



Brief Communication

# Optimal Contribution Selection in Highly Fecund Species With Overlapping Generations

Matthew G. Hamilton<sup>®</sup>

From the CSIRO Aquaculture, CSIRO Agriculture and Food, Castray Esplanade, Hobart, Tasmania 7001, Australia (Hamilton) and Current address: WorldFish, Jalan Batu Maung, Batu Maung, 11960 Bayan Lepas, Penang, Malaysia (Hamilton).

Address correspondence to M. G. Hamilton at the address above, or e-mail: [M.Hamilton@cgiar.org](mailto:M.Hamilton@cgiar.org).

Received July 14, 2020; First decision July 11, 2020; Accepted November 27, 2020.

Corresponding Editor: John Stommel

## Abstract

Optimal contribution approaches to parental selection in closed breeding populations aim to maximize genetic gains, while restraining long-term inbreeding. The adoption of optimal contribution selection (OCS) in highly fecund outcrossing species presents a number of challenges not applicable to species of low fecundity (e.g., livestock) for which they were developed. This is particularly true if overlapping-generations or rolling-front breeding strategies are applied, in which case the number of individuals per family in juvenile (i.e., sexually immature) age groups is not necessarily known but is likely to be large. In these circumstances, conventional OCS procedures must be modified or a large number of dummy individuals defined, making computations onerous. Here, an approach to OCS is presented that involves the use of “between-family relationship matrices” instead of “between-individual relationship matrices.” The method is applicable to breeding programs involving highly fecund outcrossing species with overlapping generations, including circumstances where the number of juveniles per family is unknown but large.

**Subject areas:** Quantitative genetics

**Key words:** selective breeding program, parental selection, restricting inbreeding, genetic gain, between-family relationship matrix

Managers of closed selective breeding populations of outcrossing species generally aim to maximize genetic gains while restraining long-term inbreeding. This represents an optimization problem that can be solved, or approximated, using “optimal contribution selection” (OCS) methods, including Lagrange-multiplier procedures (Meuwissen and Sonesson 1998; Tang et al. 2008; Dagnachew and Meuwissen 2016), semi-definite programming (Pong-Wong and Woolliams 2007) and second-order cone programming (Yamashita et al. 2018). Optimal contribution selection methods are dynamic rules used to select parents by maximizing the mean additive genetic value of selected parents while constraining the mean relationship (or, equivalently, co-ancestry)—a predictor of future inbreeding—among members of a breeding population (Meuwissen 1997; Meuwissen and Sonesson 1998; Hinrichs et al. 2006; Pong-Wong

and Woolliams 2007; Skaarud et al. 2011; Kerr et al. 2015; Chapuis et al. 2016; Dagnachew and Meuwissen 2016). A breeding population is herein defined as the collection of all individuals that have the potential to contribute genes to future generations (Kerr et al. 2015). For a concise overview of the history, methods, and application of OCS in selective breeding programs, refer to Woolliams et al. (2015).

Overlapping-generations or rolling-front, as opposed to discrete-generation, breeding strategies are adopted in many outcrossing species to distribute workload more evenly, more fully utilize infrastructure, and to ensure the best individuals can be used as parents as soon as they reach sexual maturity (Borrallho and Dutkowski 1998; Kube et al. 2012; Kerr et al. 2015). An overlapping-generations approach also allows parental contributions from failed families, or additional contributions from underrepresented candidates of high

genetic value, to be captured in later rounds of parental selection (i.e., selection rounds).

The adoption of OCS approaches in highly fecund species (e.g., many aquaculture and forestry species) presents a number of challenges that are not applicable in terrestrial animals, for which the methods were initially developed (Hinrichs et al. 2006; Skaarud et al. 2011; Kerr et al. 2015). These challenges are exacerbated where overlapping generations breeding strategies are applied (Meuwissen and Sonesson 1998; Kerr et al. 2015).

Firstly, when overlapping generations are present, it is necessary to account for varying parental contributions from different age groups to the future breeding population. This issue was addressed by Meuwissen and Sonesson (1998) with the use of an  $r$  vector, the elements of which reflect current plus expected future contributions of each age group until they die divided by the generation interval. This approach has been widely adopted in the implementation of OCS with overlapping generations. However, conjecture remains as to the most appropriate means of determining  $r$ —refer to Meuwissen and Sonesson (1998), Grundy et al. (2000), and Kerr et al. (2015) for discussion of approaches used to determine parental contributions from different age groups.

