Euphytica (2023) 219:48
https://doi.org/10.1007/s10681-023-03174-5

Optimizing multi-environment testing in potato breeding: using heritability estimates to determine number of replications, sites, and years for field trials

Rodomiro Ortiz · Fredrik Reslow · José Huicho · Ramesh R. Vetukuri · José Crossa

Received: 27 September 2022 / Accepted: 8 March 2023

Abstract Multi-environment trials (METs) of potato breeding clones and cultivars allow to precisely determine their performance across testing sites over years. However, these METs may be affected by the genotype × environment interaction (GEI) as noted in tuber yield. Furthermore, trials are replicated several times to optimize the predictive value of the data collected because knowledge on spatial and temporal variability of testing environments is often lacking. Hence, the objectives of this research were to use components of variance from METs to estimate broad sense heritability (H^2) based on best linear unbiased predictors and use these estimates to determine the optimum number of sites, years, and replications for testing potato breeding clones along with cultivars. The data were taken from METs in southern and northern Sweden comprising up to 256 breeding clones and cultivars that underwent testing using a simple lattice design of 10-plant plots across three sites over 2 years. Percentage starch in the tuber flesh had the largest H^2 in each testing environment (0.850–0.976) or across testing environments (0.905–0.921). Total tuber weight per plot also exhibited high H^2 (0.720–0.919) in each testing environment or across them (0.726–0.852), despite a significant GEI. Reducing sugar content in the tuber flesh had the lowest, but still medium H^2 (0.426–0.883 in each testing environment; 0.718–0.818 across testing environments). The H^2 estimates were smaller when their variance components were disaggregated by year and site, instead of lumping them as environments. Simulating H^2 with genetic, site, year, site × year, genetic × site, genetic × year, genetic × site × year, and residual variance components led to establish that two replicates at each of two sites in 2-year trials will suffice for testing tuber yield, starch and reducing sugars. This article provides a methodology to optimize the number of testing size and years for METs of potato breeding materials, as well as tabulated information for choosing the appropriate number of trials in same target population of environments.

Keywords *Solanum tuberosum* · Broad-sense heritability · Cultivar development · Early stage selection · Multi-site trials · Variance components

R. Ortiz (✉) · F. Reslow · R. R. Vetukuri
Department of Plant Breeding, Swedish University of Agricultural Sciences (SLU), P.O. Box 190, 23436 Lomma, Sweden
e-mail: rodomiro.ortiz@slu.se

J. Huicho · J. Crossa
Centro Internacional de Mejoramiento de Maíz y Trigo (CIMMYT), Carretera México-Veracruz Km. 45, El Batán, 56237 Texcoco, Edo. de Mexico, Mexico

J. Crossa
Colegio de Postgraduados, 56230 Montecillos, Edo. de México, Mexico

Published online: 29 March 2023

Introduction

Multi-environment trials of potato breeding clones along with released cultivars grown by farmers aim to get accurate results for target productivity and quality traits, which calls for maximum control of unexplained variability within a data sample, e.g. due to soil heterogeneity (Terman et al. 1967) or weather. Variability often decreases when adding replications, testing sites or years, using an appropriate experimental design or considering spatial data analysis for adjusting plot results.

Uniformity trials were used for determining shape and size for field experiments in potato (Justesen 1932). Plot shape effect seems to be minor in trials testing potato hybrids (Stockem et al. 2022). Fisher (1970) argued the degree of precision of a trial for estimating any mean depends on the replication number. In this regard, Kalamkar (1932) noted that increasing plot size decreased trial efficiency; i.e., more replications of smaller plots are better than a small number of larger plot, especially if larger plot sizes do not reduce significantly the trial variability. Furthermore, Caligari et al. (1985) indicated that the most efficient design for yield trials in potato may include a single drill or plant with as many replicates that can be managed by a breeder. However, as noticed by Bos (1983), increasing replication number, decreases the number of testing accessions in a trial, thereby counterbalancing the improvement of selection response expected from more intensive germplasm testing. Furthermore, Aikman and Langton (1983) indicated that replications had a marginal effect under high selection intensity for low heritability traits.

Although the experiment accuracy depends on the number of both testing accessions and replications, it seems that the optimum plot size for assessing total tuber weight ranges from eight to 12 hills or plants, for breeding clones (Bisogninda et al. 2006), and hybrids (Stockem et al. 2022), respectively. Guard rows are often included when the tuber yield of one plot affects that of the adjacent plot (e.g. in fertilizer trials), may be also used for cultivar testing (Mountier 1964), but at increasing cost. Nevertheless, Knight (1924) demonstrated that replicated single rows provide reliable results in potato's field experiments. Blocking improves the efficiency of potato cultivar trials (Mountier 1985). Lattice or incomplete block designs are preferred when the breeding clones or cultivars included for testing are large (> 20).

Enhancing accuracy in germplasm testing leads to an enlarged heritability, which results in increasing the expected response to selection. We may use heritability estimates along with field plot techniques (Vallejo and Mendoza 1992) to improve multi-environment testing of potato breeding clones along with released cultivars, particularly when the genotype-by-environment interaction affects productivity and quality traits in this crop (Yildirim and Çalişkan 1985). For example, heritability estimates confirmed the efficiency of unilateral sexual polyploidization for multi-trait selection and progeny testing in potato breeding (Ortiz et al. 1991). Hence, the objective of this research was to determine the minimum number of replications, testing sites and years for potato multi-environment trials based on the use of broad-sense heritability estimates. In this way, we will be able to optimize potato breeding trial efficiency in the cultivar pipeline.

Materials and methods

Data from multi-site trials over years of the Svenska potatisförädling run by the Swedish University of Agricultural Sciences (SLU, Alnarp, Sweden) were used for this research. The trials included up to 256 breeding clones and released cultivars grown by EU farmers (https://hdl.handle.net/11529/10548617) that underwent testing at Skåne (Helgegården and Mosslunda) and Norrland (Umeå) regions of Sweden in 2020 and 2021 (Table 1). The trials used simple lattice designs with two replications of 10-plant plots. Helgegården and Mosslunda are potato producing

Table 1 Number of advanced potato breeding clones and cultivars planted at three sites in Sweden over 2 years

Site	Year	Advanced breeding clones	Cultivars
Helgegården	2020	32	137
	2021	47	209
Mosslunda	2020	47	209
	2021	47	209
Umeå	2020	47	209
	2021	47	209

sites near Kristianstad (56° 01′ 46″ N 14° 09′ 24″ E) in southern Sweden, while Umeå (63° 49′ 30″ N 20° 15′ 50″ E) is in northern Sweden. In each site crop husbandry practices were the same as those used in potato farming. Fungicide sprays against the oomycete *Phytophthora infestans* were made only in Helgegården to avoid late blight throughout the growing season. This treatment was used to achieve tuber yield potential at this testing site.

