Aquaculture 553 (2022) 738058

Contents lists available at ScienceDirect

Aquaculture

journal homepage: www.elsevier.com/locate/aquaculture

Early selection to enhance genetic gain in a rohu (*Labeo rohita*) genetic improvement program

M.G. Hamilton [a,*], W. Mekkawy [a,b], Md. Badrul Alam [a], John A.H. Benzie [a,c]

[a] *WorldFish, Jalan Batu Maung, 11960, Bayan Lepas, Penang, Malaysia*
[b] *CSIRO Livestock & Aquaculture, CSIRO Agriculture and Food, Castray Esplanade, Hobart, Tasmania 7001, Australia*
[c] *School of Biological Earth and Environmental Sciences, University College Cork, Cork, Ireland*

ARTICLE INFO

Keywords:
Total additive genetic value
Estimated breeding value (EBV)
Multiple-stage selection
Genetic group model
Labeo rohita

ABSTRACT

Multiple-stage selection can, in certain circumstances, be adopted to enhance genetic gain in genetic improvement programs while avoiding the cost of assessing and maintaining all individuals for extended periods. The primary objective of the current study was to determine if multiple-stage selection can be used to enhance genetic gains in a family-based rohu (*Labeo rohita*) genetic improvement program managed by WorldFish in Bangladesh. In this program, families are maintained in separate hapas (i.e. nets) during the nursing phase of culture. At the conclusion of nursing, a sample of fingerlings from each family are weighed and tagged. After tagging fish from all families are pooled and grown out. For the purpose of this study, at the conclusion of nursing first-generation (G1) rohu families, a random sample of fingerlings and a non-random sample of visually-identified large fingerlings were selected (generally 36 individuals of each) for tagging from each of 135 families. Given that not all individuals were assessed for weight at the time of early selection (i.e. time of non-random sampling), genetic parameters and breeding values for harvest weight were initially estimated using data from the randomly-sampled individuals only. Unbiased EBVs for non-randomly sampled G1 individuals were subsequently computed from individual yield deviations, parental estimated breeding values and estimated genetic parameters. It was found that i) the total additive genetic value for mean harvest weight for non-randomly-sampled individuals was greater than that for randomly-sampled individuals, and ii) across families, a positive trend was evident– at low selection intensities, at least – between within-family non-random selection intensity and within-family differences in total additive genetic value between randomly and non-randomly selected individuals. Furthermore, the harvest weight of Generation 2 (G2) families made from non-randomly-sampled G1 parents (221 g) was greater than those made from randomly-sampled G1 parents (190 g). These findings indicated that early selection (i.e. visual identification and non-random sampling of larger individuals at an early age) of G1 individuals enhanced genetic gain in harvest weight. Accordingly, consideration should be given to the adoption of this practice in carp genetic improvement programs of the nature tested.

1. Introduction

Multiple-stage selection – in which selection within a population occurs at more than one point in time – can conceptually be adopted in family-based (i.e. pedigree-based) genetic improvement programs to enhance genetic gain while avoiding the cost of assessing and maintaining all individuals for extended periods (Pollak et al., 1984; Van Raden et al., 1984). Under multiple-stage selection, unbiased estimated breeding values (EBVs) can be computed using best linear unbiased prediction (BLUP), assuming that all individuals are assessed for trait/s under selection. However, it is not always feasible – due to time, biological and/or financial constraints – to assess all individuals prior to selection, particularly in the case of first-stage selection in highly fecund aquaculture and plant species (Hamilton, 2020).

In some circumstances it is possible to readily identify and select phenotypically superior individuals for a trait without measurement (e.g. visual evaluation of size), thus negating the need to assess (e.g. weigh) all individuals. However, such selection prior to assessment yields non-randomly sampled measurement records that deviate from the assumptions underpinning commonly adopted analytical methods – for example, best linear unbiased prediction (BLUP) (Henderson, 1975; Henderson and Quaas, 1976; Robinson, 1991) and restricted maximum

* Corresponding author.
E-mail addresses: M.Hamilton@cgiar.org (M.G. Hamilton), Wagdy.Mekkawy@csiro.au (W. Mekkawy), J.Benzie@cgiar.org (J.A.H. Benzie).

https://doi.org/10.1016/j.aquaculture.2022.738058
Received 25 November 2021; Received in revised form 14 February 2022; Accepted 17 February 2022
Available online 19 February 2022

likelihood (REML) (Gilmour et al., 2014; Harville, 1977) – and result in biased estimated breeding values (EBVs) and genetic parameters (Sonesson, 2005). One approach to avoid such biases, is to retain and measure a random sample of individuals in addition to retaining and measuring non-random putatively superior individuals. Unbiased EBVs and genetic parameters and can then be estimated by excluding non-randomly-sampled individuals from the analyses. Unbiased EBVs for non-randomly sampled individuals (i.e. phenotypically superior individuals selected prior to measurement) can subsequently be derived from their individual yield deviations, their parental EBVs and estimated genetic parameters (Sections 3.3.1 and 5.2.3 of Mrode, 2013) – noting that fixed and random effects required to calculate yield deviations, as well as parental EBVs and estimated genetic parameters, can be generated using unbiased observations from randomly sampled individuals.

Rohu is a globally important aquaculture species with 2.0 Mt. produced each year (FAO, 2020). In Bangladesh, rohu is a preferred carp species for consumption and culture, with approximately 0.3 Mt. produced annually (Alam, 2001; DoF, 2020). The objective of the current study was to determine if early selection, based on visual assessment of size at the conclusion of the nursing phase of rohu (*L. rohita*) culture, can be adopted to enhance genetic gains in harvest weight.

