

RESEARCH ARTICLE

Open Access



# Selective advantage of implementing optimal contributions selection and timescales for the convergence of long-term genetic contributions

David M. Howard<sup>1\*</sup> , Ricardo Pong-Wong<sup>1</sup>, Pieter W. Knap<sup>2,3</sup>, Valentin D. Kremer<sup>4</sup> and John A. Woolliams<sup>1</sup>

## Abstract

**Background:** Optimal contributions selection (OCS) provides animal breeders with a framework for maximising genetic gain for a predefined rate of inbreeding. Simulation studies have indicated that the source of the selective advantage of OCS is derived from breeding decisions being more closely aligned with estimates of Mendelian sampling terms ( $\hat{a}$ ) of selection candidates, rather than estimated breeding values (EBV). This study represents the first attempt to assess the source of the selective advantage provided by OCS using a commercial pig population and by testing three hypotheses: (1) OCS places more emphasis on  $\hat{a}$  compared to EBV for determining which animals were selected as parents, (2) OCS places more emphasis on  $\hat{a}$  compared to EBV for determining which of those parents were selected to make a long-term genetic contribution ( $r$ ), and (3) OCS places more emphasis on  $\hat{a}$  compared to EBV for determining the magnitude of  $r$ . The population studied also provided an opportunity to investigate the convergence of  $r$  over time.

**Results:** Selection intensity limited the number of males available for analysis, but females provided some evidence that the selective advantage derived from applying an OCS algorithm resulted from greater weighting being placed on  $\hat{a}$  during the process of decision-making. Male  $r$  were found to converge initially at a faster rate than female  $r$ , with approximately 90% convergence achieved within seven generations across both sexes.

**Conclusions:** This study of commercial data provides some support to results from theoretical and simulation studies that the source of selective advantage from OCS comes from  $\hat{a}$ . The implication that genomic selection (GS) improves estimation of  $\hat{a}$  should allow for even greater genetic gains for a predefined rate of inbreeding, once the synergistic benefits of combining OCS and GS are realised.

## Background

Selection theory [1, 2] states that sustained genetic gain ( $\Delta G$ ) is obtained from creating a covariance between the Mendelian sampling terms ( $a$ ) and long-term genetic contributions ( $r$ ) of selection candidates. This in turn provides the framework for an effective solution for the management of genomic diversity, due to the relationship between  $r$  and rate of inbreeding ( $\Delta F$ ). Various

methods [3–9] have been proposed for managing genetic resources using this framework, which has become known as optimal contributions selection (OCS). OCS has been shown to generate greater  $\Delta G$  for a predefined  $\Delta F$ , and to generate lower  $\Delta F$  for a given  $\Delta G$ , compared to earlier methods that used truncation selection [10, 11].

So far, the source of the selective advantage provided by OCS has only been investigated in theoretical and simulation studies [7, 12, 13], with the conclusion that it arises from placing greater selection pressure on estimates of Mendelian sampling terms of selection candidates ( $\hat{a}$ ) than in truncation selection. This study is the first to test

\*Correspondence: d.howard@ed.ac.uk

<sup>1</sup> The Roslin Institute and R(D)SVS, The University of Edinburgh, Edinburgh, UK

Full list of author information is available at the end of the article



these predictions in a real population by assessing the weightings placed on estimated breeding value (EBV) versus  $\hat{a}$ , both before and during OCS. Given the findings of Avendaño et al. [12] and Grundy et al. [7], the following three hypotheses were tested: (1) OCS places more emphasis on  $\hat{a}$  compared to EBV for determining the initial selection of parents; (2) OCS places more emphasis on  $\hat{a}$  compared to EBV for determining which parents make a non-zero  $r$ ; and (3) OCS places more emphasis on  $\hat{a}$  compared to EBV for determining the magnitude of  $r$ . Timescales for the convergence of  $r$  of selected parents were also examined.

