Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Quantifying the impact of missing phenotypic data on genetic variance and selection reliability in horticultural clonal breeding trials

Katie M O'Connor

Katie.OConnor@dpi.qld.gov.au

Queensland Government

Ky L Mathews

Queensland Government

Jodi M Neal

Queensland Government

Research Article

Keywords:

Posted Date: November 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11571043/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Quantifying the impact of missing phenotypic data on genetic variance and selection reliability in horticultural clonal breeding trials

Katie O'Connor[1,2*], Ky Mathews[1] and Jodi Neal[1]

[1*]Department of Primary Industries, Queensland Government, Nambour, Australia, 4560, Queensland, Maroochy Research Facility, 47 Mayers Road.

[2]ARC Training Centre in Predictive Breeding, The University of Queensland, St Lucia, Australia, 4067.

*Corresponding author(s). E-mail(s): Katie.OConnor@dpi.qld.gov.au;
Contributing authors: Ky.Mathews@dpi.qld.gov.au;
Jodi.Neal@dpi.qld.gov.au;

Abstract

Key message: Over 25% missing phenotypic data adversely affects estimates of genetic variance components, heritability and ranking stability in the Australian Strawberry Breeding Program.

Missing phenotypic data is common in horticultural programs due to seasonal and environmental challenges, low yielding genotypes and poor quality fruit. Missing data can impact estimates of variance components, and thus genetic variance, heritability, and stability of genotype rankings. The Australian Strawberry Breeding Program (ASBP) aimed to identify a threshold at which missing data adversely affected genetic parameter estimations, and when the time and labour investment in phenotyping various fruit quality traits was outweighed by imprecise and inaccurate data. We performed a data masking sensitivity study to introduce missing data in increments of 5% from 0–50% over 1000 iterations for five fruit quality traits for a single time point in a real breeding trial dataset. Estimates of genetic parameters were more adversely affected when missing data was introduced at the genotype level compared to the plot level due to the loss of entire experimental units compared to a reduction in replication. We then performed multi-year trial analysis for a subset of missing data increments. Across most traits, 25% missing phenotypic data was determined as the upper threshold at which accurate genotype ranking, genetic variance and heritability were

not observably different from those without any introduced missingness. From this study, the ASBP are choosing to avoid phenotyping fruit quality traits when there are more than 25% missing plots. This study presents a methodology for determining the level of missingness acceptable to other breeding programs or large phenotyping projects.

Keywords: Missing data, variance components, simulation, ranking stability, strawberry (*Fragaria* × *ananassa*)

1 Introduction

Missing phenotypic data are common in plant breeding trials and can arise for myriad reasons (Rubin, 1976; Piepho et al, 2014; Hartung and Piepho, 2021; Aguate et al, 2019). The nature of this missingness plays a critical role in determining appropriate modelling assumptions. Data missing completely at random (MCAR) can arise entirely by chance, is unrelated to any observed or unobserved data, and is not design-dependent (Piepho et al, 2003). For instance, plot-level missingness may be MCAR due to weather and mechanical damage, or irrigation system malfunctions (Galli et al, 2018). In contrast, data missing at random (MAR) occurs when the probability of missingness depends only on observed data, for example, where low-yielding families may not produce fruit at the time of measurement. Similarly, data in multi-environment trials (METs) may be missing due to drop-outs of accessions over successive generations due to selection (Aguate et al, 2019; Smith et al, 2021; Hartung and Piepho, 2021), which is also an example of MAR given that all data used in the selection process are used in the analysis (Little and Rubin, 2019; Piepho and Möhring, 2006). Data missing not at random (MNAR), however, occurs when the missingness is related to the unobserved value itself. An example of MNAR is where certain accessions may succumb to disease or pest pressure and can therefore not be measured for other traits, leading to the loss of entire experimental units and resulting in structured missing data at the accession (genotype) level.

Plant breeding programs commonly analyse data for multiple accessions across multiple blocks, trials, seasons and environments. Linear mixed models (LMMs) are versatile and widely used for the selection of accessions in plant breeding programs (Piepho and Möhring, 2006; Piepho et al, 2008; Smith et al, 2005). An advantage of LMMs is their ability to handle incomplete or unbalanced data (Henderson, 1975). LMMs are used to estimate variance components, which are in turn used to estimate genetic (total, additive and non-additive) variances, reliability, heritability, and predicted genetic effects (e.g. Best Linear Unbiased Predictors, BLUPs) (Curnow, 1961; Robinson, 1991; Schmidt et al, 2019; Piepho et al, 2008; Oakey et al, 2007). These parameters are essential for making informed selection decisions, and can affect the ranking of accessions (Robertson, 1959). The usual approach to estimating variance parameters in a linear mixed model is residual maximum likelihood (REML) (Patterson and Thompson, 1971). REML assumes that data are MAR or MCAR, rather than

MNAR, which could bias estimates of variance components and subsequent BLUPs (Rubin, 1976; Little and Rubin, 2019; Piepho and Möhring, 2006).

Several studies have investigated the impact of missing data on variance components in plant breeding programs, with some finding that genetic estimates remained stable up to a point but were biased with high levels of missing data (Bauer et al, 2006; Galli et al, 2018; Aguate et al, 2019). In practice, missing phenotypic data are sometimes imputed (e.g. Crain et al, 2018; Arciniegas-Alarcón et al, 2020), or whole trait measurements are removed (e.g. Smallwood et al, 2019; Klápt et al, 2022) prior to analysis. However, many studies do not address clonal or partially replicated trials, nor do they provide practical guidance for managing missing data or define operational thresholds for what constitutes "too much" missingness. There is also a notable lack of studies that examine the trade-off between the effort invested in phenotyping and the risk of unreliable selection outcomes at high levels of missing phenotypes.

Missing data are a common occurrence in clonal horticultural breeding programs due to variable yield, seasonality, and the volume of traits requiring phenotyping. The Australian Strawberry Breeding Program (ASBP) is one such program that develops new varieties of strawberry (*Fragaria* × *ananassa*) across three production nodes, namely subtropical (south-east Queensland), temperate (southern Australian states and Stanthorpe in Queensland), and Mediterranean (near Perth, Western Australia). Each year, clonal trials are conducted to evaluate yield, fruit quality, and plant architecture across multiple stages of selection. As in other crops, operational constraints often lead to partially-replicated trials and unbalanced data (Cullis et al, 2006). These constraints include unequal plant numbers across genotypes at the design phase, low fruit numbers early in the season, consistently low-yielding genotypes or families, and the exclusion of poor-quality fruit from trait assessments. Given the weekly harvest schedule and the labour-intensive nature of phenotyping, a clear decision threshold for when to phenotype certain traits would be valuable for the ASBP. Furthermore, a methodology with which to assess this threshold would apply to other horticultural crops.

To support this need, the ASBP sought to determine how much missing phenotypic data can be tolerated for reliable estimates of genetic parameters. Thus, the objectives of this study are to: (i) impose increasing levels of missing phenotypic data through data masking across a range of traits; (ii) assess the impact of missingness on replication of genotypes, total, additive, and non-additive genetic variances, heritability, reliability, and ranking stability; (iii) determine a threshold of missingness that is acceptable at a single time point; and (iv) assess the implementation of this threshold in a MET analysis. To address these objectives, we used a real multi-year breeding trial dataset with randomised partially-replicated incomplete block designs, pedigree, and connectivity of genotypes between years. Given this data, true genetic values are unknown; the focus is therefore on the robustness of genetic estimates under different missingness scenarios relative to the complete observed dataset. This study aims to provide a framework for determining pragmatic stop/go thresholds for phenotyping in plant breeding programs, and will quantify practical robustness limits of variance, heritability, and selection reliability under imposed missingness.

2 Materials and methods

2.1 Plant material and experimental design

The subtropical clonal trials of the ASBP, hosted at the Maroochy Research Facility in Nambour, Queensland, provide a motivating example for this study. Clonal plants were vegetatively propagated from runners (stolons) derived from mother stock and established as plug plants in soil cells. Clones were planted in six-plant plots in late March, with weekly harvests beginning in late April and continuing through to the end of August. Each week, all fruit at $\geq$ 70% ripeness were harvested and pooled within each plot. Harvested fruit were then counted and weighed. Fruit quality assessments were performed 1–4 times throughout the season, such that different traits were often measured in different weeks. In the ASBP, the terms 'accessions', 'clones' and 'genotypes' are interchangeable and define a genetic entity. For consistency with quantitative genetic terminology the term 'genotype' is used in this paper.

A multi-year multi-week dataset from 2020 to 2024 was used, consisting of 643 genotypes and 1161 plots total (Table 1). There was one trial conducted each year using a partially-replicated experimental design with pedigree information (Cullis et al, 2006, 2020). Each year was considered an environment, and there were multiple measurements for each trait in each year. The complete data set is a longitudinal MET consisting of 5 years, with 101-200 genotypes evaluated in 240-446 plots each year, with partial-replication varying from 55% to 99.6%, Table 1. The number of genotypes in common between years ranged from 8 to 50, Table 1.

Pedigree information going back to founders for the 643 genotypes in the five year MET dataset contained 8,093 records. All genotypes in the MET dataset have pedigree information. Strawberry is an out-crossing crop with only 3.6% of genotypes in this pedigree a result of selfing. For genetic studies, it is considered functionally diploid (Hardigan et al, 2020). A coefficient of inbreeding value of 0.5 was assumed to reflect the heterozygosity expected in this out-crossing crop. The mean of the inbreeding coefficients (calculated as per Meuwissen et al (1992)) of the 643 genotypes was 0.1245, with a range from 0.000 to 0.3164.

There were two data masking activities performed in this study: the first for a single harvest week and the second for the complete MET dataset (see Section 2.3 for details). The 2024 trial was used in the first data masking exercise as it employed the most recent trait protocols and represents current practice. The 2024 clonal trial evaluated 135 genotypes from all stages in the breeding program together in 299 six-plant plots. These plots were arranged in 13 rows (beds) and 23 columns divided into 4 incomplete blocks. The blocks were grouped into early (stage 2) and advanced (stage 3 and above) stage blocks. Hence, the block sizes were uneven: 103 and 104 plots in early stage blocks and 46 plots in the two advanced blocks. Twenty-five advanced and check genotypes were allocated to both early and advanced stage blocks to avoid pseudo-replication.

Table 1: Summary of the Australian Strawberry Breeding Program's subtropical clonal trials conducted from 2020-2024. The numbers of plots, beds, columns, stage blocks, blocks, genotypes in common between years, and genotype percentage partial replication (p-rep) are presented.