2. Methods

The study was conducted as part of the WorldFish Rohu Genetic Improvement Program (WFRGIP), one of three carp genetic improvement programs managed by WorldFish in Bangladesh (Hamilton et al., 2019a; Hamilton et al., 2019b; Hamilton et al., 2021). The founders of the WorldFish Rohu Genetic Improvement Program (WFRGIP) breeding population were sourced as spawn from three Bangladeshi rivers: the Halda, the Jamuna and the Padma (i.e. lower Ganges) (Hamilton et al., 2019b). These fish were reared for two years prior to spawning to generate a base population – herein referred to as 'Generation 0' or 'G0'. Generation 0 was spawned from founders in 2014, and the first (G1) and second (G2) selected generations were spawned in 2016 and 2018, respectively. All data collection, and fish husbandry, reported in this study was undertaken as part of the routine operations of the WorldFish Carp Genetic Improvement Program in accordance with the Guiding Principles of the Animal Care, Welfare and Ethics Policy of the WorldFish Center (WorldFish, 2004).

To produce each generation, multiple spawn batches were completed in which full-sib families were made and maintained in separate upwelling hatching jars with mesh at the top to prevent the loss of eggs (i.e. one family per jar). Approximately 30 h after hatching, fry were transferred to be nursed in one hapa (i.e. net) constructed of saree cloth (~300 μm mesh) per family. Hapas were regularly cleaned and mesh size was progressively increased during nursing – initially to mosquito netting (~1.4 mm) and ultimately to polyfilament netting (~5 mm). In the case of G0, in addition to one hapa per family, fry from each family were reared in one small earthen pond. At the conclusion of nursing (i.e. once large enough), randomly sampled individual fingerlings were tagged with passive integrated transponder (PIT) tags.

To examine the potential to enhance genetic gains through multiple-stage selection, after tagging 36 randomly-sampled individuals from each G1 family, an additional 36 phenotypically-superior (i.e. large) individuals from each G1 family were visually identified (i.e. non-randomly sampled) prior to tagging and measurement. Only families with greater than 72 individuals available for tagging at the conclusion of nursing were included in analyses. Once tagged, fingerlings were assigned, at random, to grow out ponds for progeny testing.

Parents of G1 and G2 families were selected based on estimated total additive genetic values – the sum of the estimated genetic group (i.e. river of origin - Halda, Jamuna and Padma) effect and estimated breeding value (EBV) for each individual (Wolak and Reid, 2017) – for harvest weight (i.e. body weight at the conclusion of progeny testing) and readiness to spawn (i.e. ripeness). Constraints were also placed on parent selection to ensure that the vast majority of families contributed to the next generation. For each spawn batch, mate allocation was undertaken after parent selection, adopting an assortative mating approach (Saura et al., 2017), with each parent contributing to only one full-sibling family. Within the WFRGIP, additional 'negatively selected' and 'control' G1 and G2 families were generated and maintained as separate 'lines'. The control line was maintained targeting a mean total additive genetic value (Wolak and Reid, 2017) of zero, approximating the additive genetic value of the unimproved base population, and was comprised of 43 and 44 full-sibling families in G1 and G2 respectively (Fig. 1). Refer to Table 1 for a summary of G0 (base population with no selection), G1 (positively selected) and G2 (positively selected and control) fish used in the current study.

Prior to analysis, weight data were rescaled so that randomly sampled fish had a mean of zero and phenotypic standard deviation of one, within each grow-out pond in the case of harvest weight. Preliminary analyses were undertaken by fitting the following univariate model using ASReml (Gilmour et al., 2014):

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z_1 h} + \mathbf{Z_2 Qg} + \mathbf{Z_2 a} + \mathbf{e} \quad (1)$$

where $\mathbf{y}$ is the vector of trait observations, $\mathbf{b}$ is a vector of fixed effects with its design matrix $\mathbf{X}$, $\mathbf{h}$ is a vector of random hapa effects – confounding common nursery environment and full-sibling family (i.e. specific combining ability) effects – with its design matrix $\mathbf{Z_1}$, $\mathbf{g}$ is a vector of random genetic group effects (Quaas, 1988; Swan et al., 2015) with its design matrix $\mathbf{Z_2}$, $\mathbf{Q}$ is a m x g matrix, where m is the number of individuals in the pedigree and g is the number of genetic groups, fitted in ASReml using the '!G' qualifier to read the pedigree-derived estimated proportional contribution of each genetic group to the genome of each individual (Gilmour et al., 2014; Wolak and Reid, 2017), $\mathbf{a}$ is a vector of additive genetic effects, and $\mathbf{e}$ is the vector of random residual terms. The model included as fixed effects in $\mathbf{b}$ the overall mean, spawning batch and, in the case of harvest weight only, grow-out pond. In addition, age at tagging (days), age at harvest (days; harvest weight only) and the count of surviving fish per family at tagging were included as covariates in $\mathbf{b}$. These covariates were standardised to have a mean of zero and standard deviation of one. In the analysis of G0 harvest weight, full-sibling family was initially fitted as a random effect – as full-sibling families were replicated across hapas and earthen ponds – but resulted in an estimated full-sibling family variance of zero and was subsequently removed from the model.

It was assumed that the joint distribution of the random terms was multivariate normal, with the following means and (co)variances:

$$\begin{bmatrix} \boldsymbol{h} \\ \mathbf{g} \\ \mathbf{a} \\ \mathbf{e} \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{G_h} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G_g} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{G_a} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{R} \end{bmatrix} \right) \quad (2)$$

where $\mathbf{G_h}$ is a (co)variance matrix corresponding to $\mathbf{h}$ ($\mathbf{G_h} = \sigma_h^2\mathbf{I}$), $\mathbf{G_g}$ is a (co)variance matrix corresponding to $\mathbf{g}$ ($\mathbf{G_g} = \sigma_g^2\mathbf{I}$), G_a is the (co)variance corresponding to $\mathbf{a}$ ($\mathbf{G_a} = \sigma_a^2\mathbf{A}$), $\mathbf{R}$ is a (co)variance matrix corresponding to $\mathbf{e}$ ($\mathbf{R} = \sigma_e^2\mathbf{I}$), $\mathbf{0}$ is a null matrix, σ_h^2 is the hapa variance, σ_g^2 is the genetic group variance, σ_a^2 is the additive genetic variance, σ_e^2 is the residual variance, $\mathbf{I}$ is an identity matrix, and $\mathbf{A}$ is the additive (i.e. numerator) relationship matrix accounting for putative sibship among founders identified in Hamilton et al. (2019b) (Fig. 1) – founders were obtained as spawn from rivers and samples of populations taken in the early life stages of highly fecund species, such as rohu, are prone to the over representation of siblings (Peterman et al., 2016).