## Methods

### Population analysed

The commercial population analysed in this study consisted of 115,428 pigs comprising a single closed nucleus line that was bred continuously between 12th April 1997 and 16th August 2012, with pedigree information available for all individuals. The line was developed and also expanded over this period, leading up to the cohort studied here. EBV that were calculated using a weighted index of traits were used at the time of selection, but the EBV used for assessment in this study were only available at a single time point (16th August 2012). Individuals without a calculated EBV on themselves or on their parents were excluded to ensure that the estimated Mendelian sampling term,  $\hat{a}$ , could be calculated for each individual. This left a total of 107,895 individuals (54,881 males and 53,014 females) for analysis, with the earliest individual being born on 23rd February 1999, which is denoted as time  $t = 0$ .

Woolliams et al. [1] defined the long-term genetic contribution,  $r_{i(j)}$ , of an ancestor  $i$  to an individual  $j$  as the proportion of the genes of  $j$  that are expected to be derived by descent from ancestor  $i$ . Using this definition of  $r_{i(j)}$ , each individual  $i$  had its long-term contribution ( $r_i$ ) calculated as follows:

$$r_i = 1/n \sum_j r_{i(j)}, \quad (1)$$

where the sum  $j$  was over all individuals born in the 2012 cohort and  $n$  was the total number of animals in the cohort. Generation intervals were based on the definition of Bijma and Woolliams [2], such that the total number of generations at time  $t$  was defined as  $g(t) = \sum_{i \text{ born} \leq t} r_i$ , where the sum is over all individuals  $i$  born from time 0 up to and including time  $t$ . Individuals of generation  $g$  were defined as all individuals born at time  $t$  for which  $g \leq g(t) < g + 1$ . Using this definition, 10.4 generations were born between 23rd February 1999 and 16th August 2012, representing a generation interval of  $L = 1.29$  years,

with the base generation ( $g = 0$ ) consisting of individuals born between 23rd February 1999 and 21st May 2000.

### Time scales for convergence of contributions

The value of  $r_i$  for individual  $i$  in a closed population will converge towards an asymptote over time [1]. To ensure the validity of the results from this study,  $r_i$  was required to approach convergence. To test this convergence, contributions were examined as a function of time after birth for all individuals selected as parents and born during 1999. The final contributions for this 1999 cohort were those obtained from Eq. (1), i.e. their average contribution to the 2012 cohort, i.e. 13 years later. Contributions of the 1999 cohort to the cohorts born in each of the intermediate years, 2000–2011, were also calculated, denoted as  $r_i(u)$ , where  $u$  denotes time after birth, with  $u = 1, 2, \dots, 12$  years. For each intermediate year, the following linear regression model was used to determine the degree of convergence in that year:

$$r_i - \bar{r} = \beta[r_i(u) - \bar{r}] + \varepsilon_i, \quad (2)$$

where  $\varepsilon_i$  is an error term and both the response and explanatory variables are adjusted by  $\bar{r}$ , which is the mean of the final contributions for all the individuals selected as parents and born during 1999. At convergence,  $\beta \rightarrow 1$  and  $E[\varepsilon_i^2] \rightarrow 0$ . This model was applied separately to contributions of males and females.

A similar methodology to that applied annually was used to assess the convergence of  $r_i$  on a generational scale using all individuals selected as parents in the base generation  $g$ , i.e.  $g = 0$ . Contributions of individuals in the base generation to each intermediate generation were calculated for generations 1 through 8 as  $r_i(g)$ , where  $1 \leq g \leq 8$ , and final contributions were calculated as those made to  $g = 9$ , i.e. the final complete generation. This final generation consisted of individuals born between 6th December 2010 and 5th May 2012.

### Grouping of individuals according to the selection method

An OCS algorithm was introduced for selection in this population in 2002 and was based on the approach described by Newman et al. [14] and Kinghorn [15]. An iterative algorithm was applied to calculate a response surface, with the aim of achieving maximal  $\Delta G$  for a given  $\Delta F$ . Prior to the introduction of this algorithm, selection had been by truncation with ad hoc restrictions aimed at controlling inbreeding. The full transition to this method of selection was completed by the end of 2005 and was used throughout the remainder of the period studied.