Secondly, in highly fecund species, the number of individuals in the breeding population is not always known, particularly in juvenile (i.e., sexually immature) age groups (Kerr et al. 2015). In addition, in some species, pedigree reconstruction (Kube et al. 2012) and, increasingly, genomic selection (Sonesson et al. 2012; Tsai et al. 2016) methods are adopted. In these cases, families of unlabeled individuals are generally pooled at a young age, and at the time selection decisions are made, neither the original number of individuals nor the retained number of individuals per family may be known in juvenile age groups (Skaarud et al. 2011; Kerr et al. 2015). One imperfect but simple approach to address this issue in the application of OCS is to generate dummy individuals for juvenile families with unknown numbers of individuals (Kerr et al. 2015).

Thirdly, in the case of overlapping generations for highly fecund species, the number of individuals that must be considered is potentially large, and solving the optimization problem can be computationally demanding (Meuwissen and Sonesson 1998; Hinrichs et al. 2006; Kerr et al. 2015). In highly fecund aquaculture and forestry species, a parent may have tens or hundreds of thousands of viable progeny. Various means of reducing the resulting computational burden have been proposed (Hinrichs et al. 2006; Skaarud et al. 2011). For example, Hinrichs et al. (2006) detailed modifications to Meuwissen's (1997) Lagrangian multiplier method, that substantially reduced the computational task for highly fecund species by 1) requiring inversion an additive relationship ( $A$ ) matrix of dimensions equal to the number of parents, rather than the number of candidates, and 2) adopting the Sherman-Morrison formula to negate the need for repeated inversion of  $A$ . This method can be extended to overlapping generations but still requires all individuals in the population to be known, or dummy individuals generated, in circumstances where only family-level information is available for juvenile age groups.

Fourthly, in selective breeding programs involving highly fecund species, it is generally necessary to apply a preselection step or steps to identify a small number of “candidate parents” from a large number of juvenile individuals. This issue is complicated further in the context of forestry species—due to their long-lived nature and ability to be cloned. Kerr et al. (2015) addressed these issues in the context of *Pinus* and *Eucalyptus* genetic improvement programs, by

adopting classes defined by reproductive status (i.e., “new progeny,” “juvenile,” and “candidate parent”), rather than age, in the application of OCS.

Here, an approach to OCS is presented that does not require the number of individuals in each family to be known. This method is directly applicable to highly fecund species with overlapping generations and uniform numbers of individuals per family, but has applications in other circumstances (e.g., where the number of juveniles per family is unknown but large). It involves the replacement of between-individual relationship matrices with “between-family relationship matrices” (Chapuis et al. 2016) in the definition of, and solution to, the optimization problem.

## Methods

### Between-Family Relationship Matrices

If full-sib families are large and the number of individuals per family is uniform, the average ( $\bar{A}$ ) of the elements of a between-individual additive (numerator) relationship matrix ( $A$ ) can be approximated by the average of a between-family relationship matrix (Supplementary Material 1):

$$\bar{A} \approx \frac{1' A_f 1}{n_f^2} \tag{1}$$

given

$$A_f = \begin{bmatrix} b_1 & c_{12} & \cdots & c_{1n_f} \\ c_{21} & b_2 & \cdots & c_{2n_f} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n_f 1} & c_{n_f 2} & \cdots & b_{n_f} \end{bmatrix} \tag{2}$$

where  $\bar{A}$  is the average of the elements of  $A$ ,  $n_f$  is the number of families,  $A_f$  is a between-family relationship matrix (Chapuis et al. 2016),  $b_x$  is the relationship between 2 individuals from family  $x$  and  $c_{xy}$  is the relationship between an individual from family  $x$  and an individual from family  $y$ . It follows that the elements of  $A_f$  represent twice the Wrights inbreeding coefficient ( $F$ ; Wright 1922) of the progeny of a cross between an individual from family  $x$  with a different individual from family  $y$ .

