From the
Institute of Crop Science
University of Hohenheim
Biostatistics Unit
Prof. Dr. H.-P. Piepho

The development of phenotypic protocols and adjustment of experimental designs in *Pelargonium zonale* breeding

Dissertation
submitted in fulfillment of the regulation to acquire the degree
‘Doktor der Agrarwissenschaften’
(Dr. sc. agr./Ph.D. in Agricultural Sciences)
to the
Faculty of Agricultural Sciences

presented by
Heike Molenaar
from
Keez
2018

The thesis was excepted as a doctoral thesis (Dissertation) in fulfilment of the regulations to acquire the doctoral degree “Doktor der Agrarwissenschaften” by the Faculty of Agricultural Sciences at University of Hohenheim on July, 11th, 2018.

Date of the oral examination: October, 8th, 2018

Examination Committee

Chairperson of the oral examination Prof. Dr. Bennewitz

Supervisor and Reviewer	Prof. Dr. Piepho
Co-Reviewer	Prof. Dr. Debener (Leibniz University Hannover)
Additional examiner	Prof. Dr. Würschum

Table of contents

1 General introduction

Breeding strategies for new ornamental cultivars lag behind those developed for commodity crops (Debener 2001; Debener 2002), although the floriculture industry has production values of several billon US dollars per year worldwide (Jain and De Klerk, 1998; USDA, 2015). The focus of breeders has remained by convention on visual traits such as floral traits (Onozaki *et al*., 2001), *e.g*. petal color and form or flower size. In contrast, production-related traits, which have great economic importance in *Pelargonium zonale* (Molenaar *et al*., 2017), have not been well characterized nor are they yet relevant for variety registration. This thesis investigates *P. zonale* breeding and illustrates the importance of phenotyping and efficient statistical data analysis towards in laying the foundation for marker-assisted selection (MAS).

1.1 P. zonale

Pelargonium spp. are the most popular potted plant worldwide (Deroles *et al*., 2002, p. 181) with a production value of approximately 3.2 billion US dollars in year 2015 solely on the American market (USDA, 2016). The genus *Pelargonium* comprises about 250 species and is native to southern Africa, including South Africa and Namibia (Becher *et al*., 2000). Hybrids derived from the wild species *Pelargonium* × *hortorum* Bailey, section *Ciconium*, resulted in the major *P. zonale* cultivars, commonly grown in Europe and North America (Becher *et al*., 2000). Little is known about the *P. zonale* genome. The genome is mostly tetraploid, but also diploids are known. The chromosome number is nine, however, the total genome size is unknown (CCGS, 2010). To the author's knowledge three studies have been carried out, in which the genetic variability of the genus *Pelargonium* was described by the use of different marker systems, however, markers in these studies were not implemented in MAS for a breeding program: *i*) Becher *et al*. (2000) applied microsatellite markers for cultivar identification in *Pelargonium* spp. with the purpose of protecting breeder's rights, *ii*) Renou *et al*. (1997) used 'Random Amplification of Polymorphic DNA' (RAPD) markers to describe the relatedness of 34 cultivars in *Pelargonium* and were able to characterize the variability of phenotypes with the genotypic information and *iii*) Palumbo *et al*. (2007) described the genetic

variability of 46 *Pelargonium* accessions by the use of 'Target Region Amplification Polymorphism' (TRAP) markers.

1.2 The need for novel breeding strategies

The ongoing competition between ornamental breeders on the floriculture market is influenced by rising energy costs and increased regulations of pesticides and fungicides (Debener, 2001). Thus, the importance of quantitative traits related to production efficiency increases. However, ornamental breeders still focus mainly on consumer traits and use of experimental designs in ornamental breeding is largely neglected.

The implication of the present work for *P. zonale* production is a potential reduction of 20 % in the number of stock plants required for propagation. This improvement would translate to a savings of 250,000 stock plants, equivalent to 130,000 m^2 greenhouse area, 50,000 m^3 water, above 1 ton of fertilizer as well as above 350 m^3 substrate per year, (Robert Boehm, *personal communication, June 2016*, Selecta One). In order to improve *P. zonale* for stem cutting productivity as well as to be able to compete and to meet demands imposed by the market, improved breeding strategies are required.

1.3 Phenotyping in ornamental breeding context

Phenotyping protocols are a prerequisite for breeding. The "International Union for the Protection of New Varieties of Plants" (UPOV) provides phenotyping protocols for a wide range of ornamental species allowing a characterization of mostly qualitative traits affecting consumer satisfaction, but none for production-related traits. The former traits are scored either as nominal, e.g. anthocyanin coloration on the outer sepal side, or more frequently ordinal, *e.g.* consisting of either three categories (mostly growth description), five or nine (mostly color shade description of petals or sepals). For the identification of the main color, the color chart provided by the "Royal Horticultural Society" (RHS) is frequently used by breeders. Although phenotyping equipment is rarely used in ornamental breeding, efforts are being made to design an application software (APP) to perform the color shade identification within seedling generations allowing a faster and more precise phenotyping to increase the breeding efficiency (Dominik Losert, *personal communication, April 2018*, Selecta One).

1.4 Statistical analysis in a two-phase context

The statistical analysis aims to strip down the observed phenotypes as much as possible to their genotypic values to detect differences between individuals. The statistical analysis follows the design of an experiment, i.e., it adheres to the principle "analyze as randomized". Two-phase experimental designs are encountered in ornamental breeding as a result of the breeding scheme, which is often clone breeding, and multiplication methods (Boxriker *et al*., 2017, Boxriker *et al*., 2018, Molenaar *et al.*, 2017, Molenaar *et al*., 2018a). Commonly, the two phases in ornamental breeding are: *Phase 1*) the cultivation of stock plants in the greenhouse and *Phase 2*) subsequent assessment of harvested plant material from phase one either in the greenhouse or in the lab.

The statistical analysis in a two-phase context is based on linear mixed models set up in phase-specific order, because usually for each phase a separate experimental layout is used. The precision of estimates depends not solely on the phenotyping, but also on the experimental layout. For example, breeders often consider the A-optimality criterion (John and Williams, 1995) in generating single phase experiments. This criterion minimizes the average pairwise variance among treatment means by obtaining the best combination of treatments within blocks so that each pairwise treatment comparison occurs at least once or equally often depending on the treatment number and block sizes. As this criterion minimizes the average pairwise variance, it maximizes the precision of selection for the given resources.

1.5 Consideration of pedigree information

It is known that replicated experiments are more precise with regard to estimated treatment effects than un-replicated trials (Singh and Singh, 2015). In clone breeding the first generation (seedling generation) is always tested an un-replicated design, i.e., each individual is represented by a single plant, because it is the initial stock plant cultivated from the seed of the previous year. Experimental designs that are suitable for testing un-replicated genotypes are augmented designs, as applied in *P. zonale* (Molenaar *et al*., 2018b). In augmented designs, block effects are estimated solely by the use of checks, which might not capture all environmental variation for the estimation of the treatments effects, reducing precision. In later clonal generations, replicated testing of individual genotypes is possible. However, during the primary selection of seedlings, the population size is drastically reduced from thousands to a maximum of around 200 genotypes. Then genotypes may be tested in replicated designs, for

example in resolvable incomplete block designs, as applied in *D. caryophyllus* L. (Boxriker *et al*., 2017; Molenaar *et al*., 2018a) or in randomized complete block designs as the number of individuals is more and more reduced. Higher precision of estimated treatment effects can be expected, because the block effects are estimated from the individuals included in all replicates. But, as the population size is reduced, the genetic variability is reduced, too. Therefore, differences between individuals become more difficult to detect. A way to increase the precision of estimating treatment effects or to increase the genotypic variability and hence to detect genotypic differences in spite of reduced genetic variation, is to exploit the information of relatives by a family-index (Lush, 1947).

1.6 Objectives

The main objectives of this thesis were to enhance *P. zonale* breeding by the introduction of two-phase experimental designs and to establish phenotypic protocols for production-related traits, in particular for root formation to provide a solid phenotypic foundation towards MAS. The specific objectives were:

1) To establish scoring protocols for production-related traits
2) To introduce the use of two-phase experimental designs considering real production conditions and breeding practice for stem cutting production in *P. zonale*,
3) To quantify the increase in effectiveness of selection due to the introduction of measures described under (1) and (2) by simulating the expected response to selection for production-related traits,
4) To explore potential pragmatic approaches for generating improved two-phase experimental designs with regard to the following two questions:
    a. Is there a disadvantage in leaving treatments in the same randomized order from the first phase when transferring samples to the second phase?
    b. Instead of generating a separate layout for each phase, can the design be optimized across both phases, such that the mean variance of a pairwise treatment difference can be decreased across both phases compared to two independent designs?
5) To introduce the BLUP-based selection for phenotypic selection in *P. zonale* breeding by evaluating the efficiency of this method in terms of heritability.

1.7 Outline of the thesis

The present work is conceived as a cumulative thesis, where each chapter is a journal article and each article is framed as a case study. Chapter 2 covers the initial introduction of two-phase experimental designs in *P. zonale* breeding and establishes phenotyping protocols to score production-related traits. In Chapter 3, alternative methods of generating two-phase experimental designs are explored. Chapter 4 presents methods for maximizing the use of available data by exploiting family information in un-replicated trials. A general discussion of the major findings follows in Chapter 5. In Chapter 6 the main conclusions of the thesis are presented and an outlook of the present work is given and Chapter 7 gives a summary.

2 Selection for production-related traits in *Pelargonium zonale*: improved design and analysis make all the difference[1]

[1]This chapter is published as:
Molenaar H, Glawe M, Boehm R, Piepho HP (2017). Selection for production-related traits in Pelargonium zonale: improved design and analysis make all the difference. Horticulture Research, 4:1-9.

OPEN

Citation: Horticulture Research (2017) 4, 17004; doi:10.1038/hortres.2017.4

www.nature.com/hortres

ARTICLE

Selection for production-related traits in *Pelargonium zonale*: improved design and analysis make all the difference

Heike Molenaar[1], Martin Glawe[2], Robert Boehm[2] and Hans-Peter Piepho[1]

Ornamental plant variety improvement is limited by current phenotyping approaches and neglected use of experimental designs. The present study was conducted to show the benefits of using an experimental design and corresponding analysis in ornamental breeding regarding simulated response to selection in *Pelargonium zonale* for production-related traits. This required establishment of phenotyping protocols for root formation and stem cutting counts, with which 974 genotypes were assessed in a two-phase experimental design. The present paper evaluates this protocol. The possibility of varietal improvement through indirect selection on secondary traits such as branch count and flower count was assessed by genetic correlations. Simulated response to selection varied greatly, depending on the genotypic variances of the breeding population and traits. A varietal improvement of over 20% is possible for stem cutting count, root formation, branch count and flower count. In contrast, indirect selection of stem cutting count by branch count or flower count was found to be ineffective. The established phenotypic protocols and two-phase experimental designs are valuable tools for breeding of *P. zonale*.

Horticulture Research (2017) **4**, 17004; doi:10.1038/hortres.2017.4; Published online 22 February 2017

INTRODUCTION

The improvement of plant cultivars is reflected by the response to selection in a breeding program. Response to selection, in its simplest form, is defined as the difference between the mean phenotypic value of progenies of selected parents and the mean phenotypic value of the whole parental generation before selection.[1] The better the phenotyping, the better is the response to selection.

For more than a century, selection in field crops has been evolving as phenotyping approaches and experimental design have improved. Today's phenotyping techniques have broadened the focus from hand measurements of single-plant traits or destructive analysis towards non-destructive, holistic and high-throughput phenotyping in the field.[2] Such phenotyping platforms include three-dimensional time-of-flight cameras, laser distance sensors, hyperspectral imaging, infrared thermometers, ultrasonic sensors and multi-spectral crop canopy sensors that can measure, for example, canopy temperature and spectral reflectance and plant crop height of wheat plots,[3] biomass accumulation[4] or can be used to investigate photosynthesis, nutrient uptake, and plant growth and development.[5]

By comparison, ornamental breeding still relies more heavily on the 'breeder's eye' for judging if one cultivar is better than another. Reasons are: (i) phenotyping is limited largely to relatively easily scored traits like petal and leaf color or growth type (see International Union for the Protection of New Varieties of Plants (UPOV), TG/28/9 Corr.) and (ii) the traits phenotyped are relevant to plant variety protection and thus prioritized by ornamental breeders, in contrast to traits which are not listed by UPOV. There are other no less economically important production-related traits, however, for which, to our knowledge, UPOV does not provide protocols. Presently, these traits are improved through cultivation practices or post-harvest treatments and not through breeding efforts. For example, root growth is generally improved by application of hormones.[6]

Currently there are also large differences between crop and ornamental breeding with respect to the use of experimental designs and statistical analysis for phenotypic selection. Efforts to optimize designs in crop breeding date back more than a century.[7] Improvements were first made accounting for the appropriate sample size to achieve the desired level of precision in estimates of effects and power of experiments. In addition, the need for replicates over time or within or over locations became clear and proposals were also made to randomize the allocation of treatments to experimental units.[7] In 1930s, these findings were laid down in Fisher's well-known book on experimental design.[8] On the basis of these principles more complex designs were soon developed,[7] and more recently two-phase experimental designs[9] were introduced. Such designs are needed when an experiment is conducted in more than one phase. For example, in the first phase plants of a crop may be raised in a field experiment. In the second phase, samples from the field plots are then taken to the lab for analysis.[10] Two-phase designs have the property that the observational unit changes from one phase to the next.[10] Further, phases may overlap.[10] By using two-phase experimental designs it is possible to account for environmental effects on experimental units in previous experimental phases, which might influence a response when measuring the trait in a later experimental phase. Typically, such designs are used in cereal breeding. In this respect again, ornamental breeding is still lagging behind, although two-phase experimental designs are highly suitable for breeding ornamentals. For example, in *Pelargonium zonale,* a mother stock is established to harvest stem cuttings in the first phase, whereas in the second phase the genotypes are tested for root formation by rooting harvested stem cuttings. Despite the two-phase nature

[1]University of Hohenheim, Institute of Crop Science, Biostatistics Unit, Stuttgart 70599, Germany and [2]Klemm+Sohn GmbH & Co. KG, Stuttgart 70378, Germany.
Correspondence: H Molenaar (Heike.Molenaar@uni-hohenheim.de)
Received: 1 November 2016; Revised: 1 February 2017; Accepted: 3 February 2017

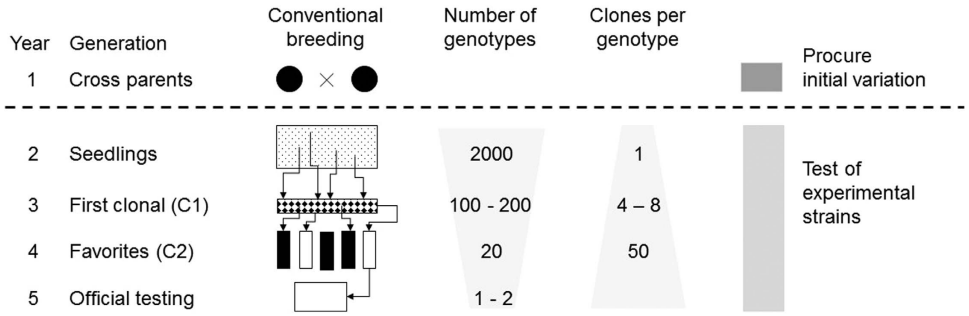

Figure 1. Current breeding scheme of *P. zonale*: from the intial parental crossing in year 1 to the official testing of the best lines in year 5, where the number of genotypes decreases, and in parallel, the number of clones per genotypes is increased.

of this experimental setup, two-phase experimental designs have not been used so far in ornamental breeding.

Our objectives for improving phenotypic selection in *P. zonale* breeding were: (i) to establish scoring protocols for production-related traits, (ii) to introduce the use of two-phase experimental designs in ornamental breeding practice; and (iii) to quantify the increase in effectiveness of selection due to the introduction of measures described under (i) and (ii) by simulating the expected response to selection for production-related traits.

MATERIALS AND METHODS

Current breeding trials

Crosses of promising parental strains are made in year one of a breeding program. The 100–200 most promising candidates are selected from an unreplicated trial in year 2. Petal color, growth type and early prematurity are traits of primary interest. In year 3, selected candidates are tested under field conditions for assessment of petal color maintenance or drought tolerance, using four to eight clones of each candidate. In year 4 follows a production test (PT) accounting for real production conditions, which consists of two phases. In phase one (P1), the establishment of stock plants from which stem cuttings are harvested and the stem cutting count (SCC) is recorded. In phase two (P2), genotypes are assessed for rooting percentage, using the harvested stem cuttings of step one. Rooting percentage is defined as the number of rooted cuttings divided by the initially planted number of stem cuttings of one clone of a genotype in one tray. Up to 50 clones of one genotype are investigated. In the current protocol, a single clone of a genotype, placed on one tray, represents the observational unit of the trial, where clones of the same genotypes are placed next to each other in the greenhouses to have direct phenotypic comparisons. In statistical terms, real replicates of genotype are lacking as well as adherence to any other design principle, such as randomized allocation to experimental units, which would allow the application of statistically founded selection decisions. But efficient selection is of utmost importance in year 4, since selected clones are subjected to official variety testing (Figure 1).

Experimental procedure of the current production test

To establish the stock plants, stem cuttings of selected genotypes are planted individually in paper pots (19 mm diameter, 33 mm height) filled with 80 % sterilized coco peat fibers and 20 % styroballs for aeration. The rooting takes 4 weeks under moderate climate conditions (15–28 °C) and irradiance between 20 and 25 klx depending on weather conditions. Fertigation starts in the third week after planting with a standard 2.5: 1 (N: K) menu containing the following nutrients (in mmol l^{-1}): 21.0 NO_3^-, 3.5 SO_4^{-2}, 3.0 H_3PO_4, 1.4 NH_4^+, 9.0 K, 7.0 Ca, 3.3 Mg, 25.0 Fe, 6.0 Zn, 25.0 B, 2.0 Cu and 2.0 Mo. A sufficient amount of Mn is contained in the soil and made available to plants by keeping the pH level below 6.0. In week 4, rooted cuttings are then repotted in ~17.3 cm diameter bags with a volume of 3 l filled with 80 % (inert) pumice and 20 % coarse coco peat fibers to cultivate the stock plants. Stock plants are pinched once to stimulate branching and again afterwards if necessary. After 18 weeks of growth, stem cuttings are harvested and counted. Cuttings must be $\leqslant 6$ cm in length, have two to four leaves of which one is fully developed, and may not have flower buds or open flowers. To score genotypes for rooting percentage, all harvested stem cuttings of a genotype and different stock plants are planted in a column-wise fashion onto the same trays (Easypot, 25/39, 35 mm height, HAWITA Gruppe GmbH, Vechta, Germany, three rows with 13 paper pots each), where always a single stem cutting is planted per paper pot. The climate conditions are moderate: 18 °C temperature during planting and otherwise 18–24 °C and irradiance approximately 20 klx. Two hours after planting, plants are misted for 24 h, after which misting is reduced over a period of about 2 weeks depending on weather conditions. Spray misting is carried out every 16 s when irradiance levels exceeded 20 klx.

A two-phase experimental design for *Pelargonium zonale* breeding

To improve the current PT, two experiments were conducted introducing two-phase experimental designs. Initially, the two phases of each of the two experiments were defined maintaining the context of the current PT steps: In P1, the cultivation of stock plants of genotypes, which was done in location 1, and in P2, the rooting of plant material, which was performed in location 2. Both phases took place in greenhouses and did not overlap. The cultivation procedures followed the current PT, whereas the planting manner was changed.

Two-phase experiment I

Two-phase experiment (TPE) I was conducted in 2013/14. Five hundred genotypes were scored for SCC on eleven dates, flower count (FC) and branch count (BC) on two dates during P1 as well as for root formation (RF) on three dates during P2 (Table 1). Three hundred and fifty genotypes belonged to an internal collection and 150 were new breeds.

In the first phase, an α-design[11] was used and generated by CycDesigN 4.0 (VSN-International, https://www.vsni.co.uk). The four cultivation tables in the greenhouse represented the four replicates. Each replicate in P1 comprised 167 incomplete blocks with three experimental units (EU1) each, except that one had only two EU1. On each EU1 a pair of stock plants was placed.

In the second phase, a conventional experimental design could not be used, because of fast quality decline of stem cuttings and therefore the necessity to work efficiently. However, to adhere to randomization, the packaging of stem cuttings for transfer from location 1 to location 2 was exploited.

Therefore, the total experimental space, represented by m rooting tables, was divided into four regions. The replicates were assigned systematically to the regions. Further, $t = 36$ trays were laid out on each rooting table. On each tray there were 39 paper pots arranged in three rows with 13 paper pots each.

It is noted, that all trays of a replicate did not necessarily fit on one rooting table, indicated by regions shaded in gray in rooting tables in P2, which correspond to replicates shaded in the same gray of cultivation tables in P1 in Figure 2. Further, the incomplete blocks from P1 did not necessarily fit on a single tray in P2.

The trays were divided into areas, which represented the experimental units in P2 (EU2). In each area were planted all the cuttings for a genotype from the replicate. The size of an area varied depending on the number of stem cuttings for the genotype and replicate allocated to it.

Further, for each area, the pots were filled in row-wise order on a tray. One area follows on from the previous area subject to the restriction that

Table 1. Timeline of the TPE I and II in years 2013/14 and 2014/15, where in two phases genotypes were assessed for SCC, FC, BC and RF

TPE	*Year*	*Week*	*Phase*			
			1		*2*	
			SCC	*FC*	*BC*	*RF*
I	2013	41	x			
		43	x			
		46	x			
		50	x			
	2014	3	x			x
		7	x			
		9	x			x
		10	x			
		11	x			
		12	x			
		18	x			x
		26		x	x	
		34		x	x	
II	2014	35	x			x
		40	x			x
		45	x			x
		50	x			
	2015	3	x			x

Abbreviations: BC, branch count; FC, flower count; RF, root formation; SCC, stem cutting count; TPE, two-phase experiment.

all the paper pots for an area were on the same trays. One paper pot was left free between areas for a better differentiation of genotypes after 4 weeks rooting.

The genotypes were allocated randomly to the areas as mentioned above by exploiting the packaging order. Harvested stem cuttings of each genotype and replicate were packed in small bags, such that each bag contained all stem cuttings from EU1 in P1 and put into cartons. Genotypes within replicates of P1 were kept together. In location 2, small bags were randomly drawn out of the cartons and planted in areas. Thus, stem cuttings from each EU1 in P1 were allocated to exactly one EU2 in P2.

Two-phase experiment II

TPE II was conducted in 2014/15 with 504 genotypes. One hundred and eighteen genotypes belonged to the internal collection and 356 to new breeds. In addition, 30 randomly chosen genotypes of TPE I were tested again. The SCC was assessed on five dates during P1 and RF was tested on four dates (Table 1). The experimental design in P1 of TPE II was modified to a resolvable row-column design to account better for a spatial trend detected in TPE I. The row-column design was generated using CycDesigN 4.0. The four replicates were represented by the four planting tables, where each replicate comprised six columns and 84 rows (Figure 2). In P2, the same approach was used as in TPE I in P2. The losses per genotype and the losses of stock plants were much higher than in TPE I.

Phenotypic protocols

SCC was assessed as the number of stem cuttings per plant for each pair of stock plants (EU1) and genotype in P1. All stem cuttings were either observed by pinching or obtained at harvest time.

The RF of stem cuttings of genotypes was described with six ordered categories after four weeks of growth (Figure 3) in P2. For each area, we counted the number of plants in categories S0 (dead) to S5 (extraordinary). From these counts we computed the sum of rooted cuttings assigned to S4 and S5, so that a single response value was obtained per area (EU2).

Secondary traits of SCC. FC was defined as the number of flowers per plant for each pair of stock plants (EU1) and genotype in P1 after eight and 12 weeks growth.

BC was defined as the number of all branches per plant for each pair of stock plants (EU1) and genotype evolved after 8 and 12 weeks growth.

Statistical analysis

Single time-point analysis. SCC, FC, BC and the count of rooted cuttings assigned to categories (S4+S5) of RF were analyzed using a linear mixed model (LMM), where the randomization-based models in both phases were used for determining the terms in the model.[12] The model notation followed by Piepho *et al.*,[13] where the colon separates fixed effects on the left-hand side from the random effects on the right-hand side. The 'dot' operator (•) in a term A•B defines combinations of levels of its constituent factors A and B.

Phase one model

To analyze SCC, BC and FC the model was successively setup as follows. The treatment model considering the randomized tier [12] was

$$\mathrm{GEN}, \tag{1}$$

where GEN denotes the genotypes (treatment factor). The randomization-based model considering the unrandomized tier[12] was

$$\mathrm{REP} + \mathrm{REP.IB} + \mathrm{REP.IB.PAIR}, \tag{2}$$

where REP denotes the replicates represented by cultivation tables comprising a full set of genotypes, REP.IB the incomplete blocks nested within the replicates and REP.IB.PAIR, the EU1. Incomplete blocks were modeled as random since the block order was permuted during randomization. The full model obtained by combining the treatment and randomization-based model for design effects was

$$\mathrm{GEN} + \mathrm{REP} : \mathrm{REP.IB} + \underline{\mathrm{REP.IB.PAIR}}, \tag{3}$$

where the underlined term designates the residual error. The full model was augmented by a covariate, A, the number of stock plants per EU1 and genotype, because due to cultivation problems, some stock plants were missing at random. Further, a column (post-blocking) factor within replicates was added to better account for environmental effects. The model in analyzing SCC, BC and FC was

$$\mathrm{A} + \mathrm{GEN} + \mathrm{REP} : \mathrm{REP.IB} + \mathrm{REP.COL} + \underline{\mathrm{REP.IB.PAIR}}. \tag{4}$$

Phase two model

To analyze the RF of stem cuttings assigned to categories (S4+S5) in P2, first the randomization-based model for P2 was set up as

$$\mathrm{REGION} + \mathrm{REGION.AREA}, \tag{5}$$

where REGION denotes the experimental space to which systematically a replicate was assigned and REGION.AREA the EU2 to which the genotypes were randomly assigned. REP and REGION as well as REP.IB.PAIR and REGION.AREA were totally confounded terms as genotypes were kept together replicate-wise from P1 to P2 and the stem cuttings per experimental unit of P1 were held together and assigned to one area in P2. Thus, effects REGION and REGION.AREA do not need to be added explicitly to the model, as they are implicitly accounted for by the effects REP and REP.IB.PAIR, respectively. However, post-blocking was needed in P2, as variable environmental conditions between the rooting tables and between the trays occurred. To capture those variations, two post-blocking factors RTABLE and TRAY were defined. The former denotes rooting tables, each comprised of an incomplete set of genotypes, and the latter denotes trays, each comprised of multiple areas and which is nested within RTABLE. To exploit the inter-RTABLE and inter-TRAY information, both post-blocking factors were designated as random. The model for RF analysis was

$$\mathrm{A} + \mathrm{GEN} + \mathrm{REP} : \mathrm{REP.IB} + \mathrm{REP.COL} + \mathrm{RTABLE} + \mathrm{RTABLE.TRAY} + \underline{\mathrm{REP.IB.PAIR}}. \tag{6}$$

All statistical analysis was conducted with SAS 9.4 (SAS Institute Inc., Cary, NC, USA, 2014).