Preliminary univariate analyses used weight data from randomly sampled individuals only – noting there were no non-randomly sampled G0 individuals. These analyses revealed evidence of heteroscedastic residuals. Accordingly, weight data were square root transformed prior to final analyses (Hamzah et al., 2014). Transformed weight data were, again, rescaled such that randomly-sampled fish had a mean of zero and

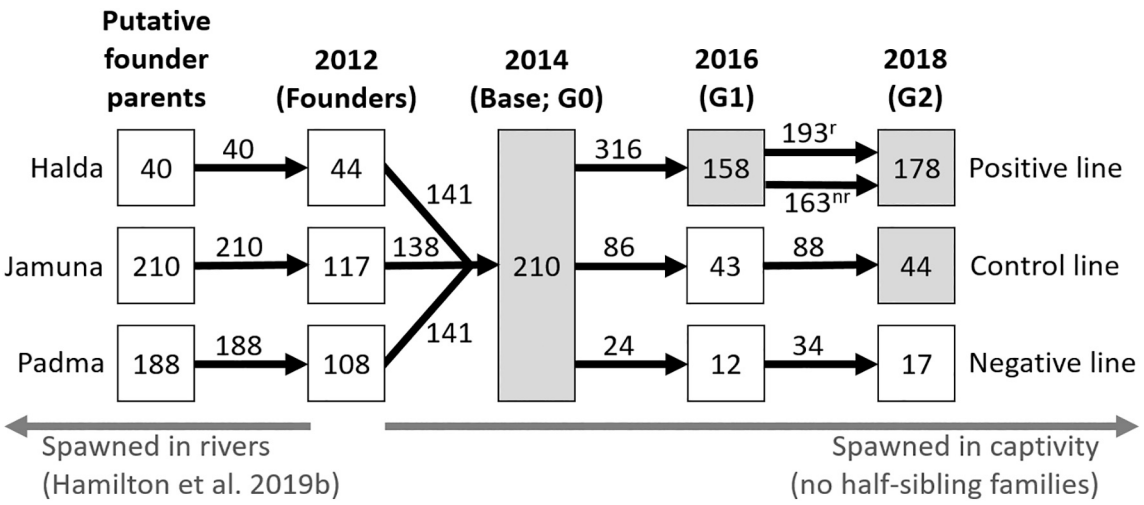

Fig. 1. Worldfish Rohu Genetic Improvement Program cohorts by generation. Founders were sourced as spawn from the Halda, Jamuna and Padma rivers (i.e. three genetic groups; Hamilton et al., 2019b). Three lines were created from the first selected generation (Generation 1; G1): selected for high harvest weight (positive line), selected to have a mean total additive genetic value of zero for harvest weight (control line) and selected for low harvest weight (negative line). The number of full-sibling families are shown in boxes and the number of parents used to generate families are shown adjacent to arrows. Harvest weight data used in the current study were from shaded cohorts. A total of 193 randomly-selected (r) parents and 163 non-randomly selected (nr) parents were used as parents of the G2 positive line.

phenotypic standard deviation of one (within each grow-out pond for harvest weight).

To estimate genetic parameters, a trivariate analysis, extending the univariate model (Eq. (1)), was undertaken with G0 harvest weight, G1 tagging weight and G1 harvest weight for randomly-selected animals as response variables (Table 1). Genetic correlations between traits were estimated, with (co)variance matrices $\mathbf{G_h}$, $\mathbf{G_g}$, $\mathbf{G_a}$ and $\mathbf{R}$ defined as follows – noting that $\sigma_{g1}{}^2$ to $\sigma_{gn}{}^2$ were fixed to their estimated values from univariate analyses due to difficulties with REML convergence:

$$\mathbf{G_h} = \begin{bmatrix} \sigma^2_{h_1}\mathbf{I} & \cdots & \sigma_{h_{1,n}}\mathbf{I} \\ \vdots & \ddots & \vdots \\ \sigma_{h_{n,1}}\mathbf{I} & \cdots & \sigma^2_{h_n}\mathbf{I} \end{bmatrix} \quad (3)$$

$$\mathbf{G_g} = \begin{bmatrix} \sigma^2_{g_1}\mathbf{I} & \cdots & \sigma^2_{g_{1,n}}\mathbf{I} \\ \vdots & \ddots & \vdots \\ \sigma^2_{g_{n,1}}\mathbf{I} & \cdots & \sigma^2_{g_n}\mathbf{I} \end{bmatrix} \quad (4)$$

$$\mathbf{G_a} = \begin{bmatrix} \sigma^2_{a_1}\mathbf{A} & \cdots & \sigma_{a_{1,n}}\mathbf{A} \\ \vdots & \ddots & \vdots \\ \sigma_{a_{n,1}}\mathbf{A} & \cdots & \sigma^2_{a_n}\mathbf{A} \end{bmatrix} \quad (5)$$

$$\mathbf{R} = \begin{bmatrix} \sigma^2_{e_1}\mathbf{I} & \cdots & \sigma_{e_{1,n}}\mathbf{I} \\ \vdots & \ddots & \vdots \\ \sigma_{e_{n,1}}\mathbf{I} & \cdots & \sigma^2_{e_n}\mathbf{I} \end{bmatrix} \quad (6)$$

where the subscripts 1 and n refer to traits one to n, $\sigma_h$denotes the hapa covariance, $\sigma_g$denotes the genetic group covariance, $\sigma_a$denotes the additive genetic covariance, $\sigma_e$denotes the residual covariance and all other terms are as previously described. Standard errors of parameters were estimated from the average information matrix, using a standard truncated Taylor series approximation (Gilmour et al., 2014). For each trait, the within-genetic-group narrow-sense heritability (h^2) was estimated as:

$$\widehat{h}^2 = \frac{\widehat{\sigma}^2_a}{\widehat{\sigma}^2_h + \widehat{\sigma}^2_a + \widehat{\sigma}^2_e}. \quad (7)$$