### Defining the Optimization Problem

Optimal contribution selection approaches generally select parents to maximize  $c_t' EBV_t$  subject to constraints on sex-age group contributions and the average relationship—refer to Meuwissen and Sonesson (1998) and Woolliams et al. (2015) and note that the method of Meuwissen and Sonesson (1998) is explicitly articulated for the case of separate sex-age classes by Tang et al. (2008). In general terms, the optimization problem can be expressed, using the notation of Tang et al. (2008), as:

$$\text{Maximize: } c_t' EBV_t \tag{3}$$

Subject to:

$$Q_1' c_1 = s_1 = s - Q_2' c_2 \tag{4}$$

and

$$C_{t+1} \leq (r_1 + r_3)^2 \overline{\overline{A_{t+1(a,a)}}} + 2(r_1 + r_3) \overline{\overline{A_{t+1(a,b)}}} r_b + r_b' \overline{\overline{A_{t+1(b,b)}}} r_b \tag{5}$$

given

$$\overline{\overline{A_{t+1(a,a)}}} = c'_1 A_{11} c_1 + 2c'_1 A_{12} c_2 + c'_2 A_{22} c_2 \quad (6)$$

$$\overline{\overline{A_{t+1(a,b)}}} = c'_1 A_{t+1(1,b)} J + c'_2 A_{t+1(2,b)} J \quad (7)$$

$$\overline{\overline{A_{t+1(b,b)}}} = J' A_{t+1(b,b)} J \quad (8)$$

where a subscript of 1 denotes available candidates for which contributions are to be optimized—herein referred to as Class 1 candidates (Meuwissen 1997; Woolliams et al. 2015)—and a subscript of 2 denotes available candidates with fixed (i.e., previously committed or mated) contributions—herein referred to as Class 2 candidates— $c_t$  is a vector of genetic contributions of all individuals at time  $t$  (i.e., contributions from Class 1 and Class 2 candidates in the selection round undertaken at time  $t$ ) to age group 1 at time  $t+1$ ;  $EBV_t$  is a vector of estimated breeding values for individuals at time  $t$ ;  $Q$  is an incidence matrix relating individuals to sex-age groups, other than age 1 at time  $t+1$ ;  $s$  is a contribution vector of sex groups, in which case  $s = [0.50.5]'$ , or sex-age groups (if the elements of  $r$  are to be fixed) in which case contributions of males and females groups must both sum to 0.5;  $C_{t+1}$  is a scalar representing the constraint on the average of relationships between members of the population at time  $t+1$ ;  $r_1$  and  $r_3$  are weight scalars denoting the sum of all current and future contributions (until death at age  $q$ ) of age 1 males and females, respectively, divided by the sex-specific generation interval;  $A$  is a between-individual additive genetic relationship matrix at time  $t$ ;  $A_{t+1(1,b)}$  is a between-individual additive genetic relationship matrix of dimensions equal to the number of Class 1 candidates by the number of existing individuals to be retained;  $A_{t+1(2,b)}$  is equivalent to  $A_{t+1(1,b)}$  for Class 2 candidates;  $J$  is a matrix with rows representing all individuals in age groups 2 to  $q$  at time  $t+1$  and columns representing the corresponding sex-age groups (the  $j$ th column of  $J$  consists of zeros except for the  $n$  elements that correspond to the animals in sex-age group  $j$  which equal  $1/n$ , where  $n$  is the number of animals in sex-age  $j$ );  $r_b$  is a vector and equals  $[r'_2 r'_4]'$  where  $r_2$  and  $r_4$  are weight vectors denoting the long-term contributions of sex-age groups, other than age 1, for males and females, respectively, divided by the sex-specific generation interval. Note that individuals in age groups 1 to  $q-1$  at time  $t$  are identical to age groups 2 to  $q$  at time  $t+1$  and thus represent “existing individuals to be retained”;  $(r_1 + r_3) \overline{\overline{A_{t+1(a,a)}}}$  represents the average relationship between new progeny;  $2(r_1 + r_3) \overline{\overline{A_{t+1(a,b)}}} r_b$  represents the average relationship between new progeny and existing individuals to be retained; and  $r'_b \overline{\overline{A_{t+1(b,b)}}} r_b$  represents the average relationship between existing individuals to be retained—refer to Tang et al. (2008) for further clarification (e.g., the distinction between  $A$  matrices with single bars and double bars). Further note that to apply OCS in circumstances where genomic selection is adopted, pedigree-based relationship matrices ( $A$ ) can be replaced with a marker-based (genomic;  $G$ ) relationship matrices (Sonesson et al. 2012; Woolliams et al. 2015).

