of A and B alleles in each locus of each population:

$$Allele\ frequency\ of\ allele\ A:$$

$$\frac{[(2 * \#\ of A\ in \text{ the population}) + (\#\ \text{of AB}\ in\ the\ population)]}{[(2 * total\ number\ of\ individuals\ in\ the\ population) - (2 * NAs)]}$$

$$Allele\ frequency\ of\ allele\ B:$$

$$\frac{[(2 * \#\ of B\ in \text{ the population}) + (\#\ \text{of AB}\ in\ the\ population)]}{[(2 * total\ number\ of\ individuals\ in\ the\ population) - (2 * NAs)]}$$

The frequency of the minor allele in the O_1sel generation was tracked in the final selection O_6 and the status this particular allele could either have remained minor, been eliminated, became reference or even fixed, although fixation of a minor allele is not expected after applying a strict 3.7% of selection.

4.3.6. F- statistics to identify positive signatures of selection

For the purpose of detecting signatures of selection, only the founding population (O_1sel) and most recent selection generation (O_6) were compared. F_{ST} values based on allele frequencies of the O_1sel and O_6 generations were calculated using the *hierfstat* R package (Goudet, 2005). This package allows the calculation of hierarchical F-statistics from either haploid or diploid genotyping data, using the algorithm as published by Yang (1998).

The function that was used estimated per locus and per population rarefied allelic counts, allelic frequencies, observed heterozygosity, genetic diversity, and F_{ST} values. It

is prerequisite to create a data file with the genotype of the individuals in the populations along with the trait of selection.

The F_{ST} values estimated following the function:

$$Fst = Dst/Ht\ ,$$

where:

- Dst is the amount of gene diversity among samples and is calculated:

$$Dst = Ht - Hs$$

- Ht is the overall gene diversity and is calculated:

$$Ht = 1 - \Sigma_i p_i^2 + Hs/(\tilde{}nnp) - Ho/(2\tilde{}nnp)$$

$$n\tilde{} = np/\Sigma_k\ ^{1/nk}$$

$$\bar{}p_i\ ^2 = \Sigma_k\ p_{ki}^{2/np}$$

  where, np is the number of samples and pki the proportion of homozygote i in a sample k

- Hs is the within population gene diversity (expected heterozygosity) and is calculated:

$$Hs = \tilde{}n/(\tilde{}n - 1)[1 - \Sigma_i p_{ki}^2 - Ho/2\tilde{}n]$$

4.3.7. Linkage disequilibrium (LD) decay between pairs of markers

The LD decay was calculated for each generation by the v2.9. version of the package *sommer* based on a marker matrix and a genetic map including distances in cM (Laidò *et al*., 2014). LD values between all pairs of linked markers. The r^2 critical value was estimated and plotted against the genetic distance between markers. Finally, the rate of

LD decay was represented by a smooth line drawn using second-degree locally weighted polynomial regression (LOESS).

4.3.8. Genome-wide association analysis following a multiple locus mixed linear model (MLMM)

Only one phenotype was available – unreplicated individual plant seed yield. GWAS scans for seed yield following a multiple locus mixed linear model (MLMM) implemented in the R package *GAPIT* was performed (Segura et al., 2012). The MLMM included both fixed and random effects and combines information about relationships among genotypes. A kinship matrix (K) is used to relate variance and covariance between individuals. The MLMM is described by the function:

- $Y = X\beta + Zu + e$ where,
- Y is the vector of observed phenotypes
- β is an unknown vector containing fixed effects, including the genetic marker, population structure (Q), and the intercept
- u is an unknown vector of random additive genetic effects from multiple background QTL for individuals/lines
- X and Z are the known design matrices
- e is the unobserved vector of residuals

The output scans were presented in the form of Manhattan plots, with marker ordering according to the map described in section 4.3.2.

More GWAS analysis options like Mixed Linear Model (MLM) (Zhang et al., 2010) and Fixed and random model Circulating Probability Unification (FarmCPU) (Liu et al., 2016) were also explored using *GAPIT*, however none of the methods detected candidate loci for the assessed trait.

4.4. Results

4.4.1. Quantifying intergeneration changes in genetic loss of maternal lineages at different possible selection intensities

Maternal genotypes share overlapping subsets of alleles, meaning that when one maternal genotype is lost, it does not necessarily follow that all the alleles contained therein are lost. Pedigree records of the genotypes were kept after the O_2 selection cycle, as described in previous chapter. Before genotyping data were available, as genotyping was conducted only after the O_6 generation had been completed, a pedigree graph was generated as a means to set expectations of how different selection intensities could impact the population diversity and illuminate how severe or not 3.7% percent of selection was in terms of losing genetic diversity. By ranking all the individuals from O_4 and O_6 in descending order of seed yield and calculating at each point as we progress down through the list the cumulative number of ancestral O_2 maternal genotypes that are represented in each generation, the pedigree record graph indicates that the 3.7% level of selection (the actual cut-off used in practice) maintained 60.7% of O_2 genotypes in the O_4 cycle and 35.7% in the O_6 (Figure 4.3.). At one extreme, keeping just over a quarter of the population each generation would have resulted in retention of all maternal lineages together with their associated allelic diversity. However, there is a steep drop in the number of maternal lineages retained as we fall below 12.5% selection intensity. In agreement with the findings of the simulations on Chapter 3, the graph depicts that the greater the selection intensity the less the diversity in the subsequent generations.

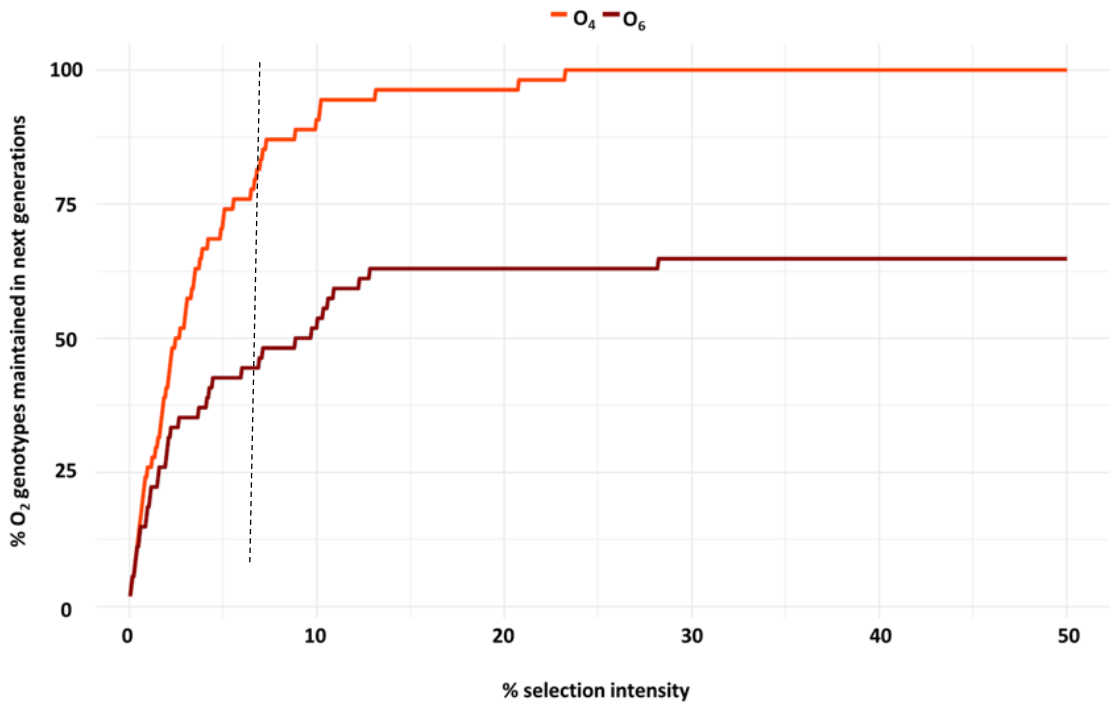


Figure 4.3. The effect of the selection intensity on the maintenance of the genetic diversity as represented by number of O_2 maternal lineages retained. 100% of genotypes is the total number of lineages present in the O_2 generation. The 3.7% selection intensity actually used in practice is represented by the vertical black dashed line.

4.4.1.1. Quantification of loss of diversity through selection

Once genotyping data were available, it was possible to directly quantify the loss of diversity in each selective step of the MRSS. The fact that allele losses were consistently less in the selected compared to the unselected sets indicates that the selection criterion applied (high individual plant seed yield) was indirectly selecting for high heterozygosity. The significance of heterozygosity-driven hybrid vigour as an indirect target of selection will be discussed later; here, it should be noted that allelic diversity is thus maintained somewhat higher than expected in each generation of the breeding scheme. Figure 4.4. shows that between O_1 and O_2 generations, 5% of unique alleles that were private to some of these lost genotypes of O_1sel generation are gone from the population. Another 5.5%

was lost from O_2 to O_4 generation with the rate of allele loss (2.8%) appearing to diminish in the third selection cycle (O_4 to O_6). This analysis indicated that the raw number of alleles retained steadily decreases under the influence of selection, in agreement with simulations of genic diversity loss shown in Chapter 3.

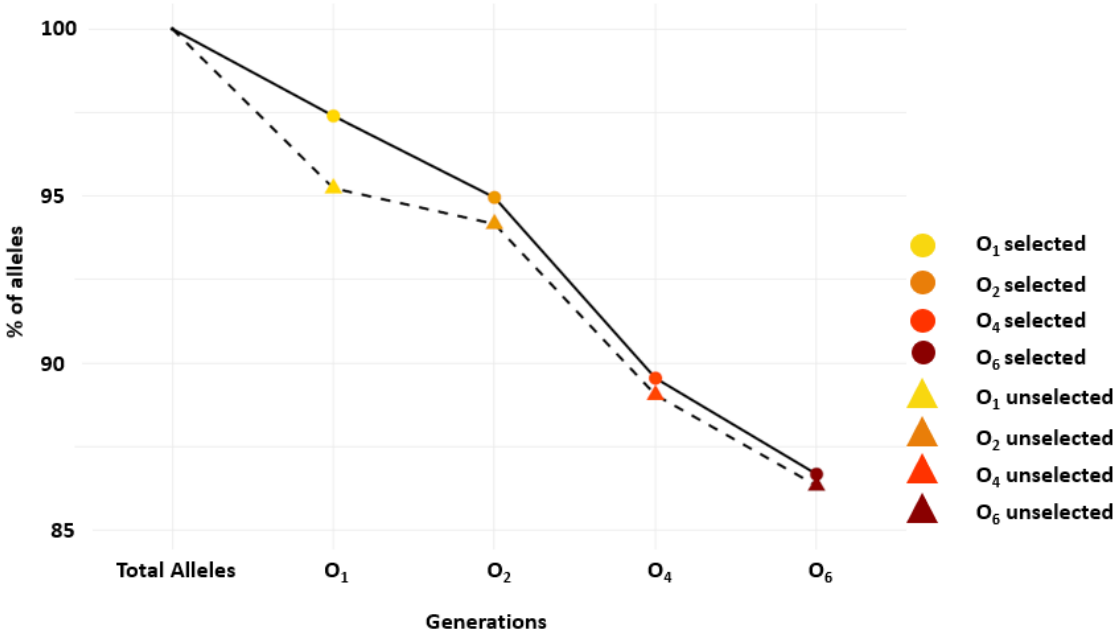


Figure 4.4. Total Allele number in cohorts of genotyped individuals over generations of selection expressed as a percentage of the total number of alleles in the O_1 gene pool (100%). The continuous line refers to the allele decrease of the selected individuals and the dashed line to the allele decrease of unselected samples that were genotyped.

4.4.2. Quantifying intergeneration changes based on genomic analysis of pedigree lineages

4.4.2.1. Quantifying inbreeding versus outcrossing based on maternal genomic data pedigree

The next question regarding the evolutionary forces at play in a "nature-assisted" recombining population that can be addressed using molecular data is the effective outcrossing rate. Through selection, alleles have been lost, but not at a huge rate. It was shown in Chapter 3 that high outcrossing rates would delay the loss of allelic diversity for the same population size and selection threshold by ensuring that the same alleles were spread over a greater number of individuals. Thus, diversity in a population at a given point in time is profoundly driven by the rate of outcrossing. In the MRSS implemented here, a certain level of outcrossing was ensured in each generation by placing captive bumblebee colonies within the glasshouse/cage environment at the onset of flowering, but the effectiveness of those bumblebees in driving a high rate of outcrossing is uncertain as this depends on the health and size of the colony, the effect of environmental parameters such as temperature and rainfall on foraging behaviour, the synchrony of flowering within the population, overall and relative attractiveness of the flowers and the flower visitation behaviour of the pollinator species used (*Bombus terrestris* in this case), which is documented to conduct nectar robbing as well as frontal visitation (Marzinzig et al, 2018). None of these parameters are in the full control of the experimenter.

As mentioned in Chapter 2, the first selection round was carried out to maximize heterozygosity after genotyping the initial population with 40 SNPs. A definitive calculation of heterozygosity and inbreeding coefficient through all the generations could be later calculated based on array-based genotyping and these results are displayed in

Figure 4.5. The key result in this Figure is that apart from the artificially high heterozygosity in O_{1sel}, which can be explained by the fact that O_{1sel} consisted wholly of F_1 hybrids between deliberately contrasting inbred parents, in further selections after O_2, from which point onwards all outcrossing was bee-assisted, result in increased heterozygosity and decreasing inbreeding coefficient. The inbreeding coefficient is the inverse of heterozygosity and refers to the probability that two alleles are identical by decent at a given locus, so where the general trend between O_2 and O_6 is towards increasing heterozygosity, there is a mirroring trend for the inbreeding coefficient to decline. The increasing heterozygosity between O_2 and O_6 suggests that there is enough outbred material available to make it possible to select high merit heterozygotes, though further investigation is needed to establish whether outcrossing was extremely high, so there was no option when selecting but to select for hybrid individuals, or whether outcrossing was low -only a few outcrosses happened but with those few outcrossed individuals having outstanding phenotypic merit and were selected even though the rate of inbreeding was high.

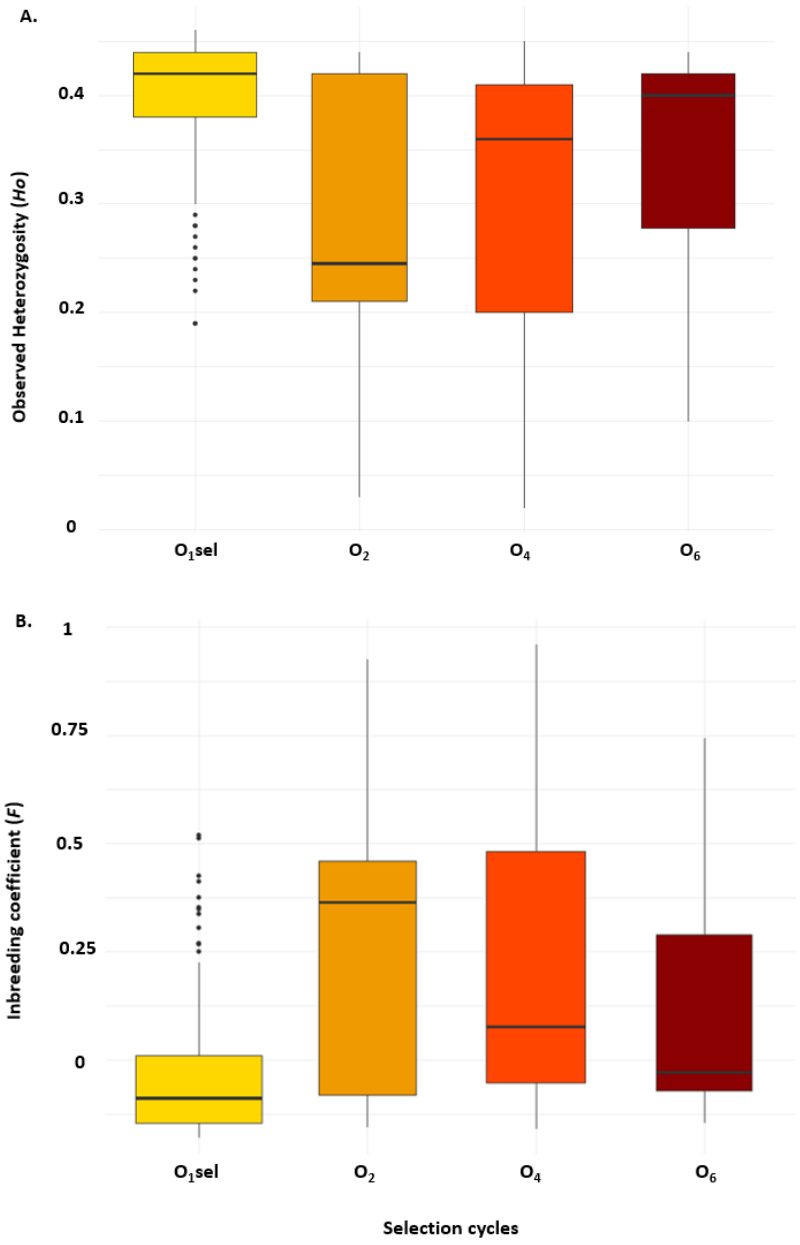


Figure 4.5. Box and whisker plots of **A.** Observed Heterozygosity (H_O) and **B.** Inbreeding coefficient (F) for the cohort of selected plants in each selection cycle. Median values are shown as horizontal black lines, the coloured boxes upper and lower quartile ranges, vertical line as the range of data outside upper and lower quartile (excluding outliers) and outliers as black dots.