Year	Plot	Bed	Column	Block	Stage Block	Genotype	p-rep	In-common genotypes			
								2021	2022	2023	2024
2020	240	16	15	3	0	101	95.0	32	11	10	8
2021	416	26	16	3	0	198	76.3		44	19	8
2022	400	25	16	3	0	180	73.9			50	18
2023	345	15	23	4	2	200	55.5				41
2024	299	13	23	4	2	135	95.6				

2.2 Phenotypic data

Many fruit quality traits are assessed in strawberry breeding programs. Here we focus on five key traits that contribute to eating quality. Firmness was assessed on 4–6 fruit per plot using an Agrosta 100 Field Digital Firmness Tester, and a mean was calculated. Flavour was scored using a 1–9 sensory scale based on a combined sample of cheeks from 3 fruit per plot (1 = very poor, 9 = excellent). Gloss was scored visually using a 1–9 ordinal scale (1 = very dull, 9 = very glossy). Soluble solids concentration (SSC; °Brix) was measured using a handheld refractometer on juice from the manual pulping of 3–6 fruit per plot, depending on fruit size. Titratable acidity (TA) was determined by acid–base titration and reported as citric acid equivalents using the same fruit samples as SSC (Table 2).

Table 2: Method of measurement and scale of traits.

Trait	**Measurement**	**Scale**
Firmness	Firmness score	1–100; 1 = very soft, 100 = very firm
Flavour	Sensory score	1–9; 1 = very poor, 9 = excellent
Gloss	Visual score	1–9; 1 = very dull, 9 = very glossy
Soluble solids concentration (SSC)	°Brix	2.6–14.3[a]
Titratable acidity (TA)	% citric acid	0.27–1.59[a]

[a] Observed phenotypic range across all records from 2020–2024.

A key focus of this study is to identify thresholds of data sufficiency to decide whether phenotyping should proceed for a given harvest week. Thus, despite multiple within-year measurements being available (Tables 3 and S1), a single harvest week in the 2024 trial, for each trait, was used in the first masking study (Table 4). To ensure that missingness effects were evaluated independently of baseline data sparsity, both traits and time points were deliberately selected to maximise initial data completeness.

Table 3: Number of harvest weeks for each trait in each year of the multi-environment trial dataset, and the number of weeks with greater than 15, 25 and 35% missing plot level data.

Year	Firmness	Flavour	Gloss	SSC	TA
2020	3	3	1	3	3
2021	3	2	1	3	3
2022	2	1	2	3	3
2023	3	1	2	2	2
2024	2	2	3	2	2
Total	12	9	9	13	13
> 15% missing	7	4	3	7	7
> 25% missing	3	1	1	2	2
> 35% missing	0	1	0	1	1

In the observed 2024 trial data, SSC had the lowest percentage of missing plots (4.7%), while Gloss had the highest (8%) (Table 4). For genotype-level missingness, TA had the lowest percentage (2.2%), and Firmness had the highest (3%). The trait with the fewest unique values was Gloss (6 unique values), which was consistent with its assessment on a discrete 1–9 ordinal scale.

Table 4: Percentage of missing data at the plot and genotype levels for quality traits measured in 2024 for harvest weeks with a minimum number of missing data. Harvest Week 1 is the first week in April.

Trait	Harvest Week	No. Unique Values	% Missing Plots	% Missing Genotypes
Firmness	19	37	7.7	3.0
Flavour	23	7	5.0	2.2
Gloss	12	6	8.0	3.0
SSC	23	56	4.7	2.2
TA	23	210	6.4	2.2

2.3 Data masking

Missing observations in the ASBP arise from both plot-level factors (e.g., wildlife damage, irrigation variability) and genotype-dependent limitations (e.g., insufficient or poor-quality fruit), and these may be considered as unstructured and structured missingness, respectively. Consequently, the missingness mechanism is best described as a mixture of missing at random (MAR) and missing not at random (MNAR), with the latter predominant in cases where trait assessment was not possible due to genotype performance.

To assess the impact of missing phenotypic data, we developed a custom function to systematically introduce missingness in increments of 5% (including the existing missing data) up to a maximum of 50% at both plot and genotype levels (pcNA 5–50%). Genotype-level masking included setting all available plot records (1–4) for a

genotype to missing (NA). This dual-level approach (plot vs genotype) reflects realistic breeding scenarios, such as random plot failure (MCAR) or genotype-level non-fruiting (MNAR). Missing values were introduced using the **sample()** function in base R (R Core Team, 2025), with phenotypic observations randomly set to NA. Masking the phenotypic data in this manner preserved the original design and variance structures for subsequent modelling.

The effect of missingness was first assessed at the selected harvest week (Table 4) for each trait to determine the level of missingness that is statistically acceptable. For each trait ($t = 5$), plot/genotype level ($l = 2$) and increment of missing data ($i = 10$) combination data masking was applied and a LMM fitted to the resulting dataset. This was repeated for 1000 independently-sampled iterations, and genetic variance components estimated and the stability of genetic parameter estimates assessed. This resulted in 100005 simulations, which includes the 0% level of missingness where there was only one dataset for each trait.

To determine the impact of the identified missingness threshold on the ASBP's usual selection decision process, a data masking simulation was performed using the MET dataset of each trait. A baseline model was fitted using the complete dataset (no masked values). Then, 1000 iterations were performed in which phenotypes were masked at the identified threshold (15%, 25% and 35%) at the plot level for all time points in each trait dataset. This resulted in 15000 simulations, and 15005 records in total which included estimates for the complete dataset. This simulation explores the worst-case scenario as the level of missingness is not this high for all time points (Table S1). The actual level of missingness in the complete dataset, across all traits, ranges from 3–52% at the genotype level, and 5–50% at the plot level for individual weeks. The number of weeks across all traits with more than 15% missing ranges from 3 (Flavour) to 7 (Firmness, SSC and TA), Table 3.

2.4 Statistical models

The statistical models for the single time point (harvest week) and MET longitudinal models are described in turn.

2.4.1 Linear mixed model: single time point

The linear mixed model (LMM) for $\boldsymbol{y}_h$, the n_h-vector of observations at the plot level for each trait at each time point h:

$$\boldsymbol{y}_h = \boldsymbol{X}_h\boldsymbol{\tau}_h + \boldsymbol{Z}_{g_h}\boldsymbol{u}_{g_h} + \boldsymbol{Z}_{p_h}\boldsymbol{u}_{p_h} + \boldsymbol{e}_h, \tag{1}$$

where $\boldsymbol{\tau}_h$ is the vector of t fixed effects with associated $n_h \times t_h$ design matrix $\boldsymbol{X}_h$, and includes an overall mean. $\boldsymbol{u}_{g_h}$ is a v_h-vector of random genotype effects with $n_h \times v_h$ design matrix $\boldsymbol{Z}_{g_h}$, $\boldsymbol{u}_{p_h}$ is a q_h-vector of random non-genetic peripheral effects with $n_h \times q_h$ design matrix $\boldsymbol{Z}_{p_h}$ and $\boldsymbol{e}_h$ is the n_h vector of residuals. For the 2024 trial, there were two non-genetic peripheral effects, related to the experimental design structure: *stage block* and *block*. The former describes the grouping of genotypes by their breeding stage (early or advanced) into stage blocks, and the latter the near-resolvable replicate blocks. These non-genetic effects were assumed mutually independent with a mean of

zero and variances of $\sigma^2_{p_{i1}}\boldsymbol{I}_2$ and $\sigma^2_{p_{i2}}\boldsymbol{I}_4$, respectively. The vector of residuals, $\boldsymbol{e}_h$, was assumed to have an identical, independent variance structure, $\sigma^2_i\boldsymbol{I}_{n_h}$.

The total genetic effects, $\boldsymbol{u}_{g_h}$, were modelled with pedigree information as per Oakey et al (2007), and are partitioned into additive and non-additive effects. The resulting genetic variance is $\sigma_{a_h}\boldsymbol{A}+\sigma_{i_h}\boldsymbol{I}_{v_h}$, where σ_{a_h} and σ_{i_h} are the additive and non-additive genetic variances, respectively. The numerator relationship matrix, $\boldsymbol{A}$, was formed from the ancestral pedigree information back to founders using the methods of Meuwissen et al (1992) assuming a diploid genetic structure and an inbreeding coefficient of 0.5.

2.4.2 Linear mixed model: multi-year longitudinal

The multi-year dataset consists of one trial conducted each year at the same location in alternating field blocks. There were $p = 5$ environments (years) with $v = 643$ genotypes evaluated across all years (not all genotypes were evaluated in each year). Each of the j trials has n_j plots arranged in a rectangular array of c_j columns by r_j beds (rows). The total number of plots was $n_p = \sum_{j=1}^{p} n_j = 1161$. Let w_j be the number of harvest weeks in the jth year, and $w = \sum_{j=1}^{p} w_j$. Note that each trait was phenotyped a different number of times per year, varying from 1 to 3 (see Table 3). Let the n-vector of phenotypic data be given by $\boldsymbol{y} = (\boldsymbol{y}_1^\top, \boldsymbol{y}_2^\top, \ldots, \boldsymbol{y}_p^\top)^\top$, where $\boldsymbol{y}_j = (\boldsymbol{y}_{j1}^\top, \boldsymbol{y}_{j2}^\top, \ldots, \boldsymbol{y}_{jw_j}^\top)^\top$ is the n_j-vector for year j and $\boldsymbol{y}_{jk}$ is the n_{jk}-vector for week k in year j. The length of $\boldsymbol{y}$ is then: $n = \sum_{j=1}^{p}\sum_{k=1}^{w_j} n_{jk}$.

The LMM for $\boldsymbol{y}$ (ordered as beds by columns within weeks, within years) can be written as:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\tau} + \boldsymbol{Z}_g\boldsymbol{u}_g + \boldsymbol{Z}_p\boldsymbol{u}_p + \boldsymbol{e}, \tag{2}$$

where $\boldsymbol{\tau}$ is a vector of fixed effects with associated design matrix $\boldsymbol{X}$, and includes an overall mean. $\boldsymbol{u}_g$ is a vpw-vector of random genotype by year by week effects with $n \times vpw$ design matrix $\boldsymbol{Z}_g$, $\boldsymbol{u}_p$ is a vector of random non-genetic peripheral effects with design matrix $\boldsymbol{Z}_p$, and $\boldsymbol{e}$ is the n vector of residuals.