### Incorporating Between-Family Relationship Matrices

If full-sib families are large and the number of individuals per family is uniform, the average of the between-individual relationship matrix for existing individuals to be retained (i.e.,  $\overline{\overline{A_{t+1(b,b)}}}$ ) can be approximated by the average of the corresponding between-family relationship matrix. That is:

$$J'_f A_{t+1(b,b)_f} J_f \approx \overline{\overline{A_{t+1(b,b)}}} = J' A_{t+1(b,b)} J \quad (9)$$

given

$$J_f = [L, L] \quad (10)$$

where  $A_{t+1(b,b)_f}$  is a between-family relationship matrix with dimensions equal to the number of families in existing sex-age groups to be retained ( $n_f$ )  $L$  is a matrix with  $n_f$  rows and  $q-1$  columns representing age groups 2 to  $q$  at time  $t+1$ . The  $j$ th column of  $L$  consists of zeros except for the  $n$  elements that correspond to the families in sex-age group  $j$  which equal  $1/n$ , where  $n$  is the number of families in sex-age group  $j$ .

Furthermore, if full-sib families are large and the number of individuals per family is uniform, the average relationship matrix between candidates and existing individuals to be retained (i.e.,  $\overline{\overline{A_{t+1(a,b)}}}$ ) can be approximated as  $c'_1 A_{t+1(1,b)_f} J_f + c'_2 A_{t+1(2,b)_f} J_f$ . That is:

$$A_{t+1(1,b)_f} J_f \approx A_{t+1(1,b)} J = M_1 \quad (11)$$

$$A_{t+1(2,b)_f} J_f \approx A_{t+1(2,b)} J = M_2 \quad (12)$$

where  $A_{t+1(1,b)_f}$  is a relationship matrix of dimensions equal to the number of Class 1 candidates by  $n_f$ , with elements equal to the relationship between the candidate and a different individual from the family; and  $A_{t+1(2,b)_f}$  is the equivalent of  $A_{t+1(1,b)_f}$  for Class 2 candidates. The optimization problem (Equations 3–8) can then be re-expressed incorporating between-family relationship matrices (Supplementary Material 2). This problem can, in turn, be solved using the Lagrange multiplier methods (Meuwissen and Sonesson 1998; Hinrichs et al. 2006); or other approaches, such as semi-definite programming (Pong-Wong and Woolliams 2007) and second-order cone programming (Yamashita et al. 2018).

### Validation

Comparison of  $\overline{\overline{A_{t+1(b,b)}}}$ , calculated as  $J'_f A_{t+1(b,b)_f} J_f$ , with  $J' A_{t+1(b,b)} J$  was undertaken using simulated data. Simulations involved the construction of a pedigree 20 selection rounds deep from a base population of unrelated individuals, with 1280 individuals per selection round and random selection of parents. Within selection rounds, parents were randomly mated. A fixed age structure was assumed—40% males from age group 2 (i.e., selection round 19), 10% females from age group 2, 10% males from age group 3 (i.e., selection round 18) and 40% females from age group 3—corresponding to a  $s$  vector of (0.1 0.4 0.0 0.4 0.1 0.0) and a  $r$  vector of (0.045 0.227 0.227 0.143 0.179 0.179). Thus age group 1 (i.e., selection round 20) was assumed to represent a juvenile age group from which no candidates were available. For simplicity, all candidate parents were assumed to be Class 1. Six scenarios were modeled: 10, 40 and 160 families per selection round with either equal numbers of individuals per family or unequal numbers of individuals per family (Table 1 and Figure 1). To examine the impact of the use of a between-family relationship matrices to estimate  $\overline{\overline{A_{t+1(b,b)}}}$  (i.e., to validate Equation 9), for each of 100 simulated pedigrees of each of the 6 scenarios,  $J' A_{t+1(b,b)} J - J' A_{t+1(b,b)_f} J_f$  was calculated and expressed as a percentage of the corresponding elements of  $J' A_{t+1(b,b)} J$ . Furthermore, for each of the same 100 simulated pedigrees, the  $M_1$  the matrix was generated to validate Equation 11. Retaining only one row (i.e., individual) per family, for each column of  $M_1$ , the intercept, slope and  $R^2$  of a simple linear regression were calculated,

fitting the column of  $M_1$  derived from  $A_{t+1(1,b)}J$  as the dependent variable and the column of  $M_1$  derived from  $A_{t+1(1,b)}J_f$  as the explanatory variable.

An R package (R Core Team 2020) entitled “OptContR,” incorporating the function “get.c,” was developed to implement and validate the use of between-family relationship matrices in OCS. This package, a simple spreadsheet-based example of implementation and the code used for simulations, is available at <https://github.com/mghamilton/OptContR>.