4.4.2.2. Deduction of outcrossing rates in selected versus unselected cohorts using genomic data.

This question could not be addressed directly since consecutive generations were not genotyped, but could be addressed at least qualitatively by placing each individual on a scatterplot representing both the actual observed heterozygosity of an individual and the calculated net allele gain compared to the individual's grandmother in the previous selection cycle as shown in Figure 4.6. The expectation is that individuals resulting exclusively from inbreeding would be confined to the bottom left of the plot, showing low absolute levels of heterozygosity and high rates of allele loss compared to their grand-maternal progenitor from the previous selection cycle whereas those individuals resulting from the intermating of unrelated individuals would tend towards the top right quadrant of the plot. The fact that net allele gain is calculated over two generations means that offspring individuals cannot be categorically said to have inbred or outbred, and indeed a large number may have outcrossed in one generation and inbred in another thus giving an intermediate overall result. Nonetheless, we can see that the two measures are broadly correlated, as expected, and that there is clearly a mixed economy of inbreeding and outcrossing. Separate examination of the unselected and selected cohorts in O_2 and O_4 generations tell subtly different stories in these two selection cycles. The centroid position of the O_4 unselected cohort (blue diamond) is well inside the bottom left (greater influence of inbreeding) quadrant, whereas that of the O_4 selected cohort (green diamond) is within the top right (greater influence of outcrossing) quadrant, suggesting that the O_2 and O_3 generations that led to the genotyped O_4 unselected sample was tended on average to produce part inbred progeny, but those individuals who were more likely the result of outcrossing were enriched in the selected cohort. Whereas in the O_6 generation, the selected and unselected cohort centroids split in the same direction but to a far lesser

extent than in O_4. This might reflect two separate factors: first, that having selected to a great extent for high heterozygosity in the previous cycle, crosses that significantly increase allele count are now fewer and the relative advantage enjoyed by the most heterozygous individuals has somewhat eroded. It also cannot be excluded that the outcrossing rate varies significantly from generation to generation simply due to factors outside the control of the breeder. Overall, however, as the majority of the selected genotypes are on balance likely to have resulted from one or more outcrosses and taking the overall trend towards increasing observed heterozygosity through consecutive cycles of selection shown in Figure 4.5, the molecular data supports the conclusion that heterosis or hybrid vigour is a strong factor in determining which individuals make the phenotypic selection threshold. In Chapter 3, high (75%) and low (30%) outcrossing rates were modelled; in practice, it looks like the distribution of heterozygosity shown in Figure 4.6 results from outcrossing rates at the low end of the modelled range.

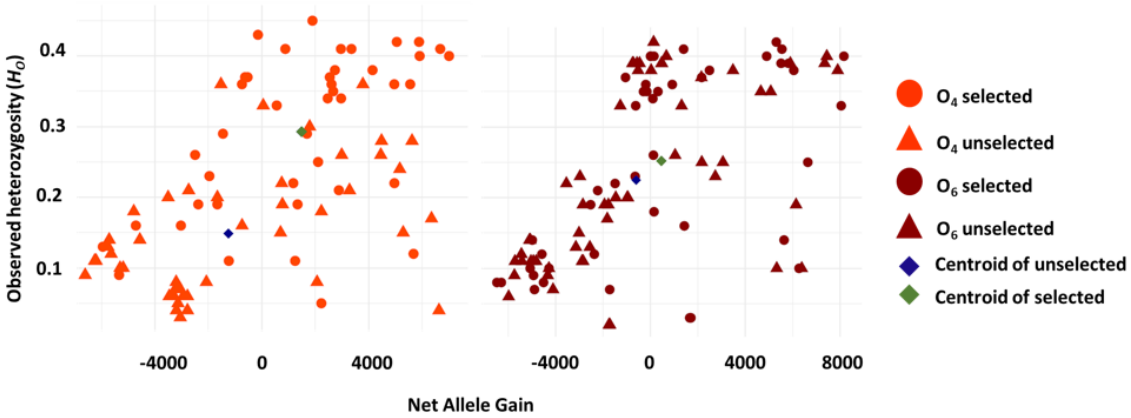


Figure 4.6. Plots of net allele gain versus observed heterozygosity H_O of genotyped individuals from the O_4 and O_6 generation. Generations are color-coded and Selection status represented by different shapes according to the scheme described in Table 4.1.

As the inbreeding cycles, conventionally included in a recurrent selection scheme (see Figure 1.2), were intentionally dismissed in this modified accelerated scheme, the

question of whether the MRSS permitted directional selection on specific loci at the same time as efficiently picking products of outcrossing was the next thing to be assessed.

4.4.3. Genomic responses to short-term selection

4.4.3.1. General changes in allele frequency

Globally, it is shown above that a majority of selected individuals are outcrossing, which means the selection could be purely determined by heterosis if its effect was stronger than differences in phenotype caused by the independent assortment of multiple mostly small additive genetic effects. In this section, evidence for underlying shifts in frequency of specific loci that could be targets of selection is explored.

Selection is considered to be a major force driving changes in the allele frequencies. Positive selection can increase the frequency of rare or newly introduced favourable alleles to the point of fixation. When a high selection intensity is practised, such as the 3.7% applied here, it is inevitable that some rare alleles will be eliminated not due to their phenotypic effect but to the impact of taking a small sample of individuals from a large population so greater caution is needed in interpreting the sudden elimination of rare alleles. Figure 4.7. shows change in marker allele frequency (of the allele stated as minor in the O_1sel generation) between the beginning and the end of the scheme for the mapped markers (ranked per chromosome and position) and Figure 4.8. for the unmapped (ranked by descending ΔAF value). There exist both increases in minimum allele frequency and decreases, some quite substantial. Depending on where we are in the genome, there are alleles that are rare and becoming more frequent as well as alleles that are frequent becoming rarer.

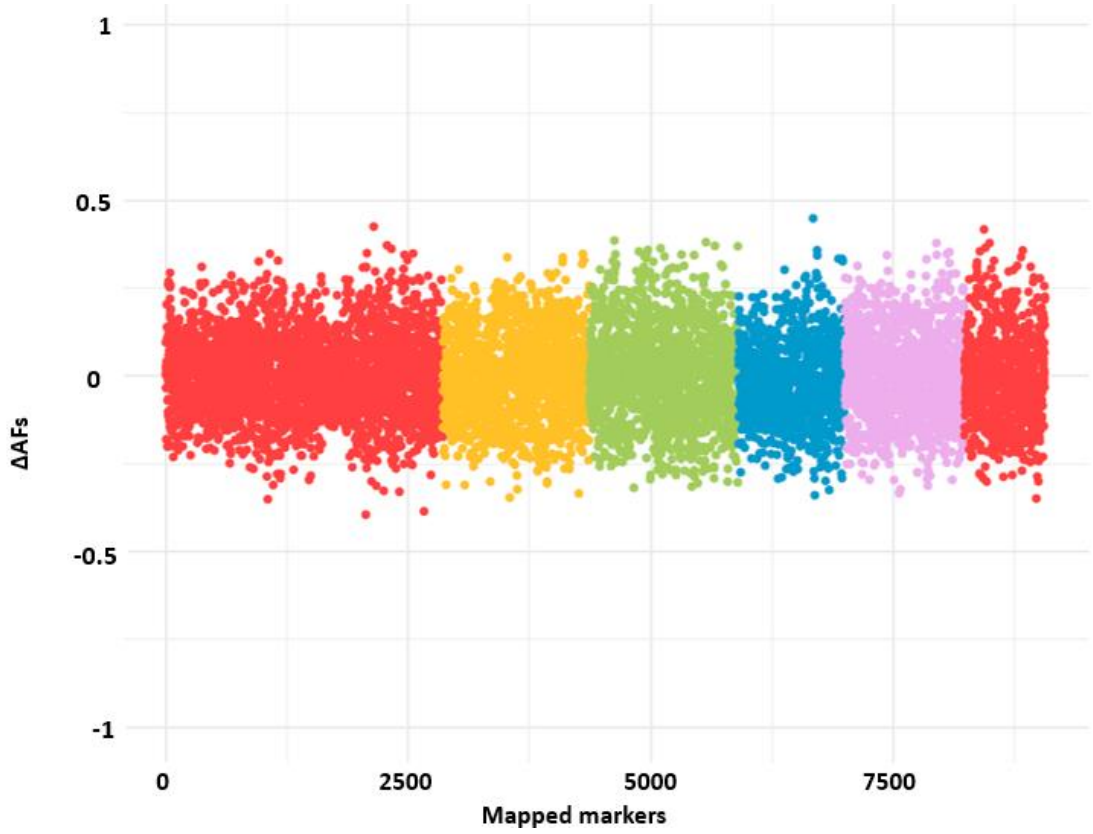


Figure 4.7. ΔAFs calculated the difference in frequency of the what is minor allele in O_1sel and its frequency in the latest O_6 generation. The graph presents the 10,784 mapped markers by chromosome and chromosomal position. Each colour block represents a different chromosome.

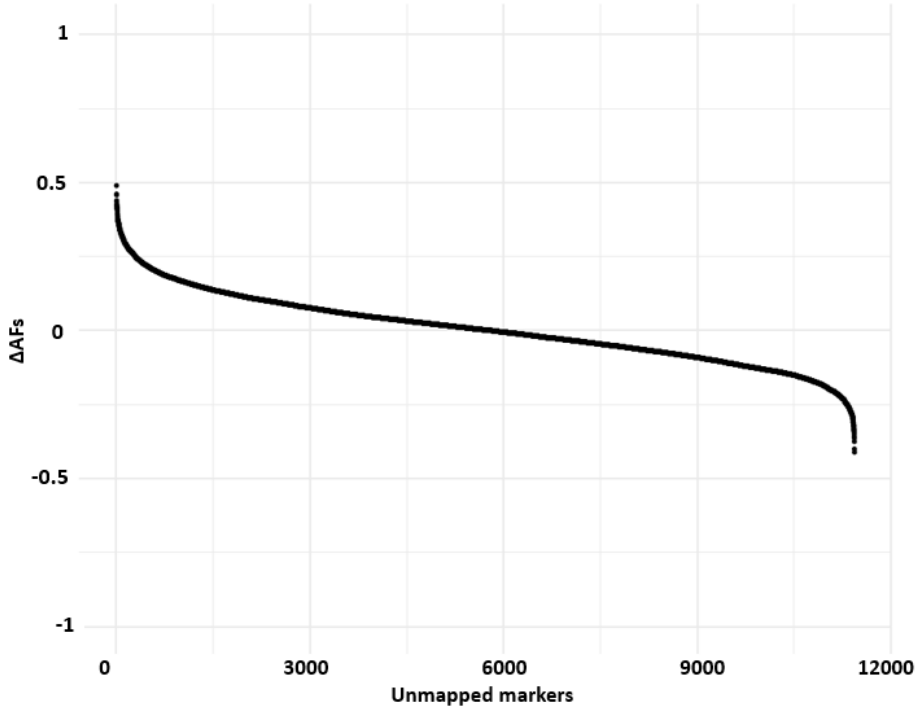


Figure 4.8. Change in Allele Frequency (ΔAF) calculated the difference in frequency of what is minor allele in O_1sel and its frequency in the latest O_6 generation. The graph presents the unmapped markers ranked by descending ΔAF order.

Although the changes on the allele frequency validate the recombinations that happened in the population during outcrossing, these changes could not at this point be related to adaptation but more to the selection intensity. A test that shows that this loss of alleles is not happening randomly everywhere in the genome but that there are some rare alleles becoming more frequent by selection meaning that slowly the population can develop a better genetic potential is the next step of the analysis.

4.4.3.2. Genomic locations bearing the signature of selection

As shown above, a certain number of alleles are lost in each generation but this could happen purely by drift, or in other words random loss as the effective population size is diminished, whereas we are seeking locus-specific genetic variance that determines adaptation to environment and yield and hence changes frequency selectively. To

investigate this, F statistics of individual mapped SNPs were scanned for positional signal or evidence that certain loci are being preferentially positively or negatively selected.

Typical thresholds used to declare significant signatures of selection were either the top 1% of the F_{ST}s or F_{ST}s which are three standard deviations above the mean. Calculating both values, the sharper selection is made when choosing the top one percentile (F_{ST}>0.111) than choosing F_{ST}s differing more than three standard deviations (3 SD) from the mean (F_{ST}> 0.094). The frequency distribution of the F_{ST}s of individual loci in the MRSS genotype data is shown on Figure 4.9. The predominance of near-zero F_{ST} values indicates that most loci (as expected) are neutral and do not show evidence of selection between the O_1sel and O_6 generations.

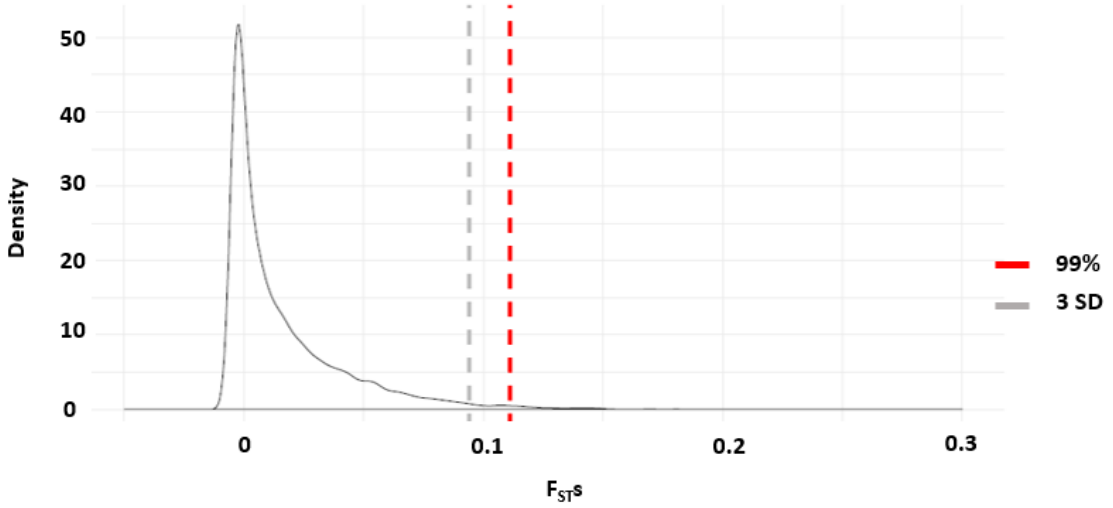


Figure 4.9. The density plot depicts the distribution of the 10,784 individual-marker F_{ST}s based on comparison of first genotyped founding generation (O_1sel) and latest selection (O_6) generations. Zero F_{ST} means no differentiation between the O_1sel and O_6 generation. The black dashed line stands for the 99th percentile and the grey dashed line for the 3 SD. Negative F_{ST}s do not have a biological explanation and can be considered as 0.

Using the stricter top 1 percentile threshold, 205 loci were associated with positive selection for the trait of interest, of which 82 have known map locations (F_{ST}s of mapped

markers are plotted in Figure 4.10.). F_{ST}s from 0.15 to 0.25 are considered to signify large genetic differentiation and values above 0.25 very large. Table 4.2. presents the top 20 markers in F_{ST} values.

Table 4.2. List of the markers with the top 20 F_{ST} values. In this table there is information regarding their observed (H_o) and expected (H_s) heterozygosity and F_{ST} value.