Checking model assumptions

Independence of residuals, normal distribution of random effects (including the residual error) and variance homogeneity are important

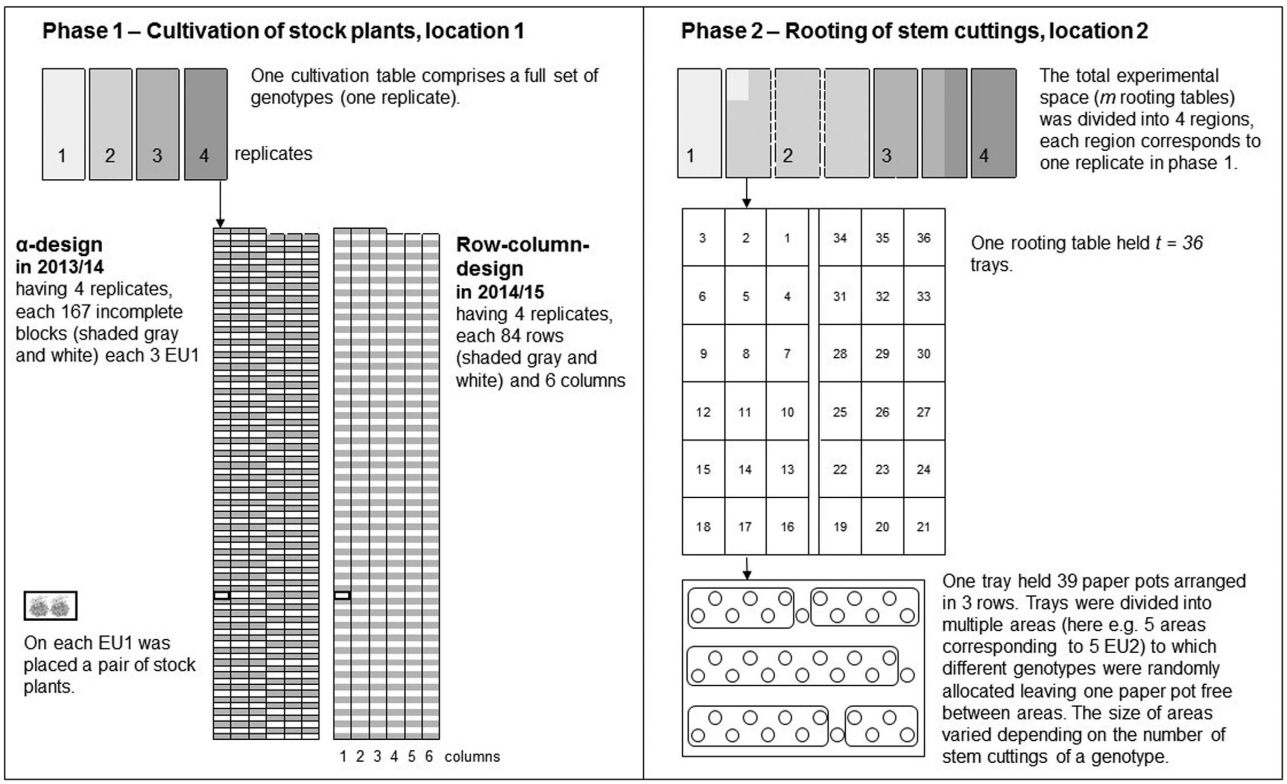

Figure 2. The two-phase experimental design intorduced in *P. zonale* breeding: P1, cultivation of stock plants for obtaining the SCC in location 1; P2, the rooting of stem cuttings to test the root formation in location 2. In P1, and α-design in 2013/14 and row-column design in 2014/15, were used. Each cultivation table represented on replicate having 500 planting positions arranged either in 167 incomplete blocks with three experimental units (EU1) each in 2013/14 or, in year 204/15 in 84 rows and six columns. On each EU1 a pair of stock plants of a genotype was placed in P1. In P2, the total experimental space represented by *m* rooting tables (at maximum 9) was divided into four regions to which the replicates were systematically assigned. Regions shaded in gray in rooting tables in P2 correspond to replicates shaded in gray of cultivation tables in P1. Eeach rooting table held 36 trays at maximum. One tray contained 39 paper pots arranged in three rows. The trays were divided into areas, representing an experimental unit in P2 (EU2), to which different genotypes were randomly allocated. The size of areas varied depending on the numbers of stem cuttings for a genotype. The planting of stem cuttings followed a row-wise order.

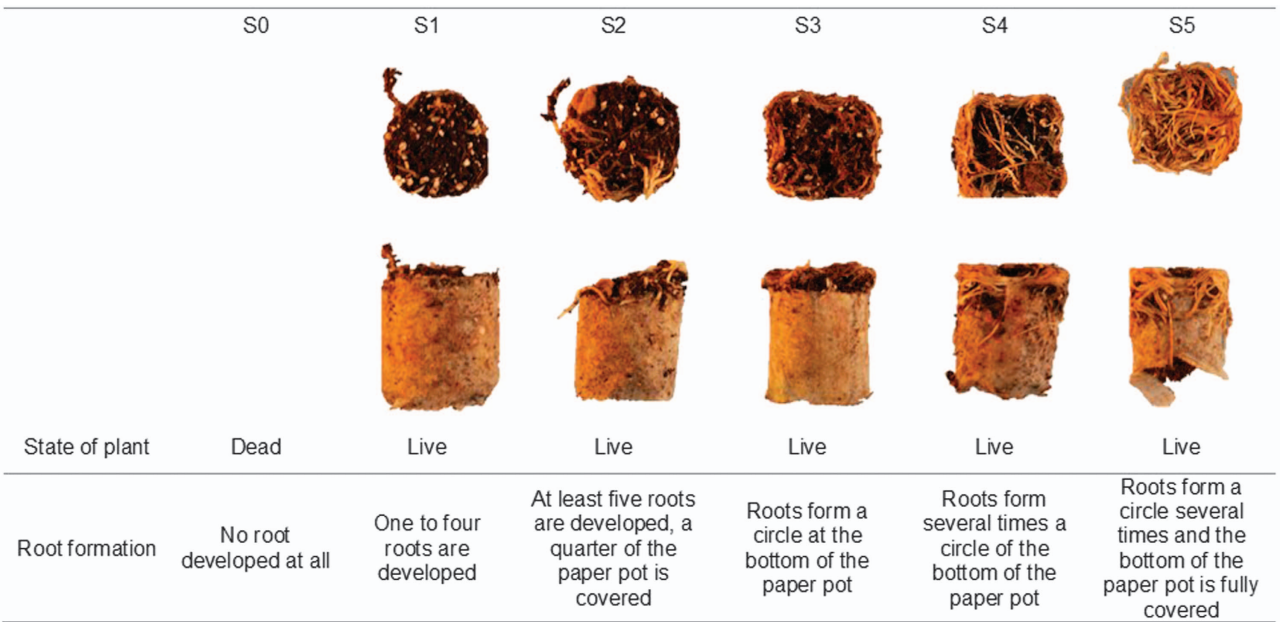

Figure 3. Ordinal categories of root formation ranging from S0 (dead) to S5 (extraordinary rooted).

assumptions for LMM. To check these LMM assumptions, studentized residuals were investigated, which are independent of scale.[14] A studentized residual is defined as $\frac{\hat{e}_i}{\sqrt{\text{Var}[\hat{e}_i]}}$, where $\hat{e}_i$ is the *i*-th estimated raw residual and $\sqrt{\text{Var}[\hat{e}_i]}$ the estimated s.d. of the *i*-th raw residual.[15] To check normality, the studentized residuals were plotted against the normal scores in quantile–quantile plots (Q–Q-plots). To check for any unaccounted variance homogeneity, studentized residuals were plotted against the predicted value.[16] Note that the LMM may entail a model allowing for heterogeneity of variance. If the model is well specified, the studentized residuals should display no remaining heterogeneity of variance. Normal distribution of random genotypic effects was checked using standardized best linear unbiased predictors (BLUPs)[17] $\frac{\hat{g}_j}{\sqrt{\text{Var}[\hat{g}_j]}}$, where $\hat{g}_j$ is the *j*-th estimated genotypic BLUP and $\text{Var}\left[\hat{g}_j\right]$ its unconditional variance. These standardized BLUPs were plotted against the normal scores in Q–Q-plots.

Model selection and fitting for repeated measurement analysis

For the traits SCC, BC, FC and counts of rooted cuttings assigned to (S4+S5) of RF repeated measurements were taken on the same plants at different harvest dates. A salient feature of repeated measurements is serial correlation among observations made on the same unit. To account for the repeated measurements nature of the data, the models (4) and (6) were expanded by a repeated factor T for time, by concatenating each factor with the repeated factor T as follows:[18,19]

$$\text{A} + \text{T} + \text{T.GEN} + \text{T.REP} : \text{T.REP.IB} + \text{T.REP.COL} + \underline{\text{T.REP.IB.PAIR}} \quad (7)$$

and

$$\begin{aligned}&\text{A} + \text{T} + \text{T.GEN} + \text{T.REP} : \text{T.REP.IB} + \text{T.REP.COL}\\ &+\text{T.RTABLE} + \text{T.RTABLE.TRAY} + \underline{\text{T.REP.IB.PAIR}}.\end{aligned} \quad (8)$$

For all random effects of model (7) serial correlations of observations were assumed. The best fitting variance–covariance structure was selected based on the smallest value of the Akaike information criterion (AIC).[20] The AIC is defined as minus twice the REML log-likelihood plus twice the number of variance parameters.[21] In model (8), serial correlations were only assumed for random effects defined for P1. The random effects defined for P2 were assumed to be independent, because at each single time-point genotypes were randomly allocated to areas. But still the repeated factor was concatenated with block factors of P2, because genotypes were systematically allocated to the same region, including the same rooting table, especially during RF assessment in TPE II, and seldom to the same area.

For selected variance–covariance structures, variance components of all model effects were estimated and used to predict the response to selection as well as to estimate the genotypic means for correlating estimates over experiments.

Response to selection

Because data were unbalanced, the expected response to selection for SCC, FC, BC and RF was simulated using the fitted LMM [22] as

$$R_q = \frac{\sum_{i \epsilon S_q} g_i}{\#(S_q)} \quad (9)$$

and

$$R = Q^{-1} \sum_{q=1}^{Q} \text{R}_\text{q}, \quad (10)$$

where Q is the number of simulation runs, R_q the predicted mean of the next generation, S_q the set of genotypes selected based on BLUPs of the true genetic values and $\#(\text{S}_\text{q})$ the size of the selected fraction. The central idea of this approach is to jointly simulate the genotypic effects (g_i) and their BLUPs ($\hat{g}_i$) for a given experimental design. If we collect genetic effects and their BLUPs into a vector **w**, we may do a Cholesky decomposition of var(**w**) as $\text{var}(\mathbf{w}) = \mathbf{\Omega} = \mathbf{\Gamma}\mathbf{\Gamma}'$. To simulate **w** from a multivariate normal distribution with zero mean and variance–covariance matrix **Ω**, determined from the bits and pieces of the mixed model equations,[22] a vector **z** of standard normal deviates is simulated that has the same length as **w**. A simulated realization of **w** is then obtained from $\mathbf{w_{sim}} = \mathbf{\Gamma z}$, so that the variance of the simulated data equals exactly the variance of the given data, $\text{var}(\mathbf{w_{sim}}) = \mathbf{\Gamma}\mathbf{\Gamma}' = \mathbf{\Omega}$. The simulation was repeated 10 000 times. For each simulation run, the best values of BLUPs are selected to obtain the mean of the next generation based on the simulated true genetic values (g_i). The predicted means of the next generation are then averaged over all 10 000 simulation runs to obtain the expected selection response.

Genetic correlation between traits

Genotypic correlations[23] between the totals of SCC, FC and BC were obtained in TPE I using the equation[24]

$$r_{gij} = \frac{\hat{\sigma}_{Gij}}{\hat{\sigma}_{Gi}\hat{\sigma}_{Gj}}, \quad (11)$$

where $\hat{\sigma}_{Gij}$ is the estimated genotypic covariance between traits *i* and *j* and $\hat{\sigma}_{Gj}$ and $\hat{\sigma}_{Gj}$ are the estimated genotypic standard deviations for traits *i* and *j*, respectively. To estimate the genotypic variances and covariance, multivariate LMMs were fitted. In order to develop a multivariate model, model (4) was first extended by factor M, which identifies the three traits:

$$\begin{aligned}&\text{M} + \text{M.REP.IB} + \text{M.REP.COL} + \text{M.REP} + \text{M.A} : \text{M.GEN}\\ &+\underline{\text{M.REP.IB.PAIR.}}\end{aligned} \quad (12)$$

Nested structures between M and design factors were declared as fixed effects to alleviate the computational burden. The genotype factor was then considered as random. The vector g_i of genetic effects for the *i*-th genotype for the *T* different traits was assumed to be multivariate normal with $g_i \sim MVN(0, \sum_g)$, where $\sum_g$ is given by $\sum_g = \boldsymbol{D_g R_g D_g}$ with $\boldsymbol{D_g}$, the diagonal matrix with genetic standard deviations for the *M* different traits on the diagonal and $\boldsymbol{R_g}$ a $T \times T$ genotypic correlation matrix. Similarly, the vector e_{ij} of errors of the *j*-th observation on the *i*-th genotype was assumed to be multivariate normal with $e_{ij} \sim MVN(0, \sum_e)$, where $\sum_e = \boldsymbol{D_e R_e D_e}$ with $\boldsymbol{D_e}$ the diagonal matrix with standard deviations on the diagonal and $\boldsymbol{R_e}$ a $T \times T$ error correlation matrix.

Correlations of adjusted genotypic means over experiments

The precision assessment of the phenotyping approach based on the estimation of the Pearson correlation of the adjusted genotype means between the two experiments for genotypes assessed in both experiments for SCC and rooted cuttings assigned to categories (S4+S5) of RF.[25] First, a repeated measurement analysis of each experiment was conducted selecting a variance–covariance structure for serial correlation of observations based on smallest AIC and then the genotype main effects for both traits were obtained. Second, the estimated genotype main effects were correlated between the TPE I and TPE II. The presence of genotype × time interaction will diminish the correlation, when genotype × time interaction is present.

RESULTS

Checking model assumptions

The overall impression from plots of studentized residuals versus predicted values revealed that the variance–covariance model was appropriate but at the same time there was some departure from normality caused by outliers (Supplementary Figures 1 to 22). Removing outliers according to manually set trait-specific thresholds supported by the subject knowledge of the experiments (Table 2), approximate normality could be achieved and the plots of studentized residuals against the predicted means showed no non-normalities. Standardized genotypic BLUPs also showed approximate normality (Supplementary Figures 23 to 44).

Model selection and fitting

The best model fit according to AIC was achieved for all traits with the unstructured variance–covariance structure for serial correlations of observations, except for RF of TPE I, where the smallest AIC was obtained for compound symmetry (Table 3). The variance components for selected variance–covariance structures presented in Table 4 were used to simulate the response to selection. Zero variance components of block factors mean that there was no correction due to those block factors during the estimation of effects. The largest variance for each trait is bold faced.

Table 2. Thresholds for labeling outliers while residual outliers of trait analysis of SCC, RF (count of rooted cuttings assigned to *S4+S5*), BC and FC

Trait	*Threshold*
SCC	3.0
RF	3.25
BC	2.5
FC	3.0

Abbreviations: BC, branch count; FC, flower count; RF, root formation; SCC, stem cutting count.

Simulated response to selection

The simulated responses to selection for SCC, RF, FC and BC can be read from Table 5 as explained for SCC, at the first time-point of phenotyping, $l = 1$, obtained in TPE I. The breeding population mean (μ) of SCC was 9.10 with a genotypic variance $\left(\sigma_g^2\right)$ of 3.98. When selecting the 40 best genotypes ($p = 40/n$) out of the breeding population containing $n = 497$ genotypes, the mean of the following generation would be increased by about three stem cuttings. Thus, the next-generation mean is expected to be 12.16 SCC. Numerical comparisons of predicted response to selection between time-points of the experiment and over experiments for the same traits are not meaningful, because n varied. The selected fraction $p = i/n$ out of n has been defined by $i = 1, 5, 10, 20, 40$ for all traits.

For SCC and RF, greater response to selection was observed during TPE I compared with TPE II as means and genotypic variance of these two breeding populations differed perceptibly. Selection of genotypes out of the breeding population of TPE I resulted in a population mean increase by two SCC at minimum in single time-point analysis when considering a selection intensity of $p = 40/n$, whereas a selection of the best individual in the breeding population of TPE II would increase the population mean of the next generation by three SCC at maximum. When selecting for RF at a selection intensity of $p = 40/n$ in the breeding population of TPE I, the population mean can be doubled in the next generation in the best case, at time-point $l = 3$. Selecting of genotypes in the breeding population of TPE II, the next-generation mean would be only increased by two-third of the breeding population mean. For BC and FC, which were phenotyped only during TPE I, similar results were found. At $p = 1/n$ and time-point $l = 2$, the population mean of the following generation is increased by approximately six branches or flower counts per plant (Table 5).

Genetic correlations of SCC, FC and BC

The obtained correlations between the totals SCC, FC and BC were in all cases in the low positive range. The total BC was found to have the highest genetic correlation with the total FC ($r_{gij} = 0.2905$). Marginally smaller was the genetic correlation between the total BC and the total SCC ($r_{gij} = 0.2886$), where the totals SCC and FC were found to have the smallest genetic correlation ($r_{gij} = 0.1512$).

Pearson correlations of adjusted genotypic means over experiments

The Pearson correlation for SCC of adjusted genotypic means over the two experiments ($r = 0.37$) was not found to be significantly different from zero ($P = 0.1301$), whereas the Pearson correlation for rooted cuttings assigned to (S4+S5) of RF over the two experiments ($r = 0.56$, $P = 0.0132$) was approximately twice as high as for the SSC. The genotype × time interaction (GEN.T) was highly significant in both experiments for SCC (GEN.T: TPE I, $P < 0.0001$ and TPE II, $P = 0.0088$) and for RF (GEN.T: TPE I, $P < 0.0001$ and TPE II, $P < 0.0001$).

Table 3. Model selection based on AIC for variance–covariance structures (VC, AR(1): first-order autoregressive model, CS, UN) for repeated measurement analysis of SCC, RF, FC and BC

Trait		*Variance–covariance structure*	*AIC*	
			TPE I	*TPE II*
SCC	Count of rooted cuttings assigned to categories (*S4+S5*)	VC	20319	11197
		AR(1)	20276	11147
		CS	20273	11100
		UN	**19815**[a]	**10817**
RF		VC	15902	11588
		AR(1)	15899	11541
		CS	**15897**	11518
		UN	15899	**11437**
FC		VC	3781.54	—
		AR(1)	3751.74	—
		CS	3751.7	—
		UN	**3741.61**	—
BC		VC	3398.55	—
		AR(1)	2822.18	—
		CS	2802.0	—
		UN	**2801.23**	—

Abbreviations: AIC, Akaike information criterion; BC, branch count; CS, compound symmetry; FC, flower count; RF, root formation; SCC, stem cutting count; TPE, two-phase experiments; UN, unstructured; VC, variance components. [a]Smallest AIC is bold faced.

DISCUSSION

Our results show that there is great potential for varietal improvement of production-related traits in *P. zonale*. With the use of the developed phenotypic protocols, two-phase experimental design and its phase-specific analysis in the traits we analyzed, at least 20 % less stock plants would be needed to produce the same amount of stem cuttings as in the past. For example, given the test population mean and genotypic variance for SCC (TPE I, $l = 3$), 10 stock plants were needed to produce in total 80 stem cuttings. After selection with the lowest selection pressure ($p = 40/n$), only eight stock plants are needed to produce the same total (Table 5). This potential reduction of 20% less stock plants would mean in the final stage of stem cutting production that 250 000 stock plants can be saved resulting in a saving of 130 000 m^2 greenhouse area, 50 000 m^3 water, above 1 tonne of fertilizer as well as above 350 m^3 substrate per year. By significantly improving genotypes for production-related traits the production becomes economically more efficient.

The simulated response to selection

The prediction of response to selection assumes the same prerequisites as LMMs do.[22] In checking those prerequisites, studentized residuals were investigated, suitable to detect outlying observations.[26] Trait-specific thresholds were set based on the normal ranges observed in the greenhouse to remove outliers. In comparison to other methods for removing outliers, this is a simple method, and was preferred here, because little is improved by more complicated methods.[27]

The largest genotypic variances, in relation to the total variance, were obtained in analyses of SCC, FC and BC totals. As a result the largest simulated response to selection was obtained for these

Table 4. Variance components of genotypic and design effects of single time-points (*l*) (GEN: genotypic variance, REP: replicate variance, REP.IB: row variance, REP.COL: column variance, RTABLE: rooting table variance, RTABLE.TRAY: tray variance, ERROR: residual error variance)

			Phase 1				*Phase 2*		
TPE	*Trait*	*l*	*GEN*	*REP*	*REP.IB*	*REP.COL*	*RTABLE*	*RTABLE.TRAY*	*ERROR*
I	SCC	1	3.98	0.05	0.77	0.46			**5.67**
		2	4.63	0.23	0.56	0.15			**9.01**
		3	1.93	0.1	0.12	0.02			**2.49**
		S_3[a]	**26.62**	0	1.62	1.26			23.49
		S_{11}[b]	**131.74**	0.53	0	8.87			117.38
		RP[c]	2.43	0.19	0	0.15			**6.9**
	RF	1	1.59	0.83	0	0.55	0.03	0.77	**6.7**
		2	2.17	0.97	0.06	0.03	1.41	0.49	**3.52**
		3	4.38	0.65	0.02	0.08	0.22	0.47	**6.26**
		RP[c]	1.74	0.99	0.03	0.17	0.26	0.62	**4.85**
	BC	1	**4.06**	0	0	0.49			3.34
		2	**4.49**	0	0	0.52			4.33
		S[d]	**27.67**	0	0	2.32			3.35
		RP[c]	5.61	0	0.72	0.31			**6.54**
	FC	1	3.67	0	0.03	0.95			**4.45**
		2	**7.14**	0	0.69	2.66			7.06
		S[d]	20.5	**53.79**	0	17.53			44.39
		RP[c]	2.86	0	0.001	0.63			**6.52**
II	SCC	1	0.08	0.04	0.1	0.03			**0.91**
		2	0.82	0	0.07	0.36			**2.19**
		3	0.3	1.1	0.27	0.09			**1.35**
		4	0.52	0.23	0.01	0.14			**3.81**
		S_4[e]	4.03	3.19	0	0.7			**13.99**
		RP[c]	0.17	0.56	0	0.3			**2.32**
	RF	1	0.28	0.01	0.03	0.01	0.03	0.05	**1.26**
		2	0.79	0.78	0.06	0.24	0	0.18	**2.64**
		3	0.28	0.42	0.13	0.05	0.03	0.11	**1.93**
		4	0.83	0.41	0.08	0.25	0	0	**4.06**
		RP[c]	0.36	0.36	0.01	0.06	1.42	0.09	**2.63**

Abbreviations: AIC, Akaike information criterion; BC, branch count; CS, compound symmetry; FC, flower count; RF, counts of rooted cuttings assigned to S4+S5 of root formation; SCC, stem cutting count; TPE, two-phase experiment. [a]Total over $l = 1, 2, 3$ time-points. [b]Total over $l = 1, \ldots, 11$ time-points. [c]The variance components obtained by smallest AIC obtained by models (7) and (8) of repeated measurement analysis. In Supplementary Tables 1 to 6 are all estimated variance components obtained by by the repeated measurement analysis. [d]Total over $l = 1, 2$ time-points. [e]Total over $l = 1, 2, 3, 4$ time-points. The largest variance component for each trait is bold-faced.

traits. The simulated response to selection in analyses of single time-points and repeated measurement were several fold lower for the same population. This was due to the relatively smaller genotypic variances obtained in analyses of single time-points and repeated measurements. Thereby, the simulated responses to selection of SCC obtained by repeated measurement analysis could be directly compared with the analyses of totals, where the simulated responses to selection obtained by repeated measurement analysis were multiplied by the number of observational time-points (*l*).

Experimental designs in breeding practice

Experimental designs were developed which adapted the current ornamental breeding practice based on consideration of experimental design theory and practicality. For example, the approach in P2 of randomization was established to enable efficient working as well as maintain cutting quality and to provide flexibility for the sizes of areas within regions which varied according to the number of stem cuttings per genotype harvested. Biases of genotypic estimates could be avoided, which would have been caused without randomization due to heterogeneous conditions reflected by variance components of design effects.[28,29]

Further, post-blocking factors were introduced, which represented the physical units of production facilities especially in P2 allowing the consideration of sources of variation[30] such as border effects caused by other cultivars, shades, heaters and fans in greenhouses.

The arrangement of clones was modified from current breeding practice for theoretical considerations. Clones are usually tested in a group-wise arrangement, the goal of which is to allow a simple scoring of the uniformity and stability of genotypes. However, we embedded the clones in the two-phase experimental layout as real replicates of genotypes (treatments) to allow estimation of variation[30] and an unbiased estimation of genotypic effects, which is of more importance than simple scoring.

Environmental effects and sources of errors

Variable environmental conditions are known to affect endogenous phytohormone levels in stock plants.[31] This can influence the biosynthesis of leaf chlorophyll, color pigments and rooting of cuttings either positively or adversely.[31] Blocking is a key strategy to control such variable conditions by making the conditions within blocks more equal than across blocks for testing treatments. In some cases, the residual error was not related at all to variable environmental conditions in the blocking factors, which were then estimated to be zero. These were in particular the replicate and row effects in analyzing SCC, BC and FC.

Some variable environmental conditions will not have been captured by the blocking structure and so will have been incorporated in the error. Some such environmental conditions were: first, varying seasonal temperatures in both experiments

Table 5. Predicted response to selection of the two TPE for assessed traits (SCC, RF: counts of rooted cuttings assigned to S4+S5 of root formation, FC, BC) for single time-point (*l*), total (S) and RP analysis for various selected fractions (*p*) for given population sizes (*n*)

TPE	*Trait*	*l*	*μ*	*p*					n
				1/n	*5/n*	*10/n*	*20/n*	*40/n*	
I	SCC	1	9.1	5.04	4.35	3.97	3.54	3.06	497
		2	6.46	5.21	4.5	4.11	3.67	3.17	496
		3	8.82	3.59	3.12	2.84	2.53	2.19	497
		S_3[a]	24.46	13.93	12.05	10.99	9.82	8.48	497
		S_{11}[b,c]	64.64	31.17	26.98	24.63	21.98	18.99	499
		RP[c]	8.12	2.36	1.92	1.68	1.41	1.13	497
	RF	1[d]	3.14	2.51	2.16	1.97	1.75	1.51	483
		2	4.09	3.58	3.09	2.82	2.51	2.17	485
		3	4.95	5.02	4.33	3.95	3.52	3.03	496
		RP[d]	3.69	2.52	2.18	1.99	1.77	1.53	497
	FC	1[c]	4.54	4.37	3.74	3.39	2.98	2.51	346
		2[c]	6.6	6.26	5.36	4.85	4.27	3.6	363
		S[c,e]	16.49	9.14	7.85	7.11	6.27	5.31	364
		RP	5.53	3.26	2.79	2.52	2.22	1.88	351
	BC	1	7.91	4.93	4.22	3.81	3.35	2.83	342
		2	8.04	6.09	5.24	4.76	4.23	3.61	347
		S[e]	15.74	14.72	12.58	11.37	9.99	8.41	336
		RP[f]	7.93	6.38	5.46	4.94	4.35	3.57	348
II	SCC	1	2.34	0.37	0.32	0.29	0.25	0.21	348
		2[c]	4.2	1.02	0.86	0.78	0.69	0.58	382
		3	3.84	1.25	1.08	0.98	0.86	0.73	372
		4	4.97	1.14	0.97	0.87	0.77	0.65	390
		S_4[c,g]	15.6	3.93	3.34	3.01	2.64	2.23	390
		RP	3.85	0.78	0.66	0.6	0.53	0.45	394
	RF	1	1.66	0.95	0.81	0.73	0.64	0.54	349
		2[h]	3.02	1.74	1.48	1.34	1.18	1	373
		3	1.85	0.87	0.74	0.67	0.58	0.49	372
		4[h,i]	3.8	1.63	1.39	1.25	1.1	0.93	372
		RP	2.61	1.24	1.06	0.96	0.84	0.72	377

Abbreviations: AIC, Akaike information criterion; BC, branch count; CS, compound symmetry; FC, flower count; RF, root formation; RP, repeated measurement; SCC, stem cutting count; TPE, two-phase experiment; UN, unstructured; VC, variance components. [a]Total over $l = 1, 2, 3$ time-points. [b]Total over $l = 1, \ldots, 11$ time-points. [c]Estimates obtained without REP.IB in model (4). [d]Estimates obtained without REP.IB in model (6). [e]Total over $l = 1, 2$ time-points. [f]Estimates obtained without REP.COL in model (4). [g]Total over $l = 1, 2, 3, 4$ time-points. [h]Estimates obtained without RTABLE in model (6). [i]Estimates obtained without RTABLE.TRAY in model (6).

across single time-points influencing the regeneration capability. Seasonal temperature increase may increase leaf tissue dehydration levels of *P. zonale* during the rooting period,[32] which is known to reduce the regeneration capability of stem cuttings.[33] Second, varying day lengths across single time-points affecting the rooting. Day length is known to have an effect on rooting in other horticultural crops such as *Dahlia*.[34] Furthermore, *P. zonale* is a short-day plant, which means its reproductive cycle, including vegetative and floral growth regulation, is affected by day length. Third, varying cutting storage length and conditions were present between harvest and planting. The standard storage duration of 4 days between harvest and planting has in our experience no negative effect on rooting. However, we noticed a negative effect on rooting and stock cultivation when the time between cooling chain and planting of stem cuttings lasted longer than 20 min and stem cuttings were subjected to temperatures over 25 °C when planting during summer periods. Serek *et al.*[35] found an inhibition of rooting in terms of a reduced number and length of roots as well as reduced dry mass of roots of *P. zonale* cuttings after a short-term storage of already 3 days. In Serek's[35] study; however, a precise definition of the control treatment is lacking. Mutui *et al.*[36] also found no adverse storage effect (4 days in the darkness) on rooting percentage, even though the length of roots and the number of roots per cutting were reduced. Fourth, varying pruning practices and watering are also likely to affect physiological processes. Pruning was variable due to alternating personnel who made different decisions regarding what constitutes a harvestable shoot. Watering varied in that there were differences in total water amount given between time-points, although within time-points, no spatial effects resulting from irrigation were observed. The effect of less water, or drought stress before phenotyping made roots poorly visible and differentiation difficult, which resulted in outlying observations especially in TPE I at $l = 2$. An excess of water inhibited the development of roots resulting in a downgrading of RF of genotypes.