To optimise selection decisions, the ASBP constructs its MET datasets following the contemporary group approach of Smith et al (2021), which minimises selection bias. This method requires the v-vector of genetic effects, $\boldsymbol{u}_g$, to be partitioned into $\boldsymbol{u}_{g_k}$ of length $v_k = 613$ and $\boldsymbol{u}_{g_d}$ of length $v_d = 30$. The $\boldsymbol{u}_{g_k}$ are those genotypes of interest when making selection decisions, and the $\boldsymbol{u}_{g_d}$ are those that are not. The observations for the not-of-interest genotypes were retained in the data to maintain the integrity of the spatial layout and experimental design. However, the $\boldsymbol{u}_{g_d}$ effects were fitted as fixed effects, i.e., they were included in $\boldsymbol{\tau}$ to remove them from the estimated genetic effects of interest.

The random effects in Equation (2) were assumed to follow a Gaussian distribution with zero mean and variance:

$$\begin{pmatrix} \boldsymbol{u}_{g_k} \\ \boldsymbol{u}_p \\ \boldsymbol{e} \end{pmatrix} = \begin{bmatrix} \boldsymbol{G}_{g_k} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{G}_p & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{R} \end{bmatrix} \tag{3}$$

$\boldsymbol{G}_p = \oplus_{h=1}^{w} \boldsymbol{G}_{p_h}$ is a block diagonal matrix with each block scaled by a separate variance component for each effect in each year-week combination.

The $\boldsymbol{u}_{g_k}$ effects were partitioned into additive and non-additive effects as follows:

$$\boldsymbol{u}_{g_k} = \boldsymbol{u}_{a_k} + \boldsymbol{u}_{i_k} \tag{4}$$

The observations for each year-week were treated as separate traits as described in Smith et al (2007), and thus the variance matrix of $\boldsymbol{u}_{g_k}$ is given by:

$$\boldsymbol{G}_{g_k} = \boldsymbol{G}_{yw_a} \otimes \boldsymbol{A} + \boldsymbol{G}_{yw_i} \otimes \boldsymbol{I}_{v_k} \tag{5}$$

where $\boldsymbol{G}_{yw_a}$ and $\boldsymbol{G}_{yw_i}$ are modelled as separate factor analytic models (Smith et al, 2001) and $\boldsymbol{A}$ is the numerator relationship matrix. As described in Smith and Cullis (2018), the total genetic effects are partitioned into the additive and non-additive effects and each of these can be modelled with a different factor analytic model. The final model for each trait was determined using model fitting statistics, the Akaike and Bayesian Information Criteria (AIC and BIC), and the percentage total genetic variance explained (Smith and Cullis, 2018).

The residual matrix, $\boldsymbol{R}$, was modelled such that the years are modelled independently but plots within a year are spatially and temporally correlated. Correlated heterogeneous (CORH-ID) and antedependence (ANTE-ID) models were used to model the temporal relationship of plots between weeks. A separable autoregressive process in both the bed (row) and column directions, following Gilmour et al (1997), modelled the spatial effects (HET-AR1). Note that the temporal models constrain the spatial model parameter estimates to be the same for all weeks within a year.

Estimation of fixed and random effects in the LMM was achieved by solving the mixed model equations using the REML algorithm implemented in ASReml-R (The VSNi Team, 2023; Butler et al, 2023). The BLUPs for each genotype were obtained from the final model, which is different in structure for each trait (Table S2).

The distributional assumptions of normality, variance homogeneity and idependence were investigated and determined to be met for all traits.

2.5 Evaluation criteria for single time points

Four evaluation criteria were calculated in each data-masking iteration to quantify the impact of missing phenotypic data on genetic parameter estimation and selection outcomes. These criteria were chosen because they inform whether single-trait, single-time-point data are suitable for downstream analyses, including subsequent MET modelling. In practice, if a trait assessed in a given harvest week shows negligible genetic variance then the ASBP investigates potential causes. The team may exclude that trait-time point from downstream analyses, or may repeat phenotyping to replace with it more robust data. The evaluation criteria, and reason for selection, were:

1. Average replication, defined as the mean number of non-missing plot observations per genotype, to confirm that the data masking algorithm worked as expected.

2. Total genetic variance (σ_g^2), estimated from the pedigree model as the sum of additive and non-additive genetic variances. Total genetic variance estimates change with the type and level of missingness. We use the distribution of the total genetic variance estimates to determine the level of appropriate missingness by comparing with the point estimate from the complete dataset.
3. Percentage of additive genetic variance, relative to total genetic variance. Additive genetic variance is dependent on the MNAR and is used to assess an acceptable missingness threshold by comparing the point estimates from the complete dataset with the distribution of the estimates from the datasets with imposed missingness.
4. Reliability (model-based), as $1 - \frac{PEV}{\sigma_g^2}$, where PEV is the prediction error variance (i.e., the squared standard error of each genotype) (Mrode, 2005). Reliability was calculated at the genotype level and averaged within each iteration, and reflects the proportion of the true genetic variance captured by the prediction.

2.6 Identification of missingness thresholds

Selection overlap was used to identify a threshold of acceptable levels of missing data. Selection intensity was set at 30%, reflecting the average percentage of genotypes retained in ASBP clonal trials each year across all advanced stages. For each scenario, the top 30% of genotypes were "selected" based on BLUPs from both the complete dataset and masked datasets with imposed missingness. Selection overlap was defined as the percentage of selected genotypes shared between the complete and missing datasets. For each trait and missingness level (plot and genotype), mean selection overlap across iterations was calculated. A threshold was defined as the smallest proportion of missing data at which mean overlap fell below 80%. This threshold was chosen as it was deemed an acceptable threshold to ensure selection decisions remain stable under missing data.

2.7 Evaluation criteria for multi-environment trial datasets

For each iteration, the following quantities were calculated, as well as their minimum, mean and maximum across all iterations.

1. Narrow-sense heritability, calculated as per Falconer and Mackay (1996):

$$\hat{h}^2 = \frac{\hat{\sigma}_a^2}{\hat{\sigma}_g^2 + \frac{\hat{\sigma}_e^2}{r}}$$

   where $\hat{\sigma}_a^2$ is the additive genetic variance, $\hat{\sigma}_g^2$ is the total genetic variance, $\hat{\sigma}_e^2$ is the residual variance, and r is the average replication (mean number of non-missing plots per genotype). It is related to genetic gain and thus understanding its sensitivity to different levels of missingness is important.
2. Ranking stability, measured as the Spearman's rank correlation between genotype BLUPs from each simulated dataset and those from the complete dataset. The distribution of this statistic provides a measure of change in selection that the level of missingness introduces.

Table 5: Relationship between imposed missingness (pcNA, %) and realised missingness at the alternate experimental level. Values are means across all traits.

Imposed missingness (pcNA, %)	Genotype-level missingness (%) (plot-level imposed)	Plot-level missingness (%) (genotype-level imposed)
0	2.5	6.4
5	2.5	8.9
10	3.0	13.9
15	4.0	18.9
20	5.3	23.1
25	7.2	28.1
30	9.4	33.1
35	12.1	38.1
40	15.3	42.4
45	19.1	47.4
50	23.5	52.3

3. Reliability, as described in Section 2.5.

3 Results

3.1 Determining a missingness threshold

This study masked increasing levels of missing phenotypic data (5–50% in 5% increments) through random sampling at both the plot and genotype levels. The plot-level aspect is analogous to MCAR, whilst genotype-level missingness is MNAR. The imposed missingness behaved as intended, with a clear positive relationship between imposed percentage of missing (pcNA) and realised missingness metrics. Across all traits, realised plot-level missingness increased from an average of 8.9 at pcNA = 5% to 52.3 at pcNA = 50%. In comparison, realised genotype-level missingness increased more gradually, from 2.5 to 23.5 (Table 5).

Average replication (non-missing observations per genotype) decreased substantially as the percentage of missing data at the plot level increased (Fig S1). This is expected, as when missingness increases at the plot level, individual plots are removed across many genotypes, hence the number of observations per genotype, i.e. replication, declines. In comparison, average replication remained relatively stable for genotype-level missingness. As entire genotypes are dropped (all plots set to missing), the remaining genotypes retain their original replication, thus, the number of genotypes in the dataset decreases but the level of replicates per genotype remains constant. Each LMM for all traits and simulation iterations converged and there were no singularities for any term, even with decreasing levels of average replication where this might be problematic.

For all traits, the mean total genetic variance remained relatively stable with increasing introduced missingness. The genetic variance estimate was lowest for TA and highest for Firmness, due to the scale in which these traits are measured. The stability of genetic variance under missingness demonstrates robustness in variance component estimation in the presence of missing data using REML. However, the

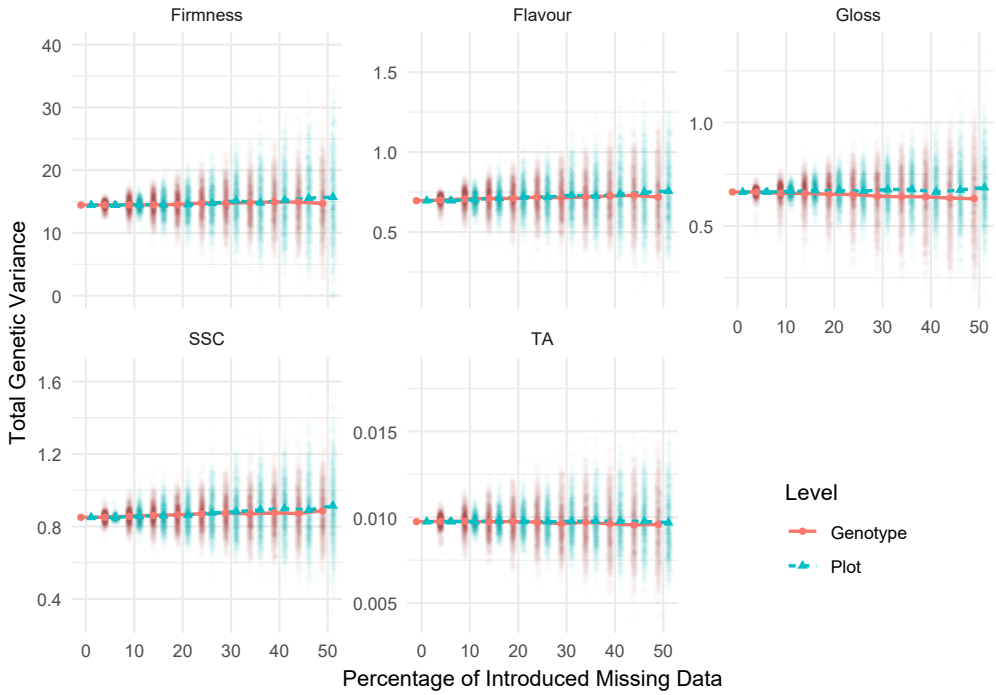

Fig. 1: Relationship between percentage of introduced missing data and total genetic variance at the genotype and plot levels.

range of estimates increased markedly with missingness at both the plot and genotype levels, particularly from 30–50% (Fig. 1).