To estimate total additive genetic values for harvest weight, a bivariate analysis was undertaken – tagging weight was excluded as a response variable to ensure EBVs for harvest weight were not affected by tagging weight measurement data. For each non-randomly sampled individual (i), a vector of estimated breeding values (EBVs) for G0 and G1 harvest weight ($\widehat{\boldsymbol{a}}_i$) was then generated as follows (Mrode, 2013; Mrode and Swanson, 2004):

$$\widehat{\mathbf{a}}_\mathbf{i} = \mathbf{W_1PA} + \mathbf{W_2YD} \quad (8)$$

$$\mathbf{PA} = (\widehat{\mathbf{a}}_\mathbf{s} + \widehat{\mathbf{a}}_\mathbf{d})/2 \quad (9)$$

$$\mathbf{YD} = \mathbf{y} - \mathbf{X}\widehat{\mathbf{b}} - \mathbf{Z_1}\widehat{\mathbf{h}} - \mathbf{Z_2Q}\widehat{\mathbf{g}} \quad (10)$$

$$\mathbf{W_1} = (\mathbf{DIAG})^{-1}2\begin{bmatrix} \widehat{\sigma}^2_{a_1} & \cdots & \widehat{\sigma}_{a_{1,n}} \\ \vdots & \ddots & \vdots \\ \widehat{\sigma}_{a_{n,1}} & \cdots & \widehat{\sigma}^2_{a_n} \end{bmatrix}^{-1} \quad (11)$$

$$\mathbf{W_2} = \mathbf{I} - \mathbf{W_1} \quad (12)$$

$$\mathbf{DIAG} = \mathbf{Z}'\begin{bmatrix} \widehat{\sigma}^2_{e_1} & \cdots & \widehat{\sigma}_{e_{1,n}} \\ \vdots & \ddots & \vdots \\ \widehat{\sigma}_{e_{n,1}} & \cdots & \widehat{\sigma}^2_{e_n} \end{bmatrix}^{-1}\mathbf{Z} + \begin{bmatrix} \widehat{\sigma}^2_{a_1} & \cdots & \widehat{\sigma}_{a_{1,n}} \\ \vdots & \ddots & \vdots \\ \widehat{\sigma}_{a_{n,1}} & \cdots & \widehat{\sigma}^2_{a_n} \end{bmatrix}^{-1}2\alpha_{par} \quad (13)$$

where $\mathbf{Z}$ is a square matrix with zeros in off-diagonals that relates records for each trait with individual i; α_{par} is the mean of the elements of $\boldsymbol{A}^{-1}$ (the inverse of the additive relationship matrix) between individual i and its parents; $\mathbf{I}$ is an identity matrix; $\widehat{\mathbf{a}}_\mathbf{s}$ and $\widehat{\mathbf{a}}_\mathbf{d}$ are vectors of EBVs for the sire and dam of individual i; $\mathbf{PA}$ is a vector of average parental EBVs; and $\mathbf{YD}$ is a vector of yield deviations – yield deviations represent the yields (i.e. harvest weight) of each animal adjusted for all effects other than breeding values and errors (Mrode, 2013; Mrode and Swanson, 2004). All other parameters are as previously described but excluding all elements not relating to individual i. Estimated total additive genetic effects for G1 harvest weight (Wolak and Reid, 2017) were subsequently computed according to the following formula:

$$\widehat{\mathbf{u}} = \mathbf{X}\widehat{\mathbf{b}}_{\boldsymbol{mean}} + \mathbf{Z_2Q}\widehat{\mathbf{g}} + \widehat{\mathbf{a}} \quad (14)$$

Table 1
Summary data for Generation 0 (G0), G1 and G2 Worldfish Rohu Genetic Improvement Program families included in the current study. Standard deviations are in parentheses.

	Generation 0	Generation 1	Generation 2	
	Base population	Selected	Selected	Control
Spawning				
Spawn batches	4	8	5	
Family count per batch	39–57	6–25	10–38	1–21
Nursing				
Family count	210	135[a]	126[b]	44
Conclusion of nursing				
Tagging age (days)	186–223	242–255	217–228	217–226
Tagging measurement date	Jan-Feb 2015	Feb-Mar 2017	Feb-Mar 2019	
Tagging count				
• Randomly sampled	11,863	5074	6042	2189
• Non-randomly sampled	0	4672	0	0
Tagging weight mean (g)				
• Randomly sampled	13.5 (7.9)	9.9 (9.1)	9.9 (10.1)	9.1 (8.9)
• Non-randomly sampled	–	11.7 (8.4)	–	
Progeny testing				
Grow out ponds	7[c]	8[d]	8[e]	
Conclusion of progeny testing				
Harvest age (days)	445–489	553–604	463–504	
Harvest measurement date	Oct-Nov 2015	Jan 2018	Nov 2019	
Harvest count				
• Randomly sampled	10,248	4656	5366	1977
• Non-randomly sampled	0	4384	0	0
Harvest weight mean (g)				
• Randomly sampled	545 (131)	340 (201)	203 (133)	164 (107)
• Non-randomly sampled	–	377 (219)	–	–

[a] Of these, seven had no non-randomly sampled individuals.
[b] Only data for families with both parents from the 135 retained G1 families were retained.
[c] Four monoculture and three polyculture ponds all fed supplementary formulated fish feed.
[d] Four monoculture and four polyculture ponds, four ponds fed supplementary formulated fish feed and four ponds with no supplementary feeding.
[e] Eight monoculture ponds – four ponds fed supplementary formulated fish feed and four ponds with no supplementary feeding.

where $\hat{\mathbf{u}}$ is a vector of estimated total additive genetic effects and $\hat{\mathbf{b}}_{mean}$ is a vector of trait (i.e. G0 harvest weight, G1 tagging weight and G1 harvest weight) mean fixed effects. Elements of $\hat{\mathbf{u}}$ were then rescaled (i.e. divided by $\sqrt{\sigma_a^2}$ corresponding to the trait) so as to be expressed in additive genetic standard deviation units. See Supplementary Materials 1 for a worked example implementing Eqs. (8) to (13).