### Results

For scenarios with large families and a uniform number of individuals per family, minimal bias was evident in the diagonal elements of  $\overline{A_{t+1(b,b)}}$ . That is, values derived from  $J_f' A_{t+1(b,b)} J_f$  were only marginally less than those derived from  $J' A_{t+1(b,b)} J$ . However, bias increased as the number of individuals per family decreased (compare Table 1, a with Table 1, e). As expected, no bias was evident in off-diagonals (Table 1, a). However, in scenarios where the number of individuals per family was not uniform, substantial bias was evident (Table 1, b, d, and f).

For the column of  $M_1$  corresponding to age 1 (i.e., juvenile males, no bias was evident when  $M_1$  was computed as  $A_{t+1(1,b)}J_f$  (Figure 1). In the case of age 2 males (i.e., mature animals, where individuals are represented in both rows and columns of  $A_{t+1(1,b)}$ ), minimal bias was evident where the number of individuals per family was large and uniform, with the mean of intercepts approximately zero, slopes approximately 1—although always greater than 1—and  $R^2$  approximately 1. In contrast, in scenarios where the number of individuals per family was small and/or where the number of individuals per family was not uniform, substantial bias was evident.

Equivalent trends were evident for columns of  $M_1$  corresponding to females (Supplementary Material 3).

### Discussion

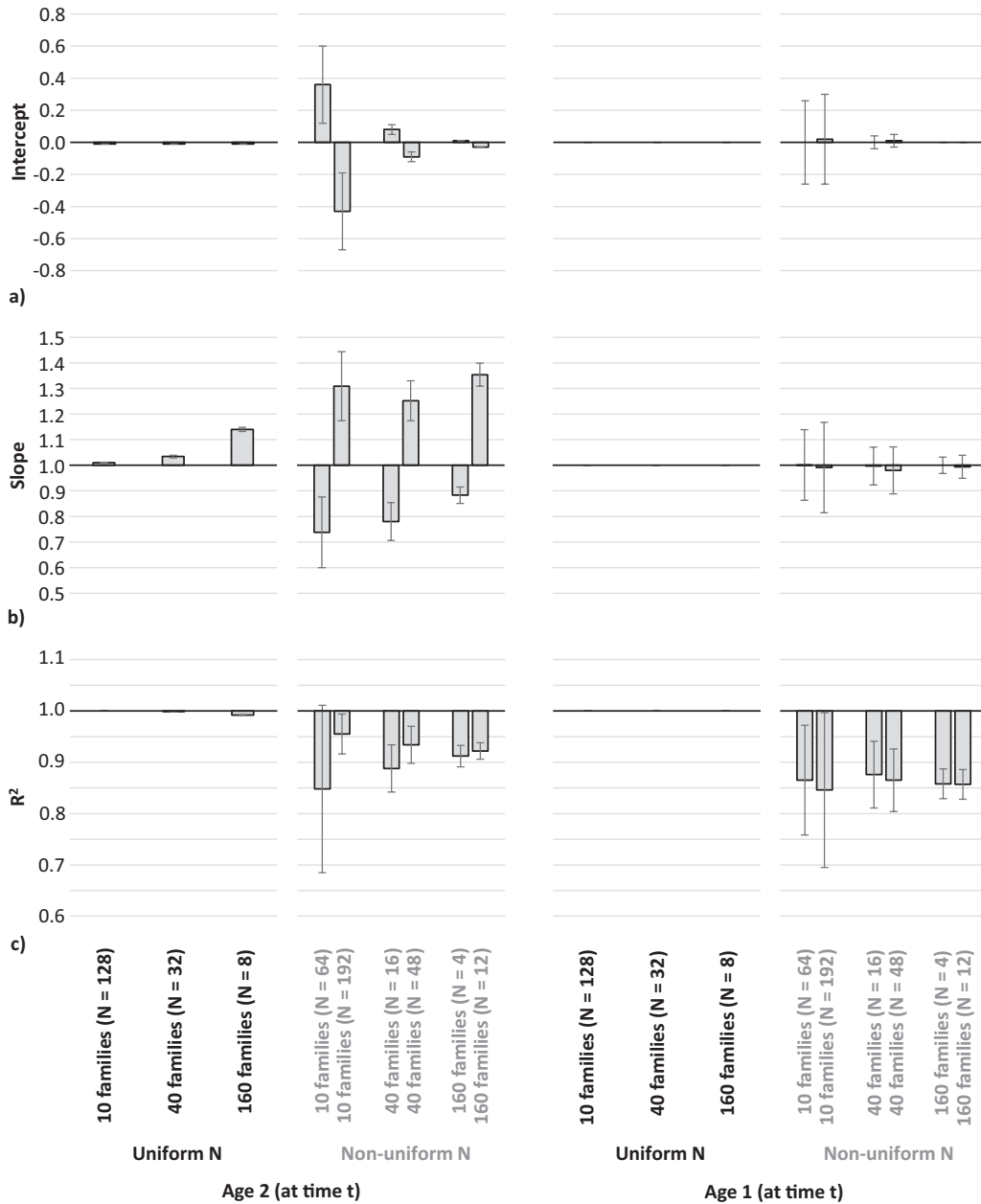
Downward bias in the elements of  $\overline{A_{t+1(b,b)}}$  and  $M_1$  results in under estimation of both the average of relationships between existing individuals to be retained (i.e.,  $r_b' \overline{A_{t+1(b,b)}} r_b$ ) and the average of relationships between new progeny and existing individuals to be retained (i.e.,  $2(r_1 + r_3)(c_1' M_1 + c_2' M_2) r_b$ ). In terms of the constraint on  $C_{t+1}$  (Equation 5), these biases act to relax the constraint on the average of relationships between new progeny (i.e.,  $(r_1 + r_3)^2 \overline{A_{t+1(a,a)}}$ ). However, in circumstances where the number of individuals per family is large and uniform, biases in the elements of  $\overline{A_{t+1(b,b)}}$ , when computed as  $J_f' A_{t+1(b,b)} J_f$  (Table 1, a), and  $M_1$ , when computed as  $A_{t+1(1,b)}J_f$  (Figure 1), are small and thus, in the application of OCS methods, unlikely to materially impact on parental selections in any given selection round or long-term genetic gains and inbreeding.