Marker_name	Chromosome	Position	Fst	Ho	Hs
AX-416763100	IV	37.9	0.218	0.176	0.357
AX-416817913	I	161.8	0.211	0.144	0.273
AX-416816429	I	134.7	0.184	0.232	0.395
AX-181485155	I	124.3	0.182	0.138	0.342
AX-416814716	VI	16.0	0.179	0.228	0.393
AX-416765270	IV	38.7	0.165	0.158	0.286
AX-416771687	III	32.9	0.154	0.115	0.271
AX-416809201	I	59.7	0.151	0.164	0.337
AX-181448721	VI	16.4	0.148	0.223	0.377
AX-181446894	IV	56.0	0.146	0.048	0.196
AX-181461332	III	93.1	0.146	0.269	0.391
AX-416800569	III	18.5	0.146	0.254	0.427
AX-416787025	III	73.8	0.145	0.293	0.396
AX-416808830	III	70.4	0.143	0.235	0.424
AX-181450280	VI	66.4	0.143	0.200	0.354
AX-416723473	IV	73.3	0.143	0.123	0.248
AX-181157268	VI	17.7	0.140	0.295	0.430
AX-416722989	II	81.7	0.140	0.192	0.335
AX-181453941	V	52.1	0.140	0.271	0.431
AX-416814986	V	37.8	0.139	0.151	0.314

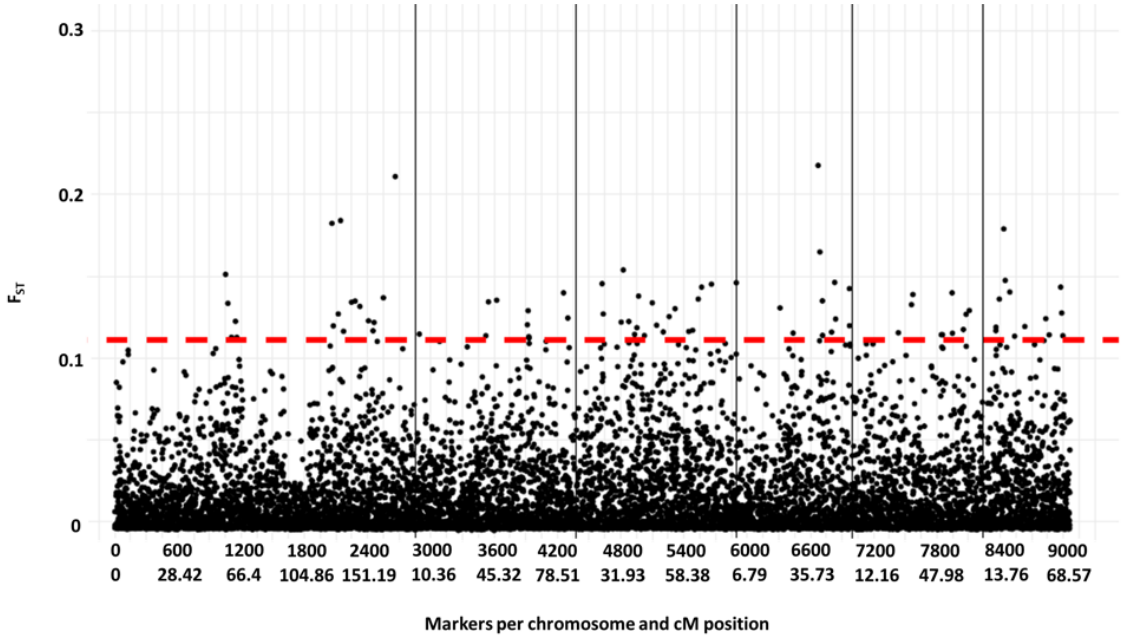


Figure 4.10. F_{ST} values (y-axis) of the 10,784 mapped Markers (x-axis) plotted along the six chromosomes in ascending inter-chromosomal position (cM). The red dashed line represents the top 1% threshold and is set at 0.111.

4.4.4. SNP-trait association study

4.4.4.1. Linkage disequilibrium analysis

Genome-wide association studies (GWAS) essentially measure linkage disequilibrium (LD) between a trait and a marker, and the GWAS resolution is determined by how steeply the LD decays with distance between markers. LD is also of interest as it is a measure of the extent to which physically linked loci have recombined historically, so aside from the interest in exploring prospects for locating genetic determinants of a phenotype via GWAS, the patterns of LD observed and any changes over time can give another perspective on the extent of outcrossing in breaking up and recombining founder haplotypes. Linkage disequilibrium analysis was performed for each generation and pairwise LD using squared allele frequency correlations (r^2) was estimated. Figure 4.11. suggests that there is a neat decay of LD over time which becomes sharper as we go

through the generations (LOESS curve crosses 0.1 at O_1-12.2 cM, O_2-12.1 cM, O_4- 10.2 cM and O_6-10 cM).

As stated above, if the selection was made on individuals of the same family only, genetic diversity would have been reduced and LD would be higher. In a case of absolute inbreeding, no LD change would be expected after the first few generations. A condition of bringing the decay rate faster is the continuous outcrossing and the fact the selection comes from a relatively even-handed way across all the products of recombination and not offspring of one desirable family or closely related individuals only.

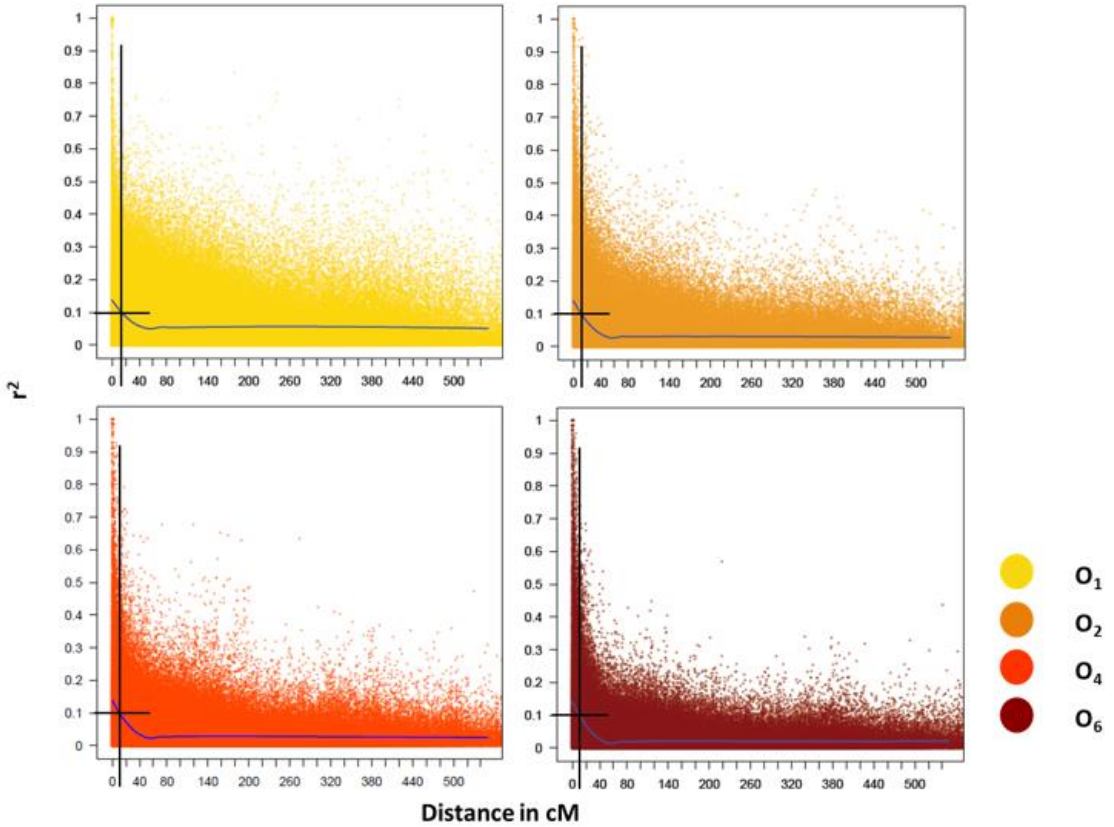


Figure 4.11. Overview of the r^2 parameter of LD of the intrachromosomal pairs in the four generations O_1, O_2, O_4 and O_6. The scatterplots show the distribution of the r^2 by genetic distance in cM. The dark blue LOESS curve indicates the decay curve.

4.4.4.2. Genome wide association study (GWAS) to identify associations between loci and yield

After detecting those loci bearing the signature of selection, the question is to find their meaning. It is assumed that there could exist a major effect QTL somewhere in the region that has been positively selected and is associated with those SNPs. With GWAS could be possible to detect associations between marker and yield, which is the only phenotype involved in the selection.

Running a genome-wide scan for SNPs associated with IPSY in the O_6 generation (combined panel of selected and unselected individuals), no significant QTL associated was detected (Figure 4.12). The GWAS result was expected as not only is yield a multigenic inheritance trait, meaning many loci are expected to explain its variation, but power to detect was limited by a combination of too few individuals in the panel and an error-prone, unreplicated phenotype (IPSY).

In the result as shown on Figure 4.12., there is nothing stunning in the manhattan plot, but at the same time there are coincidences where same markers that are turning up with relatively high -$\log_{10}$(p) values in the GWAS scan and with high F_{ST} (Figure 4.13.) fact that supports the positive signature findings.

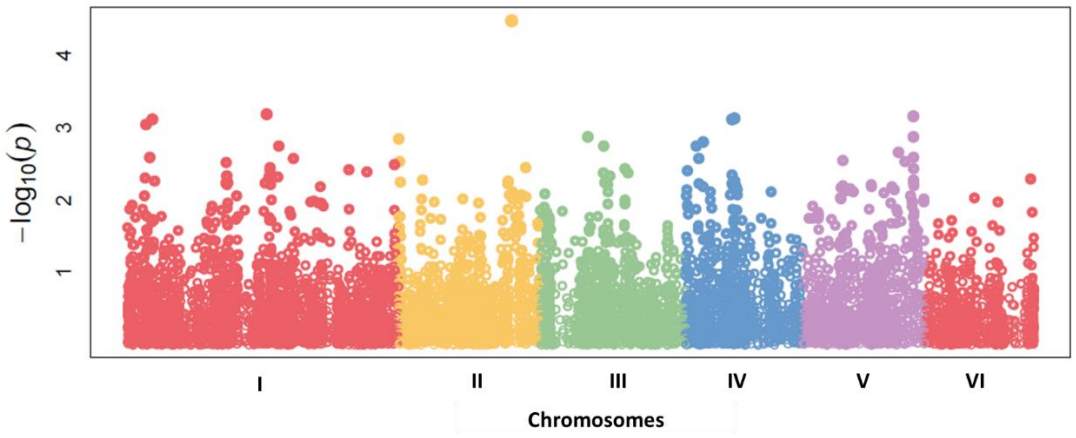


Figure 4.12. Genome-wide assocation scan for IPSY. The X-axis is the genomic position of the SNPs in chromosomal order, and the Y-axis is the negative log base of the P-values.

SNPs with stronger associations with the trait should have a larger Y-coordinate value. In case that was true, a horizontal line above which the associated SNPs were plotted, would have been drawn by the software.

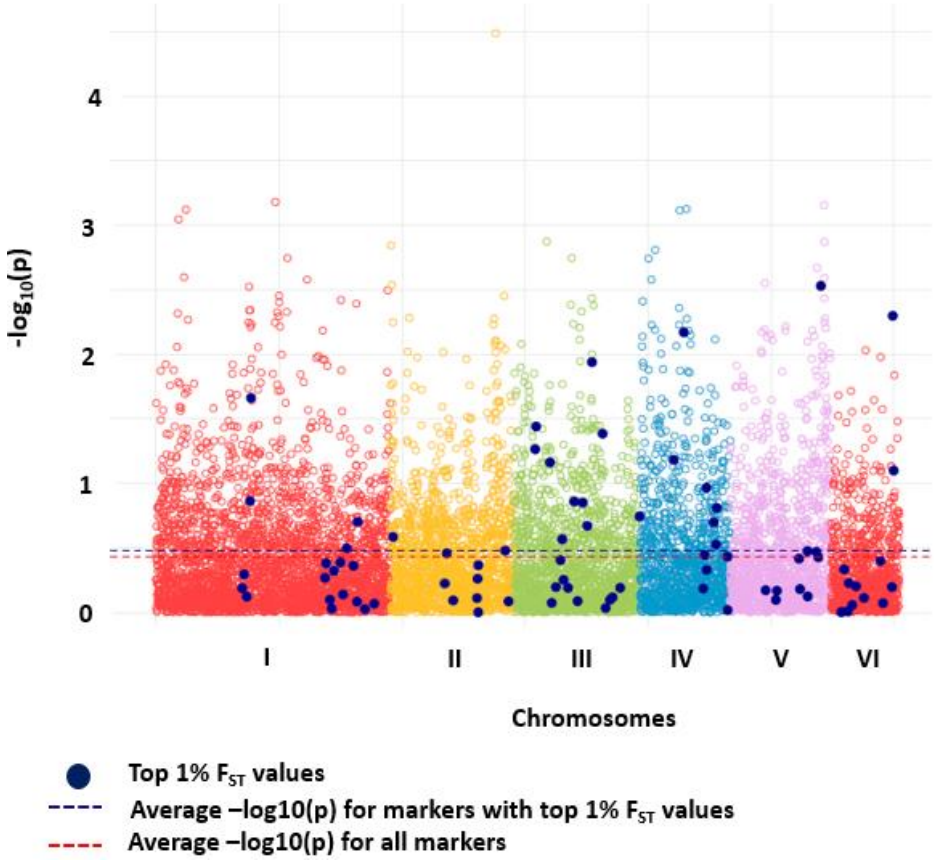


Figure 4.13. Overlay of GWAS scan for IPSY (open-coloured circles) and top 1% FST (closed blue circles). The X-axis is the genomic position of the SNPs in chromosomal order, and the Y-axis is the negative log base of the P-values. SNPs presented with dark blue colour dots are presented the SNPs with the top 1% F_{ST} values. The blue dashed line is the average –log10(p) for markers with top 1% FST values (0.48) and the red dashed line the average –log10(p) for all the mapped markers (0.43).

Plotting the average -log10(p) for the 99% of the markers (red dashed line) and of the top 1% high F_{ST} markers (blue dashed line), we can see that the two averages are lining quite apart, supporting that as expected when selecting on yield the markers that

end up being associated with yield turn out to also be statistically more likely markers that have been under selection. At the same time, selection has been going on over several generations (the difference of the first and last generation was measured by the F_{ST} test) and it is not a coincidence that the selected individuals are contributing to yield – which was expected as yield was the selection criterion, but that also results in a higher probability of existing loci with higher -log10(p) values, even though there is no marker above a GWAS threshold proving association. Table 4.3. presents the -log10(p) of the above average top 1% markers.

Table 4.3. List of above average top 1% F_{ST} mapped markers. In this table there is information regarding their chromosomal position, their observed (H_o) and expected (H_s) heterozygosity and -log10(p) values.

Marker name	Chromosome	Position	Fst	-log10(p)
AX-416735583	1	63.7648	0.1225	0.863297
AX-416787936	1	64.0021	0.1127	1.663007
AX-416724144	1	146.9837	0.1316	0.498632
AX-416781696	1	155.2121	0.1218	0.701637
AX-181484992	2	1.1325	0.1147	0.58535
AX-416800569	3	18.4993	0.1455	1.264729
AX-416773992	3	22.5462	0.127	1.440968
AX-416972787	3	32.0085	0.122	1.16268
AX-181493016	3	37.1481	0.1186	0.567334
AX-181492742	3	41.8879	0.1338	0.860236
AX-416747731	3	45.5698	0.116	0.851302
AX-416972685	3	49.5164	0.1254	0.67133
AX-181152640	3	54.5898	0.1302	1.940007
AX-181182382	3	61.8776	0.1163	1.385946
AX-181461332	3	93.1027	0.1461	0.744391
AX-416752957	4	26.915	0.1307	1.183303
AX-416740886	4	30.9604	0.1153	2.17125
AX-181494579	4	41.4847	0.135	0.966796
AX-181454574	4	53.5112	0.1159	0.698455
AX-181446894	4	55.9938	0.1463	0.526243
AX-181493153	4	57.0096	0.1239	0.809431
AX-416763296	5	65.7867	0.1291	2.530868
AX-181484985	6	67.7562	0.1276	2.298719
AX-181482492	6	68.3651	0.1137	1.099616

4.5. Conclusion

This one-person breeding programme had to be simple in its content and its implementations. For these reasons, the necessity of coming up with a radical scheme was embraced and hypothesised that maybe the structure of the classical recurrent selection programme is “obsolete” if molecular tools can be used to track the progress. In a classical recurrent selection scheme, the main reason for a time-consuming selfing step prior to each phenotypic evaluation and selection is to ensure that when selection happens plants are not selected purely on the basis of outcrossing per se. The SNP genotyping of our longitudinal series of samples spanning three selection cycles was relatively high resolution compared to other studies using RAPD or AFLP markers on the same crop (Link *et al.*, 1995; Zeid, Schön and Link, 2003). and the resulting genome-wide and population-wide profiles allowed us to determine empirically what level of outcrossing can be achieved andwhether there are specific loci that changed in frequency in response to selection, to complement the phenotyping of the population bulks.