Other considerations for selection

Selection on production-related traits should be reconsidered because the current indirect method of selection for SCC and FC, based on overall impression of the growth type and branching, is ineffective due to low correlation between these traits. One possibility is to count and assess stem cuttings for RF of selected genotypes in the seedling generation when they are vegetatively propagated for the first clonal generation (Figure 1). A selection of SCC and RF at single time-points has been found effective as there was sufficient genotypic variance (Table 5). Even better would be selection across single time-points, because the number of stem cuttings per plant increases with the plant's age, and the ability to sustain stem cutting production over time is genotype-dependent. Therefore, the total SCC per genotype is a promising trait for selection.

Efficient selection of genotypes depends greatly on the phenotyping procedure. Phenotyping platforms for investigating biomass,[4] which would be comparable to SCC, or X-ray computed tomography coupled with image-analyzing software packages[37] to assess root formation were not affordable. Other, less costly, methods for phenotyping root traits, such as counting the number of roots or measuring their length.[35,36] would have been too labor and time intensive for populations of the size considered here. Therefore, in P2, a scoring procedure for RF was established that extends the assessment of rooting percentage.[36] In contrast to rooting percentage, defined as the proportion of rooted cuttings obtained from the total number of planted cuttings, RF allows the quality of each rooted cutting to be assessed. Further, rooting percentage was not found suitable for selection, since rooting percentage was generally high and varied little between genotypes. This agrees with results of Mutui *et al.*[36] who found 100 % rooting in well-known *P. zonale* cultivars.

Throughputs of 125 stock plants in P1 and 5500 rooted cuttings in phase two per day were achieved. This makes the developed phenotyping protocol an effective and low-cost method comparable to high-throughput phenotyping procedures.

CONCLUSION

With the help of the high-throughput phenotyping procedure developed and experimental design used in this study, genotypic variation could be effectively quantified, allowing varietal improvement of over 20 %.

Difficulties in implementing the experimental design were alleviated by a non-standard randomization approach observing experimental design principles.

We found that two-phase experimental designs in *P. zonale* breeding can reduce the error variances by accounting for phase-specific factors and increase the precision of estimates of phenotypic and genotypic effects, which positively affects the response to selection.

This study serves as a guideline to use experimental design, mixed models and response to selection in *P. zonale* breeding experiments. Further, it is expected that these techniques will be equally applicable to other species that involve similar phase-wise experimental setup.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This project was partly funded by grants KF3130602 (BMWi) and PI 3777/19-1 (DFG). We thank the team from Klemm+Sohn GmbH & Co. KG who assisted in carrying out the experiments. Further, we thank Andrea Dohm and Jens Möhring for their willingness to discuss the experiments and their helpful critical comments.

REFERENCES

1 Falconer DS, Mackay TFC. *Introduction to Quantitative genetics*. Prentice Hall: London, UK, 1996.
2 Walter A, Liebisch F, Hund A. Plant phenotyping: from bean weighing to image analysis. *Plant Methods* 2015; **11**: 14.
3 Barker J, Zhang N, Sharon J, Steeves R, Wang X, Wei Y *et al.* Development of a field-based high-throughput mobile phenotyping platform. *Comput Electron Agric* 2016; **122**: 74–85.
4 Busemeyer L, Ruckelshausen A, Möller K, Melchinger AE, Alheit KV, Maurer HP *et al.* Precision phenotyping of biomass accumulation in triticale reveals temporal genetic patterns of regulation. *Sci Rep* 2016; **3**: 1–6.
5 Panguluri SK, Kumar AA (ed.). *Phenotyping for Plant Breeding*. Springer: New York, NY. 2013.
6 Mutui TM, Mibus H, Serek M. Effect of meta-topolin on leaf senescence and rooting in *Pelargonium* × hortorum cuttings. *Postharvest Biol Technol* 2012; **63**: 107–110.
7 Edmondson RN. Past developments and future opportunities in the design and analysis of crop experiments. *J Agric Sci* 2005; **143**: 27–33.
8 Fisher RA. *The Design of Experiments*. Edinburgh Oliver & Boyd: Edinburgh, Scottland, 1935.
9 McIntyre GA. Design and analysis of two phase experiments. *Biometrics* 1995; **11**: 822–828.
10 Brien CJ, Harch BD, Correll RL, Bailey RA. Multiphase experiments with at least one later laboratory phase. I. Orthogonal designs. *J Agric Biol Environ Stat* 2011; **16**: 422–450.
11 Johnl JA, Williams ER. *Cyclic and Computer Generated Designs*. London Chapman & Hall: London, UK, 1995.
12 Brien CJ, Demétrio CGB. Formulating mixed models for experiments, including longitudinal experiments. *J Agric Biol Environ Stat* 2009; **14**: 253–280.
13 Piepho HP, Büchse A, Emrich K. A hitchhiker's guide to mixed models for randomized experiments. *J Agron Crop Sci* 2003; **189**: 310–322.
14 Cook DR, Weisberg S. Residuals and influence in regression. In: Cos DR, Hinkley DV (ed.). *Monographs on Statistics and Applied Probability*. Chapman & Hall: New York, NY. 1984.
15 Littell RC, Milliken GA, Stroup WW, Wolfinger RD, Shabenberger O. SAS for mixed models. SAS Institute Inc: Cary, NC, 2006.
16 Dufner J, Jensen U, Schuhmacher E. *Statistic with SAS*. Vieweg & Teubner Verlag: Tübingen. 2004, 171–180 German.
17 Searle SR, Casella G, McCulloch CE. *Variance Components*. John Wiley & Sons: Hoboken, NJ, 1992.
18 Piepho HP, Büchse A. A mixed modelling approach for randomized experiments with repeated measures. *J Agron Crop Sci* 2004; **247**: 230–247.
19 Piepho HP, Eckl T. Analysis of series of variety trials with perennial crops. *Grass Forage Sci* 2014; **69**: 431–440.
20 Akaike H. A new look at the statistical model identification. *IEEE Trans Automat Contr* 1974; **19**: 716–723.
21 Wolfinger R. Covariance structure selection in general mixed models. *Commun Stat Simul Comput* 1993; **22**: 1079–1106.
22 Piepho HP, Möhring J. Computing heritability and selection response from unbalanced plant breeding trials. *Genetics* 2007; **177**: 1881–1888.
23 Piepho HP, Möhring J. On estimation of genotypic correlations and their standard errors by multivariate REML using the MIXED procedure of the SAS System. *Crop Sci* 2011; **51**: 2449–2454.
24 Holland JB. Estimating genotypic correlations and their standard errors using multivariate restricted maximum likelihood estimation with SAS Proc MIXED. *Crop Sci* 2006; **46**: 642–654.
25 Müller BU, Kleinknecht K, Möhring J, Piepho HP. Comparison of spatial models for sugar beet and barley trials. *Crop Sci* 2010; **50**: 794–802.
26 Schützenmeister A, Piepho HP. Residual analysis of linear mixed models using a simulation approach. *Comput Stat Data Anal* 2012; **56**: 1405–1416.
27 Bernal-Vasquez AM, Utz HF, Piepho HP. Outlier detection methods for generalized lattices: a case study on the transition from ANOVA to REML. *Theor Appl Genet* 2016; **129**: 787–804.
28 Piepho HP, Möhring J, Williams ER. Why randomize agricultural experiments? *J Agron Crop Sci* 2013; **199**: 374–383.
29 Stroup WW. Power analysis based on spatial effects mixed models: a tool for comparing design and analysis strategies in the presence of spatial variability. *J Agric Biol Environ Stat* 2002; **7**: 491–511.
30 Mead R, Curnow RN, Hasted AM. *Statistical Methods in Agriculture and Experimental Biology*. Chapman & Hall: Boca Raton, FL, 2002.
31 Kelen M, Ozkan G. Relationships between rooting ability and changes of endogenous IAA and ABA during the rooting of hardwood cuttings of some grapevine rootstocks. *Eur J Hortic Sci* 2003; **68**: 8–13.
32 Kadner R, Druege U. Role of ethylene action in ethylene production and poststorage leaf senescence and survival of pelargonium cuttings. *Plant Growth Regul* 2004; **43**: 187–196.
33 Loach K. Water relations and adventitious rooting. In: Sankhla N (ed.). *Adventitious Root Formation in Cuttings*. Dioscorides Press: Portland, OR. 1989.
34 Zimmermann PW, Hitchcock AE. Root formation and flowering of *Dahlia* cuttings when subjected to different day lengths. *Bot Gaz* 1929; **87**: 1–13.
35 Serek M, Prabucki A, Sisler EC, Andersen AS. Inhibitors of ethylene action affect final quality and rooting of cuttings before and after storahe. *Hort Sci* 1998; **33**: 153–155.
36 Mutui TM, Mibus H, Serek M. Effects of thidiazuron, ethylene, abscisic acid and dark storage on leaf yellowing and rooting of Pelargonium cuttings. *J Hort Sci Biotechnol* 2005; **80**: 543–550.
37 Kumar J, Kumar S, Pratap A (ed.). *Phenomics in Crop Plants: Trends, Options and Limitations*. Springer: New Delhi. 2015.

Supplementary Information for this article can be found on the *Horticulture Research* website (http://www.nature.com/hortres)

3 Identifying effective design approaches to allocate genotypes in two-phase designs: a case study in *Pelargonium zonale*[2]

[2]This chapter is published as:
Molenaar H, Boehm R, Piepho H-P (2018). Identifying effective design approaches to allocate genotypes in two-phase designs: A case study in Pelargonium zonale. Frontiers in Plant Science, 8:1-12.

frontiers in Plant Science

METHODS
published: 05 January 2018
doi: 10.3389/fpls.2017.02194

Identifying Effective Design Approaches to Allocate Genotypes in Two-Phase Designs: A Case Study in *Pelargonium zonale*

Heike Molenaar[1], Robert Boehm[2] and Hans-Peter Piepho[1]**

[1] Biostatistics Unit, Institute of Crop Science, University of Hohenheim, Stuttgart, Germany, [2] Klemm + Sohn GmbH & Co. KG, Stuttgart, Germany

Robust phenotypic data allow adequate statistical analysis and are crucial for any breeding purpose. Such data is obtained from experiments laid out to best control local variation. Additionally, experiments frequently involve two phases, each contributing environmental sources of variation. For example, in a former experiment we conducted to evaluate production related traits in *Pelargonium zonale*, there were two consecutive phases, each performed in a different greenhouse. Phase one involved the propagation of the breeding strains to obtain the stem cutting count, and phase two involved the assessment of root formation. The evaluation of the former study raised questions regarding options for improving the experimental layout: (i) Is there a disadvantage to using exactly the same design in both phases? (ii) Instead of generating a separate layout for each phase, can the design be optimized across both phases, such that the mean variance of a pair-wise treatment difference (MVD) can be decreased? To answer these questions, alternative approaches were explored to generate two-phase designs either in phase-wise order (Option 1) or across phases (Option 2). In Option 1 we considered the scenarios (i) using in both phases the same experimental design and (ii) randomizing each phase separately. In Option 2, we considered the scenarios (iii) generating a single design with eight replicates and splitting these among the two phases, (iv) separating the block structure across phases by dummy coding, and (v) design generation with optimal alignment of block units in the two phases. In both options, we considered the same or different block structures in each phase. The designs were evaluated by the MVD obtained by the intra-block analysis and the joint inter-block–intra-block analysis. The smallest MVD was most frequently obtained for designs generated across phases rather than for each phase separately, in particular when both phases of the design were separated with a single pseudo-level. The joint optimization ensured that treatment concurrences were equally balanced across pairs, one of the prerequisites for an efficient design. The proposed alternative approaches can be implemented with any model-based design packages with facilities to formulate linear models for treatment and block structures.

Keywords: experimental design, two-phase design, mean variance of a pair-wise treatment difference, A-optimal, dummy analysis, experimental structure, horticultural breeding, *Pelargonium zonale*

Abbreviations: EU1, the experimental unit in P1; EU2, the experimental unit in P2; IBD, incomplete block design; MVD, the mean variance of a pair-wise treatment difference; $MVD_{(F)}$, the MVD obtained by the intra-block analysis; $MVD_{(R)}$, the MVD obtained by the joint inter-block-intra-block analysis; P1, phase one; P2, phase two; VC, variance components.

OPEN ACCESS

Edited by:
Jaime Prohens, Universitat Politècnica de València, Spain

Reviewed by:
Andrew Mead, Rothamsted Research (BBSRC), United Kingdom
Elisabetta Frascaroli, Università di Bologna, Italy

****Correspondence:***
Heike Molenaar
heike.molenaar@uni-hohenheim.de
Hans-Peter Piepho
piepho@uni-hohenheim.de

Specialty section:
This article was submitted to Plant Breeding, a section of the journal Frontiers in Plant Science

Received: *03 August 2017*
Accepted: *13 December 2017*
Published: *05 January 2018*

Citation:
Molenaar H, Boehm R and Piepho H-P (2018) Identifying Effective Design Approaches to Allocate Genotypes in Two-Phase Designs: A Case Study in Pelargonium zonale. Front. Plant Sci. 8:2194. doi: 10.3389/fpls.2017.02194

INTRODUCTION

Robust phenotypic data from trials that allow an adequate statistical analysis are of utmost importance for successful varietal improvement, identification of quantitative loci, marker-assisted selection, association mapping, and genomic selection. To obtain such data, trials are laid out to best control local variability through an experimental design (Federer and Crossa, 2012). There are situations where the experiment consists of two phases, e.g., when plant material is grown in the field to obtain the yield in the first phase and in the second phase chemical analyses are conducted in the laboratory (Smith et al., 2014), in which case the environmental conditions in the field trial have an influence on the response obtained in the second phase of the experiment in the laboratory. In such situations, two-phase experimental designs are recommended. All too often, however, both the design and statistical analyses are less than optimal when the two-phase nature of the experiment is overlooked, e.g., the change of observational units from one phase to the other or an overlapping of phases (Brien et al., 2011). As a result, variation cannot be broken down into all its components, which leads to a decreased accuracy of treatment effect estimates (Curnow, 1959).

Two-phase experimental designs can be found in many research areas, for example in crop breeding programs, where plants are tested under field conditions during the first phase and collected material is processed further for chemical analysis; in clinical studies, where patients are treated first, and specimens are processed in a laboratory in the second phase; in food processing studies, when first mixtures are prepared and in a subsequent phase the mixtures are processed further to produce the final products (Brien et al., 2011); or when conducting microarray experiments, where first messenger RNA is derived from subjects that are exposed to a set of treatments and then the mRNA is used in a microarray assay to obtain the gene expression (Jarrett and Ruggiero, 2008). Even if a laboratory phase is not involved, two phases can be present, as in ornamental breeding, where in the first phase stock plants are cultivated and in the second phase harvested plant material is tested for production related traits (Molenaar et al., 2017). Both phases take place in greenhouses, which may be in different locations.

Often in planned two-phase experiments, the first phase is considered in the experimental design, while the second phase is not considered at all. For example, in cereal breeding, plant material from the field may be processed further according the "field order" resulting in a systematic allocation of treatments in the second phase or all samples of a treatment may be pooled together in the laboratory (Brien et al., 2011). Already in 1955, McIntyre (1955) described two-phase experimental designs and proposed the use of randomization in each phase.

Implementing a conventional two-phase design, where both phases use optimal designs, can pose some difficulties due to practical considerations as mentioned in a study on *Pelargonium zonale* by Molenaar et al. (2017). In that study, it would have been prohibitively labor-intensive to follow an optimized pre-defined layout, because of the elaborate process of planting thousands of stem cuttings in the second phase. As this was the first attempt to introduce a two-phase experimental design in a *P. zonale* breeding program, a compromise was made, and randomization in the second phase was carried out on site so that the requirement of randomization was met, but the design could not be optimized in view of the design used in the first phase.

This initial approach raised questions regarding options for further improving the experimental design: (i) Is there a disadvantage in leaving treatments in the same randomized order from the first phase when transferring samples to the second phase of the experiment, i.e., using exactly the same design in both phases? (ii) Instead of generating a separate layout for each phase, can the design be optimized across both phases, such that the MVD can be decreased across both phases compared to two independent designs?

Therefore, the objective of this study was to explore potential pragmatic approaches for generating improved two-phase experimental designs and thereby to answer questions (i) and (ii). Section "A Two-Phase Experiment in *P. zonale* Breeding" summarizes the former experiment of Molenaar et al. (2017), on which operational possibilities are modeled. Sections "Option 1 – Design Generation for Each Phase Separately" and "Option 2 – Design Generation across the Two Phases" present two options for generating two-phase experimental designs for each phase separately or across the two phases considering either the same or different block structures in both phases. In Section "Results", the generated designs are evaluated regarding the MVD. Sections "Discussion and Conclusion", give a discussion and conclusion to identify effective two-phase designs.

MATERIALS AND METHODS

A Two-Phase Experiment in *P. zonale* Breeding

In 2013/14, we implemented a two-phase experimental design in a *P. zonale* breeding program to assess production related traits of $v = 500$ genotypes. In Phase 1 (P1), conducted in location 1, stock plants of genotypes were cultivated, from which the stem cutting count was obtained. In Phase 2 (P2), the stem cuttings harvested from genotypes during P1 were planted to assess the root formation in location 2. Both phases took place in greenhouses.

In P1, an α-design with $r = 4$ replicates, each with $b = 167$ incomplete blocks of size $k = 3$, was used. One of the incomplete blocks in each replicate was only of size two. Each experimental unit in P1 (EU1) contained a pair of stock plants from the same genotype, for a total of six plants per incomplete block of size three. Each cultivation table accommodated a full set of genotypes, i.e., one replicate.

In P2, randomization was carried out on site as follows: First, the total experimental space of rooting tables was divided into four regions. To each region in P2 the replicates of P1 were systematically assigned. A rooting table held 36 trays and a region held 72 up to 108 trays, hence, not all trays fit necessarily on one rooting table. A tray held 39 paper pots arranged in three rows. Trays were divided into areas representing EU2. Second, all stem cuttings of a genotype and replicate in P1 were packed in a small bag to transfer the plant material from location 1 to 2, and were

randomly allocated to one area in the corresponding region to which a replicate was assigned in P2. Thus, stem cuttings from each EU1 were allocated exactly to one EU2. The sizes of areas for EU2 varied depending on the harvested stem cutting count per genotypes. The rooting tables and trays in P2 were considered as *post-blocking* factors in the analysis, which could be regarded as incomplete blocks in an IBD, for which the design was previously not optimized (**Figure 1**).

Idealized Conditions to Assume the Same Block Structure in Both Phases

We idealized the experimental conditions in several ways to different degrees in the designs to be described in this and the following Sub-section "Different Block Structures in Both Phases" for comparing different design generation scenarios. This was done in the interest of focusing on the general principles implemented in scenarios investigated in this study without having to focus on intricate specifics of the *P. zonale* study. First, we assumed that in each phase the same block structure can be used. Thus, for each pair of stock plants we presumed that no stock plants were lost and that the stem cutting counts were reduced to six stem cuttings per genotype in P1 to assess root formation in P2. Hence, the EU2 (areas) were assumed to be of equal size in P2.

Further, the physical unit of a tray should correspond exactly to one randomization unit in P2, i.e., to an incomplete block of size six. To consider the same block structure in both phases so that block units in P1 correspond to block units of the same size in P2, requires increasing the block size in P1 from $k = 3$ to $k = 6$. These idealized conditions enabled us to assume the same resolvable IBD design with $r = 4$ replicates with each $b = 84$ incomplete blocks having the same block sizes $k = 6$ in both phases (**Figure 1**). We also assumed equal block sizes and the genotype number was increased from $v = 500$ to 504. Given these design properties, two options were considered to generate the two-phase experimental design (**Table 1**): Option 1 was to

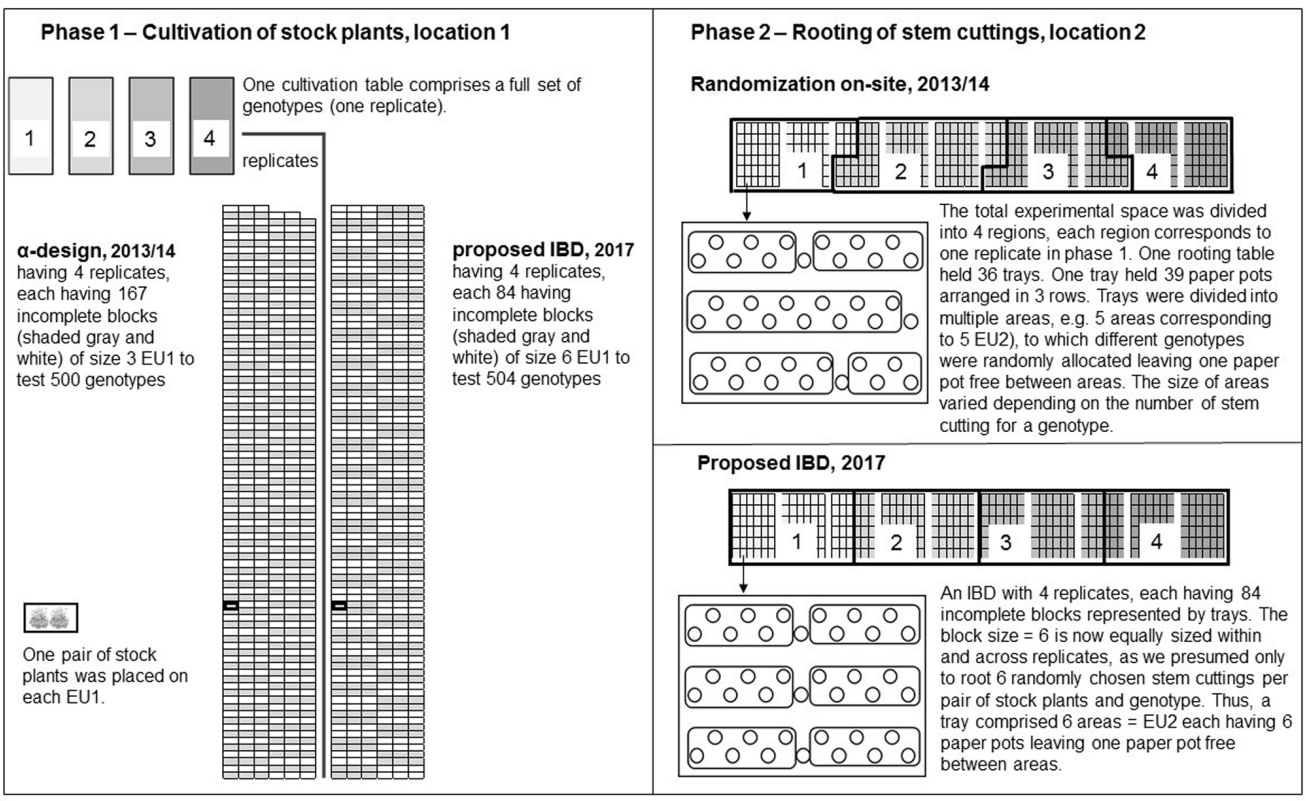

FIGURE 1 | Modified according to Molenaar et al. (2017): Two-phase experimental design accounting for idealized conditions. In comparison to 2013/14, where an α-design with four replicates, each with 168 incomplete blocks of block size 3 (one of them was only of size 2), was used to test 500 genotypes in Phase 1 (P1). We idealized the conditions in such a way, that we could use instead an incomplete complete block design (α-design) with four replicates each having 84 incomplete blocks of size 6 to test 504 genotypes. In both cases, the cultivation tables comprised a complete set of genotypes placed on 500 and 504 planting positions, respectively. On each experimental unit, a pair of stock plants was placed. In Phase 2 (P2), when randomization was carried out on site in 2013/14, we divided the total experimental space of rooting tables into four regions to which the four replicates in P1 were systematically assigned. Regions shaded in gray in rooting tables in P2 correspond to replicates shaded in gray of cultivation tables in P1. Each rooting table held 36 trays and a tray contained 39 paper pots arranged in three rows. The trays were divided into areas comprising different numbers of paper pots to which genotypes were randomly allocated. The number of paper pots, i.e., the size of areas (experimental units in P2), varied depending on the numbers of stem cuttings for a genotype. In accordance with our idealized conditions assuming that only six stem cuttings of a pair of stock plants and genotype are rooted, we proposed to use a pre-specified IBD with four replicates each having 84 incomplete blocks of size 6 in P2. Thus, the regions, to which the replicates in P1 were assigned, the incomplete blocks represented by the trays and the EU2 were now of equal size.

TABLE 1 | Overview of designs with the same block structure in both phases†.

Option	Description	Scenario	Code in Supplementary Presentation 1	Figure in Supplementary Presentation 2
1 – Design generation separately for each phase	Exactly the same design in both phases	*I*	1	A
	New randomization of genotypes to IB within replicates of P2	*II*	2	B
2 – Design generation across the two phases	Generating a single design with eight replicates and splitting these among the two phases	*III*	3	C
	Separation of block structures using phase-specific dummy coding	*IV*	4	D
	Design generation in two steps: (i) allocating blocks of P2 to incomplete blocks of P1; (ii) allocating of genotypes to IB of both P1 and P2	*V*	5	E1 and E2

†*In each phase, an incomplete block design was used with v = 504 genotypes, r = 4 replicates, b = 84 incomplete blocks of size k = 6. Designs were generated for each phase separately or across phases and for different scenarios within these options.*

generate a design for each phase separately (Section "Option 1 – Design Generation for Each Phase Separately") and Option 2 was to generate a design across phases by simultaneously accounting for the block structure of both phases (Section "Option 2 – Design Generation across the Two Phases"). During design generation using either Option 1 or Option 2, the complete replicates from P1 were kept intact in P2, except for *Scenario III*.

General Approach

The general approach for generating two-phase experimental designs by either Options 1 or 2 was model-based. Therefore, first a treatment model was defined and second a block model for each of the two phases. Such model-based approaches can be implemented in various software packages, e.g., *dae* (Brien, 2017), *DiGGer* (Coombes, 2009), or *OD* (Butler, 2013) in R or the OPTEX procedure in SAS (SAS Institute Inc., 2014). We implemented our approaches with OPTEX, and provided all relevant codes (**Supplementary Presentation 1**).

Option 1 – Design Generation for Each Phase Separately

Scenario I – Transmitting the experimental layout of P1 to P2

A resolvable IBD was generated for P1 for the specifications given above (Code 1, Figure A in **Supplementary Presentation 2**). The experimental layout of P1 was transmitted exactly to P2. In doing so, treatments in P2 were left in the same order as used in P1.

Scenario II – New randomization of genotypes to incomplete blocks within replicates in P2

In contrast to *Scenario I*, in P2 a separate resolvable IBD was generated (Code 2, Figure B).

Option 2 – Design Generation across the Two Phases

Scenario III – Generating a single design with eight replicates and splitting these among the two phases

We generated a resolvable IBD with $r = 8$ replicates (Code 3, Figure C), where all other design parameters remained unchanged ($v = 504$, $b = 84$, $k = 6$), and split these replicates equally among the two phases. Since genotypes were randomized to incomplete blocks across the eight replicates, the design was optimized across the two phases in terms of the number of concurrences per treatment pair. Each complete replicate in P1 was transferred intact to one replicate in P2.