In circumstances where the number of individuals per family is not uniform, the impact of adopting between-family relationship matrices on the computation of  $\overline{A_{t+1(b,b)}}$  and  $M_1$  is less predictable and biases more extreme (Table 1, b, d, and f; Figure 1). However, by observing that there are no significant biases in the columns of  $M_1$  corresponding to juvenile age groups (age 1 in Figure 1 and Supplementary Material 3), “combined individual-and-between-family relationship matrices” may be adopted to minimize biases in  $M_1$ , while negating the need to know the number of individuals in each juvenile family (Supplementary Materials 4 and 5). In adopting

**Table 1.** Validation of Equation 9—mean difference (and standard deviation) between  $\overline{A_{t+1(b,b)}}$  computed as  $J' A_{t+1(b,b)} J$  and  $J_f' A_{t+1(b,b)} J_f$ , expressed as a percentage of the elements of  $J' A_{t+1(b,b)} J$ , from 100 simulated pedigrees for each of 6 scenarios (a–f)

a) 10 families with 128 individuals					b) 5 families with 64 individuals and 5 families with 192 individuals				
	Male (2)	Male (1)	Female (2)	Female (1)		Male (2)	Male (1)	Female (2)	Female (1)
Male (2)	0.42 (0.03)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	Male (2)	5.96 (2.63)	0.03 (3.36)	5.64 (2.64)	0.03 (3.36)
Male (1)	0.00 (0.00)	0.40 (0.03)	0.00 (0.00)	0.00 (0.00)	Male (1)	0.03 (3.36)	5.27 (2.77)	0.03 (3.36)	4.97 (2.78)
Female (2)	0.00 (0.00)	0.00 (0.00)	0.42 (0.03)	0.00 (0.00)	Female (2)	5.64 (2.64)	0.03 (3.36)	5.96 (2.63)	0.03 (3.36)
Female (1)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.40 (0.03)	Female (1)	0.03 (3.36)	4.97 (2.78)	0.03 (3.36)	5.27 (2.77)
c) 40 families with 32 individuals					d) 20 families with 16 individuals and 20 families with 48 individuals				
	Male (2)	Male (1)	Female (2)	Female (1)		Male (2)	Male (1)	Female (2)	Female (1)
Male (2)	1.61 (0.07)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	Male (2)	5.92 (1.98)	0.08 (1.72)	4.68 (2.01)	0.08 (1.72)
Male (1)	0.00 (0.00)	1.55 (0.06)	0.00 (0.00)	0.00 (0.00)	Male (1)	0.08 (1.72)	5.76 (1.75)	0.08 (1.72)	4.56 (1.78)
Female (2)	0.00 (0.00)	0.00 (0.00)	1.61 (0.07)	0.00 (0.00)	Female (2)	4.68 (2.01)	0.08 (1.72)	5.92 (1.98)	0.08 (1.72)
Female (1)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	1.55 (0.06)	Female (1)	0.08 (1.72)	4.56 (1.78)	0.08 (1.72)	5.76 (1.75)
e) 160 families with 8 individuals					f) 80 families with 4 individuals and 80 families with 12 individuals				
	Male (2)	Male (1)	Female (2)	Female (1)		Male (2)	Male (1)	Female (2)	Female (1)
Male (2)	6.37 (0.11)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	Male (2)	10.31 (0.82)	-0.01 (0.85)	5.48 (0.87)	-0.01 (0.85)
Male (1)	0.00 (0.00)	6.15 (0.13)	0.00 (0.00)	0.00 (0.00)	Male (1)	-0.01 (0.85)	9.73 (0.78)	-0.01 (0.85)	5.05 (0.83)
Female (2)	0.00 (0.00)	0.00 (0.00)	6.37 (0.11)	0.00 (0.00)	Female (2)	5.48 (0.87)	-0.01 (0.85)	10.31 (0.82)	-0.01 (0.85)
Female (1)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	6.15 (0.13)	Female (1)	-0.01 (0.85)	5.05 (0.83)	-0.01 (0.85)	9.73 (0.78)