The percentage of additive genetic variance was more affected by missingness than total genetic variance (Fig. 2). At low levels of missing data (5–15%), the percentage of additive genetic variance was similar between the genotype and plot levels. For Firmness and SSC, as missingness increased to 50%, additive genetic variance increased slightly for both plot- and genotype-levels. Overall, Gloss had the highest percentage of additive variance (100%), and Firmness had the lowest (39%). For many of the iterations with 20–50% missingness, 100% of the genetic variance was attributed to additive variance for all traits, whilst Firmness also had instances of 0% additive variance for these datasets (Fig. 2).

Reliability decreases as the level of missingness increases, with missingness applied at the genotype level decreasing more rapidly than at the plot level (Fig. 3). TA had the highest reliability (0.81) initially, with no introduced missing data, while Firmness had the lowest (0.41) (Fig. 3). TA also had the biggest difference between the complete dataset (pcNA = 0) to 50% introduced missingness, at the genotype level; whilst Firmness was least affected by missing data. For plot-level missingness the range in reliability increases after pcNA = 25% for Firmness, SSC and TA. For genotype-level missingness the number of reliability estimates less that the reliability estimate from

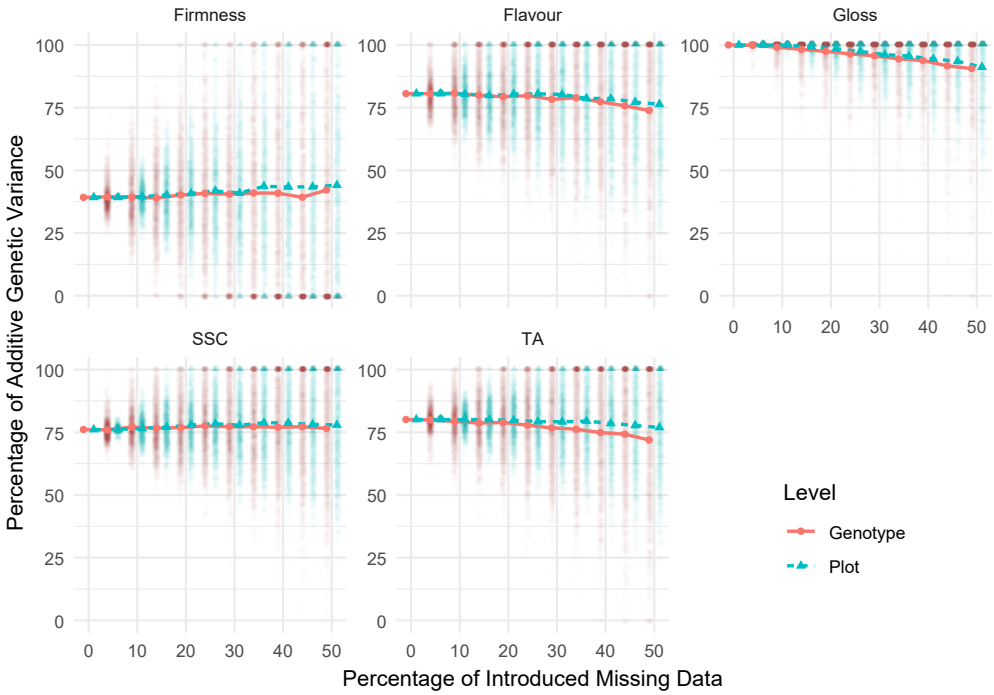

Fig. 2: Relationship between percentage of introduced missing data and the percentage of additive compared to total genetic variance at the genotype and plot levels.

the complete dataset increases once pcNA 25% for Flavour, Gloss, SSC and TA. The largest difference between genotype and plot mean reliability was for TA at 50% introduced missing data.

Table 6: Minimum level of introduced missing data (pcNA, %) at which mean selection overlap between the complete and masked datasets dropped below 80%, based on selection of the top 30% of genotypes. Values are shown for plot- and genotype-level missingness.

Trait	Plot-level missing data (%)	Genotype-level missing data (%)
Firmness	35	25
Flavour	40	30
Gloss	40	30
SSC	45	40
TA	40	30

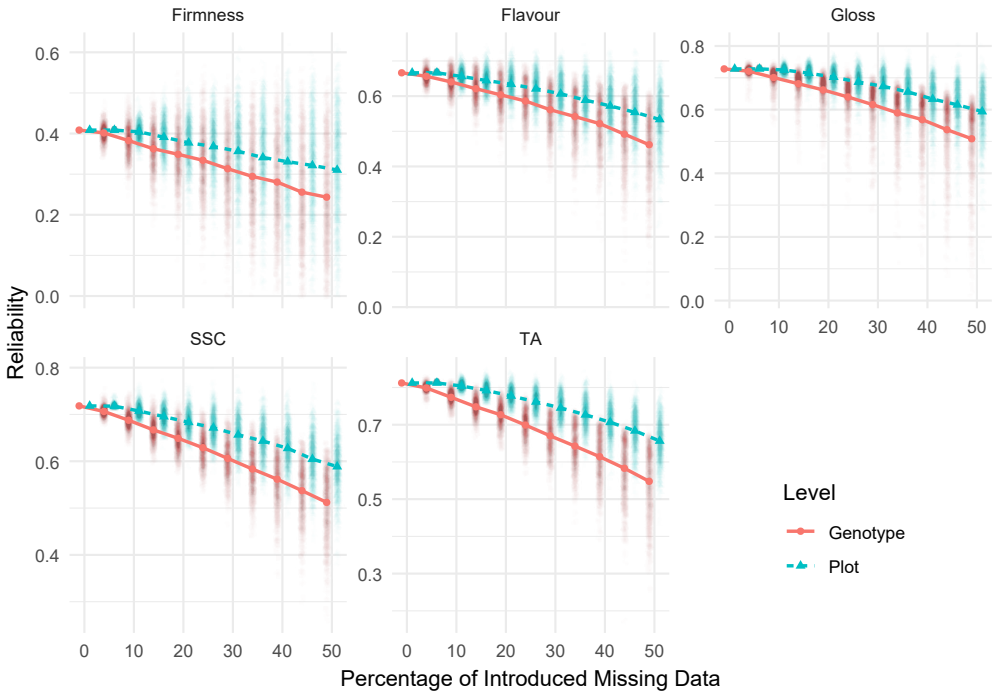

Fig. 3: Relationship between percentage of introduced missing data and reliability at the genotype and plot levels.

Selection overlap (those in the top 30%) between the complete dataset and datasets with imposed missingness declined progressively with increasing levels of missingness across all traits (Fig. 4). The minimum proportion of introduced missing data (pcNA) at which the mean selection overlap fell below 80% was 25% for all traits under genotype-level masking. Under plot-level masking, the corresponding threshold was 35% for Firmness and 40% for Gloss, SSC and TA. Across all traits, genotype-level masking resulted in lower selection overlap than plot-level masking at equivalent levels of missing data (Table 6). The reduction in overlap became more pronounced as missingness increased, indicating greater re-ranking of accession BLUPs under genotype-level than plot-level masking. Subsequently, a missingness threshold of 25% at the plot-level was determined as the starting point to investigate the effect of missingness on selection decisions in the MET dataset. To provide context of the 25% missingness level, levels of 15% and 35% missingness were also evaluated.

3.2 Application of selected threshold in a multi-year dataset

The level of missingness at the plot level in the complete MET dataset ranges from 4.7 (SSC) to 49.6 (Flavour). The real-life level of missingness is greater than 25% in six trait-week instances: one week each for Flavour and Gloss, two weeks for SSC

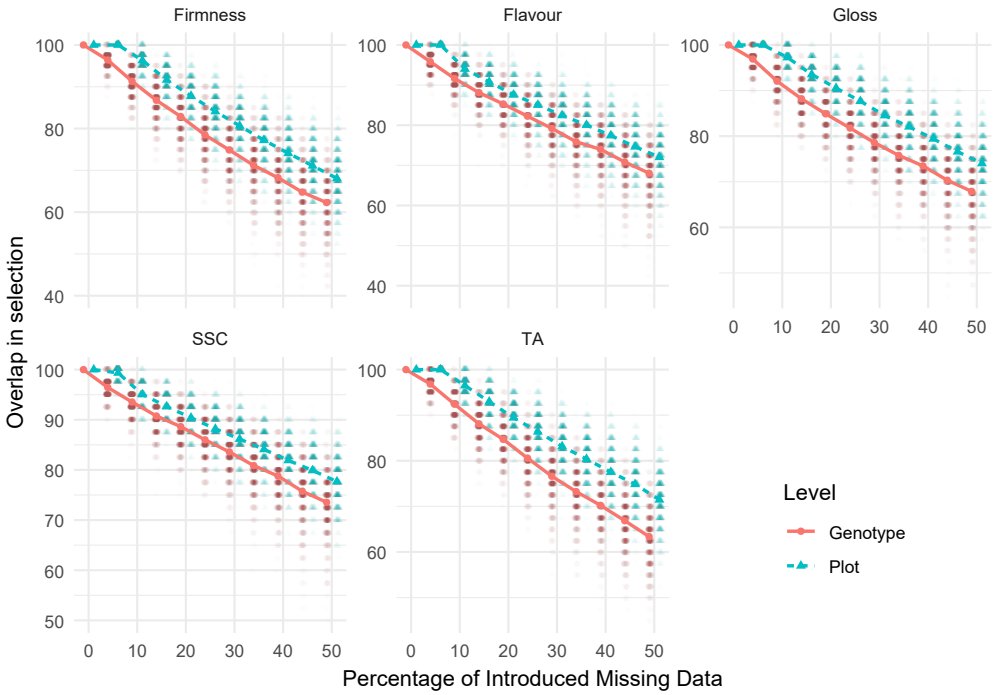

Fig. 4: Mean Overlap in selection of the top 30% of accessions under increasing levels of missing data at the genotype and plot levels.

and TA, and 3 for Firmness(Tables 3 and S1). A series of increasingly complex linear mixed models was fitted to the complete MET dataset for each trait and the best model determined by comparing the AIC, BIC and percentage of total genetic variance explained. The results for all models and traits are presented in Table S2. The best models for Firmness, Gloss and TA had the same variance structures for the year-week ($\boldsymbol{G}_{yw}$), the genetic ($\boldsymbol{G}_c$) and the residual ($\boldsymbol{R}$) variance; namely, FA1-DIAG, pedigree (PED) and heterogeneous correlation between weeks within years, and identical and independent within a week (CORH-ID). Flavour had the same ($\boldsymbol{G}_{yw}$ and $\boldsymbol{G}_c$ but $\boldsymbol{R}$ was fitted with an antedependence matrix within each year, and identical and independent within a week (ANTE-ID). For SSC, $\boldsymbol{G}_{yw}$ was fitted with an FA1 analytic model for both the additive and non-additive terms, and $\boldsymbol{R}$ was fitted with a CORH-ID model. SSC has a larger genotype $\times$ environment effect that requires a more complex model, whereas the other traits are mostly additive main effect traits (i.e., little genotype $\times$ environment interaction; data not shown). Again, there all models converged and no terms were singular in any trait-iteration combination.