Individuals were assigned to either the Pre-OCS group, the OCS group, or neither of these groups using  $g(t)$  defined above. Comparisons between and within

the Pre-OCS and OCS groups were conducted to test the three hypotheses described below. A total of 17,165 individuals born between 23rd February 1999 and 7th December 2001 were assigned to the Pre-OCS group, which represented two generations with  $g(t) < 2$ . Since the OCS algorithm was introduced in 2002, only two generations were available prior to the change to OCS. A further 25,068 individuals, born between 23rd February 2006 and 7th October 2008, were assigned to the OCS group, also consisting of two generations,  $5.6 \leq g(t) < 7.6$ . Analysing two generations in the OCS period maintained consistency with the Pre-OCS period in the analyses. In addition to  $r_i$ , the selection score ( $x_i$ ) was determined for all individuals in the Pre-OCS and OCS groups from the recorded birth of offspring ( $x_i = 1$ , if the individual had offspring, and  $x_i = 0$  otherwise); it is clear that  $x_i = 0$  implies  $r_i = 0$ . Table 1 shows the numbers of individuals in the Pre-OCS and OCS groups by sex and contribution history.

Estimates of Mendelian sampling terms for each individual  $i$ ,  $\hat{a}_i$ , were calculated as the deviation of the individual's EBV ( $EBV_i$ ) from the average EBV of its parents. In the analyses that follow, the  $EBV_i$  were calculated as deviations from the mean EBV of contemporaries, defined by all animals born within 30 days before or after birth of that individual. This on average resulted in 1497 contemporaries. Unless stated otherwise, the remaining references to EBV refer to this adjusted EBV. No adjustment for  $\hat{a}_i$  was required since it has an expectation of 0 for all animals.

The average reliability of the EBV across both sexes ( $R^2$ ) was approximately 0.16 in the pre-OCS period, increasing to 0.25 thereafter. Selection took place at off-test (~6 months of age), so both sexes had about the same volume of information available for estimating their initial breeding values. In 2012, the average  $R^2$  across both sexes was 0.28.

Inbreeding coefficients ( $F$ ) for individuals were obtained by subtracting 1 from the diagonal elements of the numerator relationship matrix [16]. Annual rates of inbreeding were calculated for the birth cohorts at year  $t$  using two methods: (i)  $\Delta F_{ped}$ , calculated as:

$$\Delta F_{ped} = (F_{t+1} - F_t)/(1 - F_t), \quad (3)$$

where  $F_t$  is the average  $F$  for all individuals in the cohort; and (ii)  $\Delta F_r$ , calculated as:

$$\Delta F_r = 1/4 \sum r_i^2, \quad (4)$$

where the sum is over all individuals in the cohort and  $r_i$  is defined by Eq. (1), regardless of convergence.

**Table 1 Number of males and females depending on selection group (Pre-OCS and OCS), selection score ( $x_i$ ), and whether or not the long-term contribution in 2012 ( $r_i$ ) was positive**

Constraint	$x_i$	$r_i$	Pre-OCS		OCS	
			Males	Females	Males	Females
-	-	-	8341	8824	13,001	12,067
>0	-	-	177 (0.021)	1444 (0.164)	111 (0.009)	1839 (0.152)
>0	>0	>0	118 (0.667)	219 (0.152)	35 (0.315)	179 (0.097)

The symbol '-' indicates that no constraint was applied. The proportion of individuals remaining from the constraint in the previous row is in brackets

### Sources of selective advantages with OCS

Bivariate regressions were used to determine whether  $EBV_i$  or  $\hat{a}_i$  provided the source of the selective advantages before and during OCS. The source of the selective advantage was examined for: (1) which individuals were initially selected as parents, (2) which parents went on to make a non-zero  $r_i$  to the population, and for (3) determined the magnitude of  $r_i$ .