To gauge the extent to which differences in estimated total additive genetic effects between randomly sampled and non-randomly sampled G1 individuals were inherited by their progeny, the mean harvest weight of G2 families (i.e. G1 progeny) by grow out pond and 'category of parents' were estimated; where categories of parents were i) control line, ii) two randomly sampled parents, iii) one randomly sampled parent and the other non-randomly sampled and iv) two non-randomly sampled parents. Means and standard errors were estimated separately for each grow out pond, using the lmer and lsmeans functions in R (Bates et al., 2015; Lenth, 2016; R Core Team, 2020), by fitting a mixed model with spawn batch and category of parents fitted as fixed effects and family as a random effect. Only data from G2 families for which total additive genetic effects were estimated for both parents (i.e. parents from G1 families with 72 or more individuals available for tagging) were included in the analysis.

3. Results

Harvest weight narrow-sense heritabilities were low to moderate, 0.17 for G0 harvest weight and 0.24 for G1 harvest weight (Table 2) and the additive genetic correlation between these traits was indistinguishable from one (Table 3). However, the additive variance for G1 harvest weight for was not significantly different from zero. The narrow-sense heritability for G1 tagging weight was moderate (0.48). The additive genetic correlation between G1 tagging weight and G1 harvest weight was favourable and not significantly different from one (Table 3), albeit only moderate (0.43; SE = 0.32).

The mean estimated total additive genetic values for G1 harvest weight of the G1 non-randomly sampled individuals was substantially greater than that for randomly-sampled individuals (0.52 cf. 0.71; Fig. 2a). However, despite their lower average estimated total additive genetic value, a substantial number of randomly sampled individuals were used as parents of G2 families (Fig. 2b). Furthermore, within families (i.e. hapas) the difference between the estimated total additive genetic value for randomly and non-randomly sampled individuals tended to increase as the selected proportion (i.e. number of non-randomly sampled individuals divided by the total number of individuals available for tagging after random sampling) decreased from one to 0.1, with no discernible trend below 0.1 (Fig. 3). It was notable that the tagging weight of fish varied substantially between families and was affected by stocking density during nursing – families from hapas with less than 300 G1 individuals had an average weight of 19.5 g and those with greater than 300 individuals weighed 6.0 g on average.

The G2 progeny of non-randomly sampled G1 parents exhibited greater harvest weight than progeny of randomly sampled individuals in all eight progeny-test ponds (Fig. 4; average grow-out pond least-squares means were 221 g for families with two non-random parents and 190 g for families with two random parents). The average G1 total additive genetic value for harvest weight was 0.96 for randomly-sampled G1 parents contributing to G2 families with two randomly-sampled parents (Fig. 4); 1.19 for randomly-sampled G1 parents contributing to G2 families with one randomly-sampled and one non-randomly-sampled parent; 1.41 for non-randomly-sampled G1 parents contributing to G2 families with one randomly-sampled and one non-randomly-sampled parent; and 1.31 for non-randomly-sampled G1 parents contributing to G2 families with two non-randomly-sampled parents.

4. Discussion

Despite a lack of statistically significant additive genetic variation in G1 harvest weight and a corresponding non-significant genetic correlation between G1 tagging and G1 harvest weight, three lines of evidence indicated that visual identification of large G1 individuals at tagging increased the average total additive genetic worth of sampled individuals for harvest weight (i.e. increased genetic gains). Firstly, the mean G1 harvest weight total additive genetic value for non-randomly-sampled individuals was greater than that for randomly-sampled individuals (Fig. 2). This suggested that individuals non-randomly sampled at a young age (i.e. tagging) were genetically superior to randomly-sampled individuals for harvest weight – although the possibly that this difference was, in part, due to residual within-hapa nursing environment effects cannot be excluded. Secondly, within families, the difference in the total additive genetic value between randomly and non-randomly sampled individuals increased as the non-randomly selected proportion of individuals decreased from approximately one to 0.1 (Fig. 3) – that is, within-families, an increase in selection intensity at tagging corresponded with an increase in genetic

Table 2
Variance components estimates for G1 harvest weight with standard errors in parentheses. Significance from zero is indicated.

Trait	Generation measured	Genetic group variance	Hapa variance	Additive variance	Residual variance	Narrow-sense heritability
Harvest weight	G0	0.110 (0.118)^{ns}	0.123 (0.014)***	0.150 (0.033)***	0.614 (0.019)	0.17 (0.04)
Tagging weight	G1	0.011 (0.045)^{ns}	0.178 (0.058)***	0.162 (0.113)^{ns}	0.001 (0.061)	0.48 (0.32)
Harvest weight	G1	0.022 (0.049)^{ns}	0.081 (0.024)*	0.111 (0.054)^{ns}	0.279 (0.028)	0.24 (0.11)

ns - not significant.
* $P < 0.05$.
*** $P < 0.001$.

Table 3
Genetic correlation ($\hat{r}_g$) (below diagonal) and hapa correlation estimates (above diagonal) with standard errors in parentheses. Significance from one is indicated.

	G0 harvest weight	G1 tagging weight	G1 harvest weight
G0 harvest weight		–	–
G1 tagging weight	0.19 (0.29)***		0.93 (0.11)^{ns}
G1 harvest weight	1.00^{b}	0.43 (0.32)^{ns}	

ns - not significant; ***$P < 0.001$; and ^{b} estimate hit the boundary of the parameter space (not significantly different from one).

gain for harvest weight. Thirdly, the harvest weight of G2 families made from two non-randomly-sampled G1 parents (221 g) was greater than those made from randomly-sampled parents (190 g; Fig. 4), indicating that putative differences in additive genetic value between randomly-sampled and non-randomly-sampled G1 parents were inherited by, and expressed in, G2 progeny. Notably, harvest weights of G2 individuals were independent of G1 nursing environment effects that, conceivably, may have biased estimates of total additive genetic value for G1 individuals. However, it is possible that differences expressed in G2 harvest weights (Fig. 4) were inflated due to the application assortative mating among G1 parents – assortative mating resulted in high-ranking randomly-sampled parents being disproportionally mated with non-randomly sampled individuals.