The modified recurrent selection scheme that was followed for three selection cycles, resulted in a highly heterozygous O_6 population. Previous studies of F1 hybrids in faba bean have shown a large heterosis of up to 95% above mid-parent values for yield (Stelling, Ebmeyer and Link, 1994; Link *et al*., 1996; Suso *et al*., 2005; Palmer *et al*., 2009; Melchinger and Gumber, 1998) so it was not surprising that our selections in each cycle were enriched for highly heterozygous individuals compared with average heterozygosity in a sample of the whole population.

The longtitudinal dimension to this population genotyping programme also revealed fluctuations in allele frequency at individual loci and detection of putative signatures of locus-specific selection, which is important as it suggests that selection on individual plant yield exerts selection pressure on individual yield-enhancing loci.

Even though GWAS is a commonly used technique for identification of qualitative and quantitative traits in grain legumes including faba beans (Sallam and Martsch, 2015; Sallam et al., 2016; Hu *et al*., 2019), in this study no QTL associated with yield were detected. This outcome was expected due to the highly heterozygous nature of the population and the polygenic inheritance of yield.

4.6. References

Bouchet, S., D. Pot, M. Deu, J.-F. Rami, C. Billot, X. Perrier, R. Rivallan, et al. 2012. Genetic Structure, Linkage Disequilibrium and Signature of Selection in Sorghum: Lessons from Physically Anchored DArT Markers. PloS One **7** (3): e33470. https://doi.org/10.1371/journal.pone.0033470

Dudley, J. W., and R. J. Lambert. 2010. 100 Generations of Selection for Oil and Protein in Corn. Plant Breeding Reviews. **24**: 79-110. https://doi.org/10.1002/9780470650240.ch5.

Goudet, J. 2005. Hierfstat, a Package for R to Compute and Test Hierarchical F-Statistics. Molecular Ecology Notes. **5**(1): 184-186. https://doi.org/10.1111/j.1471-8286.2004.00828.x.

Granato, I. S. C., G. Galli, E. Giselly de Oliveira Couto, M. Bandeira e. Souza, L. Freitas Mendonça, and R. Fritsche-Neto. 2018. snpReady: A Tool to Assist Breeders in Genomic Analysis. Molecular Breeding. **38**: 102. https://doi.org/10.1007/s11032-018-0844-8.

Hopkins, CG. 1899. Improvement in the chemical composition of the corn kernel. Ill Journal of the American Chemical Society. **21**(11):1039-1057. https://doi.org/10.1021/ja02061a012

Hu, J., F. Maalouf, H. Dong, C. Hawkins, Z. Zhang, Y. Ma, Clarice J Coyne, D Jin, Yu L-X, Z. B. M. Ahmed, A. Hamwieh, L. Abou-Khater, M. Baum. 2018. Enhancing faba bean (*Vicia faba* L.) germplasm for resilience to temperature extremes. Beirut, Lebanon. https://hdl.handle.net/20.500.11766/9260

Huang, X., N. Kurata, X. Wei, Z.-X. Wang, A. Wang, Q. Zhao, Y. Zhao, et al. 2012. A Map of Rice Genome Variation Reveals the Origin of Cultivated Rice. Nature. **490** (7421): 497–501. https://doi.org/10.1038/nature11532

Hufford, M. B., X. Xu, J. van Heerwaarden, T. Pyhäjärvi, J.-M. Chia, Reed A. Cartwright, R. J. Elshire, et al. 2012. Comparative Population Genomics of Maize Domestication and Improvement. Nature Genetics. **44** (7): 808–11. https://doi.org/10.1038/ng.2309

Kaur, S., N. Cogan, J. Forster, and J. Paull. 2014. Assessment of Genetic Diversity in Faba Bean Based on Single Nucleotide Polymorphism. Diversity. **6**(1): 88-101. https://doi.org/10.3390/d6010088.

Khazaei, H., F. L. Stoddard, R. W. Purves, and A. Vandenberg. 2018. A Multi-Parent Faba Bean (*Vicia Faba* L.) Population for Future Genomic Studies. Plant Genetic Resources: Characterization and Utilization. **16**(5): 419-423. https://doi.org/10.1017/s1479262118000242.

Kim, J.-Y., S. Jeong, K. Hyoun Kim, W.-J. Lim, H.-Y. Lee, Namhee Jeong, J.-K. Moon, and N. Kim. 2019. Dissection of Soybean Populations according to Selection Signatures Based on Whole-Genome Sequences. GigaScience **8** (12): 1-19. https://doi.org/10.1093/gigascience/giz151.

Laidò, G., D. Marone, M. A. Russo, S. A. Colecchia, A. M. Mastrangelo, P. De Vita, and R. Papa. 2014. Linkage Disequilibrium and Genome-Wide Association Mapping in Tetraploid Wheat (*Triticum turgidum* L.). PloS One. **9** (4): e95211. https://doi.org/10.1371/journal.pone.0095211

Lake, L., Y. L., J. J. Casal, and V. O. Sadras. 2016. Negative Association between Chickpea Response to Competition and Crop Yield: Phenotypic and Genetic Analysis. Field Crops Research. **196**: 409-417. https://doi.org/10.1016/j.fcr.2016.07.021.

Landry, E. J., C. J. Coyne, R. J. McGee, and J. Hu. 2017. A Modified Mass Selection Scheme for Creating Winter-Hardy Faba Bean (Vicia Faba L.) Lines with a

Broad Genetic Base. Euphytica. **213** :72. https://doi.org/10.1007/s10681-017-1843-2.

Laurie, C. C., S. D. Chasalow, J. R. LeDeaux, R. McCarroll, D. Bush, B. Hauge, C. Lai, D. Clark, T. R. Rocheford, and J. W. Dudley. 2004. The Genetic Architecture of Response to Long-Term Artificial Selection for Oil Concentration in the Maize Kernel. Genetics. **168** (4): 2141–55. https://doi.org/10.1534/genetics.104.029686

Link, W., B. Schill, A. C. Barbera, J. I. Cubero, A. Filippetti, L. Stringi, E. von Kittlitz, and A. E. Melchinger. 1996. Comparison of Intra- and Inter-Pool Crosses in Faba Beans (Vicia Faba L.). I. Hybrid Performance and Heterosis in Mediterranean and German Environments. Plant Breeding. **115**(5): 352-360. https://doi.org/10.1111/j.1439-0523.1996.tb00933.x.

Link, W., C. Dixkens, M. Singh, M. Schwall, and A. E. Melchinger. 1995. Genetic Diversity in European and Mediterranean Faba Bean Germ Plasm Revealed by RAPD Markers. Theoretical and Applied Genetics. **90** (1): 27–32. https://doi.org/10.1007/BF00220992

Liu, X., M. Huang, B. Fan, Edward S. Buckler, and Z.Zhang. 2016. Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. PLoS Genetics. **12** (2): e1005767. https://doi.org/10.1371/journal.pgen.1005767

Marzinzig, B., L. Brünjes, S. Biagioni, H. Behling, W. Link, and C. Westphal. 2018. Bee Pollinators of Faba Bean (*Vicia Faba* L.) Differ in Their Foraging Behaviour and Pollination Efficiency. Agriculture, Ecosystems and Environment. **264**: 24-33. https://doi.org/10.1016/j.agee.2018.05.003.

Melchinger, A. E., and R. K. Gumber. 2015. Overview of Heterosis and Heterotic Groups in Agronomic Crops. Concepts and Breeding of Heterosis in Crop Plants. **25**: 3. https://doi.org/10.2135/cssaspecpub25.c3.

O'Sullivan, D., D. Angra, T. Harvie, V. Tagkouli, and A. Warsame. 2019. A Genetic Toolbox for Vicia Faba Improvement. In ICLGG 2019. Dijon. https://iclgg2019.b2match.io/.

Ouellette, L. A., R. W. Reid, S. G. Blanchard, and C. R. Brouwer. 2018. LinkageMapView-rendering High-Resolution Linkage and QTL Maps. Bioinformatics. **34**(2): 306-307. https://doi.org/10.1093/bioinformatics/btx576.

Paape, T., T.s Bataillon, P. Zhou, T.J Y Kono, R. Briskine, N. D. Young, and P. Tiffin. 2013. Selection, Genome-Wide Fitness Effects and Evolutionary Rates in the Model Legume *Medicago Truncatula*. Molecular Ecology. **22** (13): 3525–38. https://doi.org/10.1111/mec.12329

Palmer, R. G., P.T. Perez, E. Ortiz-Perez, F. Maalouf, and M. J. Suso. 2009. The Role of Crop-Pollinator Relationships in Breeding for Pollinator-Friendly Legumes: From a Breeding Perspective. Euphytica. **170**: 35-52. https://doi.org/10.1007/s10681-009-9953-0.

Pankin, A., J. Altmüller, C. Becker, and M. von Korff. 2018. Targeted Resequencing Reveals Genomic Signatures of Barley Domestication. New Phytologist. **218**(3): 1247-1259. https://doi.org/10.1111/nph.15077.

Papa, R., E. Bellucci, M. Rossi, S. Leonardi, D. Rau, P. Gepts, L. Nanni, and G. Attene. 2007. Tagging the Signatures of Domestication in Common Bean (Phaseolus Vulgaris) by Means of Pooled DNA Samples. Annals of Botany. **100** (5): 1039–51. https://doi.org/10.1093/aob/mcm151

Russell, J., M. Mascher, I. K. Dawson, S. Kyriakidis, C. Calixto, F. Freund, M. Bayer, et al. 2016. Exome Sequencing of Geographically Diverse Barley Landraces and Wild Relatives Gives Insights into Environmental Adaptation. Nature Genetics. **48** (9): 1024–30. https://doi.org/10.1038/ng.3612

Sallam, A., and R. Martsch. 2015. Association Mapping for Frost Tolerance Using Multi-Parent Advanced Generation Inter-Cross (MAGIC) Population in Faba Bean (*Vicia Faba* L.). Genetica. **143** (4): 501–14. https://doi.org/10.1007/s10709-015-9848-z

Sallam, A., M. Arbaoui, M. El-Esawi, N. Abshire, and R. Martsch. 2016. Identification and Verification of QTL Associated with Frost Tolerance Using Linkage Mapping and GWAS in Winter Faba Bean. Frontiers in Plant Science. **7** (August): 1098. https://doi.org/10.3389/fpls.2016.01098

Segura, V., B. J. Vilhjálmsson, A. Platt, A. Korte, Ü. Seren, Q. Long, and M. Nordborg. 2012. An Efficient Multi-Locus Mixed-Model Approach for Genome-Wide Association Studies in Structured Populations. Nature Genetics. **44** (7): 825–30. https://doi.org/10.1038/ng.2314

Stelling, D., E. Ebmeyer, and W. Link. 1994. Yield Stability in Faba Bean, Vicia Faba L. 2. Effects of Heterozygosity and Heterogeneity*. Plant Breeding. **112**(1): 30-39. https://doi.org/10.1111/j.1439-0523.1994.tb01273.x.

Suso, M. J., L. Harder, M. T. Moreno, and F. Maalouf. 2005. New Strategies for Increasing Heterozygosity in Crops: Vicia Faba Mating System as a Study Case. Euphytica. **143**: 51-65. https://doi.org/10.1007/s10681-005-2526-y.

Twitter Coop. 2020. https://twitter.com/graham_coop/status/1159889403491180544.

Yang, R.-C. 1998. Estimating Hierarchical F-Statistics. Evolution; International Journal of Organic Evolution. **52** (4): 950–56. https://doi.org/10.1111/j.1558-5646.1998.tb01824.x

Zeid, M., C.-C. Schön, and W. Link. 2003. Genetic Diversity in Recent Elite Faba Bean Lines Using AFLP Markers. Theoretical and Applied Genetics. **107** (7): 1304–14. https://doi.org/10.1007/s00122-003-1350-9

Zhang, Z., E. Ersoz, C.-Q. Lai, R. J. Todhunter, H. K. Tiwari, M. A. Gore, P. J. Bradbury, et al. 2010. Mixed Linear Model Approach Adapted for Genome-Wide Association Studies. Nature Genetics. **42** (4): 355–60. https://doi.org/10.1038/ng.546

5. Chapter 5 - General conclusion, study limitations and future work

The overall aim of this thesis was to develop a scientifically sound small-scale breeding method suitable for niche markets. To help reflect on what has been achieved, Figure 5.1. gives an overview of the major findings of this thesis in terms of their relevance to this overarching goal. Firstly, the general introduction (Chapter 1) of this thesis sets the scene with a survey of the state-of-the-art in faba bean breeding and develops a proposal to revisit recurrent selection as an attractive breeding method for faba bean (the 'CONCEPT').

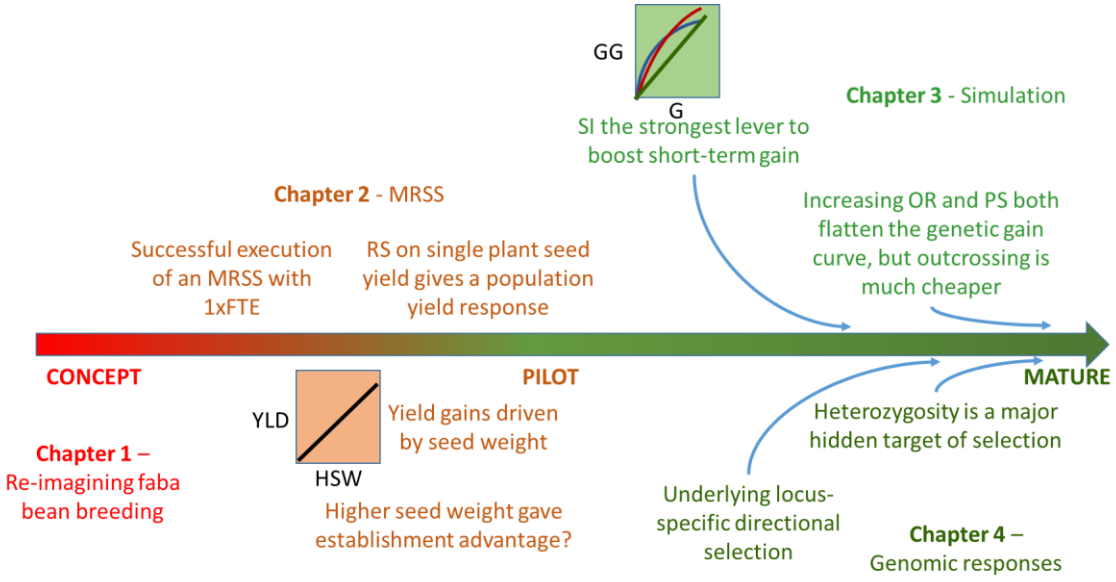


Figure 5.1. Schematic diagram showing how learning points from each Chapter feed into the elaboration of a breeding 'concept' into a 'mature' breeding programme. Each chapter of the thesis is given a shortened title and a colour code to reflect where it sits on the path (colour spectrum) from Concept (red) to Mature programme (green).

The experimental work of this thesis had a practical backbone, which was a real-world implementation of the MRSS concept (the 'PILOT), described in Chapter 2. We learn in Chapter 2 first that it is feasible to execute a recurrent selection programme staffed by a single full-time scientist. Furthermore, in a relatively short period of time

(three selective generations), it is shown that it is possible to obtain yield gain that places the selected populations ahead of both progenitors and elite contemporary common knowledge commercial varieties. The fact that this population yield response was obtained even though the RS was on individual plant seed yield is noteworthy since formally, single plant selection is not recommended as a practice mainly because a single plant can be selected by being hypercompetitive, which would be a negative trait in the population context. Experimental designs like the honeycomb selection designs (HSD) have been proposed to control the spatial heterogeneity and minimize competitive interactions among individuals and allow them to express their phenotype and phenotypic variance (Fasoulas and Fasoula 1995; Fasoula, Ioannides, and Omirou, 2019). In outcrossing species like faba bean, the additional space required by such a design would be prohibitively costly and pose a barrier to effective outcrossing, so in reality, the only option for single plant selection was to sow in agronomically realistic densities and rely on the empirical data to confirm, as it ultimately did, that selection on individual performance led to improvement of the population as a whole.