Scenario IV – Separation of block structures using phase-specific dummy coding

In the dataset defining the block structures across phases, we defined a factor identifying the two phases (Code 4, Figure D). In each phase, there were $r = 4$ replicates and incomplete blocks were nested within each of the replicates. The records for the two phases were concatenated in the dataset for design generation based on a model comprising the block effects for both phases. The clue for generating the design across the two phases was to set the factor for incomplete blocks of P1 to a single pseudo level for incomplete blocks of P2 and vice versa. By this dummy coding, the pseudo level acted as one additional block level of incomplete blocks in P1 or P2. The design was then optimized simultaneously with respect to the assignment of genotypes to the two blocking systems.

Scenario V – Design generation with optimal alignment of block units in the two phases

Instead of generating the design in phase-wise order, the design generation was conducted in replicate-wise order in two steps (Code 5, Figures E1, E2), optimizing the alignment of block structures of both phases. First, for each replicate, the incomplete blocks of P2 were allocated to incomplete blocks of P1 for each of the four replicates separately to obtain a block layout of the design across the two phases. This was achieved by formally considering blocks of P2 as the "treatment" factor and blocks in P1 as the "block" factor, thus optimizing the efficiency of block effects estimates in P2, given the block structure in P1. In this step, the allocation of genotypes to EU1 and EU2 was not yet considered. Second, the genotypes were allocated to incomplete blocks within replicates of both P1 and P2 considering all four replicates simultaneously. At this stage, the alignment of blocks in P1 with blocks in P2 was fixed at the configuration obtained in the first step, and this alignment was used as the block

model to generate an allocation of treatments to EU1 and EU2 simultaneously.

Different Block Structures in Both Phases

The assumption of a common block structure in both phases is rather idealized as in practice the phases of the experiments take place in totally different locations and different environmental conditions prevail in each phase (Molenaar et al., 2017). To account better for the environmental conditions in location 1, two different *post-blocking* factors were considered in the former analysis, which are now further considered in generating designs.

Adding a Column Factor in the First Phase

A column factor was added to the *randomization*-based model, representing columns of 84 experimental units per replicate on tables in the greenhouse. Accommodating this column factor means that different block structures are used in both phases. Hence, the operational approaches derived in Section "Option 1 – Design Generation for Each Phase Separately" were modified, using in P1 a row-column design to test $v = 504$ genotypes in $r = 4$ replicates, each arranged in $k = 84$ rows and $s = 6$ columns, whereas in P2 the resolvable IBD used previously was employed (**Table 2**).

Blocking Factor to Account for Induced Variations by Workers in the First Phase

Molenaar et al. (2017) found that in most cases the largest variance was the residual error variance, while in P1 the incomplete block variance was estimated to be zero, indicating that there was no correction due to that block factor during the estimation of effects when modeled as random. In search of sources of variation that were not explicitly taken into account so far, we found that workers harvesting the stem cuttings in P1 induced some effect. Considering that a worker can harvest stem cuttings from approximately 125 stock plants per day, an additional *post-blocking* factor "*worker-day*" was defined, which comprised eight levels (blocking strategy *a*). Levels one to eight corresponded to the positional numbers from 1 to 63, 64 to 126, 127 to 189, 190 to 252, 253 to 315, 316 to 378, 379 to 441 and 442 to 500 in the layout of EU1 in each replicate (**Figure 2**). However, during cultivation within the first 5 months stock plants were lost at random. Hence, less than 125 stock plants per "*worker-day*" were grouped together for analysis. Therefore, two other block strategies (*b* and *c*) in terms of the number of planting positions visited by a worker per day were defined. We further considered this additional block factor within a row–column design, but also as the only block factor in P1, for generating designs using either Options 1 or 2 (**Table 2**).

The Mean Variance of a Treatment Difference as a Selection Criterion

Designs generated by the procedure OPTEX are optimized for D-efficiency (*OD* or *DiGGer* packages provide algorithms for generating A-optimal designs). In plant breeding A-optimal or A-efficient designs are preferred as optimizing this criterion minimizes the average variance of genotype differences (MVD) (Hinkelmann and Kempthorne, 2005). Thus, the precision of estimates of genotype differences is increased and better phenotypic selection and varietal improvement can be achieved. Both D- and A-optimality usually lead to similar designs for comparative experiments with a single treatment factor (John and Williams, 1995), so the procedure OPTEX was a useful tool for our purposes, despite its focus on D-optimality. Thus, we computed the MVD obtained by linear mixed models either by intra-block or joint inter-block-intra-block dummy analyses (Piepho, 2015), in which the information about the precision of genotype parameters is contained in the variance-covariance matrix (Mead, 1988). Generally, the design showing the lowest MVD was preferred.

Resolvability

We ensured that all designs were resolvable, meaning that the b incomplete blocks containing k plots (EUs) can be grouped to a complete r replicate of the v treatments. Resolvability of all designs generated was verified by frequency tables for the genotype-by-replicate classifications in each phase. For a resolvable IBD, all entries in the table must be unity. If necessary, resolvability of the two-phase designs was enforced by defining effects for incomplete blocks as random effects, tuning the variance so that resolvability was achieved. Defining a block effect as random essentially allows tuning its influence on the treatment information matrix. The smaller the variance, the smaller the influence on the treatment information matrix. In OPTEX, the variance of an effect is tuned via the PRIOR option (see Pereira and Tobias, 2015, for more details). For design generation in each scenario we set the prior for replicates to zero. For all remaining block effects, the prior was set to 2016, the value corresponded to the total number of experimental units in the experiment ($504 \times 4 = 2016$ EU1 = EU2). If resolvability was not achieved the prior was increased until resolvability was achieved.

Model Set-up for Design Evaluation

As mentioned above, our approach generally requires specification of the treatment model on the one hand and the model for block effects on the other hand. The model notation used here is universally applicable in any design package allowing the specification of linear models.

We illustrate this general model set-up for either Options 1 or 2 by considering a two-phase design having the same block structure in both phases. The treatment model, representing the 'randomized-tier' (Brien and Demétrio, 2009), was

$$\text{GEN}, \tag{1}$$

where GEN denotes the genotypes.

When the designs were generated for each phase separately (Option 1), then the *randomization*-based block model for design effects was set up for each phase separately. The P1 block model was

$$\text{REP} + \text{REP.IB1} + \underline{\text{REP.IB1.PAIR}} \tag{2}$$

and the *randomization*-based model for P2 was

$$\text{REP} + \text{REP.IB2} + \underline{\text{REP.IB2.AREA}} \tag{3}$$

TABLE 2 | Overview of designs assuming different block structure in each phase[†] for the same two options as in **Table 1**.

Option	Design in		Description	Scenario	Code in Supplementary Presentation 1
	Phase 1	**Phase 2**			
1 - Design generation for each phase separately	Row–column design	Resolvable IBD		*VI*	6
	Row–column design considering the "*worker-day*"	Resolvable IBD		*VII a, b, c*	7
	Considering only the "*worker-day*" within replicates	Resolvable IBD		*VIII a, b, c*	8
2 - Design generation across phases	Row–column design	Resolvable IBD	Separation of block structures using phase-specific dummy coding	*IX*	9
			Design generation in two steps: (i) allocating blocks of P2 to rows and columns of P1; (ii) allocating genotypes to rows and columns of P1 and IB of P2	*X*	10
	Row–column design considering the "*worker-day*"	Resolvable IBD	Separation of block structures using phase-specific dummy coding	*XI a, b, c*	11
			Design generation in two steps: (i) allocating blocks of P2 to rows, columns and "*worker-day*" of P1; (ii) allocating genotypes to rows, columns and "*worker-day*" of P1 and IB of P2	*XII a, b, c*	12
	Considering only the "*worker-day*" within replicates	Resolvable IBD	Separation of block structures using phase-specific dummy coding	*XIII a, b, c*	13
			Design generation in two steps: (i) allocating blocks of P2 to "*worker-day*" of P1; (ii) allocating genotypes to the "*worker-day*" of P1 and IB of P2	*XIV a, b, c*	14

[†]*In phase 1, a row-column design with v = 504 genotypes, r = 4 replicates, z = 84 rows and s = 6 columns and in phase 2 a resolvable IBD with r = 4 replicates, b = 84 incomplete blocks of size k = 6 were assumed.*

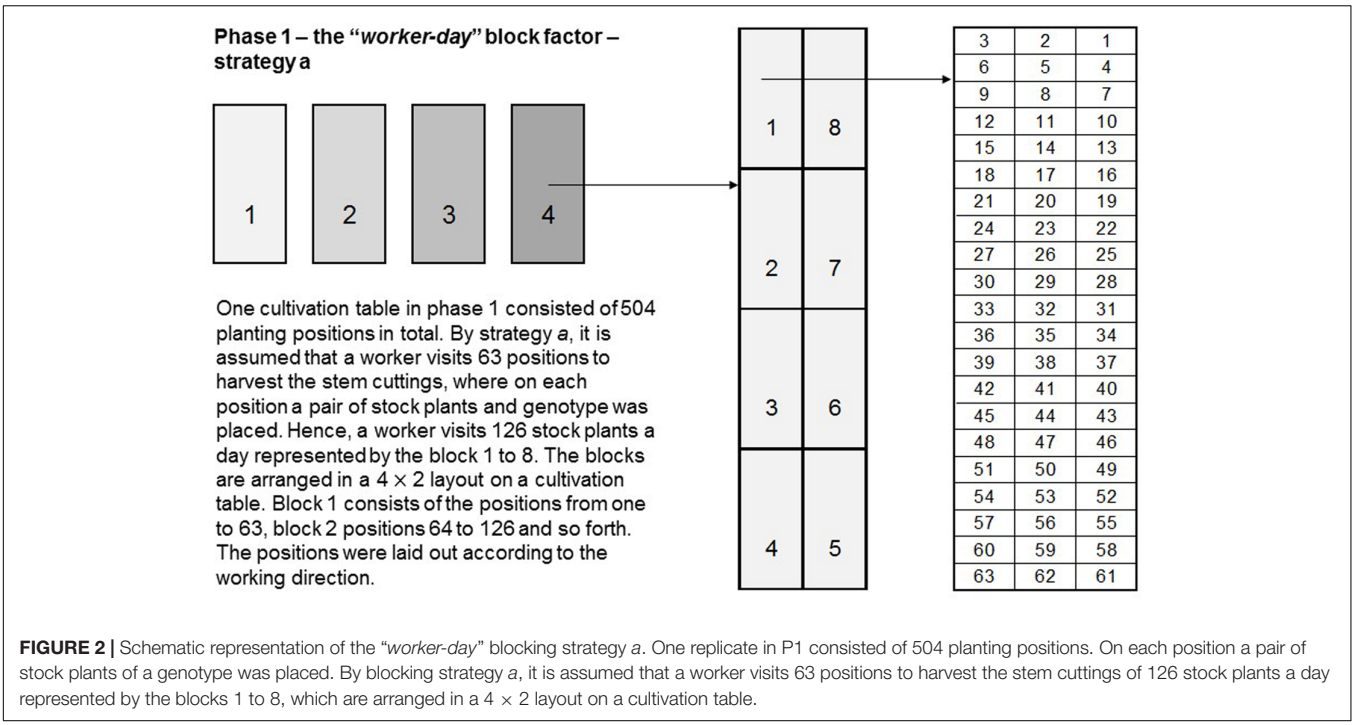

FIGURE 2 | Schematic representation of the "*worker-day*" blocking strategy *a*. One replicate in P1 consisted of 504 planting positions. On each position a pair of stock plants of a genotype was placed. By blocking strategy *a*, it is assumed that a worker visits 63 positions to harvest the stem cuttings of 126 stock plants a day represented by the blocks 1 to 8, which are arranged in a 4 × 2 layout on a cultivation table.

TABLE 3 | Full models in *Scenarios I* to *XIV*, including treatment effects, design effects in Phase 1 (P1) and Phase 2 (P2) and error terms used to estimate MVD for design evaluation.

Scenario	Model‡	Treatment effect	Design effects§					ERROR
			P1				P2	
I	2	GEN	REP	REP.IB1				REP.IB1.PAIR
II–V	4	GEN	REP	REP.IB1			REP.IB2	REP.IB2.AREA
VI, *IX*, *X*	5	GEN	REP	REP.ROW	REP.COL		REP.IB2	REP.IB2.AREA
VII, *XI*, *XII*†	6	GEN	REP	REP.ROW	REP.COL	REP.WORK	REP.IB2	REP.IB2.AREA
VIII, *XIII*, *XIV*†	7	GEN	REP			REP.WORK	REP.IB2	REP.IB2.AREA

†*For each scenario all blocking strategies a–c of "worker-day" were considered.*
‡*Model (2): IBD, which was transmitted from P1 to P2; Model (4): IBD in P1, IBD in P2; Model (5): IBD in P1 with the additional post-blocking factor column, IBD in P2; Model (6) a–c: IBD with the additional post-blocking factor column and "worker-day", where a to c indicate the different blocking strategies to account for the different numbers of positions that can be met by one worker per day. Model (7) a–c: Considering only the "worker-day" block in P1, and an IBD in P2.*
§*REP is the replicate effect, REP.WORK is the "worker-day" effect, REP.IB1 is the incomplete block effect in the first phase, REP.ROW is the row effect in the first phase, REP.COL is the column effect in the first phase, REP.IB2 is the incomplete block effect in the second phase, ERROR is the residual error.*

where REP denotes the replicates, REP.IB1, the incomplete blocks nested within replicates in P1, REP.IB1.PAIR, the residual error in P1, i.e., the EU1, on which a pair of stock plants was placed and represented the observational unit. Further, REP.IB2 denotes the incomplete blocks within replicates in P2 and REP.IB2.AREA, the residual error in P2, i.e., the EU2, from which the root formation of stem cutting was assessed. It is noted, that in model (2) and (3), REP was considered as a fixed effect, whereas incomplete blocks and the residual error were considered as random effects for generating designs.

The full model for design evaluation was obtained by augmenting the treatment model with both phase-specific *randomization*-based block models,

$$\text{GEN} + \text{REP} + \text{REP.IB1} + \text{REP.IB2} + \underline{\text{REP.IB2.AREA}} \quad (4)$$

As the replicates were kept intact from P1 to P2, only one effect was needed to define the replicates. Defining the experimental unit of the full model, the EU1 effect REP.IB1.PAIR did not need to be added explicitly either as it was implicitly accounted for by the EU2 effect REP.IB2.AREAS. This is because one EU1 was allocated to one EU2 and hence, effects of EU1 and EU2 are confounded.

When designs were generated across phases (Option 2), the block model for design generation was

$$\text{REP} + \text{REP.IB1} + \text{REP.IB2} + \underline{\text{REP.IB2.AREA}} \quad (5)$$

The full model corresponds to model (4), except for *Scenario III*, where the design was generated across the two phases by increasing the replicate number from four to eight. The model for design generation in this case was model (1), but after splitting

the eight replicates among the two phases and recoding the incomplete blocks for P1 and P2 as IB1 and IB2, respectively, the model for analysis was model (4).

Considering different block structures in both phases, e.g., in P1 a row-column design, a row-column design with the additional block factor "*worker-day*" or only the "*worker-day*" and in P2 still utilizing a resolvable IBD, the block factor of P1 (IB1) in models (2) and (4) was replaced by rows nested within replicates (REP.ROW), columns nested within replicates (REP.COL) and "*worker-day*" (REP.WORK), respectively (**Table 3**).

The Estimation of Variance Components (VC)

Variance components (VC) were required for the dummy analyses in the next step in Section "Intra-block or Joint Inter-block–Intra-block Dummy Analysis". Hence, for each block effect, including the residual error, the VC were estimated from Models (4) to (7) (**Table 3**) and the experimental data 2013/14 by taking all block effects as random. Because of the idealized conditions, IB2 was equivalent to the *post-blocking* factor TRAY in P2, whereas the other *post-blocking* factor TABLE in P2 of the past analysis (Molenaar et al., 2017) was neglected in the current analysis, because of confounding with the replicate effect (REP).

Intra-block or Joint Inter-block–Intra-block Dummy Analysis

The MVD was obtained from models (2) to (7) by a dummy analysis (**Supplementary Presentation 3**) taking block effect either as fixed or as random for each two-phase designs implemented in each scenario. The analysis with fixed blocks utilized the intra-block information and the obtained $MVD_{(F)}$ depended only on the residual error variance (fixed at the value of residual error variance obtained in the previous experiment) and the design, whereas random blocks allowed also the recovery the inter-block information (John and Williams, 1995). The variance of random block effects was set to estimated VC (**Table 4**) to obtain the $MVD_{(R)}$. Now, the $MVD_{(R)}$ depended not only on the residual error variance and the design, but also on the block variances estimated from the previous experiment. Further, the MVD was also obtained from the previous experiment applying the same models for the intra- and the joint inter-block–intra-block analysis to illustrate on the one hand the gain in precision by using the two *post-blocking* factors column and "*worker-day*" in the first phase, and, on the other hand, to compare the precision of the former experiment with its improved modifications implemented in each scenario. In analyzing *Scenario I*, the VC of IB1 and IB2 estimated by the use of model (4) (**Table 4**) were summarized and assigned to model (2) with VC of the replicate effect and residual error, estimated from model (4) too, to obtain the $MVD_{(R)}$.

RESULTS

Resolvability

In all scenarios resolvability was achieved by setting the prior value for incomplete block effects in the block model specification

TABLE 4 | Variance components of each model effect and corresponding proportions of the total variation attributable to each effect for Models (4) to (7).

	Model effect†	Model‡							
		4	**5**	**6, *a***	**6, *b***	**6, *c***	**7, *a***	**7, *b***	**7, *c***
	REP	2.6469	2.6378	2.6406	2.5939	2.5796	2.6192	2.5743	2.5621
	REP.WORK	–	–	0.2569	0.3204	0.3097	0.3131	0.3542	0.3352
	REP.IB1	0.1303	–	–	–	–	–	–	–
	REP.ROW	–	0.0505	0.0443	0.0418	0.0172	–	–	–
	REP.COL	–	0.2043	0.0823	0.0568	0.0497	–	–	–
	REP.IB2	0.5066	0.4378	0.4518	0.4516	0.4616	0.4892	0.4780	0.4854
	ERROR	3.7806	3.7315	3.5895	3.5767	3.6381	3.6524	3.6360	3.6729
Sum VC		7.0744	7.0619	7.0654	7.0412	7.0559	7.0739	7.0425	7.0556
	REP	37.5565	37.3526	37.3738	36.8388	36.5595	37.0263	36.5538	36.3130
	REP.WORK	–	–	3.6360	4.5503	4.3892	4.4261	5.0295	4.7508
	REP.IB1	1.8419	–	–	–	–	–	–	–
Proportion in %	REP.ROW	–	0.7151	0.6270	0.5934	0.2442	–	–	–
	REP.COL	–	2.8930	1.1644	0.8072	0.7039	–	–	–
	REP.IB2	7.1610	6.1995	6.3946	6.4137	6.5420	6.9156	6.7874	6.8796
	ERROR	53.4406	52.8399	50.8041	50.7966	51.5611	51.6321	51.6294	52.0565
Total		100%	100%	100%	100%	100%	100%	100%	100%

†*REP is the replicate effect, REP.WORK is the "worker-day" effect, REP.IB1 is the incomplete block effect in the first phase, REP.ROW is the row effect in the first phase, REP.COL is the column effect in the first phase, REP.IB2 is the incomplete block effect in the second phase, ERROR is the residual error.*

‡*Model (4): IBD in P1, IBD in P2; Model (5): IBD in P1 with the additional post-blocking factor column, IBD in P2; Model (6) a–c: IBD with the additional post-blocking factor column and "worker-day", where a to c indicate the different blocking strategies to account for the different numbers of positions that can be met by one worker per day. Model (7) a–c: considering only the "worker-day" block in P1, and an IBD in P2.*

to the total number of experimental units, 2016, except for *Scenarios V*, *X*, *XII a–c* and *XIV a–c*. Resolvability was realized for those scenarios by increasing the prior value to 10^6.

VCs of Random Effects

The largest VC was the residual error variance with a proportion of 53.44% of the total variance, followed by the replicate effect with 37.56 % (**Table 4**). By comparison, the variance of the incomplete block effect in P1 was small (1.8%). After adding the column *post-blocking* factor to the first phase and still using a resolvable IBD in P2 [model (5)], the residual error variance could be reduced from 53.5 to 52.8%. The proportion of variation explained by the rows was below 1%, whereas the proportion of variation explained by columns was 2.8%. The residual error variance was further reduced by accounting for the block factor "*worker-day*" in P1 to 50.8%, where simultaneously the proportion of variation explained by row and column effects was reduced to below 1% [model (6) *a–c*]. Subsequently, scenarios were considered, where the "*worker-day*" was the only block factor in P1 and the resolvable IBD was retained in P2 [model (7) *a–c*]. By doing this, the proportion of variation explained by the replicate effects was retained, the proportion of variation explained by the "*worker-day*" effect in P1, and the proportion of variation explained by the incomplete block effect in P2, were maximized. However, the proportion of variation captured by the residual error was increased again by 1% compared to model (6) *a–c*.

The Precision of the Two-Phase Design in 2013/14

In *post hoc* analysis of the previous experiment, the greatest $MVD_{(F)}$ and $MVD_{(R)}$ were observed generally in all conducted dummy analyses for model (4) (**Table 5**). By adding a column factor in the first phase [model (5)], a reduction of $MVD_{(F)}$ was achieved by more than 50%, whereas the reduction in $MVD_{(R)}$ was rather small. By considering the "*worker-day*" factor [model (6) *a–c*] in addition to the column factor, the reduction in MVD was about 0.04. When only the "*worker-day*" factor in the first phase [models (7), *a–c*] was considered, the reduction in $MVD_{(F)}$ was below 3.0, whereas the $MVD_{(R)}$ was slightly higher than the $MVD_{(R)}$ obtained by models (6), *a–c* (**Table 5**).

The Precision of Alternative Approaches

Two-Phase Designs Containing the Same Block Structure in Both Phases

By using the same pre-defined design in both phases, the smallest $MVD_{(F)}$ and $MVD_{(R)}$ were obtained for *Scenario I* (experimental layout was transmitted from P1 to P2) (**Table 6**). The $MVD_{(R)}$ were quite similar, especially between *Scenario II* to *IV*, whereas values of $MVD_{(F)}$ showed a wider range (2.4 to 3.3). Comparing the alternative approaches with the former two-phase experimental layout, the $MVD_{(F)}$ of designs implemented in each scenario was greater than the smallest $MVD_{(F)}$ obtained by model (7) *a–c*, except for *Scenario I* (**Tables 5**, **6**). Generally, a reduction in $MVD_{(R)}$ of over 0.5 was realized by every alternative two-phase design compared to the previous one (**Tables 5**, **6**). Relevant differences in MVD between options generating the design in phase-specific order (1) or across phases (2) were not observed, except for *Scenario I*.

Two-Phase Designs Containing Different Block Structures in Both Phases

Alternative two-phase designs considering in each phase a different block structure achieved a reduction especially in $MVD_{(R)}$ from about 2.09 to 1.99 in comparison to alternative approaches using the same block structure in both phases, where the minimum $MVD_{(R)}$ was about 2.09. (**Tables 6**, **7**). The reduction in $MVD_{(F)}$ was only from 2.41 to 2.35 when in the first phase the only block effect was the "*worker-day*" (*Scenario XIII a–c*). Comparing the options to generate two-phase designs in phase-specific order (1) or across the phases (2) considering different block structures in each phase, the smallest MVD were always found for Option 2 and with the approach using of a single pseudo level for incomplete blocks of P1 and P2 (**Table 7**).

DISCUSSION

We investigated several options for generating two-phase designs using a model-based design package. These options were explored for the case of an experiment with *P. zonale*, but our key

TABLE 5 | Different models for evaluating MVD for two additional blocking factors where MVD is obtained either by assuming blocks to be fixed or random.

Model†	$MVD_{(F)}$	$MVD_{(R)}$
4	9.4238	2.6767
5	4.55164	2.6147
6, *a*	4.51763	2.5499
6, *b*	4.41603	2.5270
6, *c*	4.51641	2.5399
7, *a*	2.78311	2.5541
7, *b*	2.74511	2.5303
7, *c*	2.74876	2.5425

†*Model (4): IBD in P1, IBD in P2; Model (5): IBD in P1 with the additional post-blocking factor column, IBD in P2; Model (6) a–c: IBD with the additional post-blocking factor column and "worker-day", where a to c indicate the different blocking strategies to account for the different numbers of position that can be met by a worker and a day; Model (7) a–c: considering only the "worker-day" block in P1 and an IBD in P2.*

TABLE 6 | Two options for evaluating MVD across five scenarios where MVD is obtained by assuming blocks either to be fixed or random in Model (4)† and setting block variances to values of estimated VCs‡ to obtain the $MVD_{(R)}$.

Option	Scenario	$MVD_{(F)}$	$MVD_{(R)}$
1	*I*	2.41303	2.08430
	II	3.34630	2.11680
2	*III*	3.32754	2.11686
	IV	3.33052	2.11687
	V	3.37687	2.11794

† *Model (4): IBD in P1, IBD in P2;*
‡*The VCs which were listed under the Model (4) in* ***Table 5*** *were used as block effects and residual error for random terms to obtain the* $MVD_{(R)}$.

TABLE 7 | Two options for evaluating MVD across five scenarios where MVD is obtained by assuming blocks either to be fixed or random and setting block variances to values of estimated VCs† to obtain the $MVD_{(R)}$.

Option	Model‡	Scenario	$MVD_{(F)}$	$MVD_{(R)}$
1	5	*VI*	3.36418	2.06195
	6, *a*	*VII - a*	3.27783	2.00898
	6, *b*	*VII - b*	3.25169	1.99296
	6, *c*	*VII - c*	3.29339	2.00670
	7, *a*	*VIII - a*	2.38759	2.01776
	7, *b*	*VIII - b*	2.36051	1.99974
	7, *c*	*VIII - c*	2.36826	2.01205
2	5	*IX*	3.33725	2.06106
	5	*X*	3.39264	2.06423
	6, *a*	*XI - a*	3.24335	2.0075
	6, *b*	*XI - b*	3.21642	1.99175
	6, *c*	*XI - c*	3.26363	2.00558
	6, *a*	*XII - a*	3.32675	2.01238
	6, *b*	*XII - b*	3.29614	1.9959
	6, *c*	*XII - c*	3.33943	2.00878
	7, *a*	*XIII - a*	2.37642	2.01637
	7, *b*	*XIII - b*	2.35226	1.99852
	7, *c*	*XIII - c*	2.36343	2.01112
	7, *a*	*XIV - a*	2.39635	2.01995
	7, *b*	*XIV - b*	2.36992	2.00176
	7, *c*	*XIV - c*	2.37746	2.01321

†*Variance components are given in* ***Table 5****. Those VCs were used as block effects and residual error for random terms to obtain the* $MVD_{(R)}$*, which were listed under the respective models in* ***Table 5****.*

‡*Model (5): IBD in P1 with the additional post-blocking factor column, IBD in P2; Model (6) a–c: IBD with the additional post-blocking factor column and "worker-day", where a to c indicate the different blocking strategies to account for the different numbers of position that can be visited by a worker on a day; Model (7) a–c: considering only the "worker-day" block in P1 and an IBD in P2.*

findings are applicable to other crops and two-phase experiment settings, especially with large treatment numbers in breeding. We used the OPTEX package of SAS, but other packages can be used as well.

Our results show that there is great potential for improving the two-phase design in *P. zonale* considering additional blocking factors such as "*worker-day*", using computer generated designs in both phases rather than conducting the randomization on-site, including equal block sizes in the second phase and extending the generation procedure across phases.

In detail, reductions in MVD were obtained by the use of additional block factors accounting better for environmental variation. For example, a reduction in $MVD_{(F)}$ from 9.42 to 4.41 or in $MVD_{(R)}$ from 2.67 to 2.52 was achieved by considering a column factor and the "*worker-day*" in P1 (**Table 5**). The MVD varied according to the chosen level of the "*worker-day*" factor to define the number of plants a worker may visit per day. For the *b* strategy, the smallest MVD was always obtained independently of the options, indicating that this strategy best represented a day of a worker (**Tables 5–7**).

Further differences in options were identified when different block structures in both phases were considered. In particular, the approach using a single pseudo level for incomplete blocks of P1 and P2, and different block structures in the two phases, realized always the smallest MVD.