The characteristics evaluated were total tuber yield in 10-plant plot (kg), tuber weight (kg) by size (< 40 mm, 40–50 mm, 50–60 mm, > 60 mm;) in the 10-plant plot, while percentage of starch in the tuber flesh was calculated after determining specific gravity at harvest (Schippers 1976). Potato glucose strip tests were used for measuring reducing sugars in the tuber flesh (Mann et al. 1991). Host plant resistance to *P. infestans* was evaluated over 2 years solely in Mosslunda, where the pathogen is ubiquitous and causes high late blight severity, using the area under disease progress curve (AUDPC, Fry 1978).

Analyses of the trials in each and across environments were done with META-R (Alvarado et al. 2020), which also estimated the best linear unbiased predictors (BLUPs) for the eight evaluated traits considering both the testing germplasm, sites, and years as random samples of their respective populations.

Biometrical modeling

Single-site year model

The response of the ith cultivar on the rth replicate within the bth incomplete block nested within a replicated is represented as y_{irb} in following Eq. (1):

$$y_{irb} = \mu + C_i + R_r + IB(R)_{b(r)} + e_{irb} \quad (1)$$

where μ is the overall mean, C_i is the random effect of the ith cultivar assumed to have an independent and identical distribution (iid) that is normal with mean zero and variance σ^2_C, that is, $C_i \overset{iid}{\sim} N\left(0, \sigma^2_C\right)(i = 1, 2, \ldots, I)$, and R_r is the random effect of replicates with iid normal distribution and variance $\sigma^2_R, R_r \overset{iid}{\sim} N\left(0, \sigma^2_R\right)(r = 1, 2, \ldots, R)$. The incomplete blocks nested within replicate are considered a random effect iid with normal distribution with mean zero and variance $\sigma^2_{IB(R)}$ such that $IB(R)_{b(r)} \overset{iid}{\sim} N\left(0, \sigma^2_{IB(R)}\right)(b = 1, 2, \ldots, B)$. The random residual error is $e_{irb} \overset{iid}{\sim} N\left(0, \sigma^2_e\right)$ with variance σ^2_e. The variance component estimations of this model are given in Table 2.

Multi-environment model

The response of the ith cultivar on the rth replicate within the jth environment and on the bth incomplete block nested within replicate and the jth environment is represented as y_{ijrb} in Eq. (2)

$$y_{ijrb} = \mu + C_i + E_j + R(E)_{r(j)} + IB(R, E)_{b(rj)} + (CE)_{ij} + e_{ijrb} \quad (2)$$

where the random effect of cultivar is $C_i \overset{iid}{\sim} N\left(0, \sigma^2_C\right)$ $(i = 1, 2, \ldots, I)$ with cultivar variance component σ^2_C, and the random effect of environment (location-year combination is $E_j \overset{iid}{\sim} N\left(0, \sigma^2_E\right)(j = 1, 2, \ldots, J)$ with environment variance component σ^2_E. The random effects of replicated nested within environments are described as $R(E)_{r(j)} \overset{iid}{\sim} N\left(0, \sigma^2_{R(E)}\right)(r = 1, 2, \ldots, R)$ with variance component $\sigma^2_{R(E)}$, while the random effect of incomplete block nested within replicate and environment is described as $IB(R, E)_{b(rj)} \overset{iid}{\sim} N\left(0, \sigma^2_{IB(R,E)}\right)$ $(b = 1, 2, \ldots, B)$ with variance component $\sigma^2_{IB(R,E)}$. The interaction effect of the cultivar × environment is described as $CE_{ij} \overset{iid}{\sim} N\left(0, \sigma^2_{CE}\right)$ with interaction variance component σ^2_{CE} and random residual that is defined as $e_{ijrb} \overset{iid}{\sim} N\left(0, \sigma^2_e\right)$ variance component σ^2_e. Variance components of this models are presented in Table 3.

Multi-site over years model

The response of the ith cultivar on the jth site, the mth year, the rth replicate within site and year, and the bth incomplete block nested within replicated site and year is represented is represented as y_{ijmrb} in below Eq. (3)

$$y_{i,j,m,r,b} = \mu + S_j + M_m + (SM)_{jm} + R(SM)_{r(jm)} + IB(RSM)_{b(jmr)} + C_i + (CS)_{ij} + (CM)_{im} + (CSM)_{ijm} + e_{ijmrb} \quad (3)$$

where the random effect of the site is represented as $S_j \overset{iid}{\sim} N\left(0, \sigma^2_S\right)(j = 1, 2, \ldots, J)$ with variance component

Table 2 Variance components (genetic [σ^2_G] and residual [σ^2_e]) and broad sense heritability (H^2) for potato tuber weight (kg 10-plant plot), percentage of starch in the tuber flesh, reducing sugars and host plant resistance to late blight (measured by the area under disease progress curve, AUDPC) for 2-year (1: 2020, 2: 2021) multi-environment testing as determined for breeding clones and released cultivars in three distinct sites in Sweden

Characteristic		Helgegården, Skåne (yield potential)			Mosslunda, Skåne (late-blight prone)			Umeå, Norrland (very long day-length)		
		σ^2_G	σ^2_e	H^2	σ^2_G	σ^2_e	H^2	σ^2_G	σ^2_e	H^2
Total tuber weight	1	8.318	1.760	0.904	14.088	4.389	0.865	5.623	0.989	0.919
	2	11.638	9.062	0.720	7.054	1.502	0.904	5.956	1.962	0.858
< 40 mm tuber weight	1	0.267	0.043	0.925	0.254	0.121	0.808	0.821	0.138	0.922
	2	0.077	0.087	0.639	0.120	0.055	0.814	0.732	0.172	0.895
40–50 mm tuber weight	1	1.317	0.368	0.877	0.992	0.523	0.791	1.207	0.600	0.801
	2	0.673	0.497	0.730	0.684	0.368	0.788	1.078	0.601	0.782
50–60 mm tuber weight	1	2.234	0.898	0.833	2.264	0.774	0.854	1.336	0.407	0.868
	2	2.540	1.968	0.721	1.301	0.577	0.818	1.165	0.546	0.810
> 60 mm tuber weight	1	6.656	1.431	0.903	7.087	2.097	0.871	0.868	0.320	0.844
	2	16.943	10.982	0.755	3.478	1.219	0.851	0.808	0.304	0.842
Percentage of starch in the tuber flesh	1	5.828	0.291	0.976	4.716	0.672	0.934	4.198	0.761	0.917
	2	5.158	1.821	0.850	5.987	1.298	0.902	6.716	0.739	0.948
Reducing sugars	1	0.292	0.757	0.436	0.221	0.399	0.526	0.249	0.275	0.644
	2	0.420	1.131	0.426	0.850	0.526	0.764	0.925	0.246	0.883
AUDPC	1	N/A			3000.614	422.273	0.934	N/A		
	2				1537.302	452.629	0.872			