Another important insight of Chapter 2 is that the yield gains were driven by seed weight, an effect which was consistent across both selection sites. Every successive generation had bigger and thus heavier seeds compared to its parental lines. At the evaluation trial where all the selection generations and market varieties were sown side by side a difference on plant population stood out at the outset. Overall, combining the results of both sites, market varieties and selection generations of previous cycles had smaller and lighter seeds compared to the latest selections O_6 and 2T, and at the same time, within each site, the latest selections had better establishment compared to the other entries, generating the thought that maybe the higher seed weight of the selections gives seeds a natural advantage independently of anything else (e.g. heterozygosity or merit).

Drilling 40 seeds/m^2 with heavier endosperm versus the previous selections and commercial varieties, which were sown at the same density but have less mass of endosperm, might offer a better establishment, drive an early growth and help individuals overcome extreme weather conditions as the prolonged drought accompanied by high temperatures that we had during the first months of the trial establishment.

The empirical validation of a breeding concept is by its nature retrospective and cannot encompass experimental manipulation of key selection parameters, so a simulation approach was developed in Chapter 3 to look at the theoretical responses to and interactions between population size and selection intensity (both in the control of the breeder) and outcrossing rate (not in the control of the breeder but of uncertain magnitude). The simulations carried out produced important insights that will help shape future implementations of RS in faba bean. Most notably that there was an inherent tradeoff between short-term rate of genetic gain and long-term ceiling on genetic gain. While there is every incentive to maximise short term gain by using high selection intensities, diversity loss that limits long-term potential gain could be mitigated by either attempting to achieve a higher outcrossing rate or increasing the population size, with the former probably representing the cheaper option. The outcomes of the simulations could thus guide an investor/ breeder on how to build a scheme with the maximum gain and minimum cost. The important question for further research is the understanding of how captive-bee-technology can be used most effectively to manipulate the outcrossing rate. This question could be investigated by having two identical RS subpopulations pollinated by two different pollinator species (e.g. honeybee vs bumblebee colonies) and different sizes/densities of colonies for a given population size.

Stochastic simulations are not yet commonly used in plant breeding though the approach has been used recently in sorghum to compare the effects of genomic selection

(GS) and phenotypic recurrent selection (PRS) on genetic gain and genetic variance (Muleta, Pressoir, and Morris 2019), but despite the availability of different simulation programs (Tinker and Mather 1993; Podlich and Cooper 1998; Laval and Excoffier 2004; Yabe, Iwata, and Jannink, 2017; Maurer, Melchinger, and Frisch 2008), there have been no stochastic simulations of faba bean breeding schemes prior to this work. In fact, it would not have been possible to realistically simulate different levels of outcrossing before a bespoke outcrossing function was created for this project after discussing the need with the authors of AlphaSimR. The modified AlphaSimR package was then used to look specifically at this radical proposition of recurrent selection from a population experiencing a mixed economy of inbreeding and outbreeding. Although estimates for natural outcrossing rates of faba bean vary greatly, they are likely to fluctuate within the bounds set in our simulations as it is known that hybridity increases autofertility (Drayner, 1956). The AlphaSimR simulations that were used assume that what we are selecting is additive genetic variation from a discrete number of QTLs that are capable of influencing the trait. A deficiency in these simulations is that they do not model the effect of hybrid vigour on trait expression, and further development of AlphaSimR in this regard would enable more realistic simulations of selection responses.

The importance of hybrid vigour as a determinant of individual plant performance was highlighted through the longitudinal genotyping described in Chapter 4. Here, a recently developed high density SNP genotyping array was used to make a number of inferences about genomic-level responses to selection. A steady rate of allele loss following each selective bottleneck was both expected and observed. What was more difficult to predict, but very cearly shown, was that selection caused a marked positive shift in heterozygosity. This implies that a degree of the phenotypic merit of selected individuals is not due to discrete additive QTL but to genome-wide heterozygosity and

may even mask underlying genetic variation. This concern about the undue influence of hybridity in determining individual plant performance is mitigated by two factors. First, as previously noted, the regulation of autofertility by hybridity causes heterozygosity to tend towards an equilibrium; secondly, within a closed genetic pool, even if 100% outcrossing were sustained, heterozygosity would quickly reach a maximum before slowly declining as allelic diversity is lost and differential merit would rest with underlying additive effects. In fact, the genome scan for signatures of selection showed that even after just three selective generations, it was possible to detect underlying locus specific directional selection. The reasons for selection of beneficial alleles at specific loci were beyond the scope of this study, but it should be noted that the Reading Spring Bean Panel of inbred lines drawn from the O_2 generation represent a useful resource where allelic variation segregating in the genepool can be firmly associated with specific phenotypes e.g. HSW or flowering time/duration. If large effect loci for such potentially yield-influencing traits are found, it may be possible to hypothesize the functional traits behind selective purges.

Implementing all the ideas and findings from the 'PILOT' programme points towards a potential future 'MATURE' operational scheme that could be taken forward successfully re-establishing recurrent selection (with some modifications) as a viable method in faba bean breeding.

In this study, the winter generations of the UoR MRSS workflow served only as bulking and outcrossing cycles. As the trait under selection was yield 'in the target environment', with the 'target environment' being natural field conditions from spring sowing in SE England, no selection on that could be done under GH conditions using pot-grown plants. With sufficient knowhow and resources, these winter generations could be employed to exercise complementary forms of selection e.g. to fix molecularly

understood traits like the low vicine trait (Bjornsdottir, 2020) using marker-assisted selection or to identify individuals resistant to prevalent diseases for promotion in the selection scheme by glasshouse-based pathotest screens. Simulations could be conducted to optimise the integration of different selection methods in winter as they could to investigate the potential impact of running winter generations as inbreeding cycles.

Finally, we have noted above that the success of this MRSS hinges on the validity of individual plant yield as a proxy for population yield. In future, it would seem reasonable to underpin single plant phenotyping with a better understanding of how heritable single plant yield could be made to be, for example by comparing variance of individual inbred plants spread randomly throughout a heterogenous population compared with the variance of individual plants in uniform stands of the inbred.

Overall, MRSS was successful in its own terms in achieving rapid yield gain. Even if this yield gain can be further built on and verified in multiple sites and years, there remains a significant obstacle to the exploitation of the method in mainstream breeding. The selected bulks are by definition population mixtures that would not get through the variety registration system as without significant additional work to fix key morphological characters, the population would have difficulty satisfying the distinctiveness, uniformity, and stability (DUS) criteria that all new varieties need to satisfy.

5.1. References

Björnsdotter, E., M. Nadzieja, W. Chang, L. Escobar-Herrera, D. Mancinotti, D. Angra, H. Khazaei, H. Crocoll, C. Vandenberg, A. Stoddard, F.L. et al. 2020. VC1 Catalyzes a Key Step in the Biosynthesis of Vicine from GTP in Faba Bean. bioRxiv. https://www.biorxiv.org/content/10.1101/2020.02.26.966523v1

Drayner, J. 1956. Regulation of Outbreeding in Field Beans (*Vicia faba*). Nature. **177**: 489–490. https://doi.org/10.1038/177489b0

Fasoula, D. A., I. M. Ioannides, and M. Omirou. 2019. Phenotyping and Plant Breeding: Overcoming the Barriers. Frontiers in Plant Science **10**: 1713. https://doi.org/10.3389/fpls.2019.01713

Fasoulas, A. C., Fasoula, V. A. 1995. Honeycomb Selection Designs. Plant Breeding Reviews, edited by Jules Janick. **13**(1): 87–139 https://doi.org/10.1002/9780470650059.ch3

Laval, G., and L. Excoffier. 2004. SIMCOAL 2.0: A Program to Simulate Genomic Diversity over Large Recombining Regions in a Subdivided Population with a Complex History. Bioinformatics. 20(15): 2485-2487. https://doi.org/10.1093/bioinformatics/bth264.

Maurer, H. P., A. E. Melchinger, and M. Frisch. 2008. Population Genetic Simulation and Data Analysis with Plabsoft. Euphytica. **161**: 133-139. https://doi.org/10.1007/s10681-007-9493-4.

Muleta, K. T., G. Pressoir, and G. P. Morris. 2019. Optimizing Genomic Selection for a Sorghum Breeding Program in Haiti: A Simulation Study. G3. **9** (2): 391–401. https://doi.org/10.1534/g3.118.200932

Podlich, D. W., and M. Cooper. 1998. QU-GENE: A Simulation Platform for Quantitative Analysis of Genetic Models. Bioinformatics. **14** (7): 632–53. https://doi.org/10.1093/bioinformatics/14.7.632

Tinker, N. A., and D. E. Mather. 1993. GREGOR: Software for Genetic Simulation. Journal of Heredity. **84**(3): 237. https://doi.org/10.1093/oxfordjournals.jhered.a111329.

Yabe, S., H. Iwata, and J.-L. Jannink. 2017. A Simple Package to Script and Simulate Breeding Schemes: The Breeding Scheme Language. Crop Science. **57**(3): 1347-1354. https://doi.org/10.2135/cropsci2016.06.0538.

6. Appendices

Appendix A. Glasshouse experimental conditions for O_1, O_3 and O_5 outcrossing populations:

Glasshouse information

The glasshouse experiments were conducted in the research glasshouses of the Crop and Environmental Laboratory (CEL) at the University of Reading, (located at latitude: 51°26'13.84"N and longitude: 0°56'31.96"W for the O_1 outcrossing generation and at latitude: 51°26'12.07"N and longitude: 0°56'33.14"W for the O_3 outcrossing generation), where automated irrigation systems, heating and supplementary lighting are available. The ridge of the greenhouse section was oriented West-East. The glasshouses were mechanically ventilated with automatic roof vents.

Seedling trays

The seeds were sown in 35-well nursery plant trays (Teku tray JP 3040/35 black 95 box) with dimensions 35,5 x 27,5 x 7,5cm. Since some of the founder lines were known to be winter seasonal type, those seedlings were placed for a three-week vernalisation in a walk-in cold room set at 4 °C with continuous lighting, while the spring lines were germinated on a heated glasshouse bench. All the trays were hand watered according to their needs.

Labels

All individuals were clearly labelled with plastic barcoded pot labels where trial name, plant ID and pot number was written.

Potting

For all glasshouse generations, the germinated plants were transplanted into 3Li round black plastic pots and transferred in glasshouse no 7 which was disinfected prior its use. The pots measure approximately 19cm diameter, 15.2 cm deep and weight approximately

0.07 kg each. The co-polymer made pots feature uniform irrigation, drainage system and have strong pot shoulders.

Compost

For both trays and pots compost John Innes no2 was used. John Innes composts are manufactured using peat, grid-sand and pasteurised loam in order to kill any weed seeds, propagules and eliminate pathogens. The compost pH is set around 6.5, moisture content (when packed) is typically 15-30% and conductivity is 400-600 micro siemens/cm.

Lay-out and irrigation

For the O_1 outcrossing population, the pots were equally distributed on three benches in three rows/ bench. Each bench allocated 73 pots following a completely randomised experimental design. The plants of the O_3 and O_5 outcrossing populations were placed directly on the cement ground of the glasshouse randomised in four rows. The plants were dripped irrigated with ascending water volume according to the plant growing stage and getting supported by canes and tape as they were gaining plant mass.

Temperature

The glasshouse temperature was set at 6-22°C, mimicking early spring Irish weather conditions. As a consequence, the heat was automatically on whenever the temperature dropped below 6 degrees and the ventilation activated whenever the temperature was exceeding the 22°C.

Photoperiod

In the glasshouse, for the period mid-October to mid-November a 16h photoperiod was imposed according to the solar noon and supplementary lighting system was operated when no sunlight was available during 7am to 11pm the first month and continuously 16h after that and up to harvest.

Fertiliser

The plants were weekly liquid fed (VITAFEED STANDARD, Vitax) with water soluble fertiliser which contained high levels of magnesium and some trace elements (boron, copper, iron, manganese and more). The profile of the fertiliser is 19-19-19+1.6 MgO+TE (1-1-1) and medium strength dose is recommended for weekly feeding (1.0 g/1 litre of water).

Fungicide

As plants were growing, disease symptoms and especially chocolate spot (Botrytis fabae) started to appear. Botrytis fabae is the most common fungus that affects broad beans. The plants sprayed once during the season, with a mixture of 250 g/L (25.9% w/w) tebuconazole (chemical group Triazole, Folicur) -which apart from chocolate spot controls bean rust on beans too- and 500 g/l chlorothalonil + 37.5 g/l metalaxyl-M (Phthalonitrile and Phenylamide, activity group: FRAC code M5 (chlorothalonil) + Group A1 FRAC code 4, Syngenta).

Thrip control

To control thrips, amblyline (Syngenta Bioline) biological control sachets were placed on every plant. The formulation of the hooked sachets is designed to permit a colony of *Amblyseius cucumeris* mites to be release onto to the plant daily. The mites feed on first larval stage (L1) while they lack ability to attack larger larvae or adult thrips. These sachets were replaced every 6 weeks throughout the growing season.

Pollination

At the appearance of the first young flowers, a *Bombus terrestris* beehive (Natupol by Koppert) was in the centre of the glasshouse to promote cross-pollination among the plants. As more flowers appeared, a second beehive followed. Every colony of bumblebee included a queen, workers, brood and sugar water. Specific preferable crosses were hand mediated by emasculating the mother plant and transferring pollen from the

chosen donor. These targeted crosses were marked with labels note the identity of individuals crossed and the date of the crossing.

Phenotypic records at the O_1

Record of first flower appearance, flower colour and existence of stipule spot was kept. Individuals, as expected, did not all flower at the same time, so in order to record which pods were set during the period of maximum flowering synchrony, a coloured ticket label was placed at the youngest open flowering truss every 7 days, such that each 7 day period during flowering was colour-coded. The percentage of plants flowering at the same time could thus be calculated from records of which coloured labels had been placed on each plant. The O_1Sel progeny was formed from pods formed during the 21-day period during which >85% of individuals were in flower and thus contributing to the pollen pool.

Appendix B. Isolation Cage conditions for O_2, O_4 and O_6 outcrossing generations:

The cage experiments were conducted at the protected field plot experimental of the Crop and Environmental Laboratory (CEL) of the University of Reading. The cages for the years 2017 and 2018 were built at latitude: 51°26'11.65"N - longitude: 0°56'29.36"W and latitude: 51°26'11.79"N-longitude: 0°56'29.84"W respectively. The cage dimensions were 5x12.5m long and the height is 2m. Metallic posts connected constructed the form of the cage. Insect mesh netting was set over the metallic structure to protect the crop from pests and to keep the later incoming bees limited in the crop boundaries aiming to maximum cross-pollination among the selected lines.

Cage lay-out and irrigation

The O_2 outcrossing population consisted of 1,728 bulked seeds drawn from the 93 highly heterozygous selected plants of the O_1 outcrossing generation which were allocated in 16 rows of 108 plants. On 2018 and 2019 the rainy weather condition did not permit seed sowing directly in the field, so the seeds composing the O_4 and O_6 outcrossing populations were pre- germinated into trays and transplanted as 4-5 true leaf seedlings into the cage. Bin all three years of cage experiments, seeds and seedlings were sown/ transplanted in 10cm within the rows and 20cm apart between rows. They were blocked in 4 seed beds of 4 rows each and between the blocks there were walking paths of 50cm, allowing space for phenotyping, sampling and treatments (Figure 2A1). There were also 50cm of extra working area at the two ends of the cage. The paths were covered with black weed suppressant matting. A T-tape drip irrigation line was placed on one side only of each row as appeared in the figure. To support plants while growing, wooden posts were placed at the four edges of every block where strings were supporting the plants from lodging.

