

# Improving yield prediction of synthetic cultivars in case of partial allogamy

Judith Reese<sup>1,2</sup>, Lisa Brünjes<sup>1</sup>, and Wolfgang Link<sup>1</sup>

<sup>1</sup>Department of Crop Sciences, Division of Plant Breeding Methodology, University of Göttingen, Carl-Sprengel 1, Göttingen, 37075, Germany

<sup>2</sup>Züchtung Winterraps, W. von Borries-Eckendorf GmbH & Co KG, 33818 Hovedissen, Germany

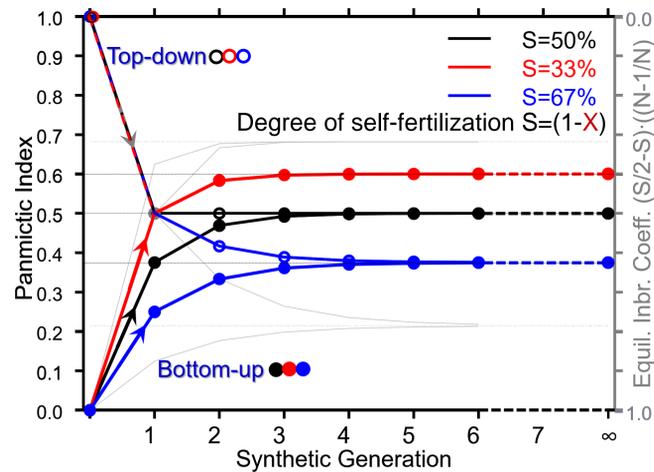
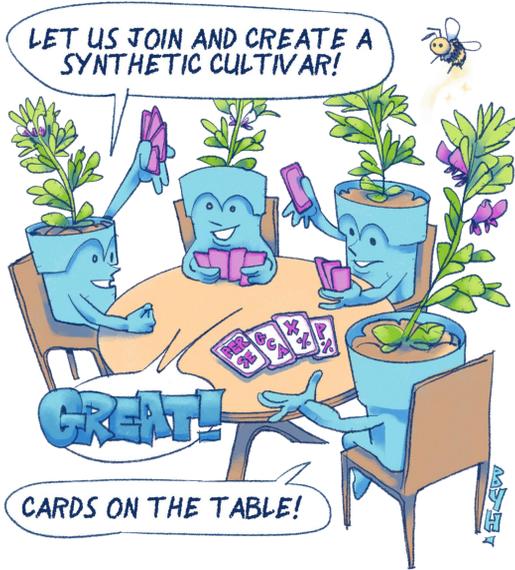
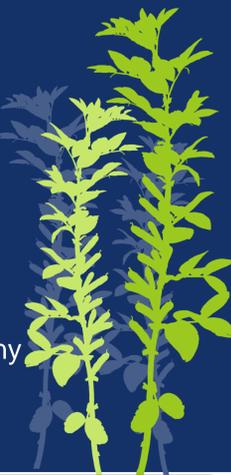


Fig. 1 Initiate a synthetic (mixing inbred lines) or **top-down** (mixing F<sub>2</sub> from manually crossed F<sub>1</sub>); to guarantee equal share contributions of the components to the synthetic cultivar)

Faba bean is partially allogamous. Breeders create synthetic faba bean cvs. usually by mixing inbred lines („Syn-0“). After three generations of partial allogamous reproduction, the inbreeding coefficient *F* is reduced to ~minimum, i.e. the Panmictic Index (=1-*F*) has increased; seed is sold to farmers. This approach is **bottom-up** (Fig.1). Here, we present algebra and results from simulations to predict performance in **bottom-up** Syn-1 using data on yield (*Y*; *per se*, GCA) and on cross-fertilization *X* and paternal outcrossing success *P* of the inbred lines (components of synthetic).

Components (usually 2<*N*<8) differ e.g. in *X* and *P*. Thus, their genetic share in generation Syn-1 ff. deviate from 1/*N*; hence the panmictic index of the cv. is lowered and less heterosis exploited. To prevent this, the synthetic may be initiated with e.g. 6 crosses among *N*=4 lines. Then, create and mix F<sub>2</sub>s of these crosses in equal shares. This F<sub>2</sub> mixture is the **top-down** Syn-1 (Fig.1). Further propagation leads to a panmictic index in Syn-4 similar to **bottom up**.