Figure 5 shows the minimum, mean and maximum values for narrow-sense heritability (h^2), Spearman rank correlation and reliability for all traits. The mean value for all traits and statistics decreases as the pcNA increases, whilst the range increases. For all traits the results from the complete dataset (pcNA = 0) and 15% missing are

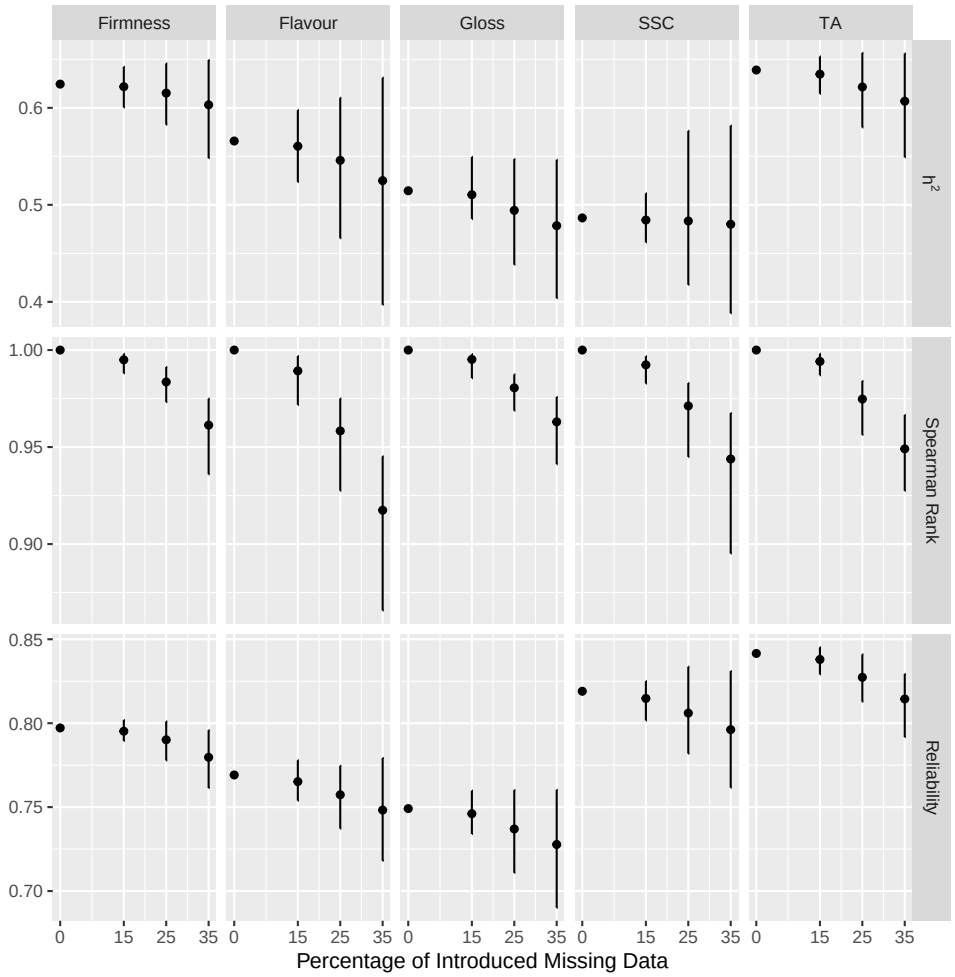

Fig. 5: Relationship between percentage of introduced missing data at the plot level and narrow sense heritability (h^2), Spearman rank correlation and Reliability for the masking study of the MET dataset.

very similar, where the complete dataset estimates are within range of those at pcNA = 15%.

The mean narrow-sense heritability changes least for SSC, and most for Flavour and the range for Flavour is also larger than other traits. The mean Spearman's rank correlation values decrease most for Flavour and least for Gloss. The range across all traits increases more from 25–35% than 15–25% missing. This suggests that the predictions change more at 35% than 25% missing. Reliability estimates were less than zero for 1.12% of the Firmness simulations, and 0.005% of the Gloss simulations. Negative reliability estimates usually occur when the genetic variance estimate is close to zero. This was the case for Firmness, with the first instance for pcNA = 30% at the

plot level, and thereafter with increasing frequency as the pcNA increased. In general the frequency was larger for plot-level missingness than genotype-level. For Gloss, negative reliabilities occurred because the residual variance estimate was higher than the genetic variance estimate. This only occurred for high levels of missingness, i.e. pcNA greater than 40, with most at 50% missing and all but one at the genotype-level. These negative values were set to NA for plotting so that the scale of the plots was not distorted. The range of reliability for pcNA = 25 and 35% for TA and 35% for Flavour do not include the complete dataset reliabilities of 0.812 and 0.666, respectively. This suggests that 35% missingness is a threshold where the results are not similar to those generated from the analysis of the complete dataset. A pcNA of 15% is acceptable for all traits and 25% for Flavour, Gloss and SSC.

The range in values for all evaluation criteria is largest for Flavour, and then SCC. Both traits have one week in the complete dataset that has 49% and 43% missing data, respectively, which could contribute to this larger variability (Table S1). TA has the same level of missingness as SSC and has less variability for all criteria. This suggests that the decision for a missingness threshold is complex and would not be the same for all traits. However, in practice it would be difficult to have different thresholds for different traits as it introduces an extra layer of complexity into the phenotyping. We recommend that a 15% missingness threshold provides the most similarity with the full dataset, and 25% missingness is acceptable.

4 Discussion

Understanding how different types and levels of missing data affect genetic evaluation is important for maintaining reliable selection decisions in breeding programs. We investigated the effect of missing data on estimates of genetic parameters, heritability and reliability in fruit quality traits in the Australian Strawberry Breeding Program. We performed data masking at a single time point for each trait at both the plot and genotype levels using a real breeding trial dataset. While simulated datasets have been used in similar studies as they offer a known truth against which statistical performance can be assessed, breeding trial datasets can never estimate true genetic values. Real trial datasets reflect biological, environmental and operational complexity under which breeding decisions are made, and these factors can be difficult to capture in simulations. Our use of real ASBP trial data instead provides an opportunity to focus on practical consequences of missing phenotypic data and to develop pragmatic recommendations directly relevant to operational conditions. A missingness threshold of 25% was determined using multiple criteria for a single time point for all traits. The identified threshold of 25% was implemented for all time points and traits in the MET analyses. This second masking study demonstrated how this level of missingness affected estimates of narrow-sense heritability and selection decisions, and informs practical application in the ASPB.

4.1 Type of missingness matters more than amount of missingness

The manner in which data are missing has a greater impact on genetic evaluation than the quantity of missing data. Loss of whole genotypes (genotype-level) has more impact than loss of individual observations (plot-level). For single time point trait data, estimates of reliability and overlap in selection decreased more rapidly with increasing levels of missingness at the genotype level compared to the plot level. Losing whole genotypes at individual time points decreases both the average replication of genotypes, and the level of genotype connectivity between time points and years. Lisle et al (2021) determined that the reliability of genetic variance parameter estimation depends on genotype replication, genotype connectivity between time points and years, and the number of genotypes evaluated in the trials, among other factors. Further, genetic relatedness and experimental design both affect the accuracy of plant breeding analyses (Cullis et al, 2020; Lisle et al, 2021). Thus, preserving representation of all genotypes may be more important than preserving every plot; reducing the level of data MNAR compared to MAR or MCAR would lead to more robust LMM analyses. The ASBP employs pedigree partially-replicated designs in its clonal trials as it is unachievable for the ASBP to have fully replicated experimental designs. This is particularly applicable in the early stage of the breeding cycle due to plant availability. The inclusion of pedigree increases the level of replication at the ancestral level, however the results in this study suggest that care be taken when there are many one-replicate genotypes in a trial.

4.2 Biological causes of genotype-level missingness and implications for MNAR

Genotype-level missingness in horticultural breeding programs is often driven by biological processes rather than random data loss. However, not all missingness is controllable in breeding programs; real breeding datasets are likely to contain a mixture of data MAR and MNAR. Specifically, missingness in longitudinal MET datasets is characterised by both the exclusion and inclusion of genotypes each year, i.e. the underlying genetic architecture, and the availability of material for observation. For example, some genotypes may not produce fruit at the tail ends (start and end) of the season. This may be due to the biological nature of flowering types, for example day neutral versus short day phenology (Barritt et al, 1982). The ASBP is actively breeding to retain only day neutral genotypes, to ensure consistent yield across the season and to ease labour pressures for growers. Thus, in subsequent generations, there should be fewer short-day genotypes and therefore fewer genotypes that do not produce any or many fruit at the season tails. Additionally, some related genotypes may be more susceptible to disease or environmental impacts, meaning yield production may be affected (Howard, 1992; Davik and Honne, 2005; Menzel, 2021). The missingness of genotype-level phenotypes is unlikely to be a unique situation in the ASBP. In papers discussing horticultural phenotypic assessments, authors rarely acknowledge missing data and instead state that fruit were measured "where available" (e.g.

Moon et al, 2018; Brewer et al, 2021; O'Connor et al, 2021). The inconsistent yielding of strawberry genotypes, and thus unavailability of fruit at certain harvests, is acknowledged in the literature (e.g. Hasing et al, 2013; Gezan et al, 2017). However, we suggest that the extent of data MNAR and the consequences of this characteristic is under-estimated in the plant breeding literature, particularly for traits which result in longitudinal MET datasets.