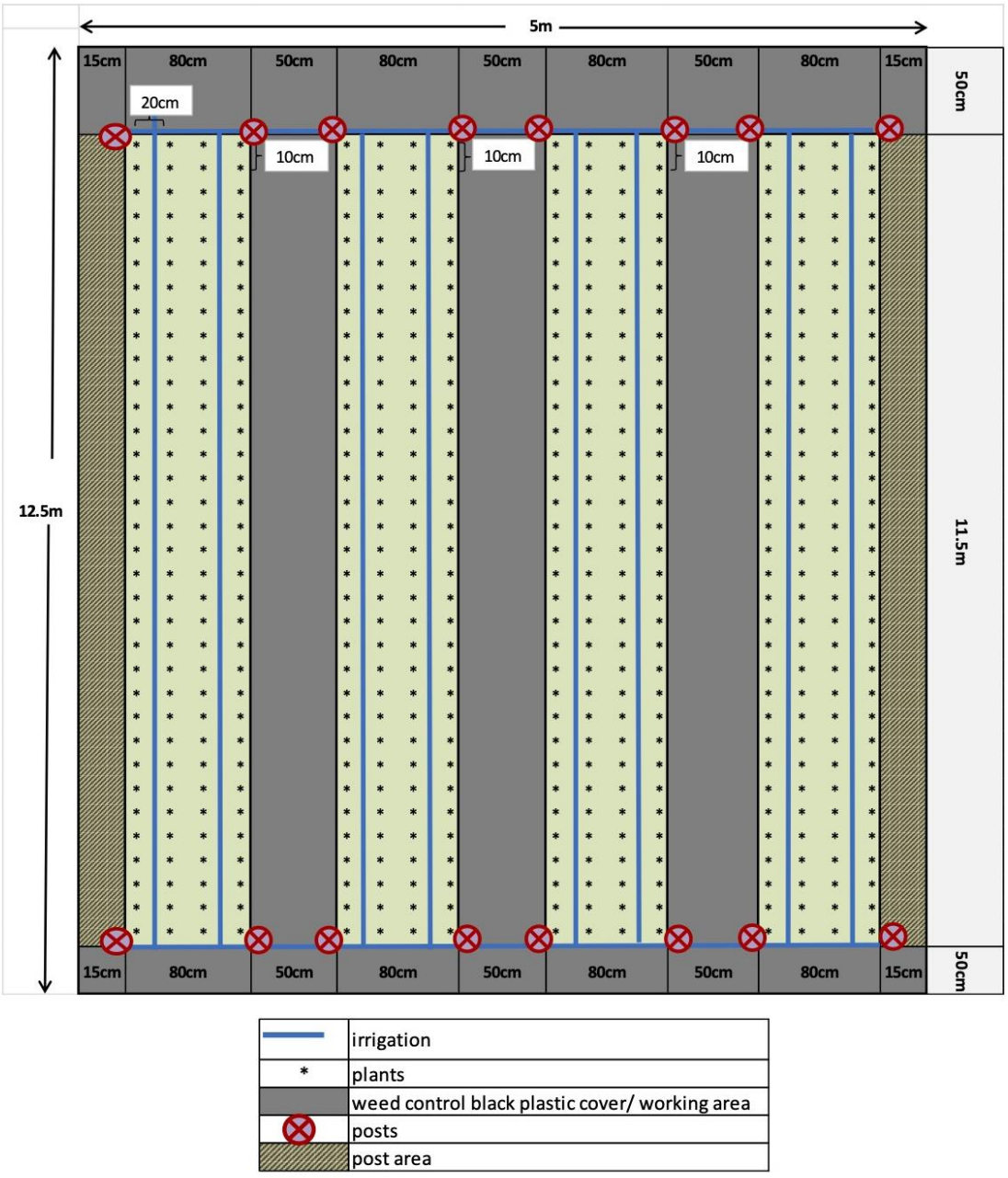


Figure 1B. Cage design showing the arrangement of the individuals in four seed beds and in 20 × 10cm distance (40 plants per m^2). Irrigation line, positions of plants, weed matting and posts are illustrated as indicated in the label above.