The MVD as the Evaluation Criterion

For the interpretation of the reduction in MVD when comparing the alternative approaches with the two-phase design in 2013/14, the idealized conditions need to be acknowledged. The reduction in the number of incomplete blocks leads to an increased number of direct genotype comparisons within incomplete blocks which also reduces the MVD.

As expected, the $MVD_{(R)}$ was always smaller than the $MVD_{(F)}$ as the estimation of $MVD_{(R)}$ is based not only on the within-block genotype differences (i.e., intra-block information) to obtain adjusted means like for the $MVD_{(F)}$, but also on the information of block sums (i.e., the inter-block information). This stresses the importance of considering the joint inter-block–intra-block analysis (Mead et al., 2012), which can be implemented by taking blocks as random, during the design evaluation (Möhring et al., 2015). Note that, when instead of the VC of the former experiment, very large values are used for the variance of block effects, while leaving the value of residual error variance unchanged during the dummy analysis, i.e., there is no inter-block information, then values of $MVD_{(R)}$ and $MVD_{(F)}$ coincide. Further, the $MVD_{(R)}$ varies depending on the values of VCs for block effects considered in dummy analyses, but the ranks of scenarios remained unchanged in the cases we investigated (**Supplementary Presentation 4**).

The Need for Randomization

Randomization is conducted to avoid systematic effects and other biases in single-phase experiments (Piepho et al., 2013). In two phase experiments, these problems exist in both phases and therefore randomization should be carried out in both phases. In *Scenario I*, we omitted randomization in the second phase and increased thereby the efficiency of analysis, because only one incomplete block adjustment was needed to estimate the genotypes effects across the two phases [compare Model (2) and Model (4)]. That is why *Scenario I* showed the smallest MVD compared to the other designs assuming in each phase the same block structure. In conclusion, it is actually advantageous to transmit the experimental layout from one phase to the other whenever possible.

Best Options for Generating Two-Phase Designs and Application to Other Breeding Trials

Two-phase designs should be generated across phases (Option 2) rather than in phase-wise order (Option 1) to guarantee the smallest MVD, which was most frequently obtained for Option 2. The only exception was *Scenario I*. A reason for the better performance of Option 2 is that the block structure in both phases is taken into account simultaneously when sets of genotypes are allocated to them. Thus, treatment concurrences occur equally often or only once across phases in an optimized two-phase design (**Supplementary Presentation 4**), which is known to be optimal in single phase experiment (John and Williams, 1995).

Further, we demonstrated in the Scenarios *XIII* to *XIV* under Option 2 that our proposed approaches for generating two-phase designs across phases can be adjusted to any block size in each phase if necessary, making our approaches relevant to a broad range of applications. For the generation of a two-phase design with eight replicates (*Scenario III*) a similar $MVD_{(R)}$ or even a smaller $MVD_{(F)}$ was found than for the approach using a phase specific dummy variable (*Scenario IV*). But *Scenario III* was not further considered, as this approach restricts block structures to be the same in both phases. However, *Scenario III* represents an option for estimating of VCs for design effects in each phase is of interest when the structure in both phases is the same.

Examples for the application of our approaches in other ornamental species are the evaluation of rooting in P1 and other phenotypic traits in P2 in *Osteospermum* or the evaluation of germination rate and flowering time in the first and second phase in *Dianthus* ssp. (Selecta one). In the former example, in each phase the same block structure was considered, whereas in the latter example, in each phase a different block was assumed.

Consideration of Worker-Days as a Block Factor

The greatest reduction in MVD was obtained when we accounted for the worker-induced variation by blocks in *post hoc* analysis of the previous experiment by models (6), *a–c* to (7), *a–c* (**Table 5**), although the reduction of error variance was relatively small (**Table 4**). This shows that workers are a source of variation and reaffirms the recommendation that known sources of variation should be captured by blocking and considered before the experiment is conducted, as precision of genotype comparisons will be increased (Mead et al., 2012).

Idealized Conditions in Practice

The notable reduction in $MVD_{(F)}$ and $MVD_{(R)}$ realized by the alternative approaches justifies the implementation of idealized conditions in the *P. zonale* breeding program, especially the use of a pre-defined layout in the second phase. Under these idealized conditions, the breeder needs to randomly select six out of the total of harvested stem cuttings per pair of stock plants and genotypes in P1, which shall be rooted in P2. The procedure of packaging genotypes remains essentially the same as in the previous experiment, where the harvested stem cuttings of each genotype and replicate are packed in small bags such that each bag contained the six randomly selected stem cuttings from the EU1 in the first phase. However, the bags are now ordered according to the planting positions in P2 and then packed into cartons, where genotypes are grouped by replicate. In P2, an efficient workflow is ensured and hence, plant quality is maintained as workers plant genotypes onto trays according the planting number.

CONCLUSION

With respect to the considered options, our results show that two-phase designs should be generated across phases (Option 2) rather than in phase-wise order (Option 1) to guarantee the smallest MVD, which was obtained for Option 2 with different block structures in both phases and the approach using a single pseudo level for incomplete blocks in P1 and P2. Increase in efficiency can be expected when the experimental layout is transmitted from P1 to P2.

With our pragmatic approaches, we could improve the present two-phase design in *P. zonale* breeding, which yields a reduction in the MVD obtained by intra-block analysis from 9.42 to about 2.35 or obtained by combined inter-block–intra-block analysis from 2.67 to approximately 1.99 by using computer generated designs in both phases rather than conducting the randomization on-site, additional block factors in P1, and extending the generation procedure across phases. This significant reduction in MVD justifies the consideration of idealized conditions in *P. zonale* breeding and indicates that the on-site randomization approach is sub-optimal. The proposed alternative approaches can be transferred to other studies that involve two-phase experimental set-ups and they can be implemented in any model-based design package with facilities to freely formulate linear models for treatment and block structures.

AUTHOR CONTRIBUTIONS

HM and H-PP developed alternative methods generating the two-phase experimental designs. HM conducted *post hoc* analysis of the experiment 2013/14, the dummy analyses and prepared the manuscript. H-PP and RB revised the manuscript. All authors discussed the results, commented on the manuscript and approved the final manuscript.

FUNDING

This project was partly funded by grants KF3130602 (BMWi) and PI 377/19-1 (DFG).

ACKNOWLEDGMENT

The authors would like to thank the team of Klemm + Sohn GmbH & Co. KG who assisted in carrying out the experiment 2013/14.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2017.02194/full#supplementary-material

PRESENTATION 1 | Coder for generating two-phase designs.

PRESENTATION 2 | Figures of approaches implemented in Scenarios I to V.

PRESENTATION 3 | Codes for dummy analyses.

PRESENTATION 4 | Concurrence of two-phase designs.

REFERENCES

Brien, C. J. (2017). *Functions Useful in the Design and Anova of Experiments*. Available at: http://chris.brien.name/

Brien, C. J., and Demétrio, C. G. B. (2009). Formulating mixed models for experiments, including longitudinal experiments. *J. Agric. Biol. Environ. Stat.* 14, 253–280. doi: 10.1198/jabes.2009s.08001

Brien, C. J., Harch, B. D., Correll, R. L., and Bailey, R. A. (2011). Multiphase experiments with at least one later laboratory phase I orthogonal designs. *J. Agric. Biol. Environ. Stat.* 16, 422–450. doi: 10.1007/s13253-011-0060-z

Butler, D. G. (2013). *On the Optimal Design of Experiments Under the Linear Mixed Model*. Ph. D. thesis, The University of Queensland, St Lucia, QLD.

Coombes, N. E. (2009). *Digger: Design Generator Under Correlation and Blocking*. Available at: http://nswdpibiom.org/austatgen/software/

Curnow, R. N. (1959). The analysis of a two phase experiment. *Biometrics* 15, 60–73. doi: 10.2307/2527601

Federer, W. T., and Crossa, J. (2012). I.4 screening experimental designs for quantitative trait loci, association mapping, genotype-by-environment interaction, and other investigations. *Front. Physiol.* 3:156. doi: 10.3389/fphys.2012.00156

Hinkelmann, K., and Kempthorne, O. (2005). *Design and Analysis of Experiments: Advanced Experimental Design*, Vol. 2, New Jersey: NJ, John Wiley & Sons. doi: 10.1002/0471709948

Jarrett, R. G., and Ruggiero, K. (2008). Design and analysis of two-phase experiments for gene expression microarrays - Part I. *Biometrics* 64, 208–216. doi: 10.1111/j.1541-0420.2007.00835.x

John, J. A., and Williams, E. R. (1995). *Cyclic and Computer Generated Designs*, 2nd Edn. London: Chapman and Hall. doi: 10.1007/978-1-4899-7220-0

McIntyre, G. A. (1955). Design and analysis of two phase experiments. *Biometrics* 11, 822–828. doi: 10.2307/3001770

Mead, R. (1988). *The Design of Experiments. Statistical Principles for Practical Application*. New York, NY: Cambridge University Press.

Mead, R., Gilmour, S., and Mead, A. (2012). *Statistical Principles for the Design of Experiments: Application to Real Experiments*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139020879

Möhring, J., Williams, E., and Piepho, H. P. (2015). Inter-block information: to recover or not to recover it? *Theor. Appl. Genet*. 128, 1541–1554. doi: 10.1007/s00122-015-2530-0

Molenaar, H., Glawe, M., Boehm, R., and Piepho, H.-P. (2017). Selection for production-related traits in *Pelargonium zonale*: improved design and analysis make all the difference. *Hortic. Res.* 4:17004. doi: 10.1038/hortres.2017.4

Pereira, C. and Tobias, R. (2015). *Catering to Your Tastes?: Using PROC OPTEX to Design Custom Experiments, with Applications in Food Science and Field Trials*. Available at: http://support.sas.com/resources/papers/proceedings15/3148-2015.pdf

Piepho, H. P. (2015). Generating efficient designs for comparative experiments using the SAS procedure OPTEX. *Commun. Biom. Crop Sci.* 10, 96–114.

Piepho, H. P., Möhring, J., and Williams, E. R. (2013). Why randomize agricultural experiments? *J. Agron. Crop Sci*. 199, 374–383. doi: 10.1111/jac.12026

Smith, A. B., Butler, D. G., Cavanagh, C. R., and Cullis, B. R. (2014). Multi-phase variety trials using both composite and individual replicate samples: a model-based design approach. *J. Agric. Sci.* 153, 1017–1029. doi: 10.1017/S0021859614000707

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

4 Phenotypic selection in ornamental breeding: It's better to have the BLUPs than to have the BLUEs[3]

[3]This chapter is published as:
Molenaar H, Boehm R, Piepho H-P (2018). Phenotypic selection in ornamental breeding: it's better to have the BLUPs than to have the BLUEs. Frontiers in Plant Science, 9:1-14.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 05 November 2018
doi: 10.3389/fpls.2018.01511

Phenotypic Selection in Ornamental Breeding: It's Better to Have the BLUPs Than to Have the BLUEs

Heike Molenaar[1]*, *Robert Boehm*[2] and *Hans-Peter Piepho*[1]*

[1] Biostatistics Unit, Institute of Crop Science, University of Hohenheim, Stuttgart, Germany, [2] Klemm + Sohn GmbH & Co. KG, Stuttgart, Germany

Plant breeders always face the challenge to select the best individuals. Selection methods are required that maximize selection gain based on available data. When several crosses have been made, the BLUP procedure achieves this by combining phenotypic data with information on pedigree relationships via an index, known as family-index selection. The index, estimated based on the intra-class correlation coefficient, exploits the relationship among individuals within a family relative to other families in the population. An intra-class correlation coefficient of one indicates that the individual performance can be fully explained based on the family background, whereas an intra-class correlation coefficient of zero indicates the performance of individuals is independent of the family background. In the case the intra-class correlation coefficient is one, family-index selection is considered. In the case the intra-class correlation coefficient is zero, individual selection is considered. The main difference between individual and family-index selection lies in the adjustment in estimating the individual's effect depending on the intra-class correlation coefficient afforded by the latter. Two examples serve to illustrate the application of the BLUP method. The efficiency of individual and family-index selection was evaluated in terms of the heritability obtained from linear mixed models implementing the selection methods by suitably defining the treatment factor as the sum of individual and family effect. Family-index selection was found to be at least as efficient as individual selection in *Dianthus caryophyllus* L., except for flower size in standard carnation and vase life in mini carnation for which traits family-index selection outperformed individual selection. Family-index selection was superior to individual selection in *Pelargonium zonale* in cases when the heritability was low. Hence, the pedigree-based BLUP procedure can enhance selection efficiency in production-related traits in *P. zonale* or shelf-life related in *D. caryophyllus* L.

Keywords: BLUP, BLUE, two-phase design, phenotypic selection, family-index selection, individual selection, ornamental breeding

OPEN ACCESS

Edited by:
Laurent Gentzbittel, National Polytechnic Institute of Toulouse, France

Reviewed by:
Hiroyoshi Iwata, The University of Tokyo, Japan
Bernardo Ordas, Consejo Superior de Investigaciones Científicas (CSIC), Spain

***Correspondence:**
Heike Molenaar
Heike.Molenaar@uni-hohenheim.de
Hans-Peter Piepho
Piepho@uni-hohenheim.de

Specialty section:
This article was submitted to Plant Breeding, a section of the journal Frontiers in Plant Science

Received: *13 April 2018*
Accepted: *26 September 2018*
Published: *05 November 2018*

Citation:
Molenaar H, Boehm R and Piepho H-P (2018) Phenotypic Selection in Ornamental Breeding: It's Better to Have the BLUPs Than to Have the BLUEs. Front. Plant Sci. 9:1511. doi: 10.3389/fpls.2018.01511

INTRODUCTION

For decades "Best Linear Unbiased Prediction" (BLUP) has been the standard selection method in animal breeding (Henderson, 1950), where the breeding values of sires are estimated based on progeny performance to select superior genotypes and to breed superior families (Robinson, 1991). More recently, this method has been used in commodity crops (Piepho et al., 2008)

and has also been applied in several clonally propagated species such as sweet potato (Borges et al., 2010), acai berry (Teixeira et al., 2012), potato (Slater et al., 2014; Ticona-Benavente and da Silva Filho, 2015), sugar cane (Barbosa et al., 2005; Zeni Neto et al., 2013), and passion fruit (Santos et al., 2015). Currently, the pedigree-based BLUP method is replaced by genomic prediction in many species (Gianola et al., 2018). In comparison to the pedigree-based BLUP, genomic prediction uses a marker-based matrix of genomic pair-wise similarities known as "genomic relationship matrix" (Van Raden, 2008; Legarra, 2016; Wang et al., 2017). Furthermore, the pedigree-based genetic variance-covariance matrix is replaced by the genomic variance (Lehermeier et al., 2017). However, marker data are severely limited in ornamental breeding programs. Thus, the pedigree-based BLUP method proposed in the present study is currently the most promising selection method to use when no marker-data is available. By this method, useful information can be obtained as to whether the trait is dependent or independent of the family background. This information is vital for selecting individuals for genotyping, because the goal of creating diversity panels is to represent the entire genetic diversity of parental populations, i.e., individuals should be selected with similar biotic or abiotic adaptation or photoperiod requirements (Singh and Singh, 2015, p. 220).

Before BLUP-based selection, selection in crop breeding was based on either simple arithmetic means or "Best Linear Unbiased Estimation" (BLUE) of genotypes, which are calculated in a mixed model context based on fixed genotype effects (Piepho et al., 2008). By contrast, BLUPs are obtained by defining the genotypes as random effects. By convention, "estimation" refers to fixed effects and "prediction" refers to random effects, even though both refer to estimators of effects in a linear mixed model. The first three letters of the acronyms BLUE and BLUP stand for *Best*, meaning they have the lowest variance, *Linear*, meaning they are linear functions of the data, and *Unbiased*. In case of BLUE, unbiased means the expected value of a mean estimate for an individual equals its true value. This is a conditional mean. By contrast, in case of BLUP the expected mean over all individuals is equal to the expected mean over all true effects. This is a marginal mean. The BLUP-based selection method predicts genetic effects more accurately than the BLUE-based method (Copas, 1983; Robinson, 1991). The gain in accuracy compared to BLUE-based selection results partly from the shrinkage property (Piepho et al., 2008), i.e., above average individual means will be shrunken downwards toward the overall mean, whereas below average individual means will be shrunken upwards toward the overall mean. The degree of shrinkage also depends on environmental variation (Hill and Rosenberger, 1985). This shrinkage property anticipates the regression to the mean observed in the selected progeny and is advantageous for selection decisions because individuals with extreme high or low performances are adjusted, which is consistent with the need for caution in making selection decisions on such extremes (Hill and Rosenberger, 1985). A further source of gain in accuracy is the facility to borrow strength from individuals in the same family (Piepho et al., 2008; Bernardo, 2010).

Currently, selection in ornamentals (Boxriker et al., 2017a,b; Molenaar et al., 2017) is based on individual performance, which is known to be a poor strategy when heritability is low. Alternatively, response to selection could be improved by considering family information. The simplest form of selection considering pedigree information is family selection, where selection is based on family means (Lynch and Walsh, 2013). A refinement of family selection is family-index selection (Lush, 1947), which incorporates the individual mean with the family mean (Lynch and Walsh, 2013). Generally, the exploitation of family information can provide greater accuracy and larger response to selection. In particular, index selection has an expected response at least as large as individual selection and even higher responses when significant effects of environmental conditions and replication of families over environments exist (Lynch and Walsh, 2013).

To our knowledge, the BLUP-based selection method has been used only in a few ornamental species so far. Huang et al. (1995) used the BLUP-based selection method to investigate the long-term genetic improvement in 16 generations of gerbera cut-flowers. In the past, software restrictions precluded directly obtaining BLUPs from the so-called "Mixed Model Equations" (MME; Henderson, 1950). Instead, facing computational constraints, Huang et al. (1995) obtained BLUPs by an indirect approach of successive averaging of genotypic effects (Misztal and Gianola, 1987) and the variance components were estimated by the derivative-free restricted maximum likelihood (Graser et al., 1987). Fogaça et al. (2012) used BLUP in daylily breeding and found that higher selection gain is expected from family selection rather than from individual selection. The BLUPs of individuals were obtained by the use of SELEGEN-REML/BLUP software (Resende, 2016).

The present work aims to demonstrate the application of BLUP-based selection in *Pelargonium zonale* and *Dianthus caryophyllus* L., two species which have the highest economic importance in the floricultural industry, and to further demonstrate the enhancement of breeding efficiency. We will briefly review the theoretical underpinnings of BLUP and individual and family-index selection. Then we compare the efficiency of strategies underlying the individual and family-index selection in terms of heritability.

MATERIALS AND METHODS

Theoretical Underpinnings of BLUE and BLUP

The context of BLUE and BLUP is the standard linear mixed model (LMM; Robinson, 1991; Piepho, 1994),

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{u} + \boldsymbol{e},$$

Abbreviations: BLUP, best linear unbiased prediction; BLUE, best linear unbiased estimation; LMM, linear mixed model; SCC, stem cutting count; CS, compound symmetry; EU1, the experimental unit in P1; EU2, the experimental unit in P2; P1, phase one; P2, phase two; RF, root formation; VL, vase life; FS, flower size; BN, bud number; SL, stem length.

where $\boldsymbol{y}$ is a vector of n observations, $\boldsymbol{\beta}$ is a vector of fixed effects, $\boldsymbol{X}$ and $\boldsymbol{Z}$ are design matrices associated with the fixed and random effects, $\boldsymbol{u}$, the vector of random effects assumed to be distributed according to $\boldsymbol{u} \sim \text{MVN}(\boldsymbol{0}, \boldsymbol{G})$ where $\boldsymbol{0}$ is a null vector and $\boldsymbol{G}$ is the variance-covariance matrix of the random effects, and $\boldsymbol{e}$ is the vector of residual errors assumed to be distributed as $\boldsymbol{e} \sim \text{MVN}(\boldsymbol{0}, \boldsymbol{R})$ with $\boldsymbol{R}$ the variance-covariance matrix of the residual errors. The distribution of observed data is assumed to be $\boldsymbol{y} \sim \text{MVN}(\boldsymbol{X\beta}, \boldsymbol{V})$, where $\boldsymbol{V}$ accounts for random effects and residual error by $\boldsymbol{V} = \boldsymbol{ZGZ}^{\mathbf{T}} + \boldsymbol{R}$.

The fixed effects (BLUEs) are estimated by $\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}^T\hat{\boldsymbol{V}}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\hat{\boldsymbol{V}}^{-1}\boldsymbol{y}$, where as the random effects (BLUPs) are predicted by $\hat{\boldsymbol{u}} = \hat{\boldsymbol{G}}\boldsymbol{Z}^T\hat{\boldsymbol{V}}^{-1}\left(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\right)$.

The BLUE and BLUP of $\boldsymbol{\beta}$ and $\boldsymbol{u}$, respectively, are best computed by solving the MME, given by (Henderson, 1950; Searle et al., 1992),

$$\begin{bmatrix} \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{G}^{-1} \end{bmatrix}\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{u}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{y} \\ \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{y} \end{bmatrix},$$

where $\boldsymbol{G}^{-1}$ and $\boldsymbol{R}^{-1}$ are the inverses of $\boldsymbol{G}$ and $\boldsymbol{R}$, respectively.

When $\boldsymbol{G}^{-1}$ tends to a zero matrix, which happens when variances in $\boldsymbol{G}$ become very large, the random effect estimates behave essentially like fixed effect estimates because the MME tend to

$$\begin{bmatrix} \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z} \end{bmatrix}\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{u}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{y} \\ \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{y} \end{bmatrix}.$$

If furthermore the residual errors are independent with homogenous variance, i.e., $\boldsymbol{R}^{-1} = \sigma^{-2}\boldsymbol{I}$, with σ^{-2} the inverse residual error variance and $\boldsymbol{I}$ an identity matrix, the MME turn into the ordinary least squares equations (Robinson, 1991),

$$\begin{bmatrix} \boldsymbol{X}^T\boldsymbol{X} & \boldsymbol{X}^T\boldsymbol{Z} \\ \boldsymbol{Z}^T\boldsymbol{X} & \boldsymbol{Z}^T\boldsymbol{Z} \end{bmatrix}\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{u}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}^T\boldsymbol{y} \\ \boldsymbol{Z}^T\boldsymbol{y} \end{bmatrix}.$$

Family-Index Selection

The basic idea of family-index selection is to obtain an index that accounts for the resemblance among individuals within a family relative to other families in the population (Lush, 1947). To depict Lush's idea, we give a selection problem in the context of ornamental breeding (modified according to Lush, 1947, pp. 242–244): Four families are considered to select the best performing individuals with respect to stem cutting count (SCC; **Figure 1**). Different strategies could be taken for selection. Individuals could be selected independently of the performance of their sibs (highest SCC). This method is known as individual selection. An alternative is for the breeder to select a complete family on the basis of family means (family selection). In the example, "Family 2" would be selected showing the highest SCC performance. Combining these two selection methods by considering both individual performances and family means in an index (family-index selection), the breeder would select "F" and "P" rather than "D" and "L." Independently of the selection method, individuals "G" and "H" will always be selected, because the family average can be high only when more than a substantial proportion of the individuals in a family are above the general population mean (Lush, 1947). Furthermore, it will almost never happen that all individuals of the superior family are superior to all members of other families (Lush, 1947). The main difference between individual, family and family-index selection consists in what is done with good individuals from mediocre families (like "D" and "P") and with intermediate individuals (like "F") or poor individuals (like "E") from better performing families (Lush, 1947), which is illustrated in the following with a particular emphasis on the partition of variance.

Continuing with the motivating example, it is assumed that for each individual "A" to "P" two observations are available and the design was completely randomized. Selection can based on the LMM

$$\boldsymbol{y} = \boldsymbol{1}_n\mu + \boldsymbol{Z}_g\boldsymbol{g} + \boldsymbol{e},$$

where $\boldsymbol{y}$ is the $(n \times 1)$ vector of SCC observations, $\boldsymbol{1}_n\mu$ is the $(n \times 1)$ vector of ones allocating the general population mean to all observations, $\boldsymbol{g}$ is the $(s \times 1)$ vector of random genetic strain effects and distributed as $\text{N}\left(\boldsymbol{0}, \sigma_g^2\boldsymbol{I}\right)$ with the genetic strain variance σ_g^2 and $\boldsymbol{I}$ the $(s \times s)$ identity matrix, $\boldsymbol{Z}_g$ is the $(n \times s)$ design matrix of random strain effects relating observations to strains and the random $(n \times 1)$ vector $\boldsymbol{e}$ distributed $N\left(\boldsymbol{0}, \sigma_e^2\boldsymbol{I}\right)$ with the non-genetic σ_e^2 variance and $\boldsymbol{I}$ the $(n \times n)$ identity matrix. Given this baseline model, the phenotypic variance is $\boldsymbol{V} = \boldsymbol{Z}_g\boldsymbol{G}\boldsymbol{Z}_g^T + \boldsymbol{R}$, where $\boldsymbol{G} = \sigma_g^2\boldsymbol{I}_{s \times s}$ and $\boldsymbol{R} = \sigma_e^2\boldsymbol{I}_{n \times n}$.

To account for the *simple* nested family structure (Piepho and Williams, 2006), i.e., for families and individuals that can be grouped by family, the genetic effect of the baseline LMM is partitioned as $\boldsymbol{g} = \boldsymbol{Z}_f\boldsymbol{f} + \boldsymbol{m}$, so that the LMM becomes

$$\boldsymbol{y} = \boldsymbol{1}_n\mu + \boldsymbol{Z}_g\boldsymbol{Z}_f\boldsymbol{f} + \boldsymbol{Z}_g\boldsymbol{m} + \boldsymbol{e},$$

where $\boldsymbol{f}$ is the $(w \times 1)$ vector of random family effects assumed to be $\text{N}\left(\boldsymbol{0}, \sigma_f^2\boldsymbol{I}\right)$, $\boldsymbol{Z}_f$ is the $(s \times w)$ design matrix of the random family effects, $\boldsymbol{m}$ is the $(s \times 1)$ vector of random effects of individuals nested within family effects assumed to be $\text{N}\left(\boldsymbol{0}, \sigma_s^2\boldsymbol{I}\right)$, and the residual term $\boldsymbol{e}$ is defined as in the baseline model. The resemblance among individuals of each family is given by the intra-class correlation coefficient, t,

$$t = \frac{\sigma_f^2}{\sigma_g^2}, \text{ where the total genetic variance is } \sigma_g^2 = \sigma_f^2 + \sigma_s^2.$$

On account of the intra-class correlation coefficient, for individuals in the same family the zeros on the off-diagonals of the variance-covariance matrix $\boldsymbol{G}$ under the baseline model are replaced with the family variance, resulting in a block diagonal $\boldsymbol{G}$ matrix with blocks corresponding to families. This structure is also known as the compound symmetry (CS) variance-covariance structure. For a single family with four individuals,

Family	INDIVIDUAL															FAMILY AVERAGE
1					A	B		C				D				8.75
2					E					F		G		H		10.25
3	I	J	K												L	4.75
4								M	N	O	P					9.5
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
	STEM CUTTING COUNT															

FIGURE 1 | A motivating example to illustrate the individual and family-index selection modified according to (Lush, 1947, pp. 242–244): Stem cutting counts of individuals from four families. The breeder is faced with the choice, e.g., between individuals "O" and "F," with the same SCC but coming from differently performing families.

the block on the diagonal of $\boldsymbol{G}$ is

$$\sigma_g^2 \boldsymbol{I}_{4\times 4} = \sigma_g^2 \begin{pmatrix} 1 & t & t & t \\ t & 1 & t & t \\ t & t & 1 & t \\ t & t & t & 1 \end{pmatrix} = \begin{pmatrix} \sigma_g^2 & \sigma_f^2 & \sigma_f^2 & \sigma_f^2 \\ \sigma_f^2 & \sigma_g^2 & \sigma_f^2 & \sigma_f^2 \\ \sigma_f^2 & \sigma_f^2 & \sigma_g^2 & \sigma_f^2 \\ \sigma_f^2 & \sigma_f^2 & \sigma_f^2 & \sigma_g^2 \end{pmatrix}.$$

The variance-covariance structure $\boldsymbol{R}$ remains unchanged. The motivating example assumes equal family sizes, which is an idealized condition. Unequal family sizes can be accounted for in BLUP (**Appendix Presentation 1** in Supplementary Material).