#### Step 1: Initial selection as parents

Bivariate regression was used to determine whether  $EBV_i$  or  $\hat{a}_i$  was the principal variable that determined the initial selection of an individual as a parent, separately for the Pre-OCS and OCS ancestors. A generalised linear model with a binomial distribution and a logit link function was fitted together with a parameter for over-dispersion, assuming a beta-binomial distribution for  $p_i$  conditional on the fixed effects. Let the probability of selection for individual  $i$  be  $\mu_i = E[x_i]$ , then if  $s(\cdot)$  is the logistic link function, the following model was fitted:

$$s^{-1}(\mu_i) = \alpha + \beta_{EBV}EBV_i + \beta_{\hat{a}}\hat{a}_i + \varepsilon_i, \quad (5)$$

where  $\alpha$  is an intercept and  $\beta_{EBV}$  and  $\beta_{\hat{a}}$  are the respective regression coefficients on  $EBV_i$  and  $\hat{a}_i$ . Changes in deviance conditional on all other model terms were used to assess the relative importance of individual model terms. This analysis was conducted separately for each sex using the 'quasi-binomial' option in the 'glm' package of R, which estimates over-dispersion from the Pearson residuals.

#### Step 2: Maintenance of contributions over time for selected individuals

To examine the maintenance of contributions over time, models were fitted only to the subset of individuals that had been selected as parents, i.e. conditional on  $x_i = 1$ , because long-term contributions are always 0 for individuals that are not selected. For this subset,  $r_i^+ = 1$  if  $r_i > 0$ , and 0 otherwise. The methods described for Step 1 were

repeated with  $r_i^+$  instead of  $x_i$ , such that  $\mu_i = E[r_i^+]$  in Eq. (5).

### Step 3: Magnitude of long-term contributions over time for selected individuals

An intrinsic property of OCS is the targeting of a desired magnitude of  $r_i$  for each individual. The source of the selective advantage for determining the magnitude of  $r_i$  was assessed using: (1) all individuals that were initially selected as parents, i.e.  $x_i = 1$ , regardless of their long-term contributions, and (2) restricted to parents with a non-zero long-term contribution ( $r_i^+ = 1$ ). The strength of the relationships of the magnitude of  $r_i$  with both  $EBV_i$  and  $\hat{a}_i$  were investigated separately for each sex in the Pre-OCS group and in the OCS group using the following linear regression model:

$$r_i = \alpha + \beta_{EBV}EBV_i + \beta_{\hat{a}}\hat{a}_i + \varepsilon_i. \quad (6)$$

Finally for the selected subgroup with  $r_i^+ = 1$ , contour plots of  $r_i$  against  $EBV$  and  $\hat{a}$  were used to visualise and aid in interpreting the relationship of  $EBV_i$  and  $\hat{a}_i$  with the magnitude of  $r_i$ . Contour plots were produced using the 'akima' package [17] of R v2.15.

## Results

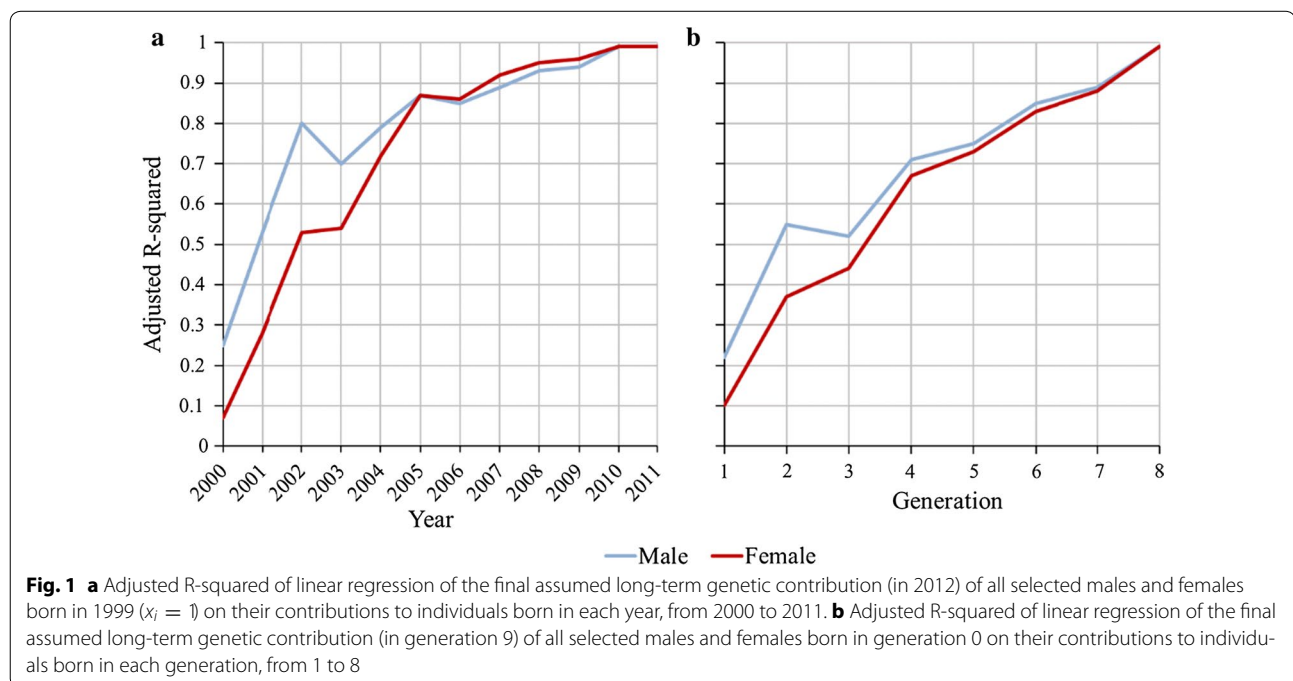
### Accumulation and convergence of contributions