Table 3 Variance components (genetic [σ^2_G], genetic × year [σ^2_{GY}], genetic × environment [σ^2_{GE}]^Z, and residual [σ^2_e]), and broad-sense heritability (H^2) for potato tuber weight (kg 10-plant plot), percentage of starch in the tuber flesh, reducing sugars and host plant resistance to late blight (measured by the area under disease progress curve [AUDPC] only in stress-prone site) estimated using 2-year multi-environmental testing at late blight-prone site, across two sites (yield potential and stressful) over 2 years in Skåne (Sweden), and across three sites (yield potential, late-blight prone, and very long days) in southern and northern Sweden

Characteristic	2 environments: Mosslunda, Skåne (late-blight prone) over 2 years				4 environments: Helgegården (yield potential) and Mosslunda (late-blight prone), Skåne over 2 years				6 environments: Helgegården (yield potential) and Mosslunda (late-blight prone), Skåne; Umeå, Norrland (very long daylength) over 2 years			
	σ^2_G	σ^2_{GY}	σ^2_e	H^2	σ^2_G	σ^2_{GE}	σ^2_e	H^2	σ^2_G	σ^2_{GE}	σ^2_e	H^2
Total tuber weight	8.365	2.263	2.939	0.818	7.802	2.938	4.289	0.856	5.206	3.707	3.312	0.853
< 40 mm tuber weight	0.140	0.048	0.088	0.753	0.113	0.056	0.079	0.825	0.202	0.179	0.107	0.839
40–50 mm tuber weight	0.527	0.327	0.440	0.658	0.475	0.404	0.444	0.752	0.363	0.614	0.498	0.716
50–60 mm tuber weight	1.265	0.537	0.670	0.744	0.598	1.477	1.066	0.543	0.577	1.203	0.862	0.679
> 60 mm tuber weight	4.313	0.986	1.653	0.826	5.976	2.857	4.117	0.829	3.198	2.823	2.807	0.819
Percentage of starch in the tuber flesh	5.263	0.140	0.976	0.944	4.916	0.349	1.119	0.956	3.873	1.501	0.970	0.921
Reducing sugars	0.318	0.266	0.470	0.559	0.366	0.104	0.736	0.756	0.340	0.176	0.555	0.818
AUDPC	1892.162	370.320	436.887	0.865	N/A				N/A			

$^Z\sigma^2_{GE}$ = genetic × environment variance; i.e., considering each site-year as an environment

σ_S^2, the random effect of the year is $M_m \overset{iid}{\sim} N(0, \sigma_m^2)$ $(m = 1, 2, \ldots, M)$ with year variance component as σ_m^2, and the random interaction effect of site × year is $(SM)_{jm} \overset{iid}{\sim} N(0, \sigma_{SM}^2)$ with interaction variance component σ_{SM}^2. The random effect of replicated nested within site and year is assumed as $R(SM)_{r(jm)} \overset{iid}{\sim} N\left(0, \sigma_{R(SM)}^2\right)$ $(r = 1,2,\ldots,R)$ with variance component of $\sigma_{R(SM)}^2$, while the random effect of the incomplete blocks nested within replicate site and year is defined as $IB(RSM)_{b(jnr)} \overset{iid}{\sim} N\left(0, \sigma_{IB(RSY)}^2\right)(b = 1, 2, \ldots, B)$ with variance component $\sigma_{IB(RSY)}^2$. The random effects of cultivar is denoted as $C_i \overset{iid}{\sim} N(0, \sigma_C^2)$ $(i = 1,2,\ldots I)$ with variance component σ_C^2, and the random effect of the interaction of cultivar × site is described by $(CS)_{ij} \overset{iid}{\sim} N(0, \sigma_{CS}^2)$ with variance component σ_{CS}^2; the random effect of the interaction of cultivar × year is assumed $(CM)_{im} \overset{iid}{\sim} N(0, \sigma_{CM}^2)$ with variance component. The random effect of the three-way interaction of cultivar × site × year is assumed $(CSY)_{ijy} \overset{iid}{\sim} N(0, \sigma_{CSY}^2)$ with variance component σ_{CSY}^2, and the random residual is described as $e_{ijyrb} \overset{iid}{\sim} N(0, \sigma_e^2)$ with variance estimation of σ_e^2. The variance components of this model are in Table 4.

Heritability estimates

BLUPs show a high predictive accuracy even when not including pedigree information (Piepho et al. 2008), and its efficiency has been already noted for selecting among segregating offspring for tuber yield and specific gravity (Ticona-Benavente and da Silva Filho 2015). The combined analyses of variance (ANOVA) over the environments were possible due to the homogeneity of variance across each of the testing environments. The variance components for the testing germplasm and environments can be estimated using the expected mean squares of the ANOVA. Broad-sense heritability (H^2), based on the plot means for each of the six-testing environment (site–year) was estimated as:

$$H^2 = \frac{\sigma_C^2}{\sigma_C^2 + \sigma_e^2/R} \quad (4)$$

in which σ_C^2, σ_e^2 and R were the genetic variance, the residual variance, and the number of replications (=2), respectively. H^2 based on the plot means across testing environments was estimated for seven tuber traits as:

Table 4 Variance components (genetic [σ^2_G], site [σ^2_L], year [σ^2_Y], site × year [σ^2_{LY}], genetic × site [σ^2_{GS}], genetic × year [σ^2_{GY}], genetic × site × year (σ^2_{GLY}), and residual [σ^2_e]), and broad-sense heritability (H^2) for potato tuber weight (kg 10-plant plot), percentage of starch in the tuber flesh, reducing sugars and host plant resistance to late blight (measured by the area under disease progress curve [AUDPC] only in stress-prone site) estimated using 2-year multi-environmental testing across three sites in Sweden