As the results of our study were in keeping with early selection (i.e. the visual identification and non-random sampling of larger individuals at tagging) increasing genetic gains in harvest weight, consideration should be given to the adoption of this practice in routine carp genetic improvement programs structured in the same way as that tested. However, rapid approaches to assess size of all individuals at tagging should be investigated (e.g. direct measurement or visual/mechanical grading into size classes) to potentially avoid the need to adopt Eqs. (8) to (13) to compute unbiased estimates of total additive genetic value (Mrode, 2013; Mrode and Swanson, 2004; Wolak and Reid, 2017) – such biases are minimised by including data from all individuals under selection in analyses (Pollak et al., 1984). To reduce the number of individuals requiring measurement and to reduce inter-hapa variability in growth rate due to differences in stocking density, thinning to a fixed number of randomly-sampled fish per hapa should also be considered. In our study, each nursing hapa was initially stocked with approximately 1.5 g of one-day-old hatchlings. However, due to variation in early survival, and thus stocking density, among hapas competition for resources and growth rate during nursing was highly variable (Fig. 3). Alternatively, issues associated with variation in environment and management among hapas could be negated entirely by nursing all families in a common environment and using molecular markers to undertake pedigree reconstruction – an approach widely adopted in aquaculture species (Flanagan and Jones, 2019; Hamilton, 2021; Kijas et al., 2019; Sahoo et al., 2017). Under such circumstances walk-back selection could be adopted to enhance genetic gains by applying selection pressure at tagging (Sonesson, 2005). Evidently, walk back selection could also be modified to avoid bias in estimates of genetic parameters and total additive genetic values by retaining an additional randomly sampled, measured and genotyped cohort at selection, and adopting Eqs. (8) to (13) (Mrode, 2013; Mrode and Swanson, 2004).

Our study adopted methods detailed in Mrode (2013) and Mrode and Swanson (2004) (Eqs. (8)–(13)) to avoid bias in estimated total additive genetic values in one generation of non-random selection in a rohu genetic improvement program. However, to optimise the implementation of this approach from a genetic, operational and economic perspective in

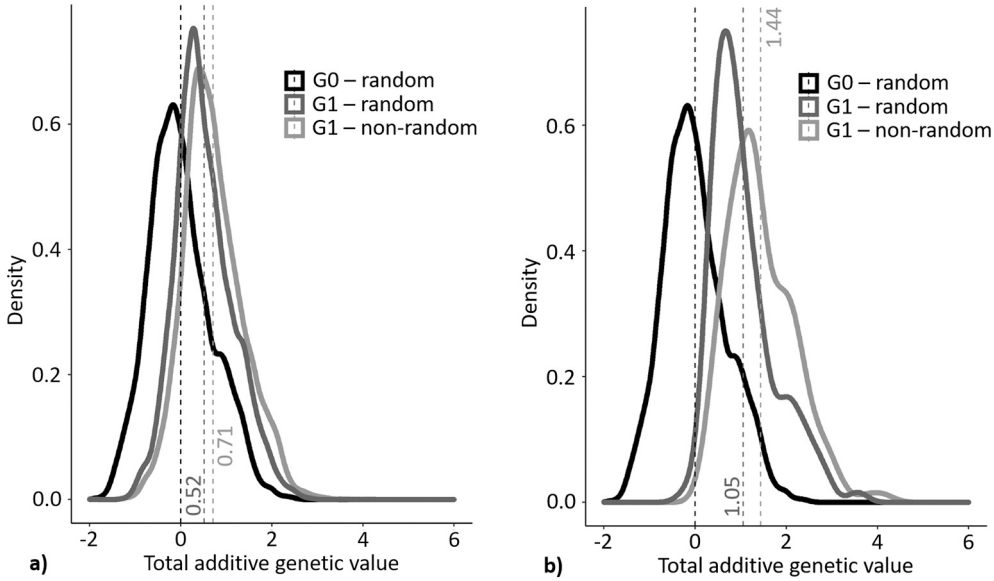

Fig. 2. Estimated total additive genetic value for G1 harvest weight expressed in additive genetic standard deviation units for a) all individuals from Generation 0 (G0) and Generation 1 (G1), b) all G0 individuals and only G1 individuals used as parents of G2. Means are shown as vertical dashed lines.

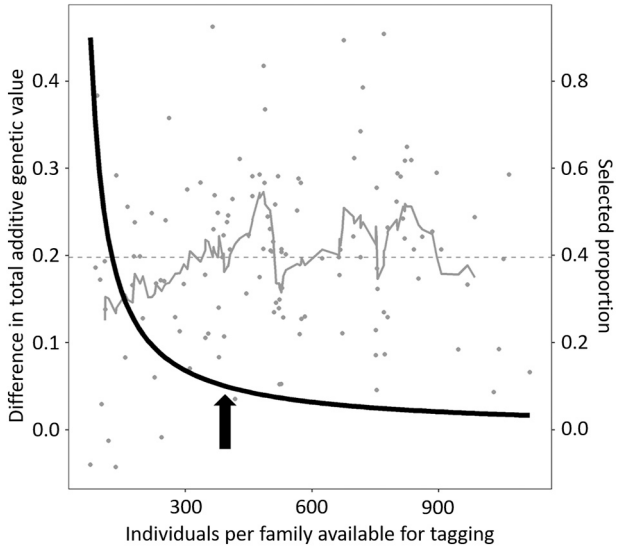

Fig. 3. Difference in the mean estimated total additive genetic value for G1 harvest weight expressed in additive genetic standard deviation units between randomly sampled and non-randomly sampled individuals within Generation 1 families (left side scale) against the number of individuals per family (i.e. hapa) available for tagging at the conclusion of nursing. Grey points represent different families (i.e. hapas), the solid grey line is the rolling mean (11-families) and the dashed grey line is the overall mean. The black line represents the number of non-randomly sampled individuals divided by the total number of individuals available for tagging after random sampling (i.e. selected proportion; right side scale). The black arrow indicates a non-randomly selected proportion of 0.1.