Algebraic decomposition of performance in generation Syn-1, showing its dependence on parameters of components (inbreds); used to study the individual parameters' importance in the prediction of **bottom-up** Syn-1\*

$$\text{Predicted Syn-1} = \mu_L + \mu_X \left(1 - \frac{1}{N}\right) (\mu_C - \mu_L) + \frac{1}{N} \sum_{i=1}^N [(\mu_X + X_i) P_{ii}] \mu_L + \frac{1}{N} \sum_{i=1}^N [1 - \mu_X \left(1 - \frac{1}{N}\right)] L_i + \frac{1}{N} \sum_{i=1}^N [(\mu_X + X_i) P_{ii}] L_i + \frac{1}{N} \sum_{i=1}^N \mu_X \left[ \left(1 - \frac{1}{N}\right) (2 \text{ GCA}_i) \right] + \frac{1}{N} \sum_{i=1}^N [X_i \left(1 - \frac{1}{N}\right) (\mu_C + \text{GCA}_i - \mu_L - L_i)] + \frac{1}{N^2} \sum_{i \neq j}^N [X_i (\text{GCA}_j) + \frac{1}{N} \sum_{i \neq j}^N [(\mu_X + X_i) P_{ij} (\mu_C + \text{GCA}_i + \text{GCA}_j)]$$

\**L*=*per se* yield of inbred lines; *C*=F<sub>1</sub>-hybrid yield; *X*=degree of cross-fertilization; *P*=paternal outcrossing success (success when siring others)

**Bottom-up (A):** Prediction of Syn-1 is not perfect, because differences between lines for *per se* yield, GCA for yield, degree of cross-fertilization *X*, paternal outcrossing success *P* are realized.

**Top-down (B):** Prediction of Syn-1 (i.e. mixture of F<sub>2</sub>) is perfect. Prediction of Syn-∞ (C) is a different ball game and can only be executed **top-down**.

Breeders predict yield of **bottom-up** created synthetics in Syn-1 via *per se* yield of their inbred components. What is the impact of knowing *per se* yield, GCA for yield, *X*, *P*, when predicting? *N*=4 component inbred lines *per* synthetic

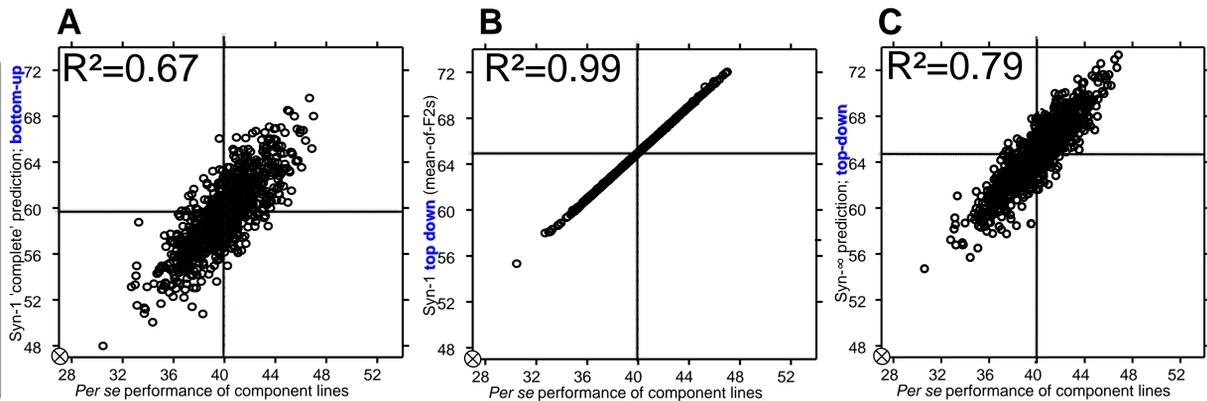
Parameters	Expl. Var (%)
<i>Per se</i> or GCA as only parameter for pred. of Syn-1	67.00
Degree of cross-fertilization <i>X</i> Added to <i>per se</i> , GCA, <i>P</i> for prediction	21.90
As only parameter for prediction	25.09
Paternal outcrossing success <i>P</i> Added to <i>per se</i> , GCA, <i>X</i> for prediction	11.10
As only parameter for prediction	14.69

*N*=4 components in Syn-0; *per se*=*per se* yield of inbred lines, GCA=their general combining ability for yield, *X*=their degree of cross-fertilization, *P*=their paternal outcrossing success (success when siring others)

Assumptions\* to simulate 3200 inbred lines and their hybrids; to predict **bottom-up** synthetics in generation Syn-1 and Syn-∞