Characteristic	σ^2_G	σ^2_L	σ^2_Y	σ^2_{LY}	σ^2_{GL}	σ^2_{GY}	σ^2_{GLY}	σ^2_e	H^2
Total tuber weight	4.5490	5.5960	1.468×10^{-6}	1.9780	2.2030	0.9571	1.3540	3.3140	0.7262
< 40 mm tuber weight	0.1727	0.2104	0.0073	0.0186	0.0715	0.0326	0.1025	0.1077	0.7230
40–50 mm tuber weight	2.489×10^{-1}	0.0307	1.594×10^{-7}	2.353×10^{-1}	0.3727	0.0901	0.2558	0.4970	0.4956
50–60 mm tuber weight	0.3154	0.9042	1.619×10^{-13}	0.0060	0.7506	0.2460	0.4675	0.8545	0.3765
> 60 mm tuber weight	2.9650	2.5600	2.907×10^{-6}	3.1790	2.2250	0.2404	0.9034	2.8140	0.7040
Percentage of starch in the tuber flesh	3.7990	6.9700	4.993×10^{-6}	7.9440	0.3813	4.220×10^{-13}	1.1160	1.0050	0.9054
Reducing sugars	0.3056	0.1522	0.4627	0.0814	0.0312	0.0996	0.0729	0.5754	0.7175

$$H^2 = \frac{\sigma_C^2}{\sigma_C^2 + \sigma_{CE}^2/E + \sigma_e^2/ER} \quad (5)$$

in which σ_{CE}^2 is the genotype × environment variance, and R and E are the number of replications and environments, respectively. Variance components and their interactions were further estimated independently for sites and years to estimate H^2 for productivity and quality traits as follows:

$$H^2 = \frac{\sigma_C^2}{\sigma_C^2 + \sigma_{CS}^2/S + \sigma_{CM}^2/Y + \sigma_{CSM}^2/LY + \sigma_e^2/LYR} \quad (6)$$

in which σ_{CL}^2, σ_{CY}^2 and σ_{GLM}^2 are the genotype × site, the genotype × year, and the genotype × site × year interactions, respectively, while L is the number of testing sites (=3) and M is the number of years (=2).

The minimum number of replications, sites and years can be determined using the estimated variance components from the data. These variance components for tuber weight, percentage of starch in the tuber flesh and reducing sugars in the tuber flesh were used to estimate H^2 assuming they were stable while the denominator coefficients L, Y and R could vary. Schutz and Bernard (1967) and Ortiz et al. (2008) used a very similar approach with the phenotypic variance (instead of H^2) estimates to examine the influence of experimental design on results in future experiments testing soybean and maize germplasm. The minimum option is given by the least number of L, Y and R that will not affect H^2 estimates. Furthermore, a curve resulting from plotting the number of environments (sites or years) or replications in the horizontal axis and H^2 estimates in the vertical axis was used to allow visualizing the critical point in which this curve starts to plateau (Duma et al. 2020); i.e., beyond this point an increase in the number of testing environments provides only a negligible gain in precision.

Results

There were highly significant differences ($P < 0.001$) among BLUPs for all productivity and quality characteristics in each and across testing environments. Helgegården had, on average, the largest tuber harvests in 10-plant plots (14.2 kg in 2021, 10.83 kg in 2022), while the lowest were, on average, in the late blight prone Mosslunda (7.915 in 2020, 6.512 in 2021), and in Umeå (7.192 in 2020, 7.567 in 2021). The AUDPC BLUPs for host plant resistance to late blight were highly significant ($P < 0.001$) among genotypes but not ($P > 0.05$) across both years at Mosslunda (234.1 in 2020, 240.9 in 2021). The percentage of starch in the tuber flesh was highest in Helgegården (above 14%), thereafter in Mosslunda (between 12 and 13.5%) and the lowest in Umeå (below 10.5%). Reducing sugars vary significantly across sites and years (ranging on average from 1 in Umeå 2021 to 3.4 in Mosslunda 2020). In Umeå, the percentage of tuber weight according to size was very similar across years: 21% for below 40 mm, 41% for 40–50 mm, 26% for 50–60 mm, and 12% for above 60 mm. Tubers were larger in the testing sites at Skåne than Norlland; i.e., on average above 2/3 of the total weight for tubers above 50 mm in Helgegården and about 3/5 in Mosslunda. The genotype by environment interaction (GEI) was also highly significant ($P < 0.001$) for all characteristics.

The highest H^2 estimates for each site (Table 2) were mostly for percentage of starch in the tuber flesh (0.85–0.98). The AUDPC due to late blight had high H^2 estimates (0.87–0.93) in both years at Mosslunda. Total tuber weight per plot had also a high H^2 in each testing environment (ranging from 0.72 in the highest yielding Helgegården 2021 to 0.92 in the low yielding Umeå 2021. H^2 estimates, on average for tubers below 40 mm or above 60 mm were greater than those for the other two tuber sizes. Reducing sugars in the tuber flesh had the lowest H^2 estimates on average.

Heritability estimates were larger for weight of tubers below 40 mm and reducing tubers in the flesh when including more testing sites (Table 3). H^2 decreases slightly for total tuber yield and percentage of starch in the tuber flesh when adding two more testing environments (Fig. 1). There were no H^2 trends according to the number of testing environments for weight of tubers with 40–50 mm and 50–60 mm sizes (Table 3), while the H^2 for AUDPC due to late blight over years (0.86) was smaller than those estimated in each year at Mosslunda.