4.3 Impacts of missingness on genetic estimates

In this study, the mean estimates of total and percentage of additive genetic variance did not vary much across different levels of missingness; however the range of estimates increased (Figures 1 and 2). At 5–20% missingness, the range of total and additive genetic variance estimates was quite narrow. For example, for plot-level missingness at 5% the minimum and maximum values for all iterations is the same for all traits except SSC which had a range of 0.821 and 0.935 for total genetic variance and 71.4 to 82% for percentage additive genetic variance. The range for these criterion for this trait increased to 0.581 to 1.26, and 51.2 to 100.0% for pcNA = 20% at the plot level. At 30% and beyond, the range of estimates for multiple parameters is much wider, particularly for additive genetic variance which is key for selecting parents in breeding programs. Curiously, the percentage of additive variance oscillated from 0 to 100% for some traits. This oscillation was observed for yield and agronomic traits between consecutive weeks in the ASBP breeding datasets (data not shown) and was a motivating factor for this study. Further investigation is required, including the appropriate number of generations of pedigree to include in kinship modelling, the modelling of genetic architecture and an understanding of the interplay between MAR and MNAR in the dataset. Using observed field trial data, (Galli et al, 2018) masked 5–30% of phenotypic records in maize hybrids. Reliability and heritability for grain yield and plant height both decreased, and standard deviations dramatically increased, with increasing levels of simulated missing plot-level phenotypic data (Galli et al, 2018). Similarly, Aguate et al (2019) used real multi-environment field trial data in wheat and observed that the genetic variance was overestimated by over 10% when more than 40% of the data were set to be missing, although the datasets with more years achieved less bias. In contrast, Bauer et al (2006) used fully simulated datasets in a self-pollinating crop with randomly imposed missingness, and found that estimates of mean genotypic value were stable up to 30% missing data across heritability levels, but declined sharply from 60 to 90%. However, although total and additive genetic variance remained relatively stable with increasing missingness, breeders do not make selection decisions based on variance components.

4.4 Selection decisions provide a practical measure of missing data impact

Selection overlap reflects real-world breeding operations and is more interpretable than variance components. Thus, selection decision criteria may provide a better measure of practical impacts of data MAR or MNAR compared to variance component stability. At 25% missing genotype-level data, overlap in selection was still 80% or more for all

traits, and even higher overlap was observed for plot-level missingness. This finding is useful for defining practical stop/go thresholds for whether or not to phenotype fruit traits within a harvest week. The high level of selection overlap at 25% missingness combined with the wider range in genetic variance estimates beyond 25% suggests that 25% missing data is an acceptable upper threshold. In simulated missing maize datasets, the correlation between adjusted phenotypes from complete and induced missingness datasets decreased as missing data increased to 30% (Galli et al, 2018). Obviously, the least missing data the better, but breeding programs still need to be able to collect data even when some plots or genotypes are missing measurable fruit. The ASBP require a pragmatic decision that is appropriate for all fruit quality traits for technicians to follow. More than one fruit quality trait is assessed each week, and it is impractical to set different thresholds for each trait. While these results provide guidance for individual harvest weeks, breeding decisions in the ASBP are based on information accumulated across environments and time. Individual harvest weeks are not analysed independently in the ASBP nor used to make selection decisions; thus, the effect of a chosen missingness threshold for a single time point must also be considered in the MET analyses.

4.5 MET analyses may partially mitigate the effects of missing data

Multi-environment trial analyses are commonly employed in breeding programs, and may be more robust to data MAR or even MNAR than single trial/time point analyses because they can exploit information across related genotypes, environments and years. Factor analytic models allow the borrowing of information to obtain more accurate and robust predictions (Smith and Cullis, 2018; Lisle et al, 2021). In addition, the ASBP employs the contemporary group methodology of Smith et al (2021) which minimises the selection bias due to genotype dropouts from one year to the next from selection decisions. When combining multiple time-points within a MET, data may be missing at different levels across years, environments, related groups, and traits. Missing data at the genotype level in this study's MET dataset exceeded 50% on rare occasions. It is clear from the simulated datasets here that 50% missing data results in very variable estimates of variance components and thus should be avoided in future in the ASBP.

4.6 Heritability and reliability are stable with low levels of missing data in METs

Despite the complex structure of the MET dataset, estimates of heritability and reliability remained relatively stable at low to moderate levels of missingness. Narrow-sense heritability is key for understanding the potential improvement in a population's genetic merit, whilst reliability gives a breeder confidence in the phenotypic data. For most traits, MET analyses with 15% missing data appeared to give stable and narrow-ranges of heritability and reliability. There were, however, very large ranges of heritability estimates for some traits at pcNA = 35%, particularly for Flavour and SSC. This may be due to the subjective nature and 1-9 rating scale used in scoring

flavour, and the large environmental impact on the accumulation of sugars in strawberries and hence changes in SSC (°Brix) values across the season. Again, whilst heritability and reliability are key factors in measuring genetic gain in breeding programs, the accurate selection of genotypes for commercialisation or use as parents remains critical.

4.7 BLUP rankings are unstable above 25% missing data

The most practically relevant measure for breeders is whether missing data change genotype rankings and selection decisions. Spearman's rank correlations between BLUPs estimated from the MET analyses from the complete dataset and the masked missingness datasets were mostly stable at pcNA = 15% . For all traits, pcNA = 25% resulted in a mean correlation of over 0.95 with a minimum correlation of 0.93 for Flavour. However, at pcNA = 35%, very large ranges in rank correlations were observed for both Flavour and SSC (Fig. 5), possibly due to the nature of these traits as previously described. Accurate ranking of genotypes for selection is paramount in breeding, and so large variation in the accuracy of rankings is undesirable. Therefore, missing data over 25% appears to have adverse impacts on fruit quality traits in the ASBP. This aligns closely with the threshold identified from the single timepoint analyses, providing support for a common operational threshold across traits and analytical approaches.

4.8 Limitations

Missing fruit quality data in the ASBP is largely driven by low or no fruit numbers (yield) available for assessment. There is thus the potential for non-random missingness to arise due to correlations among traits. Indeed, the well-documented negative correlation between strawberry yield and °Brix (Cockerton et al, 2021; Menzel, 2022; Whitaker et al, 2012) may also influence phenotypic measurements. It should also be noted that there is an important statistical difference between missing data, and zero data. Yield traits were purposefully not investigated in this study, because the absence of fruit would often be classified as true biological zero and not a missing observation. Whilst we randomly sampled genotypes to mask, a more representative approach would be to drop out low-yielding genotypes, or those susceptible to disease successively, as this would reflect real-world situations. Bivariate modelling of yield and fruit quality data was beyond the scope of this study, but could be investigated in future studies in strawberry and other crops. Additional studies could also be extended to the temperate breeding node of the ASBP, as genetic connectivity, experimental designs, flowering seasons and environmental variation will vary between breeding program. Therefore, the missing-data thresholds identified here should be interpreted as a guide for the subtropical node of the ASBP only and should not be assumed to apply directly to other crops, environments, or breeding schemes without further validation.

5 Conclusions

Missing phenotypic data, whether MAR, MCAR or MNAR, will always be an issue in plant breeding programs due to factors including environmental variability, genotype susceptibility to diseases and weather events, and human and equipment errors. The ASBP aimed to identify a threshold at which phenotyping fruit quality data remained worthwhile, balancing the time and labour required for data collection against the quality of information obtained. We found that selection decisions were more sensitive to missing data than variance component estimates, highlighting the importance of evaluating the practical breeding consequences of missingness. Some traits appeared more sensitive to high levels of missingness than others. However, a pragmatic threshold applicable across all fruit quality traits was desired. Missingness affecting entire genotypes had a greater impact than missingness affecting individual plot observations. This that the mechanism underlying data missingness may be just as important as the amount of missing data, thus reflecting the importance of maintaining genetic connectivity within analyses. As fruit availability is considerably easier to quantify at the plot level than at the genotype level during each harvest week, we sought to determine a practical plot-level threshold for data collection. A threshold of 25% missing data at the plot level was identified as appropriate for the ASBP. The methodology presented here provides a useful framework for determining acceptable levels of missing phenotypic data in other horticultural breeding programs.

Supplementary information. The online version contains supplementary material.

Acknowledgments. The authors thank present and past members of the Australian Strawberry Breeding Program and staff at the Maroochy Research Facility for their involvement in field preparation and maintenance, plant propagation, plant maintenance, fruit harvesting and processing, and data collection. We thank Dr. Daniel Tolhurst for his insightful discussions. We also thank Michael Mumford and Dr. Bernardo Blanco-Martin for discussions about missingness. AI tools, including ChatGPT and Copilot, were used to develop the data masking R function, and to assist with writing structure, clarity and flow.

Declarations

5.1 Funding

The Australian Strawberry Breeding Program has been funded by Hort Innovation using the strawberry research and development levy, with co-contributions from the Queensland Government through its Department of Primary Industries and funds from the Australian Government.

5.2 Competing Interests

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

5.3 Authors' Contributions

KO and KM conceived the study. KM and KO contributed to statistical methodology and interpretation. KO developed the imposed missing data function, with support from AI-assisted coding tools. KO and KM performed the analyses and drafted the manuscript. JN led the breeding project. All authors reviewed and approved the final manuscript

5.4 Data Availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

References

Aguate F, Crossa J, Balzarini M (2019) Effect of missing values on variance component estimates in multienvironment trials. Crop Science 59(2):508517. https://doi.org/10.2135/cropsci2018.03.0209, URL http://dx.doi.org/10.2135/cropsci2018.03.0209

Arciniegas-Alarcón S, García-Peña M, Canas Rodrigues P (2020) New multiple imputation methods for genotype-by-environment data that combine singular value decomposition and jackknife resampling or weighting schemes. Computers and Electronics in Agriculture 176:105617. https://doi.org/https://doi.org/10.1016/j.compag.2020.105617, URL https://www.sciencedirect.com/science/article/pii/S0168169920310784

Barritt BH, Bringhurst RS, Voth V (1982) Inheritance of early flowering in relation to breeding day-neutral strawberries1. Journal of the American Society for Horticultural Science 107(5):733736. https://doi.org/10.21273/jashs.107.5.733, URL http://dx.doi.org/10.21273/jashs.107.5.733

Bauer AM, Reetz TC, Léon J (2006) Estimation of breeding values of inbred lines using best linear unbiased prediction (BLUP) and genetic similarities. Crop Science 46(6):26852691. https://doi.org/10.2135/cropsci2006.01.0019, URL http://dx.doi.org/10.2135/cropsci2006.01.0019

Brewer S, Plotto A, Bai J, et al (2021) Evaluation of 21 papaya (carica papaya l.) accessions in southern florida for fruit quality, aroma, plant height, and yield components. Scientia Horticulturae 288:110387. https://doi.org/https://doi.org/10.1016/j.scienta.2021.110387, URL https://www.sciencedirect.com/science/article/pii/S0304423821004945

Butler DG, Cullis BR, Gilmour AR, et al (2023) ASReml-R Reference Manual Version 4.2. URL https://asreml.kb.vsni.co.uk/wp-content/uploads/sites/3/ASReml-R-Reference-Manual-4.2.pdf, may 27, 2023