A plot of the accumulation of  $r_i$  over time is in Additional file 1: Figure S1, starting with the first individual

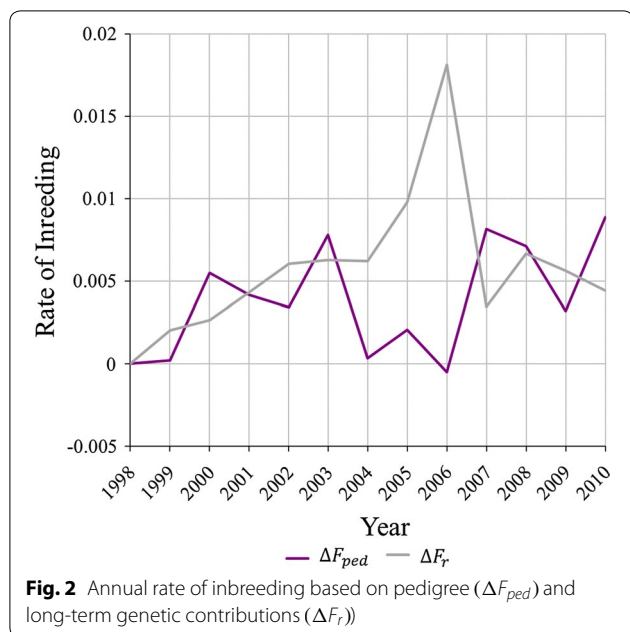
born in the Pre-OCS group until the last selected individual, which was born on 23rd September 2011. The rate of accumulation was relatively constant over time, with a levelling off of the cumulative  $r_i$  midway through 2011, which reflects individuals that were currently undergoing selection. The regression coefficient of accumulated  $r_i$  on time was  $2.14 \times 10^{-3}$  per day, equating to an average generation interval of 468 days.

Figure 1a illustrates the male and female adjusted R-squared for the model of Eq. (2) for each year. The adjusted R-squared increased rapidly at first for both males and females, with males exceeding 75% after 3 years. For males,  $r$  continued to converge but at a slower rate, reaching an adjusted R-squared of approximately 95% by 9 years. Females were slower to approach convergence, with the adjusted R-squared only exceeding 75% after 5 years, but still exceeding 95% within 9 years. For selected individuals born in 1999, the standard deviation of  $r_i$  within the 2012 cohort was  $1.05 \times 10^{-3}$ .

The adjusted R-squared for the model of Eq. (2) on a generational basis is in Fig. 1b. This trend line had greater linearity than the yearly analysis. For this analysis, males again initially converged at a faster rate than females, with both sexes achieving an average adjusted R-squared of approximately 75% in five generations. For individuals in the base generation, the standard deviation of  $r_i$  in generation 9 was  $1.09 \times 10^{-3}$ .