Individual Selection

Under a CS variance-covariance structure of $\boldsymbol{G}$, the off-diagonal elements describe the similarities of individuals within families. If the intra-class correlation coefficient, t, tends to one, the one limiting case, all individuals within the same family show about the same performance of a trait. In contrast, if the intra-class correlation coefficient, t, tends to zero, the other limiting, the performance of individuals is independent of the family background, which is exploited by individual selection.

Efficiency of Selection Methods

As shown, the degree of resemblance between individuals grouped by family can be measured by the genetic variance. The broad-sense heritability is given by

$$H^2 = \frac{\sigma_g^2}{\sigma_p^2}, \qquad \text{(Bernardo, 2010, p. 135)}$$

where σ_g^2 is the genetic variance, and σ_p^2 the phenotypic variance. This broad-sense heritability is sometimes also termed as *repeatability* (Piepho and Möhring, 2007) and is used to evaluate trials; the better a trial, the higher H^2. Similarly, H^2 can be used to determine the best breeding method (Falconer and Mackay, 1996). However, the estimation of H^2 using standard equations such as the one given above underlies strong assumptions: balanced data, non-correlated and homoscedastic genetic effects. If one of these assumptions is violated, there will not be a simple linear relationship between response to selection and selection differential, i.e., the correlation between phenotypic value and response to selection differs between genotypes (Piepho and Möhring, 2007). Different approaches (Holland et al., 2003; Cullis et al., 2006; Oakey et al., 2006; Piepho and Möhring, 2007) have been proposed for situations in which these standard assumptions are invalid.

Random and Independent Genotypes

Cullis et al. (2006) proposed to estimate the heritability when the genotypes are taken as random effects by

$$H_C^2 = 1 - \frac{\overline{v}}{2\sigma_g^2},$$

where $\overline{v}$ is the mean variance of a difference of two BLUPs and σ_g^2 is the genotypic variance. The heritability H_C^2 accounts for unbalanced data, but still genotypic effects are assumed to be independent, which would be true, if no relationship among individuals existed. However, when selection is exercised on individuals from different families, resemblance between individuals within families is present and alternative methods should be used (Oakey et al., 2006; Piepho and Möhring, 2007). The most flexible option to account for any modeled variance-covariance structure is to simulate the heritability on an entry-mean basis directly as the squared correlation of $\boldsymbol{g}$ and $\hat{\boldsymbol{g}}$ given by

$$r^2 = Q^{-1} \sum_{q=1}^{Q} r_q^2,$$

where Q is the total number of simulation runs, r_q^2 the sample correlation of the genotypic effects, g_i, and the BLUPs, $\hat{g}_i$ (Piepho and Möhring, 2007). In cases of independent genotypic effects and a balanced design, the expected squared correlation between true and predicted genotypic effects is approximately

$$\mathrm{E}\left(r^2\right) \approx \left(\frac{\mathrm{cov}\left(g_i, \hat{g}_i\right)}{\sqrt{\mathrm{var}\left(g_i\right) \mathrm{var}\left(\hat{g}_i\right)}} \right)^2,$$

where $\hat{\boldsymbol{g}} = \text{BLUP}(\boldsymbol{g})$.

The Monte Carlo standard error of the simulated heritability can be defined as

$$s.e.(r^2) = \sqrt{\frac{s^2_{r^2}}{Q}},$$

where $s^2_{r^2}$ is defined as

$$s^2_{r^2} = \frac{\sum_{q=1}^{Q}\left(r_q^2 - \bar{r}^2\right)^2}{Q-1}.$$

Selection Based on BLUPs

The use of BLUP requires the assumption of normality (Robinson, 1991), which can be graphically checked by Q-Q plots. In Q-Q plots the standardized BLUPs (Searle et al., 1992, pp. 286-287) were plotted against the normal scores in Q-Q plots. The standardized BLUPs were defined as $\frac{\hat{g}_i}{\sqrt{\text{Var}[\hat{g}_i]}}$, where $\hat{g}_i$ is the i-th estimated genotypic BLUP and var $\left[\hat{g}_i\right]$ is the unconditional variance.

Application of Family-Index Selection

In the next two sections, the application of family-index selection is illustrated by two examples in ornamental breeding. The general approach to implement the family-index selection by a LMM to select individuals across families is to define the treatment effect as the sum of the family effect and individual within family effect (FM + FM·ENTRY). Both effects are modeled as random and the genetic covariance, between individuals within a family is equal to the variance of the family effect, i.e., var(FM), whereas the covariance of individuals from different families is zero. To implement the model in a way that facilitates estimation of the genotypic value of individuals, we drop the family main effect FM and impose a CS variance-covariance structure on the FM·ENTRY effect for individuals in the same family. This implementation is equivalent to the model with independent effects FM and FM·ENTRY, but is more convenient for predicting the family-index, which may be obtained directly as the BLUP of the effect FM·ENTRY under the CS model (Piepho and Williams, 2006). For the implementation of individual selection by a LMM, only the independent term FM·ENTRY is considered. In sections Determining the best selection method by H^2 and Determining the best selection method by H^2, in which models are derived to simulate H^2 for evaluation of selection methods, it will be explicitly mentioned again which terms are crucial for the implementation of either individual or family-index selection.

Phenotypic Selection in Ornamentals: The Example of Production-Related Traits in *P. zonale* Breeding

Conventionally in the *P. zonale* breeding program of Selecta One (Stuttgart-Mühlhausen, Germany), seeds from crosses made in the first year are sown in the second year. Seedlings are selected with a focus on traits such as early flowering or petal color to reduce the population size for later tests (Figure 1 in Molenaar et al., 2017). Each selected individual is cloned (multiplied by cutting propagation) to enable replicated field trials in the third year focusing on color and flower longevity under field conditions for example. Finally, in the fourth year, candidate varieties are screened for production-related traits.

Due to recent advances in knowledge of the genetics of production-related traits (Molenaar et al., 2017), the assessment of production-related traits in 500 *P. zonale* strains has been shifted from the fourth to the second year in a new experiment, because of the great economic relevance of those traits in the breeding program (Molenaar et al., 2017). However, due to lack of time in the second year for clonal reproduction, this shift results in phenotyping of single plants for production-related traits in year two.

TABLE 1 | Parentage and size of the six *P. zonale* families evaluated in this study.

Family	Number of individuals in each family	Number of individuals in each reciprocal	Parental genotypes†		
			Paternal		Maternal
1	113	63	(A	×	b)
		50	(b	×	A)
2	51	3	(C	×	d)
		48	(d	×	C)
3	112	49	(E	×	f)
		63	(f	×	E)
4	60	26	(E	×	g)
		36	(g	×	E)
5	101	8	(E	×	d)
		91	(d	×	E)
6	63	43	(I	×	j)
		20	(j	×	I)
K_{Total}	500				

†*Uppercase letters indicate a superiority in production-related traits of the parental genotype.*

Plant Material

In 2014, twelve reciprocal crosses were made between ten heterozygous elite *P. zonale* strains to obtain six families segregating in the F_1 already (**Table 1**). Families were unrelated by pedigree, except for Families 3, 4, and 5, which all had the parent "E" in common and Families 2 and 5, which had parent "d" in common; individuals across these families were half-sibs (**Table 1**). The parental strains showed either a superior performance in production-related traits (indicated by capital letter), such as SCC or root formation (RF), or in quality traits, such as petal or leaf color. Between 10 and 113 individuals were obtained per family, amounting to 500 individuals across families.

Two-Phase Experimental Design in *P. zonale* Breeding

In 2015, the two-phase experimental design was modified to assess individuals without replication, where the phases were

as follows: Phase 1 (P1), the cultivation of stock plants of the seedling generation to obtain the SCC and Phase 2 (P2), the RF of stem cuttings (**Figure 2**). Each phase took place in a different greenhouse, but at the same location. The experimental design within each phase was an augmented design (Federer, 1956), in which the parental strains were tested with replications in incomplete blocks. The unreplicated individuals were randomly allocated to incomplete blocks. As will be described in more detail below, in P1, the experimental layout was generated by the SAS procedure OPTEX (SAS Institute Inc., 2014), whereas in P2 the randomization was carried out in the greenhouse on-site, because of biological matters of plant material.

Phase 1—augmented design −500 genotypes, 32 blocks, 8 checks

Using the OPTEX procedure (Piepho, 2015), an augmented design was generated for $c = 8$ checks, $v = 500$ genotypes in $b = 32$ blocks each of size 18 and a total 576 plots overall (**Figure 2**). Eight out of the ten parental genotypes were used as checks, except the parental strains "I" and "J." The blocks were laid out on two cultivation tables, each comprising 16 blocks, leaving 38 free plots (experimental units; EU1) for checks in addition to the 250 plots for unreplicated entries. Since 38 is not a multiple of eight, there was space to replicate checks either nine or ten times. Because of lack of cuttings, checks could not be included in the experiment in the first phase, meaning that the plots intended for checks were empty in P1.

Phase 2—augmented design generation −500 genotypes, 32 blocks, 8 checks

To conduct the randomization on-site, four rooting tables were divided into 32 regions each corresponding to a single block in P1 (**Figure 2**; the numbers within regions in P2 correspond to block numbers in P1). Each region contained a variable number of trays depending on the obtained SCC of an individual in P1. One tray contained 39 paper pots arranged in three rows, each with 13 paper pots. The trays were divided into areas, the experimental unit in P2 (EU2). The randomization on-site was as follows: First, the checks were randomly allocated to regions and to areas within regions. Within the second step, the individuals were randomly allocated to the remaining areas within regions. Blocks of P1 were packaged as a single unit for transferal from P1 to P2. Note that all trays of a region fit on the same rooting table. The areas were filled in row-wise order on a tray and one area was planted directly following the previous, subject to the restriction that all paper pots for an area were on the same tray. The size of areas (EU2) varied depending on the SCC of an individual on an EU1.

Production-Related Traits

The SCC was assessed as the number of stem cuttings per single stock plant and genotype (either check or test individual) (EU1) in P1. The RF of stem cuttings of a single stock plant and genotype (either checks or test individual) was scored after four weeks of rooting. The number of plants in categories S0 (dead) to S5 (extraordinary) for each area (EU2) was counted (Molenaar et al., 2017). From these counts we computed the sum of rooted cuttings assigned to classes S4 and S5 so that a single response value was obtained per area (EU2).

Determining the Best Selection Method by H^2

H^2 as described above was used to evaluate selection method for SCC and RF. The different selection methods were reflected by different LMM. The model for individual selection for SCC in the first phase, in symbolic form (Piepho et al., 2003; Piepho and Eckl, 2014), was

$$\mathrm{HR} : \mathrm{FM} \cdot \mathrm{ENTRY} + \mathrm{HR} \cdot \mathrm{WD} + \mathrm{HR} \cdot \mathrm{BLK} + \underline{\mathrm{HR} \cdot \mathrm{BLK} \cdot \mathrm{PLT}}, \tag{1}$$

where HR denotes the harvests, FM·ENTRY the individuals nested within families, HR·BLK the incomplete blocks nested within harvests, HR·WD the *post-blocking* factor "*worker-day*" nested within harvests, and HR·BLK·PLT, the residual error and experimental unit (plot = PLT) in P1 (EU1). The factor "*worker-day*" was defined to capture variation induced by working assistance of different people during the harvest of stem cuttings (Molenaar et al., 2018). Because of the unreplicated design, the estimation of the individuals within families-by-harvest interaction effect could not be achieved.

The model for family-index selection for SCC in P1, was an extension of model (1):

$$\begin{aligned} \mathrm{HR} \;:\; & \mathrm{FM} + \mathrm{FM} \cdot \mathrm{ENTRY} + \mathrm{HR} \cdot \mathrm{FM} + \mathrm{HR} \cdot \mathrm{WD} \\ & + \mathrm{HR} \cdot \mathrm{BLK} + \underline{\mathrm{HR} \cdot \mathrm{BLK} \cdot \mathrm{PLT}}, \end{aligned} \tag{2}$$

where FM denotes the families and HR·FM the family-by-harvest interaction. As explained above, family-index selection was implemented by fitting a CS variance-covariance structure for the sum of FM and FM·ENTRY random effects.

To evaluate the selection methods for RF in the second phase, some amendments to model (1) and (2) were necessary to account for checks, which were included in P2. Assuming individual selection, model (1) was changed to

$$\begin{aligned} \mathrm{HR} \;+\; & \mathrm{CK} + \mathrm{HR} \cdot \mathrm{CK} : \mathrm{PT} \cdot \mathrm{FM} \cdot \mathrm{ENTRY} + \mathrm{PT} \cdot \mathrm{HR} \cdot \mathrm{FM} \cdot \mathrm{ENTRY} \\ & + \mathrm{HR} \cdot \mathrm{WD} + \mathrm{HR} \cdot \mathrm{BLK} + \underline{\mathrm{HR} \cdot \mathrm{BLK} \cdot \mathrm{PLT}} \end{aligned} \tag{3}$$

where CK is a factor for checks, comprising nine levels, i.e., eight levels for the parental strains (checks) and one level for the expected value of all individuals to separate effects of checks from individuals (Piepho et al., 2006). Furthermore, to prevent random genetic effects from being fitted for checks, a dummy variable PT with PT = 0 for checks and PT = 1 for individuals was defined. The dummy variable PT was crossed with the family and individuals within family effect. Similarly, model (2) was expanded by the check factor CK and the PT dummy variable to account for family-index selection for RF in P2,

$$\begin{aligned} \mathrm{HR} \;+\; & \mathrm{CK} + \mathrm{HR} \cdot \mathrm{CK} : \mathrm{PT} \cdot \mathrm{FM} + \mathrm{PT} \cdot \mathrm{FM} \cdot \mathrm{ENTRY} + \mathrm{PT} \cdot \mathrm{HR} \cdot \mathrm{FM} \\ & + \mathrm{PT} \cdot \mathrm{HR} \cdot \mathrm{FM} \cdot \mathrm{ENTRY} + \mathrm{HR} \cdot \mathrm{WD} + \mathrm{HR} \cdot \mathrm{BLK} \\ & + \underline{\mathrm{HR} \cdot \mathrm{BLK} \cdot \mathrm{PLT}}. \end{aligned} \tag{4}$$

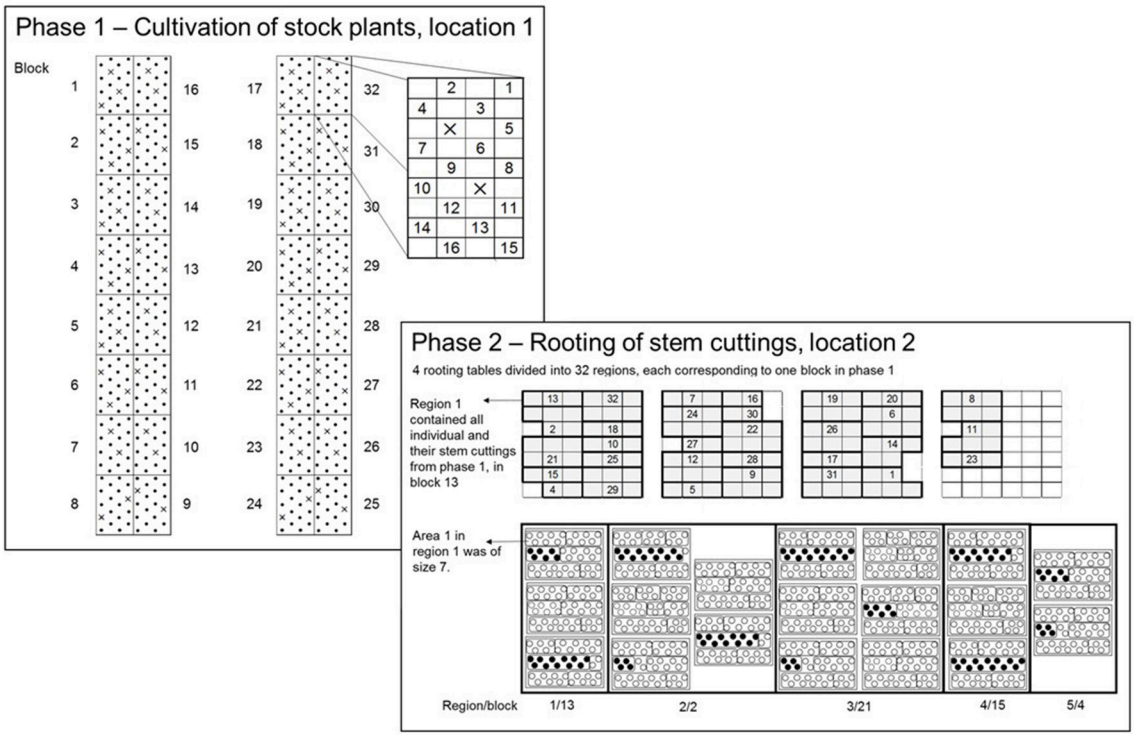

FIGURE 2 | In P1, an augmented-design for 500 individuals in 32 blocks with 8 checks was used. Each dot represents an experimental unit in P1 (EU1) to which unreplicated entries were randomly allocated. On each EU1 a single stock plant of an individual from the seedling generation was placed. The crosses within blocks 1 to 32 indicate the check plots. In P2, the total experimental space was represented by four rooting tables. The four rooting tables were divided into 32 regions corresponding to blocks in P1 to conduct the randomization on-site. The numbers within regions correspond to block numbers in P1. Each region contained a variable number of trays depending on the number of SCC per individual within a block in P1. One tray contained 39 paper pots arranged in three rows, each with 13 paper pots. The trays were divided into areas, the experimental unit in P2 (EU2). The size of areas varied depending on the numbers of stem cuttings of and individual on each EU1. The planting of stem cuttings followed in a row-wise order.

Family-index selection was implemented by fitting the CS variance-covariance structure to the sum of the PT · FM and PT · FM · ENTRY random effects.

Phenotypic Selection in Ornamentals: The Example of Vase Life Assessment and Related Traits in *D. caryophyllus* L. Breeding

This second example will illustrate the BLUP method for shelf-life and related traits in *D. caryophyllus* L., including vase life (VL) of cut flowers. The VL is one of the traits which most affects consumer satisfaction leading to repeated purchasing, and hence VL determines the economic value of a cultivar (Onozaki et al., 2001). Further, BLUP will be applied to floral traits such as flower size (FS) or number of buds (BN) and a morphology traits such as the stem length (SL). In 2016/2017 the entire seedling generation was cloned, so that each individual was tested by four replicates.

Plant Material

Five crosses were made between ten elite *D. caryophyllus* L. strains belonging either to the mini or the standard carnation type to obtain five families segregating in the F_1 already (**Table 2**). Families were assumed unrelated by pedigree. In total 176 individuals belonged to the standard type, and 328 individuals to the mini type, where three of the mini individuals were missing completely at random. The family sizes varied between 70 and 112 individuals.

TABLE 2 | Parentage and size of the three mini carnation type and two standard carnation type families in D. *caryophyllus* L. evaluated in this study.

Family	*k* individuals in each family	Parental genotypes			Type
		paternal		maternal	
1	106	(A	×	B)	*Mn*
2	110	(C	×	D)	*Mn*
3	112	(E	×	F)	*Mn*
1	106	(G	×	H)	*St*
2	70	(I	×	J)	*St*
K_{total}	504				

Two-Phase Experimental Design in *D. caryophyllus* Breeding

For the assessment of vase life and related traits, the seedlings were clonally propagated 1 year in advance so that each individual

was assessed by the use of four replications in the two-phase experimental set-up (**Figure 2**). In P1, the experimental layout was generated using CycDesigN 5.1 (VSN International, United Kingdom) and the experiment was conducted in the greenhouse, whereas in P2 the experiment was conducted in the lab, where the randomization was carried out on-site.

Phase 1 – Resolvable incomplete block design with four replicates each 61 incomplete blocks of size 9 for 504 individuals and 45 check genotypes

For each carnation type, a resolvable incomplete block design was generated by the use of CycDesigN 5.1 (VSN International, United Kingdom) with four replicates, each consisting 61 incomplete blocks of size nine. The incomplete blocks were represented by the physical units of subsurface boxes each consisting nine positions. The incomplete blocks of a replicate for both carnation types were jointly randomized, thus permitting a joint analysis of both types. Thus, a set of either nine standard or mini carnations was randomly allocated to each incomplete block. The randomization of genotypes was restricted in this way due to differences in cultivation minis with respect to flower bud removal. On each position of a subsurface box, i.e., on each experimental unit in P1 (EU1), a single stock plant of an individual was placed. Since 61 is not a multiple of 504, free positions were filled with up with another 45 check genotypes from another breeding program (**Figure 3**).

Phase 2 – randomized complete design for 504 individuals tested each with four replicates either treated or untreated

Because of biological matter (unpredictable development and maturity of flower buds of stock plants and individuals), a pre-defined design in the second phase was less suitable. That is why the total experimental space was divided up into 43 regions. Each region represented that day, on which a single stem of an individual and a replicate was placed in a vase. The vases were held by trays. Each tray comprised eight vases. A single vase represented the experimental unit in P2 (EU2). By the use of computer generated random numbers, first, the single stem of an individual and a replicate was randomly allocated to a tray within a region, and second, the stem was randomly allocated to a vase within a tray and region (EU2). The randomization of a stem was restricted, when the EU2 had been already filled with another individual's stem, in which case the stem was placed on the next empty EU2.

Interposed Transport Simulation Between the Two Phases

Two stems were harvested from a single stock plant and individual (EU1). A randomly chosen stem was assessed immediately after harvesting in the laboratory for VL, whereas the other stem was first submitted to transport simulation for 14 days (Boxriker et al., 2017b) and afterwards assessed for VL in the lab.

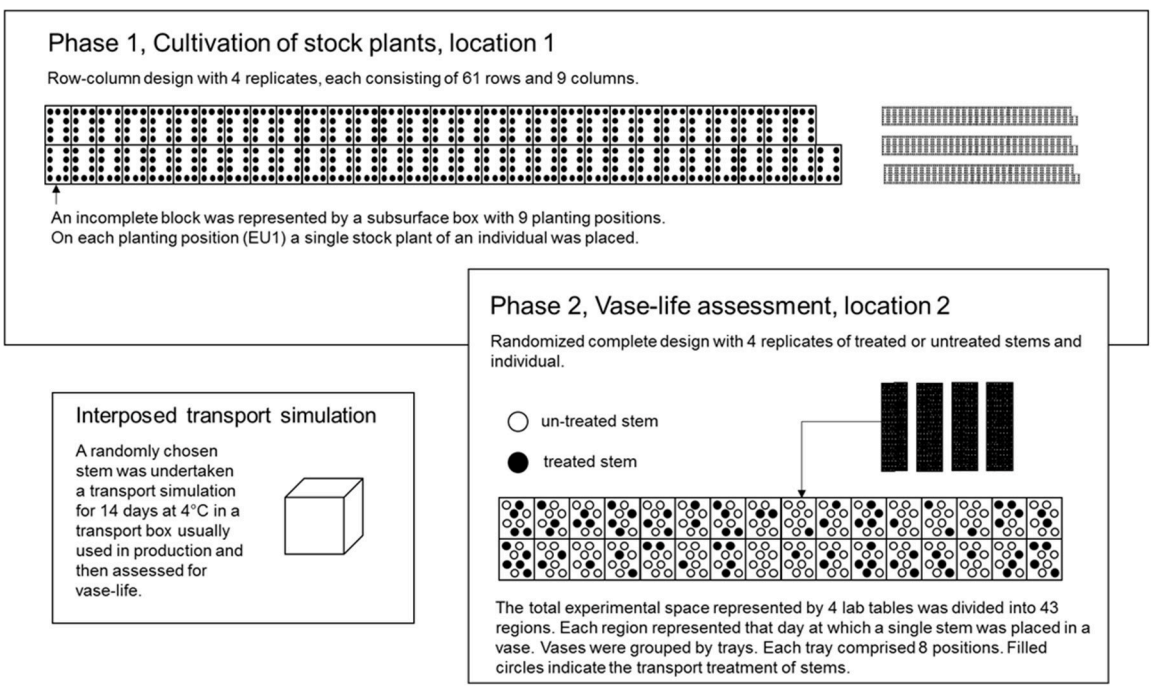

FIGURE 3 | In P1, an incomplete block design for 504 individuals and 45 check genotypes in 61 blocks of size nine was used. The incomplete blocks were represented by subsurface boxes. Each dot in an incomplete block represents a planting position and hence, an experimental unit in P1 (EU1). On each EU1 a single stock plant of an individual from the seedling generation was placed. In P2, the total experimental space was represented by four cultivation tables. The four cultivation tables were divided into up 43 regions. Each region corresponds to that day on which a single stem of an EU1 was placed in a vase, independently, of whether the stem was storage treated or not. Furthermore, each day comprised several trays, each consisting of eight vases. A single vase represented the EU2. On each EU2 a single stem from EU1 was placed.

Vase Life and Related Traits

The total number of buds (BN) was counted on a single stem from a single stock plant and individual (EU1) for the mini carnation type. The stem length (SL) was assessed as the total length in centimeters (cm) of a single stem harvested from a single stock plant and EU1. The flower size (FS) was assessed as the diameter measured in cm of a single stem harvested from a single stock plant and individual in a vase (EU2) for the standard carnation type. The FS was measured of flowers that reached the fourth floral development stage (Figure 1, p. 63 in Boxriker et al., 2017a). The VL was assessed as the flower duration in the vase (EU2) in days of two stems from a stock plant and individual of EU1. For this purpose, stems were harvested at the second floral development stage (Figure 1, p. 63 in Boxriker et al., 2017a). One stem was randomly chosen and assessed immediately after harvesting in the laboratory for VL, whereas the other stem was submitted first to transport simulation for 14 days and then assessed for VL.

Determining the Best Selection Method by H^2

On the basis of the full two-phase model, reduced models were defined to simulate H^2 for the traits BN and SL in P1, FS and VL in P2. Separate H^2 for traits SL and VL for the mini and standard carnation were simulated, because the two carnation types belong to different subspecies of *D. caryphyllus* L. with different characteristics. The H^2 assuming individual selection for SL either for mini or standard carnations was evaluated by the following model, listing fixed effects before the colon,

$$\begin{aligned}\text{REP} &+ \text{CK} + \text{STEM} + \text{TEMP} : \text{PT} \cdot \text{FM} \cdot \text{ENTRY} + \text{REP} \cdot \text{BLK} \\ &+ \underline{\text{REP} \cdot \text{BLK} \cdot \text{PLT}}, \end{aligned} \tag{5}$$

where REP denotes the replicates, CK a factor separating check genotypes, that were used to fill up empty positions in incomplete blocks and belonged either to mini or standard carnations from entries. Specifically, the CK factor comprised of 46 levels; 45 levels for the fillers and one single fixed effect to model the expected value of all individuals to separate effects of fillers from individuals. Furthermore, to prevent random effects from being fitted for fillers, a dummy variable PT with PT = 0 for fillers and PT = 1 for individuals was defined. TEMP was a covariate to account for the greenhouse temperature in P1, PT·FM·ENTRY denotes the individuals grouped by family effect, REP·BLK the incomplete blocks within replicates and the residual error <u>REP·BLK·PLT</u>. Expanding, model (5) by the term PT·FM, denoting the family effect, and implementing the CS structure for the sum of PT·FM and PT·FM·ENTRY random effects, family-index selection for SL was based on the model

$$\begin{aligned}\text{REP} &+ \text{CK} + \text{STEM} + \text{TEMP} : \text{PT} \cdot \text{FM} + \text{PT} \cdot \text{FM} \cdot \text{ENTRY} \\ &+ \text{REP} \cdot \text{BLK} + \underline{\text{REP} \cdot \text{BLK} \cdot \text{PLT}}. \end{aligned} \tag{6}$$

Individual selection for BN for the mini carnation, was considered by expanding model (5) with a *post-blocking* factor POS to account better for variation induced by drop inlets of the sub-surface boxes (Boxriker et al., 2017b),

$$\begin{aligned}\text{REP} &+ \text{CK} + \text{STEM} + \text{TEMP} : \text{PT} \cdot \text{FM} \cdot \text{ENTRY} + \text{REP} \cdot \text{BLK} \\ &+ \text{REP} \cdot \text{POS} + \underline{\text{REP} \cdot \text{BLK} \cdot \text{PLT}}. \end{aligned} \tag{7}$$

The factor POS had nine levels, each represented one planting position within a subsurface box (**Figure 2**).