Cockerton HM, Karlström A, Johnson AW, et al (2021) Genomic informed breeding strategies for strawberry yield and fruit quality traits. Frontiers in Plant Science 12. https://doi.org/10.3389/fpls.2021.724847

Crain J, Mondal S, Rutkoski J, et al (2018) Combining highthroughput phenotyping and genomic information to increase prediction and selection accuracy in wheat breeding. The Plant Genome 11(1). https://doi.org/10.3835/plantgenome2017.05.0043

Cullis BR, Smith AB, Coombes NE (2006) On the design of early generation variety trials with correlated data. Journal of Agricultural, Biological, and Environmental Statistics 11(4):381393. https://doi.org/10.1198/108571106x154443

Cullis BR, Smith AB, Cocks NA, et al (2020) The design of early-stage plant breeding trials using genetic relatedness. Journal of Agricultural, Biological, and Environmental Statistics 25(4):553–578

Curnow RN (1961) The estimation of repeatability and heritability from records subject to culling. Biometrics 17(4):553. https://doi.org/10.2307/2527855, URL http://dx.doi.org/10.2307/2527855

Davik J, Honne BI (2005) Genetic variance and breeding values for resistance to a wind-borne disease [Sphaerotheca macularis (Wallr. ex Fr.)] in strawberry (*Fragaria* Œ *ananassa* Duch.) estimated by exploring mixed and spatial models and pedigree information. Theoretical and Applied Genetics 111(2):256264. https://doi.org/10.1007/s00122-005-2019-3, URL http://dx.doi.org/10.1007/s00122-005-2019-3

Falconer DS, Mackay TF (1996) Introduction to Quantitative Genetics, 4th edn. UK Longman Group, Sussex, UK

Galli G, Lyra DH, Alves FC, et al (2018) Impact of phenotypic correction method and missing phenotypic data on genomic prediction of maize hybrids. Crop Science 58(4):14811491. https://doi.org/10.2135/cropsci2017.07.0459, URL http://dx.doi.org/10.2135/cropsci2017.07.0459

Gezan SA, Osorio LF, Verma S, et al (2017) An experimental validation of genomic selection in octoploid strawberry. Horticulture Research 4(Article 16070). https://doi.org/10.1038/hortres.2016.70

Gilmour AR, Cullis BR, Verbyla AP (1997) Accounting for natural and extraneous variation in the analysis of field experiments. Journal of Agricultural, Biological, and Environmental Statistics pp 269–293

Hardigan MA, Feldmann MJ, Lorant A, et al (2020) Genome synteny has been conserved among the octoploid progenitors of cultivated strawberry over millions of years of evolution. Frontiers in Plant Science 10. https://doi.org/10.3389/fpls.2019.01789

Hartung J, Piepho H (2021) Effect of missing values in multienvironmental trials on variance component estimates. Crop Science 61(6):40874097. https://doi.org/10.1002/csc2.20621, URL http://dx.doi.org/10.1002/csc2.20621

Hasing TN, Osorio LF, Whitaker VM (2013) Within-season stability of strawberry soluble solids content. Journal of the American Society for Horticultural Science 138(3):190197. https://doi.org/10.21273/jashs.138.3.190, URL http://dx.doi.org/10.21273/jashs.138.3.190

Henderson CR (1975) Best linear unbiased estimation and prediction under a selection model. Biometrics 31(2):423. https://doi.org/10.2307/2529430, URL http://dx.doi.org/10.2307/2529430

Howard CM (1992) Anthracnose of strawberry caused by the *Colletotrichum* complex in florida. Plant Disease 76(10):976. https://doi.org/10.1094/pd-76-0976, URL http://dx.doi.org/10.1094/PD-76-0976

Klápt J, Ismael A, Paget M, et al (2022) Genomics-enabled management of genetic resources in radiata pine. Forests 13(2):282. https://doi.org/10.3390/f13020282

Lisle C, Smith A, Birrell CL, et al (2021) Information based diagnostic for genetic variance parameter estimation in multi-environment trials. Frontiers in Plant Science 12. https://doi.org/10.3389/fpls.2021.785430

Little R, Rubin D (2019) Maximum Likelihood for General Patterns of Missing Data: Introduction and Theory with Ignorable Nonresponse. John Wiley and Sons, Ltd, https://doi.org/https://doi.org/10.1002/9781119482260.ch8

Menzel C (2021) Higher temperatures decrease fruit size in strawberry growing in the subtropics. Horticulturae 7(2):34. https://doi.org/10.3390/horticulturae7020034

Menzel CM (2022) Effect of temperature on soluble solids content in strawberry in queensland, australia. Horticulturae 8(5):367. https://doi.org/10.3390/horticulturae8050367

Meuwissen TH, Luo, Z (1992) Computing inbreeding coefficients in large populations. Genetics Selection Evolution 24(4):305–313

Moon P, Fu Y, Bai J, et al (2018) Assessment of fruit aroma for twenty-seven guava (*Psidium guajava*) accessions through three fruit developmental stages. Scientia Horticulturae 238:375383. https://doi.org/10.1016/j.scienta.2018.04.067, URL http://dx.doi.org/10.1016/j.scienta.2018.04.067

Mrode RA (2005) Linear models for the prediction of animal breeding values. Ed. 2, Wallingford UK: CABI Publishing

Oakey H, Verbyla AP, Cullis BR, et al (2007) Joint modeling of additive and non-additive (genetic line) effects in multi-environment trials. Theor Appl Genet 114(8):1319–32. https://doi.org/10.1007/s00122-007-0515-3

O'Connor KM, Hayes BJ, Hardner CM, et al (2021) Genomic selection and genetic gain for nut yield in an australian macadamia breeding population. BMC Genomics 22(1). https://doi.org/10.1186/s12864-021-07694-z, URL http://dx.doi.org/10.1186/s12864-021-07694-z

Patterson HD, Thompson R (1971) Recovery of inter-block information when block sizes are unequal. Biometrika 58(3):545–554

Piepho HP, Möhring J (2006) Selection in cultivar trialsis it ignorable? Crop Science 46(1):192–201. https://doi.org/https://doi.org/10.2135/cropsci2005.04-0038, URL https://acsess.onlinelibrary.wiley.com/doi/abs/10.2135/cropsci2005.04-0038, https://acsess.onlinelibrary.wiley.com/doi/pdf/10.2135/cropsci2005.04-0038

Piepho HP, Büchse A, Emrich K (2003) A hitchhikers guide to mixed models for randomized experiments. Journal of Agronomy and Crop Science 189(5):310322. https://doi.org/10.1046/j.1439-037x.2003.00049.x

Piepho HP, Möhring J, Melchinger AE, et al (2008) BLUP for phenotypic selection in plant breeding and variety testing. Euphytica 161(12):209228. https://doi.org/10.1007/s10681-007-9449-8, URL http://dx.doi.org/10.1007/s10681-007-9449-8

Piepho HP, Müller BU, Jansen C (2014) Analysis of a complex trait with missing data on the component traits. Communications in Biometry and Crop Science 9(1):26–40

R Core Team (2025) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL https://www.R-project.org/

Robertson A (1959) The sampling variance of the genetic correlation coefficient. Biometrics 15(3):469. https://doi.org/10.2307/2527750

Robinson GK (1991) That BLUP is a good thing: The estimation of random effects. Statistical Science 6(1). https://doi.org/10.1214/ss/1177011926, URL http://dx.doi.org/10.1214/ss/1177011926

Rubin DB (1976) Inference and missing data. Biometrika 63(3):581592. https://doi.org/10.1093/biomet/63.3.581

Schmidt P, Hartung J, Bennewitz J, et al (2019) Heritability in plant breeding on a genotype-difference basis. Genetics 212(4):9911008. https://doi.org/10.1534/genetics.119.302134, URL http://dx.doi.org/10.1534/genetics.119.302134

Smallwood CJ, Saxton AM, Gillman JD, et al (2019) Context-specific genomic selection strategies outperform phenotypic selection for soybean quantitative traits in the progeny row stage. Crop Science 59(1):5467. https://doi.org/10.2135/cropsci2018.03.0197

Smith A, Ganesalingam A, Lisle C, et al (2021) Use of contemporary groups in the construction of multi-environment trial datasets for selection in plant breeding programs. Frontiers in Plant Science 11. https://doi.org/10.3389/fpls.2020.623586

Smith AB, Cullis BR (2018) Plant breeding selection tools built on factor analytic mixed models for multi-environment trial data. Euphytica 214(8):143

Smith AB, Cullis BR, Thompson R (2001) Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. Biometrics 57(4):1138–1147. https://doi.org/https://doi.org/10.1111/j.0006-341X.2001.01138.x

Smith AB, Cullis BR, Thompson R (2005) The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. The Journal of Agricultural Science 143(6):449462. https://doi.org/10.1017/s0021859605005587, URL http://dx.doi.org/10.1017/S0021859605005587

Smith AB, Stringer JK, Wei X, et al (2007) Varietal selection for perennial crops where data relate to multiple harvests from a series of field trials. Euphytica 157(1):253–266

The VSNi Team (2023) asreml: Fits Linear Mixed Models using REML. URL www.vsni.co.uk, R package version 4.2.0.332

Whitaker VM, Osorio LF, Hasing T, et al (2012) Estimation of genetic parameters for 12 fruit and vegetative traits in the university of florida strawberry breeding population. Journal of the American Society for Horticultural Science 137(5):316324. https://doi.org/10.21273/jashs.137.5.316

Supplementary Material

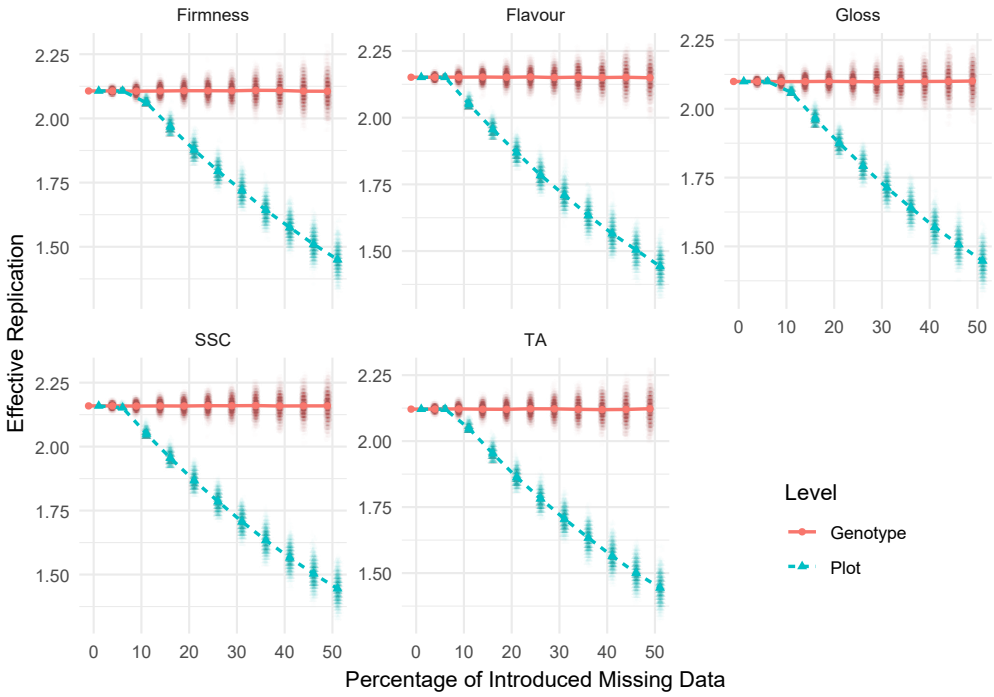

Fig. S1: Relationship between percentage of introduced missing data and average replication at the genotype and plot levels. Solid line indicates the mean across all iterations.