**Fig. 1** **a** Adjusted R-squared of linear regression of the final assumed long-term genetic contribution (in 2012) of all selected males and females born in 1999 ( $x_i = 1$ ) on their contributions to individuals born in each year, from 2000 to 2011. **b** Adjusted R-squared of linear regression of the final assumed long-term genetic contribution (in generation 9) of all selected males and females born in generation 0 on their contributions to individuals born in each generation, from 1 to 8



**Rates of inbreeding over time**

Plots of  $\Delta F_{ped}$  and  $\Delta F_r$  over time are shown in Fig. 2. The elevated value for  $\Delta F_r$  observed in 2006 was strongly influenced by a prolific male that sired 1419 offspring, which alone contributed 0.012 to  $\Delta F_r$ , representing 63.5% of the total contributions to  $\Delta F_r$  for the year. The impact of this prolific individual would not be expected on a plot of  $\Delta F_{ped}$ , since its impact is spread across future generations beyond the endpoint in Fig. 2.  $\Delta F_r$  is prospective in the sense that it measures the impact further down the pedigree arising from an individual’s gene flow [18, 19], whereas  $\Delta F_{ped}$  is retrospective in the sense that it measures the accumulated impact of changes in contributions across all previous generations.

**Table 2** Estimates of regression coefficients ( $\beta$ ) from the bivariate logistic regression of selection score on estimated breeding values (EBV) and estimated Mendelian sampling terms ( $\hat{a}$ )

	Pre-OCS		OCS	
	$\beta$ (s.e.)	F	$\beta$ (s.e.)	F
<i>Males</i>				
EBV	-0.12 (0.11)	1.09	0.34 (0.10)	11.88
$\hat{a}$	-1.24 (0.33)	13.61	0.68 (0.22)	9.43
<i>Females</i>				
EBV	0.01 (0.02)	0.16	0.16 (0.02)	82.24
$\hat{a}$	0.31 (0.08)	15.11	0.44 (0.04)	105.26

Standard errors (s.e.) are between parentheses. Approximate F-values are shown with numerator d.f. = 1 and denominator d.f. = 8338, 8821, 12,998, and 12,064 for Pre-OCS males and females, and OCS males and females, respectively

**Sources of selective advantages with optimal contributions selection**

**Step 1: Initial selection as parents**

Estimates of regression coefficients from the logistic regression of the selection score  $x_i$  on  $EBV_i$  and  $\hat{a}_i$  are in Table 2. There was no evidence of extra-binomial variation among Pre-OCS females ( $P \geq 0.05$ ) but there was significant evidence in the remaining subgroups ( $P < 0.001$ ): the estimated parameters for over dispersion were 1.02, 1.08, 3.7 and 2.3 for Pre-OCS and OCS females, and Pre-OCS and OCS males, respectively. In the Pre-OCS group,  $\hat{a}_i$  was negatively associated with selection in males ( $P < 0.001$ ), but was positively associated with selection in females ( $P < 0.001$ ). In both male and female OCS groups, both  $EBV_i$  and  $\hat{a}_i$  were significant factors ( $P < 0.001$ ) in promoting selection as a parent, with more positive values of either favouring selection, but with  $\hat{a}_i$  having greater regression coefficient estimates. The relative importance of these two terms differed slightly between male and female candidates. Estimates from univariate logistic regression are in Additional file 2: Table S1.

**Step 2: Maintenance of contributions over time for selected individuals ( $x_i = 1$ )**

Estimates of regression coefficients from the logistic regression of  $r_i^+$  on  $EBV_i$  and on  $\hat{a}_i$  in the OCS period are in Table 3. In the Pre-OCS group, there was no significant effect of  $EBV_i$  for either sex but a positive association with  $\hat{a}_i$  was observed for the females ( $P < 0.05$ ). In the OCS males,  $EBV_i$  was positively associated with maintaining  $r_i^+ = 1$  ( $P < 0.001$ ), whereas  $\hat{a}_i$  was negatively associated with maintaining a positive contribution across generations  $r_i^+ = 1$  ( $P < 0.05$ ), conditional on initial selection as a parent. Among OCS females, both  $EBV_i$  ( $P < 0.001$ ) and  $\hat{a}_i$  ( $P < 0.05$ ) were positively associated