Rows and columns represent sex-age groups at time t.



**Figure 1.** Validation of Equation 11—the (a) intercept, (b) slope, and (c)  $R^2$  of simple linear regression models fitting elements of the columns of  $M_1$  corresponding to males (with only 1 row per family retained) derived from  $A_{t+1(1,b)}J$  as the dependent variable and columns of  $M_1$  derived from  $A_{t+1(1,b)}J_f$  as the explanatory variable. The mean (and standard deviation) of the analysis of 100 simulated pedigrees, for each of 6 scenarios, is presented. For scenarios with non-uniform numbers of individuals per family, results are presented separately for small and large families. N refers to the number of individuals per family.

combined individual-and-between-family relationship matrices, columns in  $A_{t+1(1,b)}$  (and rows in  $J$ ), corresponding to mature age groups are represented by individuals—noting that the identities of individuals in mature age groups are likely to be known—and columns corresponding to juvenile age groups are represented by families. Alternatively, combined individual-and-between-family relationship matrices could be partitioned into “mature” and “juvenile” sub-matrices (Kerr et al. 2015). Combined individual-and-between-family  $A_{t+1(2,b)}$  and  $A_{t+1(b,b)}$  matrices may also be used to compute  $M_2$  and  $A_{t+1(b,b)}$ , respectively. However, some degree of inaccuracy in the elements of  $A_{t+1(b,b)}$  corresponding to juvenile age groups is unavoidable in circumstances where the number

of individuals per juvenile family is unknown (Supplementary Materials 4 and 5).

In conclusion, the use of between-family relationship matrices in the application of OCS procedures to highly fecund species reduces the size of relationship matrices—and the corresponding computational tasks—and, in the case of juvenile families, negates the need for the number of individuals to be known or an arbitrary number of dummy individuals to be generated (Kerr et al. 2015). Approximations of  $A_{t+1(b,b)}$ , and  $M$  matrices (Equations 9–12), computed using between-family relationship matrices, closely reflect those computed using individual relationship matrices in circumstances where the number of individuals per family is large and



the number of individuals per family is uniform. Such circumstances exist in some aquaculture and forestry breeding populations, in which excess individuals per family are grown in a common rearing environment until they are of a size that allows a fixed number of individuals per family to be tagged or labeled (Kube et al. 2011; Hamzah et al. 2014). In breeding programs with large numbers of individuals per family but where individuals in juvenile age groups are not individually identified—because individuals are too numerous, to minimize or simplify data storage (Kerr et al. 2015), or as a consequence of deferred genotyping and pedigree reconstruction (Kube et al. 2012)—combined individual-and-between-family relationship matrices can be adopted in the implementation of OCS procedures to minimize biases in key inputs (i.e.,  $\bar{A}_{t+1(b,b)}$  and  $M$  matrices).

## Supplementary Material

Supplementary material is available at *Journal of Heredity* online.