The family-index selection, model (6) was expanded by the *post-blocking* factor POS,

$$\begin{aligned}\text{REP} &+ \text{CK} + \text{STEM} + \text{TEMP} : \text{PT} \cdot \text{FM} + \text{PT} \cdot \text{FM} \cdot \text{ENTRY} \\ &+ \text{REP} \cdot \text{BLK} + \text{REP} \cdot \text{POS} + \underline{\text{REP} \cdot \text{BLK} \cdot \text{PLT}}. \end{aligned} \tag{8}$$

For the analysis of BN, the logarithm of the count data was used.

Individual selection in FS of standard carnation was implemented by extending model (5) with the terms DAY, DAY·VSE and STO,

$$\begin{aligned}\text{REP} &+ \text{STO} + \text{CK} + \text{STO} \cdot \text{CK} + \text{TEMP} : \text{PT} \cdot \text{FM} \cdot \text{ENTRY} \\ &+ \text{PT} \cdot \text{STO} \cdot \text{FM} \cdot \text{ENTRY} + \text{REP} \cdot \text{BLK} + \text{DAY} + \text{DAY} \cdot \text{VSE} \\ &+ \underline{\text{REP} \cdot \text{BLK} \cdot \text{PLT}}, \end{aligned} \tag{9}$$

where STO denote the transport simulation (yes/no), STO·CK the check genotype-by-transport interaction, PT·STO·FM·ENTRY the individual-by-transport interaction, DAY the block when a single stem of an individual and position was and DAY·VSE the positional effect of a vase within a day. Family-index selection was implemented by adding the family effect and the family-by-transport interaction to model (9),

$$\begin{aligned}\text{REP} &+ \text{STO} + \text{CK} + \text{STO} \cdot \text{CK} + \text{TEMP} : \text{PT} \cdot \text{FM} \\ &+ \text{PT} \cdot \text{FM} \cdot \text{ENTRY} + \text{PT} \cdot \text{STO} \cdot \text{FM} + \text{PT} \cdot \text{STO} \cdot \text{FM} \cdot \text{ENTRY} \\ &+ \text{REP} \cdot \text{BLK} + \text{DAY} + \text{DAY} \cdot \text{VSE} + \underline{\text{REP} \cdot \text{BLK} \cdot \text{PLT}}. \end{aligned} \tag{10}$$

For the sum of random effects PT· FM and PT· FM· ENTRY of model (10) the CS structure was fitted.

Individual selection in VL for mini or standard carnation was performed by,

$$\begin{aligned}\text{REP} &+ \text{STO} + \text{CK} + \text{STO} \cdot \text{CK} + \text{TEMP} : \text{PT} \cdot \text{FM} \cdot \text{ENTRY} \\ &+ \text{PT} \cdot \text{STO} \cdot \text{FM} \cdot \text{ENTRY} + \text{REP} \cdot \text{BLK} + \text{REP} \cdot \text{POS} \\ &+ \text{DAY} + \underline{\text{REP} \cdot \text{BLK} \cdot \text{PLT}}, \end{aligned} \tag{11}$$

whereas family-index selection was implemented for by

$$\begin{aligned}\text{REP} &+ \text{STO} + \text{CK} + \text{STO} \cdot \text{CK} + \text{TEMP} : \text{PT} \cdot \text{FM} \\ &+ \text{PT} \cdot \text{FM} \cdot \text{ENTRY} + \text{PT} \cdot \text{STO} \cdot \text{FM} \\ &+ \text{PT} \cdot \text{STO} \cdot \text{FM} \cdot \text{ENTRY} + \text{REP} \cdot \text{BLK} + \text{REP} \cdot \text{POS} \\ &+ \text{DAY} + \underline{\text{REP} \cdot \text{BLK} \cdot \text{PLT}}, \end{aligned} \tag{12}$$

and the CS structure was fitted for the sum of random effects PT·FM and PT·FM·ENTRY.

TABLE 3 | Evaluation of selection methods in terms of simulated heritability and corresponding standard errors.

Species	Trait	CT†	Selection methods			
			Individual selection		Family-index selection	
			r^2	s.e.(r^2)	r^2	s.e.(r^2)
P.zonale	SCC		0.3376	0.003394	0.3718	0.003783
	RF		0.4112	0.004126	0.5601	0.005672
D. caryophyllus L.	BN	*Mn*	0.6904	0.006910	0.6970	0.006986
	SL	*Std*	0.9490	0.009491	0.9447	0.009449
		Mn	0.8911	0.008912	0.8913	0.008914
	FS	*Std*	0.6458	0.006473	0.6561	0.006583
	VL	*Std*	0.7118	0.007128	0.7128	0.007147
		Mn	0.7688	0.007692	0.7752	0.007817

† *Carnation type.*

TABLE 4 | Variance components of random effects obtained from model (1) and (2) to evaluate the individual and family-index selection for stem cutting count (SCC).

Model term	Variance	
	Model (1)	Model (2)
FM	–	0.2121
HR·FM	–	0.1481
FM·ENTRY	1.0835	0.8727
HR·BLK	2.3069	2.3397
HR·WD	0.7977	0.8456
HR·BLK·PLT	5.4417	5.3129

TABLE 5 | Variance component estimates of random effects obtained from model (3) and (4) to evaluate individual and family-index selection for root formation (RF).

Model term	Variance	
	Model (3)	Model (4)
FM	–	1.0818
HR·FM	–	0.2108
FM·ENTRY	1.9071	1.7697
HR·FM·ENTRY	4.1932	3.9096
HR·BLK	0.9126	0.9508
HR·WD	2.0586	0.9006
HR· BLK· PLT	2.9356	2.9598

RESULTS

In *P. zonale*, the highest H^2 was always found for family-index selection. In *D. caryophyllus* L. the H^2 was approximately the same for individual and family-index selection for BN, SL and VL in standard carnations. H^2 was greater for family-index selection for FS in standard carnation and VL in mini carnation (**Table 3**). The results were supported by the variance component estimates and box plots of BLUPs, which are described in detail below.

Variance Component Estimates From the *P. zonale* Breeding

The genotypic variance component estimate (FM·ENTRY) for both SCC and RF was relatively low in proportion to the total variation (**Tables 4**, **5**), which is also reflected by the shrinkage property of BLUPs toward the general mean (zero reference line in Presentation 2).

In P1, by far the largest variance component was the residual error variance, followed by the block variance for analyzing the SCC (**Table 4**). Similar variance components for the residual error variance and the block effects were calculated in a former experiment in 2013/14, although genotypes in that experiment were tested in four replications (Molenaar et al., 2017). Moreover, under the assumption of family-index selection, the effect for the family-by-harvest interaction (HR·FM) was found to be small, and the residual error variance was reduced. This suggested that the unaccounted for serial correlation on the same plots within blocks and harvest might have inflated the small genotypic and the residual error variance. The worker-induced (HR·WD) and the genotypic (FM·ENTRY) variances were of comparable size, indicating the considerable effect of the person carrying out the assessment of production-related traits (Molenaar et al., 2017).

In contrast to P1, in P2 the largest variance component was calculated for the individual within family-by-harvest interaction effect (HR·FM·ENTRY) and was greater than the residual error variance (**Table 5**). The interaction between genotypes and harvests had already been observed and discussed in 2013/14 (Molenaar et al., 2017). Reasons were attributed to environmental conditions such as change in day length during the experimentation or cultivation management, in particular the watering. Similar to P1, the variance of the average "*worker-day*" effect had almost the same size as the variance of the FM·ENTRY effect. The genotypic variance for RF was approximately twice as high as for SCC and also the family effect for RF was greater than for SCC.

In box plots, the relatively low genotypic variance for SCC and RF became visible by the shrinkage property of BLUP in cases when genotypic variation was low or missing; the BLUPs are then all shrunken toward the general mean (**Appendix Presentation 2**

TABLE 6 | Variance component estimates of random effects obtained from model (5) and (6) to evaluate individual and family-index selection for stem length (SL).

Model term	Variance component estimates			
	Mini carnation		Standard carnation	
	Model (5)	Model (6)	Model (5)	Model (6)
FM	–	4.1724	–	69.3944
FM·ENTRY	56.8179	53.9425	111.18	78.0894
REP·BLK	1.6502	1.6067	1.8719	1.4208
REP·BLK·PLT	31.8940	31.9223	34.2640	34.3225

TABLE 7 | Variance component estimates of random effects obtained from model (7) and (8) to evaluate individual and family-index selection for bud number (BN).

Model term	Variance component estimates†	
	Model (7)	Model (8)
FM	-	0.0128
FM·ENTRY	0.0581	0.0486
REP·BLK	0.0030	0.0029
REP·POS	0.0349	0.0353
REP·BLK·PLT	0.0695	0.0695

† *Log-transformed .*

in Supplementary Material). As BLUP assumes a normal distribution with zero mean, the zero reference line represents the zero on the y-axis in comparing box plots for the two selection methods. Assuming individual selection the BLUPs for SCC and RF were close to zero, except for Families 1 and 5. In contrast, by accounting for family-information, a ranking between families was notable. Furthermore, the increased accuracy of BLUPs when accounting for family information is illustrated also by the shortened whiskers of boxes in box plots for family-index selection in comparison to box plots for individual selection (**Appendix Presentation 2** in Supplementary Material).

Generally, the selection based on BLUP for SCC and RF would be reasonable, because the Q-Q plots of standardized BLUPs for SCC and RF revealed that random genotypic effects were approximately normal as required (**Appendix Presentation 3** in Supplementary Material).

Variance Component Estimates From the *D. caryphyllus* L. Breeding

The genotypic variance component estimate for FM·ENTRY for BN, SL, FS, and VL was almost always relatively high in proportion to the total variation (**Tables 6–9**) and hence, large simulated H^2 were obtained for shelf-life traits in comparison to the simulated H^2 for production-related traits *P. zonale* breeding.

The different characteristics between mini and standard carnation with respect to SL became apparent in particular when family-index selection was considered. The family variance component estimate for SL of mini carnation was negligibly small in comparison to the individual variance component estimate, indicating that families of mini carnation vary less for SL than individuals vary within families. In contrast, standard carnation

TABLE 8 | Variance component estimates of random effects obtained from model (9) and (10) to evaluate individual and family-index selection method for flower size (FS).

Model term	Variance component estimates	
	Model (9)	Model (10)
FM	–	0.0249
FM·ENTRY	0.1409	0.1305
STO·FM	–	0
STO·FM·ENTRY	0	0
REP·BLK	0	0
DAY	0.0056	0.0055
DAY·VSE	0.0421	0.0730
REP·BLK·PLT	0.2347	0.2034

TABLE 9 | Variance component estimates of random effects obtained from model (11) and (12) to evaluate individual and family-index selection for vase life (VL).

Model term	Variance component estimates			
	Mini carnation		Standard carnation	
	Model (11)	Model (12)	Model (11)	Model (12)
FM	–	3.9604	–	0.7040
FM·ENTRY	4.6019	1.8754	2.4475	2.0988
STO·FM	–	0.0320	–	0.0247
STO·FM·ENTRY	0.4432	0.4439	0	0
REP·BLK	0.0275	0	0.1788	0.1736
REP·POS	0.6453	0.6829	0.6522	0.6419
DAY	0.3015	0.3443	0.0658	0.0705
REP·BLK·PLT	4.4935	4.4823	4.8872	4.8909

families differ greatly for SL, as indicated by individual variance component estimates similar to family variance component estimate. Mini and standard carnation differ greatly in size, the minis being smaller than standards. For both carnation types the variance of incomplete block effects was marginal in comparison to the genotypic variance components (FM and FM·ENTRY) or the residual error variance.

Another P1 trait was BN for the mini carnation type. Also for this trait, the family variance component estimate was much smaller than the individual variance component estimate. By far the smallest variance component estimate was found for incomplete blocks. The variance component estimate of the *post-blocking* positional effect within incomplete blocks was much greater suggesting that variation in water supply influences the development of BN per single stem and stock plant per position.

In P2, for FS in standards by far the smallest genotypic variance component estimate for FM·ENTRY was obtained and accordingly the smallest H^2 was simulated for this shelf-life trait (**Table 8**). Moreover, the FS of individuals was not affected by the interposed transport simulation, indicated by the zero variance component estimate for the random family-by-transport interaction effect (STO·FM) or individual-by-transport interaction effect (STO·FM·ENTRY). No or only a small proportion of the environmental variation was captured by

the incomplete block effect in P1 or by the day block effect and the positional effect in P2, where the residual error variance was the largest variance component estimate and the family-effect for FS was smaller than that for the positional effect.

The different characteristic between mini and standard carnation became apparent again for the VL assessed for both carnation types. In particular both carnation types varied for individual and family effects and for individual-by-transport interaction effect. A much greater genotypic (FM·ENTRY) variance component for VL was found for mini carnation, which was almost as large as the residual error variance component estimate when individual selection was considered (**Table 9**). The genotypic (FM·ENTRY) variance component estimate was half the size for that of standards when individual selection was considered. When family information was exploited, the largest genotypic variation was observed for the family effect (FM) in mini carnation indicating that the families varied strongly for VL. However, the family-by-transport interaction effect was the smallest variance component estimate when considering family-index selection for mini carnations, beside the incomplete block effect variance which was estimated to be zero. The individual effect variance was half the size of the family effect variance, however, for minis a relatively large variance for the individual-by-transport interaction effect was estimated either under individual or family index selection. This interaction effect variance was estimated to be zero for the standard carnation type, although the estimated family-by-transport interaction effect variance was similar to that for the mini carnation type. The variance of the family effect for standard carnations was much smaller in comparison to that for the individual effect. Interestingly, also the environmental conditions seemed to affect the VL differently. The effect of incomplete blocks was much smaller for the mini carnations than for the standard carnations, where the variance of the *post-blocking* effect in P1 was of similar size. But the day effect for mini carnations was much greater than for the standard carnations.

The selection based on BLUPs for shelf-life traits would be reasonable, because Q-Q plots of standardized BLUPs for shelf-life traits revealed no departure from normality (**Appendix Presentation 2** in Supplementary Material). This is also evidenced by the box plots of BLUPs of the shelf-life traits, except for VL in mini carnation for family-index selection (**Appendix Presentation 3** in Supplementary Material). The standardized BLUPs showed a bimodal distribution. However, the shrinkage property of BLUP was not as strong as for the production-related traits, because the individual effect almost always had the largest variance component estimate under individual selection. Hence, changes in ranks between individual and family-index selection was not as pronounced as for production-related traits.

DISCUSSION

Plant breeders always face the challenge to select the best individuals. Selection methods are required that maximize the use of available data (Bernardo, 2010) and greater selection gain can be expected when methods accounting for pedigree structure are employed (Piepho and Williams, 2006). The BLUP procedure achieves this by combining phenotypic data with information on pedigree relationships (Bernardo, 2010). A selection method that exploit family information is the family-index selection, which is at least as efficient as individual selection (Lynch and Walsh, 2013). This was confirmed by *D. caryophyllus* L., except for FS in standard carnation and VL in mini carnation.

When Family-Index Selection Is the Better Choice

Family-index selection is the better choice for traits with low heritability, which was confirmed for *P. zonale* (**Table 3**). The simulated H^2 for the family-index selection was always upwards from four units better than for the individual selection, which may be explained by the exploitation of relationships of relatives. When family-index selection outperformed individual selection, the total genetic variance was higher than under individual selection, whereas changes of variance component estimates of random block effects or the residual error variance were not as succinct (**Tables 4**, **5**, **7**, **9**). In relation to Lush's example, the individual effects were shrunken toward the family means rather than the overall mean by the use of the intra-class correlation coefficient (**Appendix Presentation 2** in Supplementary Material). As a result, the family means were estimated with higher accuracy and differences between families become more obvious. Best performing individuals of the superior Family 1 remained best. Best performing individuals of poor families were shrunken toward the lower family mean and remained no longer in the selected fraction. Thus, the main difference in individual and family-index selection lies in the adjustment of estimating the individual's effects depending on the estimated variance component of random individual and family effects in a breeding trial, i.e., on the intra-class correlation coefficient. By making this adjustment, the superior performance of family-index depends not only on the ratio between total genotypic and residual error variances, but also on the ratio between the family and individual variances of the total genetic variance. Pure family selection was not considered, because individual performances between families will almost always overlap (**Appendix Presentation 2** in Supplementary Material).

More on Exploiting the Information of Relatives

It is well known that floral characteristics, growth characteristics and cultivation methods differ between mini and standard carnation types. However, such strong differences of the individual performances depending on the family background were unexpected for SL and VL (**Table 3**) implying different strategies in breeding. For example in mini carnations, the performance of individuals on SL is almost totally independent of the family background ($t = 0.07$), whereas for VL a high intra-class correlation coefficient was found ($t = 0.68$). The high intra-class correlation coefficient means that individuals within families are more similar than across families. This indicates the importance of selecting parental strains used for crosses for VL

improvement, which can be further investigated by the general or specific combining ability for example. This is contrary to the dependence of individual performance for SL and VL on the family background in standard carnations, which reveals that in *D. caryophyllus* L. the two carnation types should be bred in different programs. Thus, exploiting the information of relatives the genetics underlying the traits must be observed.

BLUP in Ornamentals

In clonal breeding, the greatest genetic variability exists in the seedling generation (Figure 1 in Molenaar et al., 2017). Each seedling and individual is represented by a single plant. Experimental designs that are suitable to test unreplicated treatments are augmented designs, as applied in *P. zonale*. Block effects are estimated solely by the use of checks, which might not capture all environmental variation on the estimation of the treatments effects. In the seedling generation, the only way to increase the precision of estimating treatment effects is to exploit the information of relatives, confirmed by *P. zonale* (**Tables 3**, **4**). In the primary selection of seedlings, the population size is the drastically reduction of from thousands to a maximum of 200 individuals. From that primary selection until the official testing, selected individuals are only clonally propagated.

In the clonal generations, individual genotypes can be tested in replications, for example in resolvable incomplete block designs, as applied in *D. caryophyllus* L. or in randomized complete block designs in later breeding stages as the number of individuals is more and more reduced (Figure 1 in Molenaar et al., 2017). Higher precision of estimated treatment effects can be expected, because the block effects are estimated from the individuals included in all replicates. However, as the population size is reduced, the genetic variability is reduced, too, from the seedling to the first clonal generation. Differences between individuals might become difficult to detect. But here too, consideration of family information may improve treatment estimates. In the present study, the effect of reduced genetic variability and the use of a replicated individuals could not be investigated, as the entire seedling generation was clonally propagated before the vase life tests, which is rarely done in practice.

Furthermore, if families in clonal generations are tested in different locations, each genetically identical individual within a family can be tested across locations, and hence, each individual is replicated across locations. This allows a precise determination of genotype-by-environment interaction. By comparison, with non-clonable species, only families can be replicated across locations, but not individuals within families.

CONCLUSION

The choice of a selection method has implications for selection gain. Family-index selection was found to be at least as efficient as individual selection, surpassing the efficiency of individual selection when the heritability was low. Another important aspect for breeders is the shrinkage property of the family-index selection, where superior individuals are shrunken downwards and inferior individuals are shrunken upwards, yielding in a change of genotype ranks protecting to do false selection decision. Furthermore, exploiting the information of relatives can be used to investigate the genetics behind traits and reveal strategies for selecting parental strains for crosses. Our results support the need for separating the breeding program for *D. caryophyllus* L. into mini and standard types. The present work illustrated the use of BLUP in *P. zonale* and *D. caryophyllus* L., which are exemplary for ornamental and clonal breeding in general.

AUTHOR CONTRIBUTIONS

HM conceived and participated in the design of the study, conducted the analysis, interpreted the results and prepared the manuscript. H-PP participated in the design of the study, in its writing and editing of the manuscript and oversaw the project. H-PP and RB revised the manuscript. All authors discussed the results, commented on the manuscript and have approved the final manuscript.

ACKNOWLEDGMENTS

This project was partly funded by grants KF3130602 (BMWi) and PI 377/19-1. Further, we would like to thank the team from Klemm + Sohn GmbH & Ko KG who assisted in carrying out the experiments in 2015/16 and 2016/17.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2018.01511/full#supplementary-material

REFERENCES

Barbosa, M., Resende, M., Bressiani, J., Silveira, L., and Peternelli, L. (2005). Selection of sugarcane families and parents by Reml/Blup. *Crop Breed. Appl. Biotechnol.* 5, 443–450. doi: 10.12702/1984-7033.v05n04a10

Bernardo, R. (2010). *Breeding for Quantitative Traits in Plants*. Woodbury, MN: Stemna Press.

Borges, V., Ferreira, P. V., Soares, L., Santos, G. M., and Santos, A. M. M. (2010). Seleção de clones de batata-doce pelo procedimento REML/BLUP. *Acta Scientiarum Agronomy* 32, 643–649. doi: 10.4025/actasciagron.v32i4.4837

Boxriker, M., Boehm, R., Krezdorn, N., Rotter, B., and Piepho, H. P. (2017a). Comparative transcriptome analysis of vase life and carnation type in *Dianthus caryophyllus* L. *Sci. Hortic.* 217, 61–72. doi: 10.1016/j.scienta.2017.01.015

Boxriker, M., Boehm, R., Möhring, J., and Piepho, H. P. (2017b). Efficient statistical design in two-phase experiments on vase life in carnations (*Dianthus caryophyllus* L.). *Postharvest Biol. Technol.* 128, 161–168. doi: 10.1016/j.postharvbio.2016.12.003

Copas, J. (1983). Regression, prediction and shrinkage. *J. R. Statist. Soc. Series B.* 45, 311–354.

Cullis, B. R., Smith, A. B., and Coombes, N. E. (2006). On the design of early generation variety trials with correlated data. *J. Agric. Biol. Environ. Stat.* 11, 381–393. doi: 10.1198/108571106X154443

Falconer, D. S., and Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics*. London: Prentice Hall.

Federer, W. T. (1956). Augmented (or hoonuiako) designs. *Hawaiian Planter Record* 55, 191–208.

Fogaça, L. A., Oliveira, R. A., Cuquel, F. L., Filho, J. C. B., Vendrame, W. A., and Tombolato, A. F. C. (2012). Heritability and genetic correlation in daylily selection. *Euphytica* 184, 301–310. doi: 10.1007/s10681-011-0478-y

Gianola, D., Cecchinato, A., Naya, H., and Schön, C.-C. (2018). Prediction of complex traits: robust alternatives to best linear unbiased prediction. *Front. Genet.* 9:195. doi: 10.3389/fgene.2018.00195

Graser, H., Smith, S., and Tier, B. (1987). A derivative-free approach for estimating variance components in animal models by restricted maximum likelihood. *J. Anim. Sci.* 64, 1362–1370. doi: 10.2527/jas1987.6451362x

Henderson, C. R. (1950). Estimation of genetic parameters. *Ann. Math. Statist.* 21, 309–310.

Hill, R., and Rosenberger, J. (1985). Methods for combining data from germplasm evaluation trials. *Crop Sci.* 25, 467–470. doi: 10.2135/cropsci1985.0011183X002500030009x

Holland, J. B., Nyquist, W. E., and Cervantes-Martínez, C. T. (2003). Estimating and interpreting heritability for plant breeding: an update. *Plant Breed. Rev.* 22, 9–112. doi: 10.1002/9780470650202.ch2

Huang, H., Harding, J., Byrne, T., and Famula, T. (1995). Estimation of long-term genetic improvement for gerbera using the best linear unbiased prediction (BLUP) procedure. *Theor. Appl. Genet.* 91, 790–794. doi: 10.1007/BF00220961

Legarra, A. (2016). Comparing estimates of genetic variance across different relationship models. *Theor. Popul. Biol.* 107, 26–30. doi: 10.1016/j.tpb.2015.08.005

Lehermeier, C., de los Campos, G., Wimmer, V., and Schön, C.-C. (2017). Genomic variance estimates: with or without disequilibrium covariances? *J. Anim. Breed. Genet.* 134, 232–241. doi: 10.1111/jbg.12268

Lush, J. L. (1947). Family merit and individual merit as bases for selection. *Part I Am. Nat.* 81, 241–261. doi: 10.1086/281520

Lynch, M., and Walsh, B. (2013). *Genetics and Analysis of Quantitative Traits*. Sunderland, M. A: Sinauer Associates, Inc.

Misztal, I., and Gianola, D. (1987). Indirect solution of mixed-model equations. *J. Diary Sci.* 70, 716–723. doi: 10.3168/jds.S0022-0302(87)80063-2

Molenaar, H., Boehm, R., and Piepho, H. P. (2018). Identifying effective design approaches to allocate genotypes in two-phase designs: a case study in *Pelargonium zonale*. *Front. Plant Sci.* 8:2194. doi: 10.3389/fpls.2017.02194

Molenaar, H., Glawe, M., Boehm, R., and Piepho, H. P. (2017). Selection for production-related traits in Pelargonium zonale: improved design and analysis make all the difference. *Horticult. Res.* 4:17004. doi: 10.1038/hortres.2017.4

Oakey, H., Verbyla, A. P., Cullis, B. R., Wei, X., and Pitchford, W. S. (2006). Joint modeling of additive and non-additive (genetic line) effects in multi-environment trials. *Theoret. Appl. Genet.* 114, 1319–1332. doi: 10.1007/s00122-007-0515-3

Onozaki, T., Ikeda, H., and Yamaguchi, T. (2001). Genetic improvement of vase life of carnation owers by crossing and selection. *Sci. Hortic.* 87, 107–120. doi: 10.1016/S0304-4238(00)00167-9

Piepho, H. P. (1994). Best Linear Unbiased Prediction (BLUP) for regional yield trials: a comparison to additive main effects and multiplicative interaction (AMMI) analysis. *Theoret. Appl. Genet.* 89, 647–654. doi: 10.1007/BF00222462

Piepho, H. P. (2015). Generating efficient designs for comparative experiments using the SAS procedure OPTEX. *Commun. Biomet. Crop Sci.* 10, 96–114.

Piepho, H. P., Büchse, A., and Emrich, K. (2003). A hitchhiker's guide to mixed models for randomized experiments. *J. Agronomy Crop Sci.* 189, 310–322. doi: 10.1046/j.1439-037X.2003.00049.x

Piepho, H. P., and Eckl, T. (2014). Analysis of series of variety trials with perennial crops. *Grass Forage Sci.* 69, 431–440. doi: 10.1111/gfs.12054

Piepho, H. P., and Möhring, J. (2007). Computing heritability and selection response from unbalanced plant breeding trials. *Genetics* 177, 1881–1888. doi: 10.1534/genetics.107.074229

Piepho, H. P., Möhring, J., Melchinger, A. E., and Büchse, A. (2008). BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica* 161, 209–228. doi: 10.1007/s10681-007-9449-8

Piepho, H. P., and Williams, E. R. (2006). A comparison of experimental designs for selection in breeding trials with nested treatment structure. *Theoret. Appl. Genet.* 113, 1505–1513. doi: 10.1007/s00122-006-0398-8

Piepho, H. P., Williams, E. R., and Fleck, M. (2006). A note on the analysis of designed experiments with complex treatment structure. *HortScience* 41, 446–452.