rohu and other genetic improvement programs, further investigation is required to determine i) the most appropriate ratio of random to non-random samples within and across families – given that a smaller random sample reduces the number of individuals and relatives with retained measurement records and thus the accuracy of estimated genetic parameters and total additive genetic values (Gilmour et al., 2014; Mrode, 2013) – and ii) the implications of adopting this method of analysis over multiple generations of selection, particularly in circumstances where few or no randomly-sampled individuals are selected as parents. Furthermore, BLUP-estimated total additive genetic values are sensitive to deviations from Gaussian assumptions (Gianola et al., 2018). Accordingly, the possibility that non-randomly sampled animals selected from the extremes of the tagging weight distribution represent individuals that are disproportionately affected by environmental or non-additive-genetic effects – and thus deviate from the assumptions underpinning analyses – also requires investigation.

5. Conclusion

The harvest weight of G2 families made from non-randomly-sampled G1 parents was greater than those made from randomly-sampled G1 parents, indicating early selection – specifically, visual identification and non-random sampling of larger individuals at the time of tagging – could be used to enhance genetic gains for harvest weight in carp genetic improvement programs of the nature tested. This difference was observed despite the estimated G1 harvest weight additive genetic variance not being statistically significantly different from zero. Further study is required to validate the findings of this study in other species of carp, to optimise nursing methods (e.g. the possible adoption of thinning of randomly-sampled individuals during nursing), to determine the optimal/appropriate number and ratio of random to non-random selections at tagging and investigate the implications of adopting early selection over multiple generations.

Funding

This publication was made possible through financial support provided by USAID (Feed the Future Bangladesh Aquaculture and Nutrition Activity [grant number 72038818IO00002]); and the CGIAR Research Program on Fish Agrifood Systems (FISH), led by WorldFish and supported by contributors to the CGIAR Trust Fund.

CRediT authorship contribution statement

M.G. Hamilton: Conceptualization, Methodology, Formal analysis, Data curation, Writing – original draft, Visualization. **W. Mekkawy:** Conceptualization, Investigation, Data curation, Writing – review & editing, Supervision, Project administration. **Md. Badrul Alam:** Investigation, Resources, Data curation, Writing – review & editing, Supervision, Project administration. **John A.H. Benzie:** Conceptualization, Writing – review & editing, Supervision, Project administration, Funding acquisition.

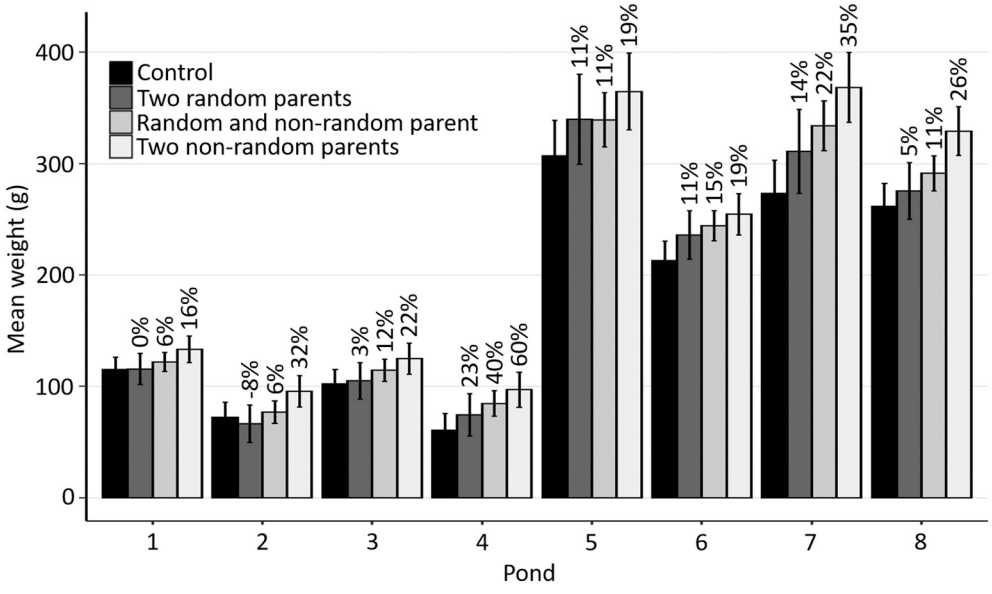

Fig. 4. Mean harvest weight and standard errors for Generation 2 families (i.e. Generation 1 progeny) by grow-out pond and category of parents – control line, two randomly sampled parents, one randomly sampled parent and the other non-randomly sampled and two non-randomly sampled parents. Ponds 1 to 4 received no supplementary formulated fish feed and ponds 5 to 8 received supplementary formulated fish feed. Estimated genetic gains, expressed as a percentage of the mean of the control line families within each pond, are indicated.

Declaration of Competing Interest

The authors declare no conflicts of interest.

Acknowledgements

The authors thank all the members of the WorldFish Carp Genetic Improvement Program technical team in Jashore for managing and measuring fish – Rahman, Md. Mustafizur; Mohmud, Md. Sultan; Sarkar, Uzzal Kumar; Roy, Aashish Kumar; Kundu, Ram Prosad; Shanta, Sirajum Monira; Kamruzzaman, Md.; Hossain, Jamal; and Hossain, Md. Tutul. We also thank the anonymous reviewers for their insights and suggestions.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.aquaculture.2022.738058.

References

Alam, A., 2001. Production, accessibility, marketing and consumption patterns of aquaculture products in Bangladesh. In: FAO (Ed.), Production, Accessibility, Marketing and Consumption Patterns of Freshwater Aquaculture Products in Asia: A Cross Country Comparison. FAO Fisheries Circular 973. Food and Agriculture Organization of the United Nations, Rome, Italy.

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67, 1–48. https://doi.org/10.18637/jss.v067.i01.

DoF, 2020. Yearbook of Fisheries Statistics of Bangladesh 2018–19. Fisheries Resources Survey System (FRSS), Department of Fisheries, Bangladesh.