Supplementary Material 1. Proof of Equation 1.docx

Supplementary Material 2. Definition of the optimization problem and its solution using between-family relationship matrices.docx

Supplementary Material 3. Extension of Figure 1 (columns of  $M_1$  corresponding to females).docx

Supplementary Material 4. Biases using combined individual-and-between-family relationship matrices.docx

Supplementary Material 5. Biases using between-individual relationship matrices with four dummy individuals per juvenile family.docx

## Funding

This work was supported by the CSIRO Agriculture and Food “Genomics platforms to assist applied aquaculture breeding” (AgSIP53) project. WorldFish, through the CGIAR Research Program on Fish Agrifood Systems (FISH), financially supported completion of the manuscript subsequent to the author’s departure from CSIRO.

## Data Availability

A simple spreadsheet-based example of implementation and the code used for simulations is available at <https://github.com/mghamilton/OptContR>.

## References

- Borralho NMG, Dutkowski GW. 1998. Comparison of rolling front and discrete generation breeding strategies for trees. *Can J For Res.* 28:987–993.
- Chapuis H, Pincet C, Colleau JJ. 2016. Optimizing selection with several constraints in poultry breeding. *J Anim Breed Genet.* 133:3–12.
- Dagnachew BS, Meuwissen THE. 2016. A fast Newton–Raphson based iterative algorithm for large scale optimal contribution selection. *Genetics Selection Evolution* 48:1–10.
- Grundy B, Villanueva B, Woolliams JA. 2000. Dynamic selection for maximizing response with constrained inbreeding in schemes with overlapping generations. *Anim Sci.* 70:373–382.
- Hamzah A, Ponzoni RW, Nguyen NH, Khaw HL, Yee HY, Mohd Nor SA. 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.
- Hinrichs D, Wetten M, Meuwissen THE. 2006. An algorithm to compute optimal genetic contributions in selection programs with large numbers of candidates. *J Anim Sci.* 84:3212–3218.
- Kerr RJ, McRae TA, Dutkowski GW, Tier B. 2015. Managing the rate of increase in average co-ancestry in a rolling front tree breeding strategy. *J Anim Breed Genet.* 132:109–120.
- Kube P, Cunningham M, Dominik S, Parkinson S, Finn B, Henshall J, Bennett R, and Hamilton M. 2011. *Enhancement of the Pacific oyster selective breeding program.* CSIRO Marine and Atmospheric Research, Australian Seafood Industries P/L, Seafood CRC, and Fisheries Research and Development Corporation (FRDC), Australia. pp. 117.
- Kube PD, Taylor RS, Elliott NG. 2012. Genetic variation in parasite resistance of Atlantic salmon to amoebic gill disease over multiple infections. *Aquaculture* 364:165–172.
- Meuwissen THE. 1997. Maximizing the response of selection with a predefined rate of inbreeding. *J Anim Sci.* 75:934–940.
- Meuwissen THE, Sonesson AK. 1998. Maximizing the response of selection with a predefined rate of inbreeding: overlapping generations. *J Anim Sci.* 76:2575–2583.
- Pong-Wong R, Woolliams JA. 2007. Optimisation of contribution of candidate parents to maximise genetic gain and restricting inbreeding using semidefinite programming - (Open Access publication). *Genetics Selection Evolution* 39:3–25.
- R Core Team. 2020. *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.
- Skaarud A, Woolliams JA, Gjoen HM. 2011. Strategies for controlling inbreeding in fish breeding programs; an applied approach using optimum contribution (OC) procedures. *Aquaculture* 311:110–114.
- Sonesson AK, Woolliams JA, Meuwissen THE. 2012. Genomic selection requires genomic control of inbreeding. *Genetics Selection Evolution* 44:127.
- Tang GQ, Li XW, Zhu L, Shuai SR, Bai L. 2008. Maximizing the selection response by optimal quantitative trait loci selection and control of inbreeding in a population with different lifetimes between sires and dams. *Asian-Australas J Anim Sci.* 21:1559–1571.
- Tsai H-Y, Hamilton A, Tinch AE, Guy DR, Bron JE, Taggart JB, Gharbi K, Stear M, Matika O, Pong-Wong R, Bishop SC, Houston RD. 2016. Genomic prediction of host resistance to sea lice in farmed Atlantic salmon populations. *Genetics Selection Evolution* 48:47.
- Woolliams JA, Berg P, Dagnachew BS, Meuwissen THE. 2015. Genetic contributions and their optimization. *J Anim Breed Genet.* 132:89–99.
- Wright S. 1922. Coefficients of inbreeding and relationship. *Am Nat.* 56:330–338.
- Yamashita M, Mullin TJ, Safarina S. 2018. An efficient second-order cone programming approach for optimal selection in tree breeding. *Optimization Letters* 12:1683–1697.