Resende, M. (2016). Software Selegen-REML/BLUP: a useful tool for plant breeding. *Crop Breed. Appl. Biotechnol.* 16, 330–339. doi: 10.1590/1984-70332016v16n4a49

Robinson, G. K. (1991). That BLUP is a good thing: the estimation of random effects. *Statist. Sci.* 6,15–51. doi: 10.1214/ss/1177011926

Santos, E. A., Viana, A. P., de Oliveira Freitas, J. C., Rodrigues, D. L., Tavares, R. F., Paiva, C. L., et al. (2015). Genotype selection by REML/BLUP methodology in a segregating population from an interspecific *Passiflora* spp. crossing. *Euphytica* 204, 1–11. doi: 10.1007/s10681-015-1367-6

Searle, S. R., Casella, G., and McCulloch, C. E. (1992). *Variance Components*. Hoboken, NJ: John Wiley &Sons. doi: 10.1002/9780470316856

Singh, B. D., and Singh, A. K. (2015). *Marker-Assisted Plant Breeding: Principles and Practices*. Heidelberg; New York, NY, Dordrecht, London: Springer. doi: 10.1007/978-81-322-2316-0

Slater, A. T., Wilson, G. M., Cogan, N. O. I., Forster, J. W., and Hayes, B. J. (2014). Improving the analysis of low heritability complex traits for enhanced genetic gain in potato. *Theoret. Appl. Genet.* 127, 809–820. doi: 10.1007/s00122-013-2258-7

Teixeira, D. H. L., Oliveira, M. S. P., Goncalves, F. M. A., and Nunes, J. A. R. (2012). Selection index for simultaneously improving fruit production components of assai palm. *Pesq. Agropec. Brasil.* 47, 237–243. doi: 10.1590/S0100-204X2012000200012

Ticona-Benavente, C. A., and da Silva Filho, D. F. (2015). Comparison of BLUE and BLUP/REML in the selection of clones and families of potato (Solanum tuberosum). *Genet. Mol. Res.* 14, 18421–18430. doi: 10.4238/2015.December.23.30

Van Raden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Diary Sci.* 91, 4414–4423. doi: 10.3168/jds.2007-0980

Wang, B., Sverdlov, S., and Thompson, E. (2017). Efficient estimation of realized kinship from single nucleotide polymorphism genotypes. *Genetics* 205, 1063–1078. doi: 10.1534/genetics.116.197004

Zeni Neto, H., Daros, E., Bespalhok Filho, J. C., Scapim, C. A., Vidigal, M. C. G., and Vidigal Filho, P. S. (2013). Selection of families and parents of sugarcane (Saccharum spp.) through mixed models by joint analysis of two harvests. *Euphytica* 193, 391–408. doi: 10.1007/s10681-013-0947-6

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

5 General discussion and future perspectives

The present work has demonstrated that the efficiency of phenotypic selection in *P. zonale* can be increased by the implementation of suitable experimental designs. Efficiency of selection can be further enhanced by optimizing the allocation of genotypes in a two-phase experimental set-up as well as by using the BLUP-method, in particular for un-replicated trails assessing the seedling generation. By increasing the efficiency of phenotyping and selection, the present work has also laid the foundation for effective marker-assisted selection in *P. zonale*.

5.1 Phenotyping protocols

Phenotyping protocols do not only serve as guidelines for repeated, accurate and widespread phenotyping by different persons for arriving at valid conclusions (Singh and Singh, 2015), such protocols are also reflective of the expected variation for a given trait. The focus in establishing a phenotyping protocol for production-related traits was on the root formation of stem cuttings which reflects the phenotypic variation present for this trait and which is amenable to high-throughput scoring of individual plants. Within the present work, an ordinal scale was defined consisting of six categories, ranging from dead (S0) to extraordinarily rooted (S5) (Molenaar *et al*., 2017). An extension of the phenotyping protocol for root formation is possible to other vegetative propagated species such as *Osteospermum* or *Dianthus caryophyllus* L. (Molenaar *et al*., 2018a).

A benefit of this protocol is that it can easily be modified to meet the norm of a one to nine scale for variety registration through the Federal Plant Variety Office or UPOV. By omitting the category S0, because this is essentially missing data, the remaining five categories (S1-S5) may be simply extended to nine by inserting a new intermediate score between each level. Then score S1 from Molenaar *et al*. (2017) is equivalent to the score 1 on the extended scale, whereas score S2 is assigned to score 3, S3 to score 5, S4 to score 7 and S5 to score 9. The phenotypic range of categories would be retained, and the new scores between the original S_i (i = 1, 2, 3,

4, 5) scores would allow more precise single plant evaluation if necessary (see for example UPOV Pelargonium, TG/109/3).

5.2 Phases of an experiment

In the present work, solely two phases were considered to define the experimental set-up used in assessing stem cutting productivity. Phase one comprised the cultivation of the stock plants, and phase two comprised the assessment of root formation in the material harvested in the first phase. However, ornamental breeding need not be limited to two phases as two-phase experiments are only a subclass of multiphase experiments (Brien, 2017). Consideration of more than two phases in ornamental breeding is possible, as will be illustrated in the following. The breeder should keep in mind potential disadvantages, such as an increase in number of block effects which must be estimated when different block structures in phases of the experiment are required, and increased complication of the design by increased number of phases, possibly resulting in additional sources of error, especially if the additional phases require several workers.

Hypothetical three-phase experiment in P. zonale breeding - A further phase in *P. zonale* breeding could be the initial rooting of stem cuttings or the sowing of seeds to obtain the stock plants, so that the experiment would contain the following phases: phase one, rooting stem cuttings or sowing of seeds, phase two, planting of plant material into pots from phase one and subsequent cultivation to obtain stock plants and phase three, the assessment of harvested plant material for root formation. The reasoning behind considering three instead of two phases is to capture the variation in plant material induced by different water or heat stress before planting in pots. In the present work, however, the variable plant conditions were expected to be equalized the longer the stock plant cultivation period took place due to the high level of standardization in cultivation techniques (Martin Glawe, *personal communication, June 2013,* Selecta One). Ultimately, the decisive reason to consider the initial rooting or sowing of seeds as a separate phase is the transmission of the observational unit, *i.e.*, the rooted stem cuttings or seedlings were ‘transformed’ into stock plants and from the initial plant material (stem cuttings or seedlings) no observations are taken.

A resistance multiphase experiment - An example for a multiphase experiment consisting of more than two phases in ornamental breeding is the evaluation of *Fusarium oxysporum ssp.* resistance in *Dianthus caryophyllus* L. (current experiment at Selecta One). The test consists of

the following three phases: Phase one, the cultivation of the stock plants and the harvest and count of stem cuttings, phase two, the rooting of stem cuttings and the assessment of root formation, and phase three, the assessment of the resistance. In this example, the rooting of plant material in the second phase is of interest for the assessment of resistance in the third phase, because the rooted cuttings of phase two are infected via the soil and remain as the observational unit in the third phase. No cultivation to stock plants follows and the root formation is assessed from stem cuttings. The stem cuttings in the third phase are discarded after the assessment for resistance. Then the experiment is repeated by harvesting, rooting and assessing stem cuttings for resistance.

As illustrated, in ornamental breeding the number of potential phases can vary and may be increased to any number. Therefore, process characterization is needed to reduce the number of phases to an optimum by identifying the most important factors that affect the responses in phases when different block structures across phases exist (Brien, 2017). In cases in which the same block structure across phases in the entire experiments exists, the number of phases of the experiment is of minor importance as no computational burden can be expected, because block effects need to be estimated only once.

5.3 Pseudo-levels for generating designs across phases

To optimize the two-phase experimental designs used in 2013 to 2015, two main approaches were found, for which the smallest MVD could be achieved. By assuming the same block structure in both phases, the first approach was to generate once an optimal single-phase design and transfer this to the consecutive phase. This has the following advantages: First, the efficiency of the statistical analysis is increased as only one set of block effects needs to be estimated for both phases (Molenaar *et al.*, 2018a). Second, this will affect the conduction of the experiment in that the genotypes have to be ordered only once according to a working number, which can be used for transferal between phases and to plant genotypes in consecutive phases.

It is a rather strong assumption to use the same block structure in both phases because these take place often in very different locations with different environment influences. Thus, different block structures in both phases were also considered in the second approach. The second optimization approach, which resulted in a low MVD was a two-phase design which

consisted in both phases of a different block structure, where the phases were coded by a pseudo-variable to generate the design across phases (Table 7 in Molenaar *et al*., 2018a).

There is as yet little research on the optimization of two-phase experimental design (Brien, 2017), but the use of pseudo-variables, also named as pseudo-factors, is long known. However, none of the authors (Yates, 1936; Monod and Bailey, 1992; Bailey and Brien, 2006) defined the phases themselves as pseudo-factors as proposed by Molenaar *et al*. (2018a). Generally, a pseudo-factor groups the levels of a factor in equally-sized groups, as an aid to ensure equal concurrences in blocks while generating the experimental design (Bailey and Brien, 2006).

5.4 Considering pedigree information

Ornamental breeding is lagging behind not only in application of experimental designs, but also in selection methods allowing an efficient estimation of breeding values. As a first step, the BLUP-procedure was proposed, which allows to consider the nested treatment structure due to the grouping of individuals by family derived from a pedigree and exploiting the genetic variances of families and individuals within families (Falconer and Mackay, 1996; Bueno and Gilmour, 2003).

Other BLUP approaches exploit pedigree information by the numerator relationship matrix based on additive genetic variances (ABLUP) or exploit the pedigree information based on the genomic relationship matrix based on genomic pair-wise similarities (GBLUP) (Van Raden, 2008; Legarra 2016; Wang *et al*., 2017), where the additive genetic variance is replaced by the genomic variance (Yang *et al*., 2010; Lehermeier *et al*., 2017). The GBLUP is also known as genomic prediction and used for implementing genomic selection (Piepho *et al*., 2012; Gianola *et al*., 2018). GBLUP has potential advantages over ABLUP, such as lower dependency on idealized evolutionary assumptions of random mating or the absence of selection and the realized similarities can be pair-specific instead of family-specific (Gianola *et al*., 2018). Under ABLUP the expected additive relationship between any pair of full-sibs is 0.5 within a family, whereas by GBLUP the realized similarity may vary over pairs of full-sibs within a family (Gianola *et al*., 2018).

As the availability of marker data is limited in ornamental breeding programs, the proposed family-based BLUP method is currently a promising selection method to use and information can be exploited, whether the trait is dependent or independent of the family background. Information about the genetics of traits is required for selecting individuals for genotyping inter

alia, because the goal of creating diversity panels is to represent the entire genetic diversity of parental populations, *i.e.*, individuals should be selected with similar biotic or abiotic adaptation or photoperiod requirements (Singh and Singh, 2015, p. 220). On the other hand individuals should show as much phenotypic variation as possible in the trait.

5.6 Future perspective: Genotyping and implementation of MAS in P. zonale breeding

Although not currently practiced, in the future, the benefits of MAS should improve *P. zonale* breeding and the competitiveness on the floricultural market. These benefits are increasing response to selection per unit time and enabling effective early selection for poorly heritable or difficult-to-phenotype traits where phenotypic testing capacity is not sufficient in early breeding stages (Lynch and Walsh, 1998, p. 456). The most important factor influencing the effectiveness of MAS, phenotyping (Xu and Crouch, 2008), was investigated within this work. The foundation within *P. zoanle* breeding was laid for MAS by phenotyping approximately 1,500 *P. zonale* strains. Following the phenotyping and statistical analysis, the 273 most important accessions from the 1,500 strains were selected, for DNA isolation and genotyped by the 'Diversity Arrays Technology' (DArT). More than 33,000 SNP markers were found in the *P. zonale* genome which will be used in future association studies.

Additionally, 120 of the 273 accessions were sequenced by 'Massive analysis of cDNA ends' (MACE). MACE is a transcriptome sequencing approach and aims to find causal genes for traits of interests. For each of the two traits, stem cutting count and root formation, 60 contrasting accessions were selected (30 accessions with a high performance and 30 accessions with a poor performance in each of the two traits). Samples of the selected accessions were taken from differentially expressed tissues of both apical meristems and buds three and six days after planting, as well as from roots three and 12 days after planting stem cuttings. From those samples, composite samples for each genotype and tissue were created. Then the RNA was isolated using the Quiagen Kit (QIAGEN, Hilden, Germany). The isolated RNA of genotypes was further bulked by combining equal amounts of RNA from genotypes. In total four composite samples were created. Two RNA bulks for stem cutting count and two RNA bulks for root formation. The RNA bulks were then sequenced via MACE. The genotypic data output after alignment and bioinformatics analysis was as follows: Reads (number of transcription of cDNA) per contig (DNA strand with SNP) were counted for each bulk and trait. As no labeling

was used for single RNA strands of genotypes, no information on individuals was available. The SNP calling was conducted by using the allele frequencies in bulks per position and its Clopper-Pearson interval (Clopper and Pearson, 1934) within five steps. Those 42 mRNA SNPs were selected showing the highest fifth index, to convert those mRNA SNPs marker into user friendly DNA based marker. The 120 accessions were then genotyped by a DNA based genotyping assay and SNPs per locus were tested for significance either by G-Tests or t-tests (McDonald 2014), where significant markers could be identified for stem cutting count. Whether those identified markers can be validated in the DArT marker data will be investigated in the future research.

The combination of efficient phenotyping and newly available markers should accelerate selection gain in ornamental breeding, resulting in potential reduction of required stock plants by 20 %, which carries obvious economic and environmental benefits.

6 Conclusion

With the help of the developed high-throughput phenotyping protocol and introduction of a two-phase experimental design, genotypic variation could be effectively quantified. Furthermore, we found that two-phase experimental designs in *P. zonale* breeding reduce error variances by accounting for phase-specific factors and increase the precision of estimates of phenotypic and genotypic effects, which positively affects the response to selection and is an absolute necessity for any marker-based analysis that exploits phenotype-genotype associations.

Because of the fast decay of stem cutting quality the randomization was initially carried out on-site in the greenhouse while observing experimental design principles. In search of optimized two-phase experimental designs, with respect to the various options considered in Molenaar *et al*. (2018), the results show that randomization on-site was sub-optimal and in both phases pre-defined designs should be used. Thereby, two-phase designs should be generated across phases (Option 2) rather than in phase-wise order (Option 1). The smallest MVD was most frequently obtained for Option 2 with different block structures in both phases and the approach using a single pseudo level for incomplete blocks in phase one and two.

With the pragmatic approaches taken in this work, the generation of two-phase designs in *P. zonale* breeding was improved, yielding a reduction in the MVD from 9.42 to about 2.35 by intra-block analysis or from 2.67 to approximately 1.99 by using computer generated designs in both phases, additional block factors in phase one and generating the design across phases. This great reduction in MVD justifies the consideration of idealized conditions in *P. zonale* breeding and indicates that the on-site randomization approach is sub-optimal. Most importantly, the idealized conditions included the assumption of the same block size, $k = 6$ in both phases such that the stem cutting count per genotype needs to be reduced to six in P1 to assess root formation in P2, which resulted in a decrease of blocks in P1 and made the use of the same block structure across phases possible.

If the use of the same block structure in both phases is feasible, a further increase in efficiency of analysis can be expected when the experimental layout is transmitted from the first to the second phase, because only one incomplete block adjustment is required to estimate genotype

effects across the two phases. Among scenarios considering the same block structure across phases, that two-phase design showed the smallest MVD, too.

Not only the experimental layout, but also the choice of selection method has implications for selection gain. Family-index selection was found to be at least as efficient as individual selection, outperforming the efficiency of individual selection when the heritability was low. Exploiting the information of relatives can be used to investigate the genetics behind traits and reveal strategies for selecting parental strains for crosses or individuals undergoing the genotyping.

7 Summary

Ornamental plant variety improvement is limited by current phenotyping approaches and the lack of use of experimental designs. Robust phenotypic data obtained from experiments laid out to best control local variation by blocking allow adequate statistical analysis and are crucial for any breeding purpose, including MAS. Often experiments consist of multiple phases like in *P. zonale* breeding, where in the first phase stock plants are cultivated to obtain the stem cutting count and in the second phase the stem cuttings are further assess for root formation.

The first analyses of rooting experiments raised questions regarding options for improving the two-phase experimental layout, for example whether there is a disadvantage to using exactly the same design in both phases. The other question was, whether a design can be optimized across both phases, such that the MVD can be decreased. Instead of generating a separate layout for each phase. Moreover, optimal selection methods that maximize selection gain in *P. zonale* breeding based on available data collected from unreplicated trials and containing pedigree information were sought.

This thesis was conducted to evaluate the benefits of using two-phase experimental designs and corresponding analysis in *P. zonale* for production-related traits, for which it was necessary to establish phenotyping protocols. To optimize the rooting experiments with their two-phase nature, alternative approaches were explored involving two-phase design generation either in phase-wise order or across phases. Furthermore, selection methods considering pedigree-information (family-index selection) or not (individual selection), were evaluated to enhance selection efficiency in *P. zonale* breeding.

The benefits of using experimental designs in *P. zonale* breeding was shown by the simulated response to selection. Alternative designs were evaluated by the MVD obtained by the intra-block analysis and the joint inter-block-intra-block analysis. The efficiency of individual and family-index selection was evaluated in terms of heritability obtained from linear mixed models implementing the selection methods.

Simulated response to selection varied greatly, depending on the genotypic variances of the breeding population and traits. However, by using efficient designs allowing adequate analysis, a varietal improvement of over 20% of stock plant reduction is possible for stem cutting count,

root formation, branch count and flower count. The smallest MVD for alternative designs was most frequently obtained for designs generated across phases rather than for each phase separately, in particular when both phases of the design were separated with a single pseudo-level. Family-index selection was superior to individual selection in *P. zonale* indicating that the pedigree-based BLUP procedure can further enhance selection efficiency in production-related traits in *P. zonale*.

The quantification of genotypic variation by phenotypic protocols and the optimized two-phase designs for estimating genotypic values were necessary and successful steps in laying the foundation for effective MAS. Phenotypic protocols effectively characterized the genetic material on an observational unit level, while the two-phase experimental designs enabled effective characterization on a genotype level by adjusting entry means using linear mixed models. The resulting adjusted entry means are the basis for future genotype-phenotype association for MAS.

8 Zusammenfassung

In der Zierpflanzenzüchtung werden die Möglichkeiten von Sortenverbesserungen nicht vollständig ausgeschöpft, bedingt durch fehlende Boniturschemata und Anwendung von experimentellen Designs. Robuste phänotypische Daten sind für jegliche züchterische Belange von größter Bedeutung, u.a. die Marker-gestützte Selektion. Dabei können robuste Daten nur von Experimenten erhoben, welche an die jeweiligen umweltspezifischen Bedingungen bestmöglich mittels geeigneter Blockstrukturen angepasst wurden. Experimente können dabei auch mehr als eine Phase beinhalten, wie beispielsweise in der Pelargonienzüchtung, wobei die erste Phase die Kultivierung der Mutterpflanzen und die Ermittlung der Stecklingsanzahl berücksichtigt und in der zweiten Phase, die Stecklinge hinsichtlich der Bewurzelung untersucht werden.

Während den ersten Auswertungen der Bewurzelungsexperimenten stellten sich zwei zentrale Optimierungsfragen hinsichtlich der Erstellung von zwei-phasigen Experimenten: „Gibt es einen Nachteil in der Nutzung desselben Designs in beiden Phasen?“ und „Kann ein zwei-phasiges Design über beide Phasen so optimiert werden, dass die mittlere Varianz eines paarweisen Mittelwertvergleiches (MVD) minimiert werden kann, anstelle für jede Phase ein eigenes Design zu generieren?“ Außerdem wurde nach optimalen Selektionsmethoden gesucht, die einen bestmöglichen Selektionsgewinn unter der Anwendung von unwiederholten Versuchen und Vorliegen von Verwandtschaftsverhältnissen ermöglichen.

Im Rahmen dieser Arbeit wurden die Anwendungsvorteile zwei-phasiger Experimente und deren Analyse in der *P. zonale* Züchtung für produktionsrelevante Merkmale untersucht, für welche zunächst Boniturschemata erstellt werden mussten. Zur Optimierung der zwei-phasigen Bewurzelungsexperimente, wurden alternative Ansätze entwickelt, die eine Designgenerierung über Phasen hinweg ermöglichen oder wie bisher für jede einzelne Phase des zwei-phasigen Experimentes eine separate Designgenerierung vorsehen. Um die Effizienz der Selektion zu steigern, wurden ferner Selektionsmethoden untersucht, welche die Verwandtschaftsverhältnisse berücksichtigen („Familien-Index“ Selektion) oder nicht („Individual“ Selektion) in der Schätzung genotypischer Effekte.

Evaluiert wurden die Anwendungsvorteile zwei-phasiger Experimente mittels des simulierten Selektionserfolges, während die alternativen zwei-phasigen Designs der Bewurzelungsexperimente mittels des MVD basierend auf „Dummy-Analysen“ unter Nutzung der Intra-Block-Information oder unter Nutzung der kombinierten Intra-Inter-Block-Information verglichen wurden. Die „Familien-Index“ und „Individual“ Selektionsmethoden wurden mittels der Heritabilität miteinander verglichen.

Der simulierte Selektionserfolg variierte erheblich in Abhängigkeit von den geschätzten genotypischen Varianzen der Zuchtpopulation und den untersuchten Merkmalen. Jedoch können unter Nutzung der zwei-phasigen Designs und adäquaten statistischen Auswertung Sortenverbesserungen in Hinblick auf einer Reduzierung der Mutterpflanzen von bis zu 20 % für die Merkmale Stecklingsanzahl, Bewurzelung, Verzweigung und Blütenanzahl wie gefordert erwartet werden. Für die alternativen Ansätze zur Generierung zwei-phasiger Experimente über Phasen hinweg wurden überwiegend die kleinste MVD ermittelt, insbesondere für den Ansatz unter Verwendung von Pseudo-Variablen zur Definition der zwei Phasen des Experimentes. Im Vergleich der Selektionsmethoden war die „Familien-Index“ Selektion besser hinsichtlich der Heritabilitäten als die „Individual“ Selektion und deutete somit eine weitere Steigerung der Selektionseffizienz hin, als diese schon durch die Anwendung von Design und Statistik erreicht worden war.

Die Quantifizierung der genotypischen Varianz mittels erstellter Boniturschemata und die Optimierung der zwei-phasigen Experimenten zur besseren Schätzung genotypischer Effekte waren notwendige und erfolgreiche Schritte als Grundlage zur Einführung der Marker-gestützten Selektion in die *P. zonale* Züchtung. Mittels den erstellten Boniturschemata konnte das genetisch untersuchte Material effektiv charakterisiert werden, während adjustierte Mittelwerte der Genotypen durch Anwendung zwei-phasiger experimenteller Layouts und deren Auswertung mittels linear gemischter Modelle optimal geschätzt werden konnten. Die adjustierten Mittelwerte der Genotypen sind die Basis für Phänotyp-Genotyp Assoziationen im Rahmen der Marker-gestützten Selektion die künftig in der Züchtung genutzt werden soll.

References

Anderson CA, Petterson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT (2010). Data quality control in genetic case-control association studies. Nature Protocols, 5:1564-1573.

Becher SA, Steinmetz K, Weising K, Boury S, Peltier D, Renou JP, Kahl G, Wolff K (2000). Microsatellites for cultivar identification in *Pelargonium*. Theoretical Applied Genetics, 101:643-651.

Boxriker M, Boehm R, Möhring J, Piepho HP (2017). Efficient statistical design in two-phase experiments on vase life in carnation (*Dianthus caryophyllus* L.). Postharvest Biology and Technology, 128:161-168.

Boxriker M, Möhring J, Piepho HP (2018). Genetic and phenotypic correlation for breeding relevant traits in *Dianthus caryophyllus* L. Postharvest Biology and Technology, 143:129-136.

Brien CJ, Bailey RA (2006). Multiple randomizations (with discussion). Journal of the Royal Statistical Society Series B-Statistical Methodology, 68:571-599.

Brien CJ (2017). Multiphase experiments in practice: a look back. Australian & New Zealand Journal of Statistics, 59(4):327-352.

Bueno JSD, Gilmour SG (2003). Planning incomplete block experiments when treatments are genetically related. Biometrics, 59:375-381.

CCGS (2010). Chromosome numbers in *Pelargonium* in: Central Coast Geranium Society, 6.6.18, http://www.geraniumsonline.com/numbers.htm, 8.8.18

Clopper C, Pearson ES (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika, 28:404-413.

Collard BCY, Mackill DJ (2008). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philosophical Transactions of the Royal Society, 363:557-572.

Debener T (2001). Molecular tools for modern ornamental plant breeding and selection. In: Van Huylenbroeck J et al. (ed.) XX EUCARPIA Symposium on new ornamentals, Acta Horticulturae, ISHS.

Debener T (2002). Molecular markers as a tool for analyses of genetic relatedness and selection in ornamentals. In: Vainstein A (ed.) Breeding for ornamentals: classical and molecular approaches. Kluwer Acedemic Publishers Dordrecht, Boston, London.

Deroles SC, Boase MR, Lee CE, Peters TA (2002). Gene transfer to plants. In: Vainstein A (ed.) Breeding for ornamentals: classical and molecular approaches. Kluwer Acedemic Publishers Dordrecht, Boston, London.

Falconer DS, Mackay TFC (1996). Introduction to quantitative genetics. Prentice Hall, London, UK.

Federer W, Crossa J (2012). Screening experimental designs for quantitative trait loci, association mapping, genotype-by-environment interaction, and other investigations. Frontiers in Physiology, 3:1-8.

Gianola D, Cecchinato A, Naya H, Schön C-C (2018). Prediction of complex traits: robust alternatives to best linear unbiased prediction. Frontiers in Genetics, 9:195.

Jain SM, De Klerk GJ (1998). Somaclonal variation in breeding and propagation of ornamental crops. Plant Tissue Culture and Biotechnology, 4:63-75.

John JA, Williams ER (1995). Cyclic and computer generated designs, 2nd Edn, Chapman and Hall, London.

Legarra A (2016). Comparing estimates of genetic variance across different relationship models. Theoretical Population Biology, 107:26-30.

Lehermeier C, de los Campos G, Wimmer V, Schön C-C (2017). Genomic variance estimates: with or without disequilibrium covariances? Journal of Animal Breeding and Genetics, 134:232-241.

Lush JL (1947). Family merit and individual merit as bases for selection. Part I. The American Naturalist, 81(799):241–261.

Lynch M, Walsh B (2013). Genetics and analysis of quantitative traits. Sinauer Associates, Inc.: Sunderland, MA, USA, p. 456.

McDonald JH (2014). G–test of goodness-of-fit. Handbook of Biological Statistics, 3rd Ed., Sparky House Publishing, Baltimore, Maryland.

Molenaar H, Glawe M, Boehm R, Piepho H-P (2017). Selection for production-related traits in *Pelargonium zonale*: improved design and analysis make all the difference. Horticulture Research, 4:1-9.

Molenaar H, Boehm R, Piepho H-P (2018a). Identifying effective design approaches to allocate genotypes in two-phase designs: a case study in *Pelargonium zonale*. Frontiers in Plant Science, 8:1-12.

Molenaar H, Boehm R, Piepho H-P (2018b). Phenotypic selection in ornamental breeding: it's better to have the BLUPs than to have the BLUEs. Frontiers in Plant Science, (in revision).

Monod H, Bailey RA (1992). Pseudofactors: normal use to improve design and facilitate analysis. Applied Statistics, 41:317-336.

Onozaki T, Ikeda H, Yamaguchi T (2001). Genetic improvement of vase life of carnation flowers by crossing and selection. Scientia Horticulturae 87:107-120.

Palumbo R, Hu J, Craig R, Locke J, Krause C, Tay D (2007). Target region amplification polymorphism (TRAP) as a tool for detecting genetic variation in the genus Pelargonium. Horticulture Science, 43(5):1118-1123

Piepho H-P, Ogutu JO, Schulz-Streck T, Estaghvirou B, Gordillo A, Technow F (2012). Efficient computation of ridge-regression best linear unbiased prediction in genomic selection in plant breeding. Crop Science, 52:1093-1104.

Renou JP, Aubry C, Serveau M, Jalouzot P (1997). Evaluation of the genetic variability in the genus Pelargonium using RAPD markers. Journal of Hoticultural Science, 72(2):229-237.

Singh BD, Singh AK (2015). Marker-assisted plant breeding: principles and practices. Springer, New Dehli, Heidelberg, New York, Dordrecht, London.

USDA (2016). Floriculture Crops 2015 summary. ISSN: 1949-0917

Van Raden PM (2008). Efficient methods to compute genomic predictions. Journal of Diary Science, 91:4414-4423.

Wang B, Sverdlov S, Thompson E (2017). Efficient estimation of realized kinship from single nucleotide polymorphism genotypes. Genetics, 205:1063-1078.

Xu Y, Crouch JH (2008). Marker-assisted selection in plant breeding: from publications to practice. Crop Science, 48:391-407.

Yates F (1936). A new method of arranging variety trails involving a large number of varieties. Journal of Agricultural Science, 26:424-455.

Acknowledgement

First of all, I am grateful to my academic supervisor Prof. Dr. Hans-Peter Piepho for giving me the opportunity to write this thesis, for his advice, suggestions, and continuous support with all manuscripts written during the thesis.

I also sincerely thank Dr. Ulrich Sander, Dr. Andrea Dohm and Dr. Robert Boehm for their contributions, support and for providing such a stimulating environment for the application of biostatistics research at Selecta One.

Special thanks to Martin Glawe for teaching me what the ‘breeder’s eye’ in ornamentals must focus on, for discussing the practicality of theory and for helping me to find the best combination of theory and practice.

I also sincerely thank all technical assistants, especially Nadja Beuttenmüller, at Selecta One who helped to carry out the extensive rooting trials.