Appendix C – Annotated AlphaSimR code for the MRSS simulations

```
rm(list = ls())
library(AlphaSimR)
#Define trait architecture parameters###
Founders=2000 ; fabachr=6 ; segSites= fabasegSites= 1000
fabaQtl=300 ; minSnpFreq= fabaSnpFreq= 0.1
a= 50 ; mean=b=0 ; c=0.3 ; d=0.1 ; e=0.3
f=0.5 #A new variable for the residual error
#Define variable parameters###
selection1=40 ; selectedSeeds1=50 ; fabaSelf1=0.25
number_of_simulations=100 ;set.seed(12345)
##########################################################################
#Simulate founder population and add trait###
founderPop = runMacs(nInd=Founders,nChr=fabachr,segSites=fabasegSites,
                     inbred=FALSE)
SP = SimParam$new(founderPop)
SP$restrSegSites(maxQtl = fabaQtl,  overlap = FALSE, minSnpFreq = fabaSnpFreq,
force = FALSE)
SP$addTraitAG(nQtlPerChr=a, mean=b, var=c, varEnv=d, varGxE = e, gamma =
TRUE, shape = 0.5)
BasePop1=newPop(founderPop)
x= meanG(BasePop1)
y= genicVarG(BasePop1)
n = number_of_simulations
x=rep(0,n); y=rep(0,n)
##########################################################################
#Scenario 1 for selection1 of 40 best plants in each cycle with 50 seeds per plant to
Generate a pop of 2000 for the next cycle#
xA1 = rep(0,n); yA1=rep(0,n) ;xA2 = rep(0,n); yA2=rep(0,n) ;xA3 = rep(0,n);
yA3=rep(0,n) ;
xA4 = rep(0,n); yA4=rep(0,n) ;xA5 = rep(0,n); yA5=rep(0,n) ;xA6 = rep(0,n);
yA6=rep(0,n) ;
```

```
xA7 = rep(0,n); yA7=rep(0,n) ;xA8 = rep(0,n); yA8=rep(0,n) ;xA9 = rep(0,n);
yA9=rep(0,n) ;
xA10 = rep(0,n); yA10=rep(0,n) ;xA11 = rep(0,n); yA11=rep(0,n) ;xA12 = rep(0,n);
yA12=rep(0,n) ;
xA13 = rep(0,n); yA13=rep(0,n) ;xA14 = rep(0,n); yA14=rep(0,n) ;xA15 = rep(0,n);
yA15=rep(0,n) ;
xA16 = rep(0,n); yA16=rep(0,n) ;xA17 = rep(0,n); yA17=rep(0,n) ;xA18 = rep(0,n);
yA18=rep(0,n) ;
xA19 = rep(0,n); yA19=rep(0,n) ;xA20 = rep(0,n); yA20=rep(0,n) ;xA21 = rep(0,n);
yA21=rep(0,n) ;
xA22 = rep(0,n); yA22=rep(0,n) ;xA23 = rep(0,n); yA23=rep(0,n) ;xA24 = rep(0,n);
yA24=rep(0,n) ;
xA25 = rep(0,n); yA25=rep(0,n) ;xA26 = rep(0,n); yA26=rep(0,n) ;xA27 = rep(0,n);
yA27=rep(0,n) ;
xA28 = rep(0,n); yA28=rep(0,n) ;xA29 = rep(0,n); yA29=rep(0,n) ;xA30 = rep(0,n);
yA30=rep(0,n) ;
for (j in 1:n)
{
  BasePop1 = setPheno(BasePop1, varE=f, p=runif(1)) #Sets a phenotype using the
GxAE model
  GenA1 =
selectOP(BasePop1,nInd=selection1,nSeeds=selectedSeeds1,probSelf=fabaSelf1,pollen
Control = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xA1[j]= meanG(GenA1)
  yA1[j]= genicVarG(GenA1)

  GenA1 = setPheno(GenA1, varE=f, p=runif(1)) #Sets a phenotype using the GxAE
model
  GenA2 =
selectOP(GenA1,nInd=selection1,nSeeds=selectedSeeds1,probSelf=fabaSelf1,pollenCo
ntrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xA2[j]= meanG(GenA2)
  yA2[j]= genicVarG(GenA2)
```

```
  GenA2 = setPheno(GenA2, varE=f, p=runif(1)) #Sets a phenotype using the GxAE
model
  GenA3 =
selectOP(GenA2,nInd=selection1,nSeeds=selectedSeeds1,probSelf=fabaSelf1,pollenCo
ntrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xA3[j]=meanG(GenA3)
  yA3[j]= genicVarG(GenA3)

  GenA3 = setPheno(GenA3, varE=f, p=runif(1)) #Sets a phenotype using the GxAE
model
  GenA4 =
selectOP(GenA3,nInd=selection1,nSeeds=selectedSeeds1,probSelf=fabaSelf1,pollenCo
ntrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xA4[j]=meanG(GenA4)
  yA4[j]= genicVarG(GenA4)

  GenA4 = setPheno(GenA4, varE=f, p=runif(1)) #Sets a phenotype using the GxAE
model
  GenA5 =
selectOP(GenA4,nInd=selection1,nSeeds=selectedSeeds1,probSelf=fabaSelf1,pollenCo
ntrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xA5[j]= meanG(GenA5)
  yA5[j]= genicVarG(GenA5)

  GenA5 = setPheno(GenA5, varE=f, p=runif(1)) #Sets a phenotype using the GxAE
model
  GenA6 =
selectOP(GenA5,nInd=selection1,nSeeds=selectedSeeds1,probSelf=fabaSelf1,pollenCo
ntrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xA6[j]= meanG(GenA6)
  yA6[j]= genicVarG(GenA6)

  GenA6 = setPheno(GenA6, varE=f, p=runif(1)) #Sets a phenotype using the GxAE
model
```

```
  GenA7 =
selectOP(GenA6,nInd=selection1,nSeeds=selectedSeeds1,probSelf=fabaSelf1,pollenCo
ntrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xA7[j]= meanG(GenA7)
  yA7[j]= genicVarG(GenA7)

  GenA7 = setPheno(GenA7, varE=f, p=runif(1)) #Sets a phenotype using the GxAE
model
  GenA8 =
selectOP(GenA7,nInd=selection1,nSeeds=selectedSeeds1,probSelf=fabaSelf1,pollenCo
ntrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xA8[j]= meanG(GenA8)
  yA8[j]= genicVarG(GenA8)

  GenA8 = setPheno(GenA8, varE=f, p=runif(1)) #Sets a phenotype using the GxAE
model
  GenA9 =
selectOP(GenA8,nInd=selection1,nSeeds=selectedSeeds1,probSelf=fabaSelf1,pollenCo
ntrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xA9[j]= meanG(GenA9)
  yA9[j]= genicVarG(GenA9)

  GenA9 = setPheno(GenA9, varE=f, p=runif(1)) #Sets a phenotype using the GxAE
model
  GenA10 =
selectOP(GenA9,nInd=selection1,nSeeds=selectedSeeds1,probSelf=fabaSelf1,pollenCo
ntrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xA10[j]= meanG(GenA10)
  yA10[j]= genicVarG(GenA10)

  GenA10 = setPheno(GenA10, varE=f, p=runif(1)) #Sets a phenotype using the GxAE
model
```

```
  GenA11 =
selectOP(GenA10,nInd=selection1,nSeeds=selectedSeeds1,probSelf=fabaSelf1,pollenC
ontrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xA11[j]= meanG(GenA11)
  yA11[j]= genicVarG(GenA11)

  GenA11 = setPheno(GenA11, varE=f, p=runif(1)) #Sets a phenotype using the GxAE
model
  GenA12 =
selectOP(GenA11,nInd=selection1,nSeeds=selectedSeeds1,probSelf=fabaSelf1,pollenC
ontrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xA12[j]= meanG(GenA12)
  yA12[j]= genicVarG(GenA12)

  GenA12 = setPheno(GenA12, varE=f, p=runif(1)) #Sets a phenotype using the GxAE
model
  GenA13 =
selectOP(GenA12,nInd=selection1,nSeeds=selectedSeeds1,probSelf=fabaSelf1,pollenC
ontrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xA13[j]= meanG(GenA13)
  yA13[j]= genicVarG(GenA13)

  GenA13 = setPheno(GenA13, varE=f, p=runif(1)) #Sets a phenotype using the GxAE
model
  GenA14 =
selectOP(GenA13,nInd=selection1,nSeeds=selectedSeeds1,probSelf=fabaSelf1,pollenC
ontrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xA14[j]= meanG(GenA14)
  yA14[j]= genicVarG(GenA14)

  GenA14 = setPheno(GenA14, varE=f, p=runif(1)) #Sets a phenotype using the GxAE
model
```

```
  GenA15 =
selectOP(GenA14,nInd=selection1,nSeeds=selectedSeeds1,probSelf=fabaSelf1,pollenC
ontrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xA15[j]= meanG(GenA15)
  yA15[j]= genicVarG(GenA15)

  GenA15 = setPheno(GenA15, varE=f, p=runif(1)) #Sets a phenotype using the GxAE
model
  GenA16 =
selectOP(GenA15,nInd=selection1,nSeeds=selectedSeeds1,probSelf=fabaSelf1,pollenC
ontrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xA16[j]= meanG(GenA16)
  yA16[j]= genicVarG(GenA16)

  GenA16 = setPheno(GenA16, varE=f, p=runif(1)) #Sets a phenotype using the GxAE
model
  GenA17 =
selectOP(GenA16,nInd=selection1,nSeeds=selectedSeeds1,probSelf=fabaSelf1,pollenC
ontrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xA17[j]=meanG(GenA17)
  yA17[j]= genicVarG(GenA17)

  GenA17 = setPheno(GenA17, varE=f, p=runif(1)) #Sets a phenotype using the GxAE
model
  GenA18 =
selectOP(GenA17,nInd=selection1,nSeeds=selectedSeeds1,probSelf=fabaSelf1,pollenC
ontrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xA18[j]= meanG(GenA18)
  yA18[j]= genicVarG(GenA18)

  GenA18 = setPheno(GenA18, varE=f, p=runif(1)) #Sets a phenotype using the GxAE
model
```

```
  GenA19 =
selectOP(GenA18,nInd=selection1,nSeeds=selectedSeeds1,probSelf=fabaSelf1,pollenC
ontrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xA19[j]= meanG(GenA19)
  yA19[j]= genicVarG(GenA19)

  GenA19 = setPheno(GenA19, varE=f, p=runif(1)) #Sets a phenotype using the GxAE
model
  GenA20 =
selectOP(GenA19,nInd=selection1,nSeeds=selectedSeeds1,probSelf=fabaSelf1,pollenC
ontrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xA20[j]= meanG(GenA20)
  yA20[j]= genicVarG(GenA20)

  GenA20 = setPheno(GenA20, varE=f, p=runif(1)) #Sets a phenotype using the GxAE
model
  GenA21 =
selectOP(GenA20,nInd=selection1,nSeeds=selectedSeeds1,probSelf=fabaSelf1,pollenC
ontrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xA21[j]= meanG(GenA21)
  yA21[j]= genicVarG(GenA21)

  GenA21 = setPheno(GenA21, varE=f, p=runif(1)) #Sets a phenotype using the GxAE
model
  GenA22 =
selectOP(GenA21,nInd=selection1,nSeeds=selectedSeeds1,probSelf=fabaSelf1,pollenC
ontrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xA22[j]= meanG(GenA22)
  yA22[j]= genicVarG(GenA22)

  GenA22 = setPheno(GenA22, varE=f, p=runif(1)) #Sets a phenotype using the GxAE
model
```

```
  GenA23 =
selectOP(GenA22,nInd=selection1,nSeeds=selectedSeeds1,probSelf=fabaSelf1,pollenC
ontrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xA23[j]= meanG(GenA23)
  yA23[j]= genicVarG(GenA23)

  GenA23 = setPheno(GenA23, varE=f, p=runif(1)) #Sets a phenotype using the GxAE
model
  GenA24 =
selectOP(GenA23,nInd=selection1,nSeeds=selectedSeeds1,probSelf=fabaSelf1,pollenC
ontrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xA24[j]= meanG(GenA24)
  yA24[j]= genicVarG(GenA24)

  GenA24 = setPheno(GenA24, varE=f, p=runif(1)) #Sets a phenotype using the GxAE
model
  GenA25 =
selectOP(GenA24,nInd=selection1,nSeeds=selectedSeeds1,probSelf=fabaSelf1,pollenC
ontrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xA25[j]= meanG(GenA25)
  yA25[j]= genicVarG(GenA25)

  GenA25 = setPheno(GenA25, varE=f, p=runif(1)) #Sets a phenotype using the GxAE
model
  GenA26 =
selectOP(GenA25,nInd=selection1,nSeeds=selectedSeeds1,probSelf=fabaSelf1,pollenC
ontrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xA26[j]= meanG(GenA26)
  yA26[j]= genicVarG(GenA26)

  GenA26 = setPheno(GenA26, varE=f, p=runif(1)) #Sets a phenotype using the GxAE
model
```

```
  GenA27 =
selectOP(GenA26,nInd=selection1,nSeeds=selectedSeeds1,probSelf=fabaSelf1,pollenC
ontrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xA27[j]= meanG(GenA27)
  yA27[j]= genicVarG(GenA27)

  GenA27 = setPheno(GenA27, varE=f, p=runif(1)) #Sets a phenotype using the GxAE
model
  GenA28 = selectOP(GenA27,
nInd=selection1,nSeeds=selectedSeeds1,probSelf=fabaSelf1,pollenControl = FALSE,
trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xA28[j]= meanG(GenA28)
  yA28[j]= genicVarG(GenA28)

  GenA28 = setPheno(GenA28, varE=f, p=runif(1)) #Sets a phenotype using the GxAE
model
  GenA29 =
selectOP(GenA28,nInd=selection1,nSeeds=selectedSeeds1,probSelf=fabaSelf1,pollenC
ontrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xA29[j]= meanG(GenA29)
  yA29[j]= genicVarG(GenA29)

  GenA29 = setPheno(GenA29, varE=f, p=runif(1)) #Sets a phenotype using the GxAE
model
  GenA30 =
selectOP(GenA29,nInd=selection1,nSeeds=selectedSeeds1,probSelf=fabaSelf1,pollenC
ontrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xA30[j]= meanG(GenA30)
  yA30[j]= genicVarG(GenA30)
}
GGA1_S1=mean(xA1);GGA2_S1=mean(xA2);GGA3_S1=mean(xA3);GGA4_S1=mea
n(xA4);GGA5_S1=mean(xA5);GGA6_S1=mean(xA6);GGA7_S1=mean(xA7);GGA8_
S1=mean(xA8);GGA9_S1=mean(xA9);GGA10_S1=mean(xA10) ;
GGA11_S1=mean(xA11);GGA12_S1=mean(xA12); GGA13_S1=mean(xA13);
```

```
GGA14_S1=mean(xA14);GGA15_S1=mean(xA15);GGA16_S1=mean(xA16);GGA17
_S1=mean(xA17);GGA18_S1=mean(xA18);GGA19_S1=mean(xA19);GGA20_S1=me
an(xA20) ;GGA21_S1=mean(xA21);
GGA22_S1=mean(xA22);GGA23_S1=mean(xA23);GGA24_S1=mean(xA24);GGA25
_S1=mean(xA25);GGA26_S1=mean(xA26);GGA27_S1=mean(xA27);GGA28_S1=me
an(xA28);GGA29_S1=mean(xA29);GGA30_S1=mean(xA30)
GVA1_S1=mean(yA1);GVA2_S1=mean(yA2);GVA3_S1=mean(yA3);GVA4_S1=mea
n(yA4);GVA5_S1=mean(yA5);GVA6_S1=mean(yA6);GVA7_S1=mean(yA7);GVA8_
S1=mean(yA8);GVA9_S1=mean(yA9);GVA10_S1=mean(yA10)
;GVA11_S1=mean(yA11);GVA12_S1=mean(yA12); GVA13_S1=mean(yA13);
GVA14_S1=mean(yA14);GVA15_S1=mean(yA15);GVA16_S1=mean(yA16);GVA17
_S1=mean(yA17);GVA18_S1=mean(yA18);GVA19_S1=mean(yA19);GVA20_S1=me
an(yA20) ;GVA21_S1=mean(yA21);
GVA22_S1=mean(yA22);GVA23_S1=mean(yA23);GVA24_S1=mean(yA24);GVA25
_S1=mean(yA25);GVA26_S1=mean(yA26);GVA27_S1=mean(yA27);GVA28_S1=me
an(yA28);GVA29_S1=mean(yA29);GVA30_S1=mean(yA30) ;
Genetic_gain_S1<-
c(GGA1_S1,GGA2_S1,GGA3_S1,GGA4_S1,GGA5_S1,GGA6_S1,GGA7_S1,GGA8_
S1,GGA9_S1,GGA10_S1,GGA11_S1,GGA12_S1,GGA13_S1,GGA14_S1,GGA15_S
1,GGA16_S1,GGA17_S1,GGA18_S1,GGA19_S1,GGA20_S1,GGA21_S1,GGA22_S1
,GGA23_S1,GGA24_S1,GGA25_S1,GGA26_S1,GGA27_S1,GGA28_S1,GGA29_S1,
GGA30_S1)
Genetic_variation_S1<-
c(GVA1_S1,GVA2_S1,GVA3_S1,GVA4_S1,GVA5_S1,GVA6_S1,GVA7_S1,GVA8_
S1,GVA9_S1,GVA10_S1,GVA11_S1,GVA12_S1,GVA13_S1,GVA14_S1,GVA15_S
1,GVA16_S1,GVA17_S1,GVA18_S1,GVA19_S1,GVA20_S1,GVA21_S1,GVA22_S1
,GVA23_S1,GVA24_S1,GVA25_S1,GVA26_S1,GVA27_S1,GVA28_S1,GVA29_S1,
GVA30_S1)

######Scenario 2 for selection of 80 best plants in each cycle with 25 seeds per plant to
generate a pop of 2000 for the next cycle#####
#Define variable parameters###
selection2=80 ; selectedSeeds2=25 ; fabaSelf2=0.25
```

```
xB1 = rep(0,n); yB1=rep(0,n) ;xB2 = rep(0,n); yB2=rep(0,n) ;xB3 = rep(0,n);
yB3=rep(0,n) ;
xB4 = rep(0,n); yB4=rep(0,n) ;xB5 = rep(0,n); yB5=rep(0,n) ;xB6 = rep(0,n);
yB6=rep(0,n) ;
xB7 = rep(0,n); yB7=rep(0,n) ;xB8 = rep(0,n); yB8=rep(0,n) ;xB9 = rep(0,n);
yB9=rep(0,n) ;
xB10 = rep(0,n); yB10=rep(0,n) ;xB11 = rep(0,n); yB11=rep(0,n) ;xB12 = rep(0,n);
yB12=rep(0,n) ;
xB13 = rep(0,n); yB13=rep(0,n) ;xB14 = rep(0,n); yB14=rep(0,n) ;xB15 = rep(0,n);
yB15=rep(0,n) ;
xB16 = rep(0,n); yB16=rep(0,n) ;xB17 = rep(0,n); yB17=rep(0,n) ;xB18 = rep(0,n);
yB18=rep(0,n) ;
xB19 = rep(0,n); yB19=rep(0,n) ;xB20 = rep(0,n); yB20=rep(0,n) ;xB21 = rep(0,n);
yB21=rep(0,n) ;
xB22 = rep(0,n); yB22=rep(0,n) ;xB23 = rep(0,n); yB23=rep(0,n) ;xB24 = rep(0,n);
yB24=rep(0,n) ;
xB25 = rep(0,n); yB25=rep(0,n) ;xB26 = rep(0,n); yB26=rep(0,n) ;xB27 = rep(0,n);
yB27=rep(0,n) ;
xB28 = rep(0,n); yB28=rep(0,n) ;xB29 = rep(0,n); yB29=rep(0,n) ;xB30 = rep(0,n);
yB30=rep(0,n) ;
for (j in 1:n)
{
  BasePop1 = setPheno(BasePop1, varE=f, p=runif(1)) #Sets a phenotyBpe using the
GxBE model
  GenB1 =
selectOP(BasePop1,nInd=selection2,nSeeds=selectedSeeds2,probSelf=fabaSelf2,pollen
Control = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xB1[j]= meanG(GenB1)
  yB1[j]= genicVarG(GenB1)

  GenB1 = setPheno(GenB1, varE=f, p=runif(1)) #Sets a phenotyBpe using the GxBE
model
```

```
  GenB2 =
selectOP(GenB1,nInd=selection2,nSeeds=selectedSeeds2,probSelf=fabaSelf2,pollenCo
ntrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xB2[j]= meanG(GenB2)
  yB2[j]= genicVarG(GenB2)

  GenB2 = setPheno(GenB2, varE=f, p=runif(1)) #Sets a phenotyBpe using the GxBE
model
  GenB3 =
selectOP(GenB2,nInd=selection2,nSeeds=selectedSeeds2,probSelf=fabaSelf2,pollenCo
ntrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xB3[j]=meanG(GenB3)
  yB3[j]= genicVarG(GenB3)

  GenB3 = setPheno(GenB3, varE=f, p=runif(1)) #Sets a phenotyBpe using the GxBE
model
  GenB4 =
selectOP(GenB3,nInd=selection2,nSeeds=selectedSeeds2,probSelf=fabaSelf2,pollenCo
ntrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xB4[j]=meanG(GenB4)
  yB4[j]= genicVarG(GenB4)

  GenB4 = setPheno(GenB4, varE=f, p=runif(1)) #Sets a phenotyBpe using the GxBE
model
  GenB5 =
selectOP(GenB4,nInd=selection2,nSeeds=selectedSeeds2,probSelf=fabaSelf2,pollenCo
ntrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xB5[j]= meanG(GenB5)
  yB5[j]= genicVarG(GenB5)

  GenB5 = setPheno(GenB5, varE=f, p=runif(1)) #Sets a phenotyBpe using the GxBE
model
```

```
  GenB6 =
selectOP(GenB5,nInd=selection2,nSeeds=selectedSeeds2,probSelf=fabaSelf2,pollenCo
ntrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xB6[j]= meanG(GenB6)
  yB6[j]= genicVarG(GenB6)

  GenB6 = setPheno(GenB6, varE=f, p=runif(1)) #Sets a phenotyBpe using the GxBE
model
  GenB7 =
selectOP(GenB6,nInd=selection2,nSeeds=selectedSeeds2,probSelf=fabaSelf2,pollenCo
ntrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xB7[j]= meanG(GenB7)
  yB7[j]= genicVarG(GenB7)

  GenB7 = setPheno(GenB7, varE=f, p=runif(1)) #Sets a phenotyBpe using the GxBE
model
  GenB8 =
selectOP(GenB7,nInd=selection2,nSeeds=selectedSeeds2,probSelf=fabaSelf2,pollenCo
ntrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xB8[j]= meanG(GenB8)
  yB8[j]= genicVarG(GenB8)

  GenB8 = setPheno(GenB8, varE=f, p=runif(1)) #Sets a phenotyBpe using the GxBE
model
  GenB9 =
selectOP(GenB8,nInd=selection2,nSeeds=selectedSeeds2,probSelf=fabaSelf2,pollenCo
ntrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xB9[j]= meanG(GenB9)
  yB9[j]= genicVarG(GenB9)

  GenB9 = setPheno(GenB9, varE=f, p=runif(1)) #Sets a phenotyBpe using the GxBE
model
```

```
  GenB10 =
selectOP(GenB9,nInd=selection2,nSeeds=selectedSeeds2,probSelf=fabaSelf2,pollenCo
ntrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xB10[j]= meanG(GenB10)
  yB10[j]= genicVarG(GenB10)

  GenB10 = setPheno(GenB10, varE=f, p=runif(1)) #Sets a phenotyBpe using the GxBE
model
  GenB11 =
selectOP(GenB10,nInd=selection2,nSeeds=selectedSeeds2,probSelf=fabaSelf2,pollenC
ontrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xB11[j]= meanG(GenB11)
  yB11[j]= genicVarG(GenB11)

  GenB11 = setPheno(GenB11, varE=f, p=runif(1)) #Sets a phenotyBpe using the GxBE
model
  GenB12 =
selectOP(GenB11,nInd=selection2,nSeeds=selectedSeeds2,probSelf=fabaSelf2,pollenC
ontrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xB12[j]= meanG(GenB12)
  yB12[j]= genicVarG(GenB12)

  GenB12 = setPheno(GenB12, varE=f, p=runif(1)) #Sets a phenotyBpe using the GxBE
model
  GenB13 =
selectOP(GenB12,nInd=selection2,nSeeds=selectedSeeds2,probSelf=fabaSelf2,pollenC
ontrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xB13[j]= meanG(GenB13)
  yB13[j]= genicVarG(GenB13)

  GenB13 = setPheno(GenB13, varE=f, p=runif(1)) #Sets a phenotyBpe using the GxBE
model
```

```
  GenB14 =
selectOP(GenB13,nInd=selection2,nSeeds=selectedSeeds2,probSelf=fabaSelf2,pollenC
ontrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xB14[j]= meanG(GenB14)
  yB14[j]= genicVarG(GenB14)

  GenB14 = setPheno(GenB14, varE=f, p=runif(1)) #Sets a phenotyBpe using the GxBE
model
  GenB15 =
selectOP(GenB14,nInd=selection2,nSeeds=selectedSeeds2,probSelf=fabaSelf2,pollenC
ontrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xB15[j]= meanG(GenB15)
  yB15[j]= genicVarG(GenB15)

  GenB15 = setPheno(GenB15, varE=f, p=runif(1)) #Sets a phenotyBpe using the GxBE
model
  GenB16 =
selectOP(GenB15,nInd=selection2,nSeeds=selectedSeeds2,probSelf=fabaSelf2,pollenC
ontrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xB16[j]= meanG(GenB16)
  yB16[j]= genicVarG(GenB16)

  GenB16 = setPheno(GenB16, varE=f, p=runif(1)) #Sets a phenotyBpe using the GxBE
model
  GenB17 =
selectOP(GenB16,nInd=selection2,nSeeds=selectedSeeds2,probSelf=fabaSelf2,pollenC
ontrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xB17[j]=meanG(GenB17)
  yB17[j]= genicVarG(GenB17)

  GenB17 = setPheno(GenB17, varE=f, p=runif(1)) #Sets a phenotyBpe using the GxBE
model
```

```
  GenB18 =
selectOP(GenB17,nInd=selection2,nSeeds=selectedSeeds2,probSelf=fabaSelf2,pollenC
ontrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xB18[j]= meanG(GenB18)
  yB18[j]= genicVarG(GenB18)

  GenB18 = setPheno(GenB18, varE=f, p=runif(1)) #Sets a phenotyBpe using the GxBE
model
  GenB19 =
selectOP(GenB18,nInd=selection2,nSeeds=selectedSeeds2,probSelf=fabaSelf2,pollenC
ontrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xB19[j]= meanG(GenB19)
  yB19[j]= genicVarG(GenB19)

  GenB19 = setPheno(GenB19, varE=f, p=runif(1)) #Sets a phenotyBpe using the GxBE
model
  GenB20 =
selectOP(GenB19,nInd=selection2,nSeeds=selectedSeeds2,probSelf=fabaSelf2,pollenC
ontrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xB20[j]= meanG(GenB20)
  yB20[j]= genicVarG(GenB20)

  GenB20 = setPheno(GenB20, varE=f, p=runif(1)) #Sets a phenotyBpe using the GxBE
model
  GenB21 =
selectOP(GenB20,nInd=selection2,nSeeds=selectedSeeds2,probSelf=fabaSelf2,pollenC
ontrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xB21[j]= meanG(GenB21)
  yB21[j]= genicVarG(GenB21)

  GenB21 = setPheno(GenB21, varE=f, p=runif(1)) #Sets a phenotyBpe using the GxBE
model
```

```
  GenB22 =
selectOP(GenB21,nInd=selection2,nSeeds=selectedSeeds2,probSelf=fabaSelf2,pollenC
ontrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xB22[j]= meanG(GenB22)
  yB22[j]= genicVarG(GenB22)

  GenB22 = setPheno(GenB22, varE=f, p=runif(1)) #Sets a phenotyBpe using the GxBE
model
  GenB23 =
selectOP(GenB22,nInd=selection2,nSeeds=selectedSeeds2,probSelf=fabaSelf2,pollenC
ontrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xB23[j]= meanG(GenB23)
  yB23[j]= genicVarG(GenB23)

  GenB23 = setPheno(GenB23, varE=f, p=runif(1)) #Sets a phenotyBpe using the GxBE
model
  GenB24 =
selectOP(GenB23,nInd=selection2,nSeeds=selectedSeeds2,probSelf=fabaSelf2,pollenC
ontrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xB24[j]= meanG(GenB24)
  yB24[j]= genicVarG(GenB24)

  GenB24 = setPheno(GenB24, varE=f, p=runif(1)) #Sets a phenotyBpe using the GxBE
model
  GenB25 =
selectOP(GenB24,nInd=selection2,nSeeds=selectedSeeds2,probSelf=fabaSelf2,pollenC
ontrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xB25[j]= meanG(GenB25)
  yB25[j]= genicVarG(GenB25)

  GenB25 = setPheno(GenB25, varE=f, p=runif(1)) #Sets a phenotyBpe using the GxBE
model
```

```
  GenB26 =
selectOP(GenB25,nInd=selection2,nSeeds=selectedSeeds2,probSelf=fabaSelf2,pollenC
ontrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xB26[j]= meanG(GenB26)
  yB26[j]= genicVarG(GenB26)

  GenB26 = setPheno(GenB26, varE=f, p=runif(1)) #Sets a phenotyBpe using the GxBE
model
  GenB27 =
selectOP(GenB26,nInd=selection2,nSeeds=selectedSeeds2,probSelf=fabaSelf2,pollenC
ontrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xB27[j]= meanG(GenB27)
  yB27[j]= genicVarG(GenB27)

  GenB27 = setPheno(GenB27, varE=f, p=runif(1)) #Sets a phenotyBpe using the GxBE
model
  GenB28 = selectOP(GenB27,
nInd=selection2,nSeeds=selectedSeeds2,probSelf=fabaSelf2,pollenControl = FALSE,
trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xB28[j]= meanG(GenB28)
  yB28[j]= genicVarG(GenB28)

  GenB28 = setPheno(GenB28, varE=f, p=runif(1)) #Sets a phenotyBpe using the GxBE
model
  GenB29 =
selectOP(GenB28,nInd=selection2,nSeeds=selectedSeeds2,probSelf=fabaSelf2,pollenC
ontrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xB29[j]= meanG(GenB29)
  yB29[j]= genicVarG(GenB29)

  GenB29 = setPheno(GenB29, varE=f, p=runif(1)) #Sets a phenotyBpe using the GxBE
model
```

```
  GenB30 =
selectOP(GenB29,nInd=selection2,nSeeds=selectedSeeds2,probSelf=fabaSelf2,pollenC
ontrol = FALSE, trait = 1, use ="pheno" ,selectTop = TRUE, simParam = SP )
  xB30[j]= meanG(GenB30)
  yB30[j]= genicVarG(GenB30)
}
GGB1_S2=mean(xB1);GGB2_S2=mean(xB2);GGB3_S2=mean(xB3);GGB4_S2=mea
n(xB4);GGB5_S2=mean(xB5);GGB6_S2=mean(xB6);GGB7_S2=mean(xB7);GGB8_S
2=mean(xB8);GGB9_S2=mean(xB9);GGB10_S2=mean(xB10)
;GGB11_S2=mean(xB11);GGB12_S2=mean(xB12); GGB13_S2=mean(xB13);
GGB14_S2=mean(xB14);GGB15_S2=mean(xB15);GGB16_S2=mean(xB16);GGB17_
S2=mean(xB17);GGB18_S2=mean(xB18);GGB19_S2=mean(xB19);GGB20_S2=mean
(xB20) ;GGB21_S2=mean(xB21);
GGB22_S2=mean(xB22);GGB23_S2=mean(xB23);GGB24_S2=mean(xB24);GGB25_
S2=mean(xB25);GGB26_S2=mean(xB26);GGB27_S2=mean(xB27);GGB28_S2=mean
(xB28);GGB29_S2=mean(xB29);GGB30_S2=mean(xB30)
GVB1_S2=mean(yB1);GVB2_S2=mean(yB2);GVB3_S2=mean(yB3);GVB4_S2=mea
n(yB4);GVB5_S2=mean(yB5);GVB6_S2=mean(yB6);GVB7_S2=mean(yB7);GVB8_S
2=mean(yB8);GVB9_S2=mean(yB9);GVB10_S2=mean(yB10)
;GVB11_S2=mean(yB11);GVB12_S2=mean(yB12); GVB13_S2=mean(yB13);
GVB14_S2=mean(yB14);GVB15_S2=mean(yB15);GVB16_S2=mean(yB16);GVB17_
S2=mean(yB17);GVB18_S2=mean(yB18);GVB19_S2=mean(yB19);GVB20_S2=mean
(yB20) ;GVB21_S2=mean(yB21);
GVB22_S2=mean(yB22);GVB23_S2=mean(yB23);GVB24_S2=mean(yB24);GVB25_
S2=mean(yB25);GVB26_S2=mean(yB26);GVB27_S2=mean(yB27);GVB28_S2=mean
(yB28);GVB29_S2=mean(yB29);GVB30_S2=mean(yB30)
Genetic_gain_S2<-
c(GGB1_S2,GGB2_S2,GGB3_S2,GGB4_S2,GGB5_S2,GGB6_S2,GGB7_S2,GGB8_
S2,GGB9_S2,GGB10_S2,GGB11_S2,GGB12_S2,GGB13_S2,GGB14_S2,GGB15_S2,
GGB16_S2,GGB17_S2,GGB18_S2,GGB19_S2,GGB20_S2,GGB21_S2,GGB22_S2,G
GB23_S2,GGB24_S2,GGB25_S2,GGB26_S2,GGB27_S2,GGB28_S2,GGB29_S2,GG
B30_S2)
Genetic_variation_S2<-
c(GVB1_S2,GVB2_S2,GVB3_S2,GVB4_S2,GVB5_S2,GVB6_S2,GVB7_S2,GVB8_
```