Parameter	Mean	Standard Deviation
Degree of cross-fertilization	50.0%	8.0%
Paternal outcrossing success* <i>N</i> =4	25.0%	8.6%
<i>Per se</i> yield of inbred lines	40 dt ha <sup>-1</sup>	5.0 dt ha <sup>-1</sup>
GCA (yield) of inbred lines	2.5 dt ha <sup>-1</sup>	2.5 dt ha <sup>-1</sup>
SCA for pairs of inbred lines	0.0 dt ha <sup>-1</sup>	0.0 dt ha <sup>-1</sup>
Yield of F <sub>1</sub> -hybrids	90 dt ha <sup>-1</sup>	7.1 dt ha <sup>-1</sup>

\*based on data from literature, e.g. Brünjes et al., 2021 (DOI 10.1007/S00122-021-03832-z)



## Take home message

*Per se* yield of inbred components explained 67% of variance between faba bean synthetics in **bottom-up** Syn-1 (*N*=4). Adding data on components' degree of cross-fertilization *X* allows to explain 67.0%+21.9%=88.9%; then-adding data on *P* explained 100%. **Top-down** allows to better (99% > 67%) predict Syn-1 from *per se* yield. Approaching the cultivar's equilibrium inbreeding coefficient (i.e. max. heterosis exploitation) is faster with **top-down** (given 33% < *X* < 67%). Increase in heterosis exploitation with **top-down** (compared to **bottom-up**) is very small.

With 3200 simulated candidate inbreds and *N*=4 inbred components per synthetic, a large number of different synthetics can be created:  $(3200 \cdot 3199 \cdot 3198 \cdot 3197) / (4!) \approx 4.361 \cdot 10^{12}$ . We did not simulate them all but only 800 (taking 800 random, non-overlapping sets of four; 3200=800·4).

Breeders prefer to 'just' mix seed of superior inbreds (Syn-0) and let such mixture propagate to Syn-1 and beyond. With partial allogamy, such synthetic reaches its inbreeding equilibrium (= inbreeding minimum) in about Syn-4. At about Syn-4, seed is sold to farmers. This is the **bottom-up** approach - as the inbreeding minimum is approached from the 'bottom' (depth) of inbreeding. **Yes**. **Bottom-up** causes deviations of the cultivar from expectation, which is based on the inbreds' *per se* yield, average level of heterosis (& maybe on degree of cross-fertilization of the inbreds); and on their number *N*. **Bottom-up** allows the components to genetically contribute unequally to their synthetic. Components with more seeds per plant in Syn-0 contribute a higher share than 1/*N* to their Syn-1. Further, components with the higher degree of cross-fertilization *X* in Syn-0 contribute a genetic share <1/*N*, because a larger share of their seeds carries their genes only via maternal gamete. Even further, on the paternal side, the components with higher paternal outcrossing success *P* contribute more than 1/*N*. Unequal contributions reduce the population's effective size, hence increase inbreeding, hence decrease share of heterosis: **decrease vigour**.

An alternative to improve the situation is the **top-down** approach: Manually cross the components to prevent unequal contributions! Here, the inbreeding equilibrium is approached **top-down** (from the 'peak' of heterozygosity). With *N*=4, one can realize 2 or 4 or 6 crosses to have equal-share of the 4 components. From these F<sub>1</sub>, produce F<sub>2</sub> and mix them (no selection; called **top-down** Syn-1). Interestingly, the Panmictic Index in F<sub>2</sub> is - accidentally - identical to the equilibrium Panmictic Index in Syn-∞ (if *X*=50% and *N*=4; realistic values in faba bean). **Message**: With *X*=50% and *N*=4, the **top-down** Syn-1 (F<sub>2</sub>-mixture) already arrived at its inbreeding minimum. Moreover: With 33% < *X* < 67% and *N*=4, **top-down** approaches the inbreeding equilibrium faster (is nearer to it in e.g. Syn-4) than **bottom-up** :-)

**Top-down** almost fully prevents unequal shares. Hence, expected performance for **top-down** Syn-∞ can indeed be formulated! This is not the case for the **bottom-up** Syn-∞; because with **bottom-up**, the components 'themselves' are still in some frequency present in Syn-1 and beyond and still influence composition and average inbreeding coefficient of subsequent generations.

How important are unequal genetic shares? Let's take one extreme scenario for *N*=4: two components have *X*=0% and two components have *X*=100%. This causes deviation from 'each-contributes 1/4': now 'two-give 5/16' (*X*=0%) and 'two-give 3/16' (*X*=100%). The average Panmictic Index of out-crossed individuals in such population (Syn-1 onwards) is less than  $(N-1)/N=0.750$ , it is 0.7344. For the synthetic, this means an approximate reduction of its equilibrium Panmictic Index from 0.50 down to 0.4895. Admittedly, a very small loss.