FAO, 2020. The State of World Fisheries and Aquaculture 2020. Sustainability in Action. FAO, Rome, Italy, p. 244. https://doi.org/10.4060/ca9229en.

Flanagan, S.P., Jones, A.G., 2019. The future of parentage analysis: from microsatellites to SNPs and beyond. Mol. Ecol. 28, 544–567. https://doi.org/10.1111/mec.14988.

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

Gilmour, A.R., Gogel, B.J., Cullis, B.R., Welham, S.J., Thompson, R., 2014. ASReml User Guide Release 4.1 Functional Specification. VSN International Ltd, Hemel Hempstead, UK.

Hamilton, M.G., 2020. Optimal contribution selection in highly fecund species with overlapping generations. J. Hered. 111, 646–651. https://doi.org/10.1093/jhered/esaa051.

Hamilton, M.G., 2021. Maximum likelihood parentage assignment using quantitative genotypes. Heredity. https://doi.org/10.1038/s41437-021-00421-0.

Hamilton, M.G., Mekkawy, W., Benzie, J.A.H., 2019a. Sibship assignment to the founders of a Bangladeshi *Catla catla* breeding population. Genet. Sel. Evol. 51, 17. https://doi.org/10.1186/s12711-019-0454-x.

Hamilton, M.G., Mekkawy, W., Kilian, A., Benzie, J.A.H., 2019b. Single nucleotide polymorphisms (SNPs) reveal sibship among founders of a Bangladeshi rohu (*Labeo rohita*) breeding population. Front. Genet. 10 https://doi.org/10.3389/fgene.2019.00597.

Hamilton, M.G., Mekkawy, W., Barman, B.K., Alam, M.B., Karim, M., Benzie, J.A.H., 2021. Genetic relationships among founders of a silver carp (*Hypophthalmichthys molitrix*) genetic improvement program in Bangladesh. Aquaculture 736715. https://doi.org/10.1016/j.aquaculture.2021.736715.

Hamzah, A., Ponzoni, R.W., Nguyen, N.H., Khaw, H.L., Yee, H.Y., Mohd Nor, S.A., 2014. Performance of the genetically improved farmed tilapia (GIFT) strain over ten generations of selection in Malaysia. Pertanika J. Trop. Agric. Sci. 37, 411–429.

Harville, D.A., 1977. Maximum likelihood approaches to variance component estimation and to related problems. J. Am. Stat. Assoc. 72, 320–338. https://doi.org/10.1080/01621459.1977.10480998.

Henderson, C.R., 1975. Best linear unbiased estimation and prediction under a selection model. Biometrics. 31, 423–447.

Henderson, C.R., Quaas, R.L., 1976. Multiple trait evaluation using relatives' records. J. Anim. Sci. 43, 1188–1197. https://doi.org/10.2527/jas1976.4361188x.

Kijas, J., Hamilton, M., Botwright, N., King, H., McPherson, L., Krsinich, A., McWilliam, S., 2019. Genome sequencing of blacklip and greenlip abalone for development and validation of a SNP based genotyping tool. Front. Genet. 9 https://doi.org/10.3389/fgene.2018.00687.

Lenth, R.V., 2016. Least-squares means: the R package lsmeans. J. Stat. Softw. 69, 1–33. https://doi.org/10.18637/jss.v069.i01.

Mrode, R.A., 2013. Linear Models for the Prediction of Animal Breeding Values, 3rd ed. CABI, Boston, MA.

Mrode, R.A., Swanson, G.J., 2004. Calculating cow and daughter yield deviations and partitioning of genetic evaluations under a random regression model. Livest. Prod. Sci. 86, 253–260. https://doi.org/10.1016/j.livprodsci.2003.09.001.

Peterman, W., Brocato, E.R., Semlitsch, R.D., Eggert, L.S., 2016. Reducing bias in population and landscape genetic inferences: the effects of sampling related individuals and multiple life stages. PeerJ. 4, e1813 https://doi.org/10.7717/peerj.1813.

Pollak, E., Van der Werf, J., Quaas, R., 1984. Selection bias and multiple trait evaluation. J. Dairy Sci. 67, 1590–1595.

Quaas, R.L., 1988. Additive genetic model with groups and relationships. J. Dairy Sci. 71, 1338–1345.

R Core Team, 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Robinson, G.K., 1991. That BLUP is a good thing: the estimation of random effects. Stat. Sci. 6, 15–32.

Sahoo, L., Meher, P.K., Das Mahapatra, K., Saha, J.N., Jayasankar, P., Das, P., 2017. A molecular tool for parentage analysis in Indian major carp, *Labeo rohita* (Hamilton, 1822). Aquac. Int. 25, 1159–1166. https://doi.org/10.1007/s10499-016-0104-z.

Saura, M., Villanueva, B., Fernández, J., Toro, M.A., 2017. Effect of assortative mating on genetic gain and inbreeding in aquaculture selective breeding programs. Aquaculture. 472, 30–37. https://doi.org/10.1016/j.aquaculture.2016.05.013.

Sonesson, A.K., 2005. A combination of walk-back and optimum contribution selection in fish: a simulation study. Genet. Sel. Evol. 37, 587–599. https://doi.org/10.1051/Gse:2005020.

Swan, A.A., Brown, D.J., van der Werf, J.H.J., 2015. Genetic variation within and between subpopulations of the Australian merino breed. Anim. Prod. Sci. 56, 87–94. https://doi.org/10.1071/AN14560.

Van Raden, P., Freeman, A., Rothschild, M., 1984. Maximizing genetic gain under multiple-stage selection. J. Dairy Sci. 67, 1761–1766.

Wolak, M.E., Reid, J.M., 2017. Accounting for genetic differences among unknown parents in microevolutionary studies: how to include genetic groups in quantitative genetic animal models. J. Anim. Ecol. 86, 7–20. https://doi.org/10.1111/1365-2656.12597.

WorldFish, 2004. Animal Care, Welfare and Ethics Policy of WorldFish Center. WorldFish, Penang, Malaysia.