Disaggregating the variance component of environment into testing sites and years led to smaller H^2 estimates (Table 4) than those when lumping them together as environments. The highest H^2 estimate,

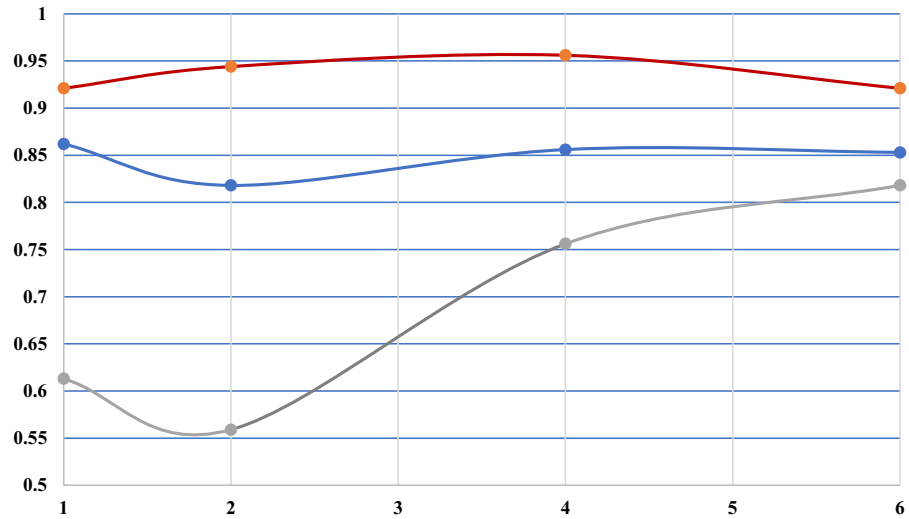

Fig. 1 Broad-sense heritability estimates according to the number of testing environments for percentage of starch in the tuber flesh in red, tuber weight (10-plant plot) in blue, and reducing sugars in grey using trials data after testing in 3 Nordic sites (northern and southern Sweden) over 2 years

after disaggregating into testing sites and years, was again for percentage of starch in the tuber flesh (0.90), while the lowest were for weight of tubers with 40–50 mm (0.50) and 50–60 (0.38) mm sizes. Medium–high H^2 were estimated for total tuber weight (0.76), and weight of tubers below 40 mm (0.72) or above 60 mm (0.74) and reducing sugars in the tuber flesh (0.72). The magnitude of the variance component for the genotype × location (σ^2_{GL}) interaction was larger than that of genotype × year (σ^2_{GY}) interaction for most tuber traits except the variable reducing sugars in the tuber flesh as measured by the sugar strip test. The variance component for the genotype × location × year (σ^2_{GLY}) was larger than the σ^2_{GL} and σ^2_{GY} for percentage of starch in the tuber flesh and weight of tubers below 40 mm and with 40–50 mm size, but smaller than the σ^2_{GL} for total tuber weight and weight of tubers 50–60 mm size and above 60 mm, and than σ^2_{GY} for reducing sugars in the tuber flesh.

Table 5 provides the results of simulating H^2 when keeping unchanged the variance components for tuber weight, percentage of starch in the tuber flesh and reducing sugars in the tuber flesh (Table 4) but varying the number of testing sites, years, and replications. It appears clearly H^2 estimates are larger that by increasing any of them but it will be most costly to run the multi-environment testing. Hence, the tabulated data allows detection of the plateau beyond which an increase in the number of testing sites, years and replications only will result in a negligible gain in the H^2 estimate. Accordingly, it seems that multi-environment trials using incomplete block designs with two replications across two sites over 2 years will suffice to estimate H^2 reliably. Table 5 further assists understanding why selection in early generations (non-replicated single hill in first clonal generation [T_1] or larger plots in second clonal generation [T_2]) does not seem to be efficient for total tuber weight because of low heritability estimates in such trials. Using trials with at least two replications or even better if testing occurs with multi-environment trials (e.g. from T_4 onwards as done by Svenska potatisförädling) at the target population of environments provides means for identifying more precisely promising breeding clones during potato cultivar development.

Discussion

Broad-sense heritability is the percentage of the phenotypic variance accounted by genetic differences due to significant variability amongst genotypes (Schmidt et al. 2019b). H^2 is also associated with the coefficient of determination (R^2) of a linear regression ($P = \mu + bG$) of the unobservable genotypic value (G_i) on the observed phenotype (P_i), or to the squared correlation between predicted phenotypic value and genotypic value. It is of further interest to plant breeding because H^2 may be used in the genetic gain (Δ_G) equation to predict response to selection

Table 5 Simulated broad-sense heritability with varying number of testing sites (L), years (Y) and replications (R) for tuber weight, percentage of starch in the tuber flesh and reducing sugars. Bold numbers indicate testing clonal selections (T_2, T_3, T_4) at one site (S = 1) in one (Y = 1),) or over two (Y = 2) and 3 years (Y = 3), respectively, using non-replicated plots (R = 1) in the field