Table S1: Summary of available and percentage missing clones and plots and mean and standard deviations for Firmness, Flavour, Gloss, SSC and TA for the Australian Strawberry Breeding Program subtropical clonal trials, 2020-2024.

			Genotypes		Plots			Standard
Trait	Year	Week	Available	% missing	Available	% missing	Mean	deviation
Firmness	2020	12	102	51.0	240	33.3	76.49	8.223
Firmness	2020	16	102	26.5	240	20.4	83.41	6.878
Firmness	2020	20	102	7.8	240	8.8	78.21	8.655
Firmness	2021	11	199	31.7	416	31.0	81.01	6.212
Firmness	2021	16	199	24.1	416	24.5	81.73	5.524
Firmness	2021	21	199	25.6	416	24.0	77.76	7.288
Firmness	2022	17	181	32.0	400	29.8	81.57	5.533
Firmness	2023	21	201	15.4	345	10.4	58.10	6.572
Firmness	2023	13	201	22.4	345	17.7	65.92	8.412
Firmness	2023	18	201	15.4	345	11.0	65.83	7.036
Firmness	2024	19	135	11.9	299	7.7	61.86	7.319
Firmness	2024	23	135	12.6	299	8.4	66.26	7.246
Flavour	2020	16	102	18.6	240	15.4	6.29	1.269
Flavour	2020	21	102	48.0	240	49.6	5.81	1.374
Flavour	2020	23	102	12.7	240	11.2	5.63	1.366
Flavour	2021	17	199	14.6	416	16.8	5.90	1.607
Flavour	2021	21	199	20.6	416	19.7	5.91	1.483
Flavour	2022	21	181	7.2	400	15.0	6.07	1.420
Flavour	2023	22	201	9.0	345	6.7	6.51	1.404
Flavour	2024	23	135	7.4	299	5.0	6.54	1.245
Flavour	2024	18	135	11.1	299	7.0	7.11	1.224
Gloss	2020	14	102	20.6	240	15.8	6.27	1.101
Gloss	2021	16	199	12.6	416	15.6	6.60	0.862
Gloss	2022	12	181	37.6	400	33.8	6.43	0.932
Gloss	2022	19	181	4.4	400	13.8	6.15	0.940
Gloss	2023	15	201	8.5	345	6.1	6.86	0.840
Gloss	2023	20	201	8.5	345	5.8	6.33	0.919
Gloss	2024	12	135	13.3	299	8.0	6.92	1.040
Gloss	2024	17	135	14.8	299	10.0	6.99	0.758
Gloss	2024	22	135	17.8	299	12.4	6.61	0.784
SSC	2020	17	102	26.5	240	17.9	8.15	1.101
SSC	2020	19	102	16.7	240	13.3	8.06	1.182
SSC	2020	24	102	5.9	240	7.1	7.81	1.360
SSC	2021	11	199	12.1	416	18.0	8.39	1.126
SSC	2021	16	199	14.6	416	16.8	7.91	1.284
SSC	2021	22	199	16.1	416	17.5	7.60	1.565
SSC	2022	12	181	51.9	400	43.8	8.70	1.690
SSC	2022	15	181	39.2	400	35.0	7.19	1.105
SSC	2022	21	181	3.3	400	13.2	6.28	1.087
SSC	2023	15	201	11.9	345	8.1	8.32	1.408
SSC	2023	20	201	10.0	345	7.5	8.49	1.416
SSC	2024	17	135	25.2	299	15.7	8.97	1.077
SSC	2024	23	135	5.9	299	4.7	8.23	1.289
TA	2020	17	102	26.5	240	17.5	0.80	0.147
TA	2020	19	102	16.7	240	13.3	0.76	0.157
TA	2020	24	102	4.9	240	6.7	0.72	0.141
TA	2021	11	199	15.1	416	19.5	0.76	0.146
TA	2021	16	199	14.6	416	16.8	0.68	0.103
TA	2021	22	199	15.6	416	17.3	0.71	0.124
TA	2022	12	181	51.9	400	43.8	0.87	0.154
TA	2022	15	181	38.7	400	34.8	0.68	0.100
TA	2022	21	181	3.3	400	13.2	0.64	0.096
TA	2023	15	201	13.4	345	9.0	0.78	0.143
TA	2023	20	201	10.9	345	8.1	0.71	0.107
TA	2024	17	135	25.2	299	16.1	0.76	0.123
TA	2024	23	135	9.6	299	6.4	0.71	0.112

Table S2: Model fitting statistics, number of total and genetic parameters (n_p and n_g), log-likelihood (LOGL), Akaike and Bayesian Information Criterion (AIC and BIC), percentage additive genetic and total genetic variance explained for Firmness, Flavour, Gloss, SSC and TA for the Australian Strawberry Breeding Program subtropical clonal trials, 2020-2024. Codes for the varianece models are: DIAG = diagonal, FAk = factor analytic of order k, FA1-DIAG = FA1 on $\boldsymbol{G}_{yw_a}$ and DIAG to $\boldsymbol{G}_{yw_i}$, FA1-F1 = FA1 on n $\boldsymbol{G}_{yw_a}$ and $\boldsymbol{G}_{yw_i}$, IDE = no pedigree, PED = pedigree, HET-ID = heterogenous identical independent distribution, HET-AR1 = heterogeneous separable autoregressive correlation in bed and column directions, ANTE-ID = antedependence between year weeks, no spatial, CORH-ID = heterogeneous correlation model between year-weeks, no spatial. For each trait the best model is highlighted in **bold** font.

Trait	Model	$\boldsymbol{G}_{yw}$	$\boldsymbol{G}_c$	$\boldsymbol{R}$	n_g	n_p	LOGL	AIC	BIC	$\bar{v}_a$	$\bar{v}_{g_k}$
Firmness	1	DIAG	IDE	HET-ID	12	24	-7326	14725	14942		
Firmness	2	DIAG	IDE	HET-AR1	12	51	-7254	14634	15015		
Firmness	3	DIAG	PED	HET-AR1	24	51	-7197	14544	14997		
Firmness	4	DIAG	PED	ANTE-ID	24	34	-7199	14515	14865		
Firmness	5	DIAG	PED	CORH-ID	24	31	-7189	14487	14819		
Firmness	**6**	**FA1-DIAG**	**PED**	**CORH-ID**	**36**	**31**	**-7030**	**14193**	**14598**	**98**	
Firmness	7	FA1-FA1	PED	CORH-ID	48	31	-7018	14194	14671	98	95
Flavour	1	DIAG	IDE	HET-ID	9	18	-1927	3908	4063		
Flavour	2	DIAG	IDE	HET-AR1	9	36	-1909	3908	4168		
Flavour	3	DIAG	PED	HET-AR1	18	36	-1888	3885	4196		
Flavour	4	DIAG	PED	ANTE-ID	18	22	-1897	3875	4105		
Flavour	5	DIAG	PED	CORH-ID	18	21	-1900	3878	4103		
Flavour	**6**	**FA1-DIAG**	**PED**	**ANTE-ID**	**27**	**22**	**-1862**	**3823**	**4105**	**82**	
Flavour	7	FA1-FA1	PED	ANTE-ID	36	22	-1854	3825	4159	75	75
Flavour	8	FA2-FA1	PED	ANTE-ID	44	22	-1852	3836	4216	86	86
Gloss	1	DIAG	IDE	HET-ID	9	18	-945	1944	2102		
Gloss	2	DIAG	IDE	HET-AR1	9	37	-914	1919	2188		
Gloss	3	DIAG	PED	HET-AR1	18	37	-841	1792	2114		
Gloss	4	DIAG	PED	ANTE-ID	18	23	-857	1795	2035		
Gloss	5	DIAG	PED	CORH-ID	18	22	-857	1794	2028		
Gloss	**6**	**FA1-DIAG**	**PED**	**CORH-ID**	**27**	**22**	**-797**	**1691**	**1978**	**84**	
Gloss	7	FA1-FA1	PED	CORH-ID	36	22	-794	1704	2044	84	84
SSC	1	DIAG	IDE	HET-ID	13	26	-2467	5012	5252		
SSC	2	DIAG	IDE	HET-AR1	13	63	-2414	4980	5449		
SSC	3	DIAG	PED	HET-AR1	26	63	-2353	4884	5433		
SSC	4	DIAG	PED	ANTE-ID	26	45	-2353	4848	5286		
SSC	5	DIAG	PED	CORH-ID	26	42	-2351	4837	5257		
SSC	6	FA1-DIAG	PED	CORH-ID	39	42	-2238	4639	5139	87	
SSC	**7**	**FA1-FA1**	**PED**	**CORH-ID**	**52**	**42**	**-2224**	**4635**	**5215**	**85**	**82**
SSC	8	FA2-FA1	PED	CORH-ID	64	42	-2215	4641	5295	92	89
TA	1	DIAG	IDE	HET-ID	13	26	5794	-11510	-11270		
TA	2	DIAG	IDE	HET-AR1	13	58	5864	-11586	-11149		
TA	3	DIAG	PED	HET-AR1	26	58	5911	-11654	-11136		
TA	4	DIAG	PED	ANTE-ID	26	40	5910	-11687	-11280		
TA	5	DIAG	PED	CORH-ID	26	37	5915	-11704	-11316		
TA	**6**	**FA1-DIAG**	**PED**	**CORH-ID**	**39**	**37**	**6055**	**-11958**	**-11490**	**95**	
TA	7	FA1-FA1	PED	CORH-ID	52	37	6062	-11946	-11398	95	92