S2,GVB9_S2,GVB10_S2,GVB11_S2,GVB12_S2,GVB13_S2,GVB14_S2,GVB15_S2,GVB16_S2,GVB17_S2,GVB18_S2,GVB19_S2,GVB20_S2,GVB21_S2,GVB22_S2,GVB23_S2,GVB24_S2,GVB25_S2,GVB26_S2,GVB27_S2,GVB28_S2,GVB29_S2,GVB30_S2)

###Parameters for Scenarios 3 to 18###
#Scenario 3 for selection of 200 best plants in each cycle with 10 seeds per plant to generate a pop of 2000 for the next cycle###
selection3=200 ; selectedSeeds3=10 ; fabaSelf3=0.25
#Scenario 4 for selection of 200 best plants in each cycle with 50 seeds per plant to generate a pop of 10000 for the next cycle###
selection4=200 ; selectedSeeds4=50 ; fabaSelf4=0.25
#Scenario 5 for selection of 400 best plants in each cycle with 25 seeds per plant to generate a pop of 10000 for the next cycle###
selection5=400 ; selectedSeeds5=25 ; fabaSelf5=0.25
#Scenario 6 for selection of 1000 best plants in each cycle with 10 seeds per plant to generate a pop of 10000 for the next cycle###
selection6=1000 ; selectedSeeds6=10 ; fabaSelf6=0.25
#Scenario 7 for selection of 400 best plants in each cycle with 50 seeds per plant to generate a pop of 20000 for the next cycle###
selection7=400 ; selectedSeeds7=50 ; fabaSelf7=0.25
#Scenario 8 for selection of 800 best plants in each cycle with 25 seeds per plant to generate a pop of 20000 for the next cycle###
selection8=800 ; selectedSeeds8=25 ; fabaSelf8=0.25
#Scenario 9 for selection of 2000 best plants in each cycle with 10 seeds per plant to generate a pop of 20000 for the next cycle###
selection9=2000 ; selectedSeeds9=10 ; fabaSelf9=0.25
#Scenario 10 for selection of 40 best plants in each cycle with 50 seeds per plant to generate a pop of 2000 for the next cycle###
selection10=40 ; selectedSeeds10=50 ; fabaSelf10=0.7
#Scenario 11 for selection of 80 best plants in each cycle with 25 seeds per plant to generate a pop of 2000 for the next cycle###
selection11=80 ; selectedSeeds11=25 ; fabaSelf11=0.7

#Scenario 12 for selection of 200 best plants in each cycle with 10 seeds per plant to generate a pop of 2000 for the next cycle###
selection12=200 ; selectedSeeds12=10 ; fabaSelf12=0.7
#Scenario 13 for selection of 200 best plants in each cycle with 50 seeds per plant to generate a pop of 10000 for the next cycle###
selection13=200 ; selectedSeeds13=50 ; fabaSelf13=0.7
#Scenario 14 for selection of 400 best plants in each cycle with 25 seeds per plant to generate a pop of 10000 for the next cycle###
selection14=400 ; selectedSeeds14=25 ; fabaSelf14=0.7
#Scenario 15 for selection of 1000 best plants in each cycle with 10 seeds per plant to generate a pop of 10000 for the next cycle###
selection15=1000 ; selectedSeeds15=10 ; fabaSelf15=0.7
#Scenario 16 for selection of 400 best plants in each cycle with 50 seeds per plant to generate a pop of 20000 for the next cycle###
selection16=400 ; selectedSeeds16=50 ; fabaSelf16=0.7
#Scenario 17 for selection of 800 best plants in each cycle with 25 seeds per plant to generate a pop of 20000 for the next cycle###
selection17=800 ; selectedSeeds17=25 ; fabaSelf17=0.7
#Scenario 18 for selection of 2000 best plants in each cycle with 10 seeds per plant to generate a pop of 20000 for the next cycle###
selection18=2000 ; selectedSeeds18=10 ; fabaSelf18=0.7

Appendix D – Modelling of cost-efficiency of different breeding scenarios

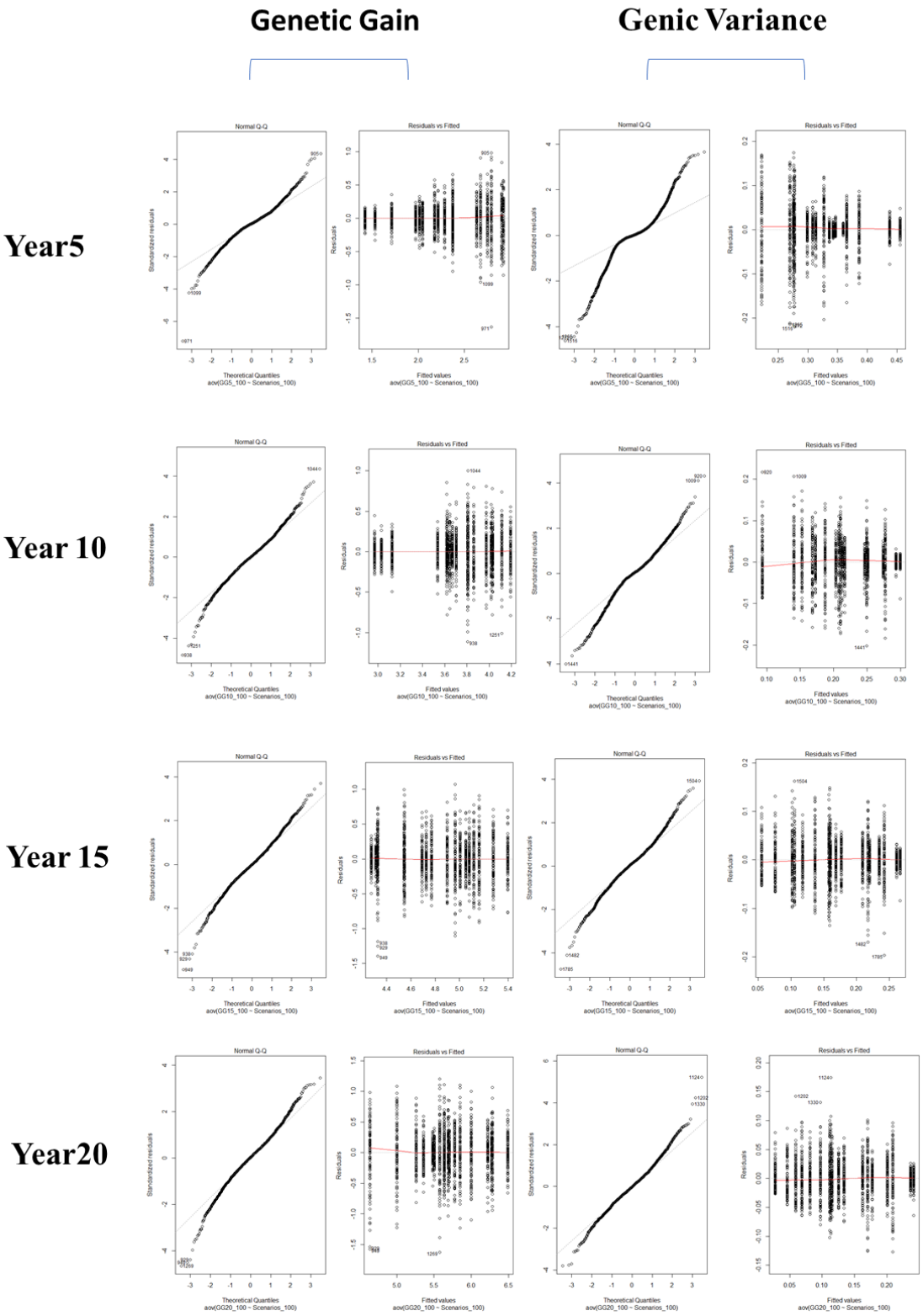
UK's national minimum wage is £8.21 plus holiday pay for age 25 and over. For sowing, harvesting and phenotyping one cage the number of working hours for one seasonal technician was calculated. To the final equation the minimum stipend of a student per year was added along with the standard cost of establishment of the cage structure considering that every 10 years the cages are being renewed.

Table 1D. Cost estimates for various activities in UoR breeding scheme and general crop requirements according to Teagasc guidelines. Inflation is not considered, so that costs in today's money of per unit gain can be compared across timepoints.

	Unit costs per 60m^2 cage (£)					
Activities	**Year 5**	**Year 10**	**Year 15**	**Year 20**	**Year 25**	**Year 30**
Seed	0	0	0	0	0	0
Fertilisers	0	0	0	0	0	0
Sprays:						
Herbicides	£2.4	£4.8	£7.2	£9.6	£12.0	£14.4
Fungicides	£2.4	£4.8	£7.2	£9.6	£12.0	£14.4
Aphicide	£0.2	£0.4	£0.6	£0.8	£1.1	£1.3
Cage cost- Seasonal technician:						
Sowing	£1,539	£3,079	£4,618	£6,158	£7,697	£9,236
Harvesting	£1,539	£3,079	£4,618	£6,158	£7,697	£9,236
Phenotyping	£6,497	£12,994	£19,491	£25,987	£32,484	£38,981
Bee colony:	£1,200	£2,400	£3,600	£4,800	£6,000	£7,200
Infrastructure*:						
Frame	£1,000	£1,000	£2,000	£2,000	£3,000	£3,000
Net	£1,000	£1,000	£2,000	£2,000	£3,000	£3,000
Labour:						
Postgraduate breeder	£75,000	£150,000	£225,000	£300,000	£375,000	£450,000
Total Cost per number of cages						
PS= 2,000	£87,781	£173,561	£261,342	£347,122	£434,903	£520,684
PS= 10,000	£438,903	£867,806	£1,306,709	£1,735,612	£2,174,515	£2,603,418
PS= 20,000	£877,806	£1,735,613	£2,613,418	£3,471,224	£4,349,030	£5,206,837

* : assuming 10 years' use

Appendix E – Homogeneity of variance and normality assumption diagnostic tests for genetic gain and genic variance in six timepoints.



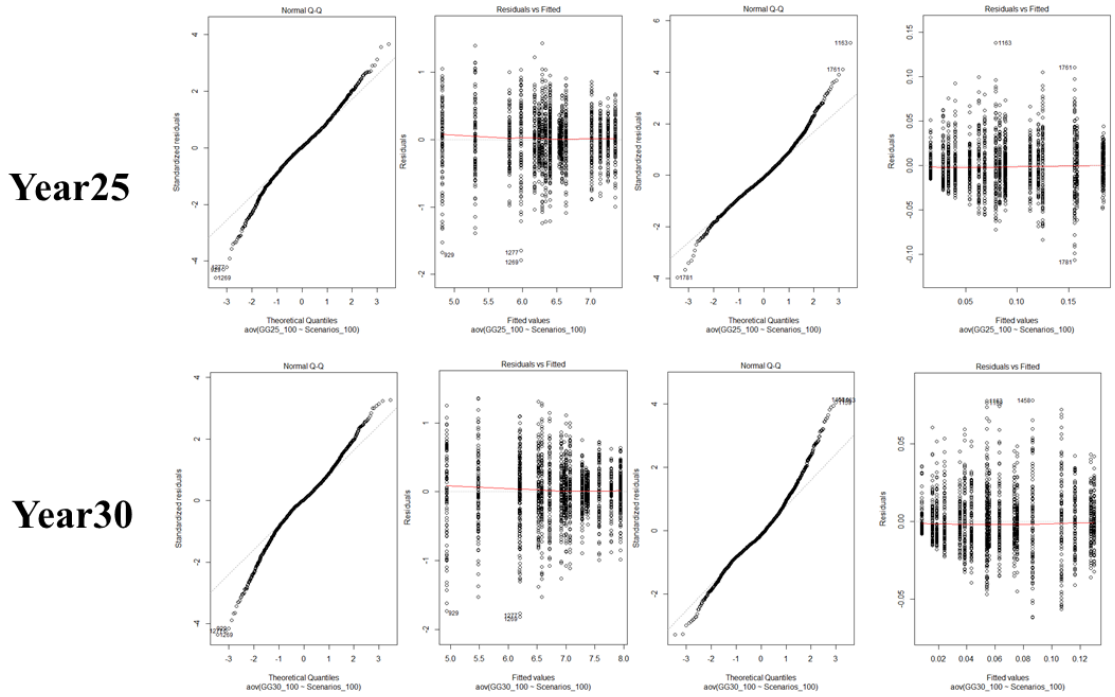


Figure 1E. Homogeneity of variance and normality assumption check for six different timepoints, year/generation 5, 10, 15, 20, 25, 30. For the normality plots, x-axis shows the quantiles of the normal distribution and y-axis the quantiles of the residuals. The dashed line represents ta 45-degree reference line. For homogeneity of variance, x-axis plots the fitted values (mean of each group) and y-axis the residuals. The red line shows that there is no evidence of relationships between residuals and fitted values.