L	Y	Total tuber weight (kg 10-plant plot)						Percentage of starch in the tuber flesh						Reducing sugars					
		1 R	2 R	3 R	4 R	5 R	10 R	1 R	2 R	3 R	4 R	5 R	10 R	1 R	2 R	3 R	4 R	5 R	10 R
1	1	**0.368**	0.424	0.447	0.460	0.468	0.484	**0.603**	0.655	0.675	0.685	0.691	0.704	**0.282**	0.383	0.436	0.468	0.489	0.539
	2	**0.476**	0.521	0.538	0.547	0.552	0.563	**0.725**	0.761	0.774	0.781	0.785	0.793	**0.430**	0.539	0.589	0.617	0.636	0.676
	3	**0.527**	0.563	0.577	0.583	0.587	0.596	**0.777**	0.805	0.815	0.819	0.822	0.828	**0.521**	0.623	0.667	0.691	0.706	0.739
	4	0.558	0.587	0.598	0.604	0.607	0.614	0.806	0.829	0.836	0.840	0.842	0.847	0.583	0.676	0.714	0.735	0.748	0.775
	5	0.578	0.603	0.612	0.616	0.619	0.625	0.825	0.843	0.850	0.853	0.855	0.859	0.628	0.713	0.746	0.764	0.775	0.798
	10	0.622	0.636	0.641	0.644	0.645	0.648	0.865	0.875	0.878	0.880	0.881	0.883	0.743	0.798	0.819	0.830	0.836	0.849
2	1	0.509	0.561	0.580	0.591	0.597	0.611	0.752	0.792	0.806	0.813	0.817	0.826	0.410	0.508	0.552	0.577	0.594	0.629
	2	0.623	*0.661*	0.675	0.682	0.686	0.694	0.841	*0.865*	0.873	0.877	0.880	0.885	0.573	*0.663*	0.699	0.719	0.731	0.757
	3	0.674	0.703	0.713	0.718	0.721	0.728	0.875	0.892	0.898	0.901	0.903	0.906	0.661	0.737	0.767	0.783	0.792	0.812
	4	0.703	0.726	0.734	0.738	0.741	0.746	0.893	0.906	0.911	0.913	0.914	0.917	0.715	0.781	0.806	0.819	0.827	0.843
	5	0.721	0.741	0.747	0.751	0.753	0.757	0.904	0.915	0.919	0.921	0.922	0.924	0.753	0.810	0.831	0.842	0.849	0.863
	10	0.761	0.771	0.775	0.777	0.778	0.780	0.928	0.933	0.935	0.936	0.937	0.938	0.841	0.875	0.887	0.894	0.897	0.905
3	1	**0.583**	0.628	0.644	0.653	0.658	0.669	**0.820**	0.851	0.861	0.867	0.870	0.877	**0.484**	0.570	0.607	0.626	0.639	0.666
	2	**0.696**	0.726	0.737	0.743	0.746	0.753	**0.888**	0.905	0.911	0.915	0.916	0.920	**0.645**	0.718	0.745	0.760	0.770	0.789
	3	**0.743**	0.766	0.774	0.778	0.781	0.786	**0.913**	0.925	0.929	0.932	0.933	0.935	**0.725**	0.785	0.807	0.819	0.826	0.840
	4	0.770	0.788	0.794	0.797	0.799	0.803	0.926	0.935	0.939	0.940	0.941	0.943	0.774	0.824	0.842	0.851	0.857	0.869
	5	0.786	0.802	0.807	0.809	0.811	0.814	0.934	0.942	0.944	0.946	0.946	0.948	0.806	0.849	0.864	0.872	0.877	0.887
	10	0.822	0.830	0.833	0.834	0.835	0.837	0.951	0.955	0.956	0.957	0.957	0.958	0.879	0.904	0.913	0.917	0.920	0.925
4	1	0.630	0.668	0.682	0.689	0.693	0.702	0.859	0.884	0.892	0.897	0.899	0.905	0.531	0.607	0.638	0.654	0.664	0.686
	2	0.738	0.764	0.773	0.777	0.780	0.786	0.913	0.927	0.932	0.935	0.936	0.939	0.688	0.748	0.771	0.783	0.790	0.805
	3	0.783	0.802	0.809	0.812	0.814	0.818	0.933	0.943	0.946	0.948	0.949	0.951	0.763	0.811	0.829	0.838	0.844	0.855
	4	0.808	0.823	0.828	0.831	0.832	0.836	0.943	0.951	0.953	0.955	0.955	0.957	0.807	0.847	0.861	0.869	0.873	0.882
	5	0.823	0.836	0.840	0.842	0.844	0.846	0.950	0.956	0.958	0.959	0.959	0.961	0.836	0.870	0.882	0.888	0.892	0.899
	10	0.856	0.863	0.865	0.866	0.867	0.869	0.962	0.965	0.967	0.967	0.967	0.968	0.900	0.919	0.926	0.930	0.932	0.936
5	1	0.661	0.695	0.707	0.713	0.716	0.724	0.884	0.905	0.912	0.916	0.918	0.922	0.565	0.632	0.658	0.672	0.681	0.698
	2	0.766	0.788	0.796	0.800	0.802	0.807	0.929	0.941	0.945	0.947	0.948	0.950	0.717	0.768	0.787	0.797	0.803	0.816
	3	0.809	0.826	0.831	0.834	0.836	0.839	0.946	0.954	0.956	0.958	0.959	0.960	0.787	0.828	0.843	0.850	0.855	0.864
	4	0.833	0.846	0.850	0.852	0.854	0.856	0.954	0.960	0.962	0.963	0.964	0.965	0.828	0.861	0.873	0.879	0.883	0.890
	5	0.847	0.858	0.862	0.863	0.865	0.867	0.959	0.964	0.966	0.967	0.967	0.968	0.854	0.883	0.893	0.898	0.901	0.907
	10	0.878	0.884	0.886	0.887	0.887	0.889	0.970	0.972	0.973	0.973	0.974	0.974	0.913	0.929	0.934	0.937	0.939	0.942

Table 5 (continued)

L	Y	Total tuber weight (kg 10-plant plot)						Percentage of starch in the tuber flesh						Reducing sugars					
		1 R	2 R	3 R	4 R	5 R	10 R	1 R	2 R	3 R	4 R	5 R	10 R	1 R	2 R	3 R	4 R	5 R	10 R
10	1	0.735	0.755	0.762	0.765	0.767	0.772	0.938	0.950	0.954	0.956	0.957	0.960	0.646	0.688	0.703	0.711	0.715	0.725
	2	0.830	0.843	0.847	0.849	0.850	0.853	0.963	0.970	0.972	0.973	0.973	0.975	0.782	0.812	0.822	0.827	0.831	0.837
	3	0.867	0.877	0.880	0.881	0.882	0.884	0.972	0.976	0.978	0.978	0.979	0.980	0.841	0.863	0.871	0.875	0.878	0.883
	4	0.888	0.895	0.897	0.898	0.899	0.901	0.977	0.980	0.981	0.981	0.982	0.982	0.874	0.892	0.898	0.901	0.903	0.907
	5	0.900	0.906	0.908	0.909	0.910	0.911	0.979	0.982	0.983	0.983	0.983	0.984	0.895	0.910	0.915	0.918	0.919	0.923
	10	0.926	0.929	0.930	0.931	0.931	0.932	0.985	0.986	0.986	0.987	0.987	0.987	0.940	0.948	0.951	0.952	0.953	0.955

Bold underlined numbers indicate testing fourth clonal selections (T_4) across three sites (L = 3) in 1 year (Y = 1), combining data of T_4 and T_5 clonal selections across three sites (L = 3) over 2 years (Y = 2), or combining data of T_4, T_5 and T_6 clonal selections across three sites (L = 3) over 3 years (Y = 3). Italic font indicates optimum trial numbers

($\Delta_G = H^2 \times S$[= mean phenotypic value of the selected genotypes as a deviation from μ]), or as a descriptive measurement to determine how useful and precise are the results from cultivar trials (Schmidt et al. 2019a). The H^2 estimates in this research were mostly high, which is not surprising because, as noted by D'hoop et al. (2011), this often occurs for variable traits in asexual crops such as potato.