While we're at it :- ) ... breeders may intentionally create synthetics with unequal share. Take *N*=4 and let the components deviate from 25% by amounts *a, b, c, d* per component. For example, create the Syn-0 with 100 seed of each of three inbreds but add 700 seed from the fourth component; hence contributions are 10%, 10%, 10%, 70% (deviations from 25% are *a, b, c*=15% and *d*=45%). The resulting increase in inbreeding,  $\Delta F$ , of the outcrossed individuals of that population is:  $\Delta F = (a+b+c)^2 + a^2 + b^2 + c^2$  (here,  $\Delta F=0.52$ , an increase from *F*=0.25 to *F*=0.77; a marked loss of the heterotic potential). For the entire synthetic, this means a reduction of its average Panmictic Index from 0.50 down to 0.15 - this now is a marked loss.

Although variation in *X* between components in **bottom-up** Syn-0 causes no marked impact on heterosis exploitation, nevertheless: Ignoring differences in *X* between components leaves ~21.9% of variation between **bottom-up** Syn-1 unexplained. Unequal shares are caused by coincidental combinations of e.g. high (or low) *per se* yield and high (or low) *X* of a component in **bottom-up** Syn-0.

With *N*=4, randomly occurring associations between *X* and *per se* yield (and GCA) have to be expected. In our basic pool of the simulated 3200 lines, that correlation was *r*=0.00, yet, it is *r*>0 in a typical quartet of components (initiating any of our 800 synthetics). In most Syn-0, there was a negative or a positive correlation, the mean absolute value was *r*=0.494. This association is the reason for the improvement of the prediction of **bottom-up** Syn-1, if *X* of the individual component is considered (in addition to *per se* yield). Such association is almost resolved with the **top-down** approach, especially if you realize all six possible cross-combinations and if yield loci are unlinked with loci for reproductive behaviour. **Nota bene**: The prediction of the **bottom-up** Syn-1 beyond Syn-1, for Syn-4 or Syn-∞, is a different game, and *R*<sup>2</sup> (using *per se* performance) would be rather -0.60 than 0.67. And: If you select among components based on their -polycross-progeny-test instead of their yield *per se* data, you are closer to bottom-up - this approach was not studied here.

If the *per se* yield was considered, GCA was implicitly considered, because the correlation between *per se* yield and GCA for yield was assumed as *r*=1 (Quantitative Genetic expectation in case of no epistasis and degree of dominance 0 ≤ *d/a* ≤ 1).

Further assumptions to be mentioned: No genetic differences in fitness, i.e. no differences in seed numbers per plant. No heterosis for *X*. Linear relationship between inbreeding coefficient and yield heterosis. No contamination in Syn-0 and later with alien pollen. The mixture of the **top down** F<sub>2</sub> bulks was defined as **Syn-1 top down**, thus equated with the Syn-generation of **bottom up** Syn-1.

Breeders may argue in favour of **bottom-up** for reasons of easier feasibility; no need to manually cross. Yet, in faba bean, propagation of inbreds via controlled selfing involves a certain effort, not much different from selfing F<sub>1</sub> to get F<sub>2</sub>. Often, line breeding precedes synthetic breeding, hence, crosses and F<sub>2</sub> may be ready-at-hand. Besides, lately breeders dispose on highly homozygous individuals to be used for Genomic Selection, hence ready for crossing - but not ready for field testing (because not propagated yet).

**This methodological discussion seems to not yet be finished.**

Brünjes, Lisa, and Wolfgang Link W, 2021: Paternal outcrossing success differs among faba bean genotypes and impacts breeding of synthetic cultivars. Theoretical and Applied Genetics 134: 2411-2427.

Kelly, John. K., and Scott Williamson, 2000: Predicting response to selection on a quantitative trait: A comparison between models for mixed-mating populations. Journal of Theoretical Biology 207: 37-56.

Link, Wolfgang, Winfried Ederer and Peter Ruckebauer, 1991: Züchtung von synthetischen Sorten bei der Fababohne: Detaillierte Untersuchungen über die Züchtungsmethode auf experimentellem und theoretischem Gebiet. BMEL, Forschungsdokumentation Nachwachsende Rohstoffe, Ausgabe 1991: 253 - 262.

Link, Wolfgang, and Lisa Brünjes, 2017: Ein Neustart zur Vorhersage der Syn-1 Generation von partiell allogamen, synthetischen Sorten mit konkretem Bezug zur Ackerbohne. Georg-August Universität Göttingen, Göttingen. Unveröffentlichtes Manuskript.

Reese, Judith, 2022: Breeding method simulations to predict partially allogamous, synthetic faba bean cultivars from their parental components (number of components, *per se* performance, GCA, degree of cross-pollination, paternal success (original in German language), MSc Thesis, Georg-August-Universität Göttingen