Defining the target populations of environments where a cultivar will be released is key in plant breeding. A large variation between sites could lead to either developing cultivars for each site or showing adaptability across sites over years. Hence, it will be necessary to know the relative magnitude of the interactions of genotypes with both sites and years to develop an efficient selection program particularly when significant GEI occurs, as often noticed in potato (Yildirim and Çalişkan 1985). Indeed, GEI leads to increasing minimum detectable differences that further reduce selection precision (Sengwayo et al. 2018). As indicated by the results, GEI was highly significant for productivity and quality traits in the multi-environment trials in Scandinavia, with the genotype × location interaction being larger than the genotype × year interaction for almost all tuber traits except the reducing sugars in the tuber flesh. It may be very suitable to select for tuber weights (total and according to sizes) and percentage of starch in the tuber flesh because genotype × location interactions are predictable, while the genotype × year are mostly unpredictable (Allard and Bradshaw 1964), which explain the significant variability noticed for reducing sugar in the tuber flesh across sites over years. Seeking stable potato cultivars for reducing sugars in the tuber flesh that perform consistently in multi-environment trials at representative sites may reduce the magnitude of the genotype × year interaction for this trait. On the other hand, multi-environment testing across sites is more important than testing over years to identify high yielding breeding clones with desired starch content in the tuber flesh for the target population of environments.

This research addresses an important topic in potato breeding; i.e., the use of an experimental design seeking to minimize the phenotypic variation (both GEI and error or residual) at a given cost (or number of plots for trials). The more environments used for trials, the lower H^2 estimates because of a large GEI, while the high germplasm diversity may

inflate H^2 (D'hoop et al. 2011). The minimum number of testing environments and replications may be, however, debatable because both depend on various factors, including the availability of planting materials. Early generation (T_1) testing uses nonreplicated 1-plant plots in the first-year field trial. There are sufficient tubers for having replicated trials in the T_3 or T_4 generation, when total tuber weight, specific gravity (as a proxy for dry matter or starch in the tuber pulp) and crisping suitability should be properly evaluated. Furthermore, the minimum number of testing sites should consider the target population of environments where the breeding clones along with cultivar checks will be included in multi-environment trials. Curves ensuing from plotting H^2 at different number of testing environments (Fig. 1) suggests that the ideal will be four for tuber weight and percentage of starch in the tuber flesh because thereafter the H^2 gain is minimal. Furthermore, a minimum of two sites over 2 years will suffice for determining accurately these traits when using simple lattice designs with two replications (Table 5) in trials of bred germplasm from T_3 (if enough planting materials available) or T_4 onwards during potato cultivar development. The saved resources resulting from reducing number of testing sites, replications and years may be used for planting more on-farm trials with advanced breeding clones (T_6 onwards), which may also provide more information about associated crop husbandry practices.

Strong selection for quantitative traits, even if they are highly heritable, based on nonreplicated small plots (1–4 plants) in the T_1 or T_2 appears to be unreliable in potato because of a significant GEI and a high error variance. Brown (1987) demonstrated that the error variance for total tuber weight of 1-plant plots was significantly greater than that of 5-plant plots. Moreover, H^2 estimates when considering potato breeding trials using non-replicated plots in one testing environment were always the lowest (Table 5). Caligari et al. (1986) indicated that the inefficiency of selection in the T_1 could be also attributed to the inaccuracy of tuber yield assessment. Hence, selection for productivity using nonreplicated breeding trials seems to be ineffective, even when considering the best breeding clones from the previous year assessment; i.e., in T_2. Trial heterogeneity in early-stage potato breeding trials calls for the use of augmented (Federer 1956) or p-rep designs (Paget et al. 2017) and spatial data analysis (Kempton et al. 1994) when using non-replicated plots, and pedigree-based BLUPs for selection of promising bred-germplasm in T_1 and T_2. As indicated by Slater et al. (2014), BLUPs that use pedigree results in increased Δ_G when having low H^2 in potato.

Ticona Benavente and Pereira Pinto (2012) indicated that family selection for tuber yield and specific gravity may be also effective in early potato breeding generations because heritability at the family level was always larger than at the breeding clone level. Inter-family variation is also more efficient than within-family variation because the former has a lower environmental effect (thus larger H^2) than among breeding clones of the same family (Simmonds 1996). Furthermore, as noted by Bradshaw et al. (1998), combining family selection in T_1 with within family selection in T_2 may lead to promising T_3 bred germplasm. This combined selection approach appears to be very appropriate when having low within family variation (Silva Melo et al. 2011); i.e., low H^2 for the desired trait among siblings.

Potato breeding trials normally involve testing of promising advances clones along with released cultivars in several environments across testing sites over years. This article provides a methodology to optimize their numbers in METs of potato breeding materials, as well as tabulated information for choosing the appropriate number of trials in same target population of environments in the cultivar development pipeline.

Acknowledgements We thank Boel Sandström and other staff of SLU in Umeå and Hushallningssallskapet staff at both Helgegården and Mosslunda for planting and managing the multi-site trials.

Author contributions RO and JC contributed to the study conception and design. Material preparation, data collection and analysis were performed by FR, JH, RRV, JC and RO. The first draft of the manuscript was written by RO and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open access funding provided by Swedish University of Agricultural Sciences. Swedish University of Agricultural Sciences (SLU) and the Swedish Research Council Formas through the Project (2019–2022) "Genomisk prediktion i kombination med högkapacitetsfenotypning för att öka potatisens knölskörd i ett föränderligt klimat".

Data availability The datasets generated during and/or analysed during the current study are available at https://hdl.handle.net/11529/10548617.

Declarations

Competing interests The authors have no relevant financial or non-financial interests to disclose.

References

Aikman DP, Langton FA (1983) Replication in initial selection trials of clonally propagated crops. Euphytica 32:821–829. https://doi.org/10.1007/BF00042163

Allard RW, Bradshaw AD (1964) Implication of genotype–environmental interaction in applied plant breeding. Crop Sci 5:503–506

Alvarado G, Rodríguez FM, Pacheco A et al (2020) META-R: a software to analyze data from multi-environment plant breeding trials. Crop J 8:745–756

Bisogninda DA, Storck L, Costa LC, Bandinelli MG (2006) Plot size variation to quantify yield of potato clones. Hort Brasil 24:485–488

Bos I (1983) The optimum number of replications when testing lines or families on a fixed number of plots. Euphytica 32:311–318. https://doi.org/10.1007/BF00021439

Bradshaw J, Dale M, Swan G et al (1998) Early-generation selection between and within pair crosses in a potato (*Solanum tuberosum* subsp. *tuberosum*) breeding programme. Theor Appl Genet 97:1331–1339. https://doi.org/10.1007/s001220051026

Brown JA (1987) Comparison between single plant plots and five plant plots for the initial selection stage of a potato breeding programme. Euphytica 36:711–718. https://doi.org/10.1007/BF00051853

Caligari PDS, Brown J, Manhood CA (1985) The effect of varying the number of drills per plot and the amount of replication on the efficiency of potato yield trials. Euphytica 34:291–296. https://doi.org/10.1007/BF00022921

Caligari PDS, Brown J, Abbott RJ (1986) Selection for yield and yield components in the early generations of a potato breeding programme. Theor Appl Genet 73:218–222. https://doi.org/10.1007/BF00289277

D'hoop BB, Paulo MJ, Visser RGF et al (2011) Phenotypic analyses of multi-environment data for two diverse tetraploid potato collections: comparing an academic panel with an industrial panel. Potato Res 54:157–181. https://doi.org/10.1007/s11540-011-9186-1

Duma SW, Shimelis H, Ramburan S, Shayanowako AIT (2020) Optimising the number of test sites, crop-years and replications to maximize post-release testing efficiencies, South African. J Plant Soil 37:61–70. https://doi.org/10.1080/02571862.2019.1646331

Federer WT (1956) Augmented (or hoonuiaku) designs. Hawaii Planter's Rec 55:191–208

Fisher RA (1970) Statistical Methods for Research Workers. Oliver & Boyd, Edinburgh, London

Fry WE (1978) Quantification of general resistance of potato cultivars and fungicide effects for integrated control of late blight. Phytopathology 68:1650–1655

Justesen SH (1932) Influence of size and shape of plots on the precision of field experiments with potatoes. J Agric Sci 22:366–372. https://doi.org/10.1017/S0021859600053685

Kalamkar RJ (1932) Experimental error and the field-plot technique with potatoes. J Agric Sci 22:373–385. https://doi.org/10.1017/S002185960005369

Kempton RA, Seraphin JC, Sword AM (1994) Statistical analysis of two-dimensional variation in variety yield trials. J Agric Sci 122:335–342

Knight HG (1924) The influence of plot size and replication on experimental error in field trials with potatoes. Bulletin N° 189. Agriculture Experimental Station, College of Agriculture, West Virginia University, Morgantown

Mann DJ, Lammerink JP, Coles GD (1991) Predicting potato crisp darkening: two methods for analysis of glucose. N Z J Crop Hortic Sci 19:199–201. https://doi.org/10.1080/01140671.1991.10421799

Mountier NS (1964) Plot size and guard rows in potato experiments. N Z J Agric Res 7:180–197. https://doi.org/10.1080/00288233.1964.10418150

Mountier NS (1985) Techniques for the design of potato cultivar trials. In: Techniques for trial design. Agronomy Society of New Zealand, Palmerston North, New Zealand, pp 29–31. https://www.agronomysociety.org.nz/files/SP3_5._Design_techniques_for_potato_cv_trials.pdf. Accessed on 14th Sept 2022

Ortiz R, Peloquin SJ, Freyre R, Iwanaga M (1991) Efficiency of potato breeding using FDR 2n gametes for multitrait selection and progeny testing. Theor Appl Genet 82:602–608. https://doi.org/10.1007/BF00226797

Ortiz R, Crossa J, Sevilla R (2008) Minimum resources for phenotyping morphological traits of maize (*Zea mays* L.) genetic resources. Plant Genet Resour 6:195–200. https://doi.org/10.1017/S1479262108994168

Paget MF, Alspach PA, Anderson JAD et al (2017) Replicate allocation to improve selection efficiency in the early stages of a potato breeding scheme. Euphytica 213:221. https://doi.org/10.1007/s10681-017-2004-3

Piepho HP, Möhring J, Melchinger AE et al (2008) BLUP for phenotypic selection in plant breeding and variety testing. Euphytica 161:209–228. https://doi.org/10.1007/s10681-007-9449-8

Schippers PA (1976) The relationship between specific gravity and percentage dry matter in potato tubers. Am Potato J 53:111–122

Schmidt P, Hartung J, Bennewitz J, Piepho H-P (2019a) Heritability in plant breeding on a genotype-difference basis. Genetics 212:991–1008. https://doi.org/10.1534/genetics.119.302134

Schmidt P, Hartung J, Rath J, Piepho H-P (2019b) Estimating broad-sense heritability with unbalanced data from agricultural cultivar trials. Crop Sci 59:525–536. https://doi.org/10.2135/cropsci2018.06.0376

Schutz WM, Bernard RL (1967) Genotype × environment interactions in the regional testing of soybean strains. Crop Sci 7:125–130

Sengwayo S, Zhou M, Labuschagne M (2018) Trends and magnitudes of genotype × environment interaction variance components for yield, quality and agronomic traits among coastal short cycle sugarcane breeding populations. S Afr J Plant Soil 35:41–50. https://doi.org/10.1080/02571862.2017.1331383

Silva Melo D, Pereira Pinto CAB, Santos Peixouto L et al (2011) Early selection of full-sib potato families. Ciênc Agrotecnol Lavras 35:1101–1109

Simmonds NW (1996) Family selection in plant breeding. Euphytica 90:201–208. https://doi.org/10.1007/BF00023859

Slater AT, Wilson GM, Cogan NOI et al (2014) Improving the analysis of low heritability complex traits for enhanced genetic gain in potato. Theor Appl Genet 127:809–820. https://doi.org/10.1007/s00122-013-2258-7

Stockem JE, Korontzis G, Wilson SE et al (2022) Optimal plot dimensions for performance testing of hybrid potato in the field. Potato Res 65:417–434. https://doi.org/10.1007/s11540-021-09526-9

Terman GL, Covell MR, Cunningham CE (1967) Effect of size of plot experimental design and replication on efficiency of potato fertilizer experiments. Am Potato J 34:59–68

Ticona Benavente CA, Pereira Pinto CAB (2012) Selection intensities of families and clones in potato breeding. Ciênc Agrotecnol Lavras 36:60–68

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

Vallejo RL, Mendoza HA (1992) Plot technique studies on sweetpotato yield trials. J Am Soc Hortic Sci 117:508–511

Yildirim MB, Çalişkan CF (1985) Genotype × environment interactions in potato (*Solanum tuberosum* L.). Am Potato J 62:371–375. https://doi.org/10.1007/BF02855608