**Table 3** Estimates of regression coefficients ( $\beta$ ) from the bivariate logistic regression of maintenance of non-zero contributions on estimated breeding values (EBV) and estimated Mendelian sampling terms ( $\hat{a}$ )

	Pre-OCS		OCS	
	$\beta$ (s.e.)	F	$\beta$ (s.e.)	F
<i>Males</i>				
EBV	0.03 (0.12)	0.07	0.96 (0.23)	17.81
$\hat{a}$	0.10 (0.22)	0.22	-0.65 (0.26)	6.45
<i>Females</i>				
EBV	-0.01 (0.06)	0.04	0.25 (0.06)	21.31
$\hat{a}$	0.24 (0.11)	4.37	0.22 (0.10)	5.07

Standard errors (s.e.) are between parentheses. Approximate F-values are also shown: with numerator d.f. = 1 and denominator d.f. = 174, 1441, 108 and 1836 for Pre-OCS males and females, and OCS males and females, respectively

**Table 4 Estimates of regression coefficients ( $\beta$ ) from the bivariate regression of the long-term contributions ( $r$ ) on estimated breeding values (EBV) and estimated Mendelian sampling terms ( $\hat{a}$ ) for all selected individuals**

	Pre-OCS		OCS	
	$\beta$ (s.e.)	F	$\beta$ (s.e.)	F
<i>Males</i>				
EBV	10.98 (5.24)	4.38	45.34 (14.37)	9.95
$\hat{a}$	-6.97 (9.13)	0.58	9.78 (21.45)	0.21
<i>Females</i>				
EBV	0.65 (0.60)	1.18	0.87 (0.51)	2.91
$\hat{a}$	0.77 (1.25)	0.39	3.21 (0.93)	11.92

Standard errors (s.e.) are between parentheses. The importance of the terms is assessed using F-values; numerator d.f. = 1, and denominator d.f. are 174, 1441, 108 and 1836 for Pre-OCS males and females, and OCS males and females, respectively

**Table 5 Estimates of regression coefficients ( $\beta$ ) from the bivariate regression of the long-term contributions ( $r$ ) on estimated breeding values (EBV) and estimated Mendelian sampling terms ( $\hat{a}$ ) for all individuals with  $r_i^+ = 1$** 

	Pre-OCS		OCS	
	$\beta$ (s.e.)	F	$\beta$ (s.e.)	F
<i>Males</i>				
EBV	15.87 (7.19)	4.87	38.19 (53.82)	0.50
$\hat{a}$	-13.92 (12.36)	1.27	53.63 (62.43)	0.74
<i>Females</i>				
EBV	4.36 (3.07)	2.02	-2.97 (5.00)	0.35
$\hat{a}$	-4.08 (5.31)	0.59	15.97 (7.46)	4.58

Standard errors (s.e.) are between parentheses. The importance of the terms is assessed using F-values; the numerator d.f. = 1, and denominator d.f. are 115, 216, 32 and 176 for Pre-OCS males and females, and OCS males and females, respectively

with maintaining a positive contribution. Estimates of univariate regression coefficients are in Additional file 2: Table S2 and show that, in these models,  $EBV_i$  and  $\hat{a}_i$  were always positively correlated with maintaining non-zero contributions. There was no evidence of extra-binomial variation ( $P \geq 0.05$ ) for any of the four sub-groups.

### Step 3: Magnitude of long-term contributions over time for selected individuals ( $x_i = 1$ )

Estimates of regression coefficients from the bivariate regression of  $r_i$  on  $EBV_i$  and  $\hat{a}_i$  conditional only on  $x_i = 1$  are in Table 4 and those conditional on  $r_i^+ = 1$  are in Table 5. The additional restriction imposed by  $r_i^+ = 1$  substantially increased the standard errors of the estimates and, as a result, estimates in Table 5 do not differ qualitatively from those in Table 4 if these uncertainties are taken into consideration. In males (conditional on

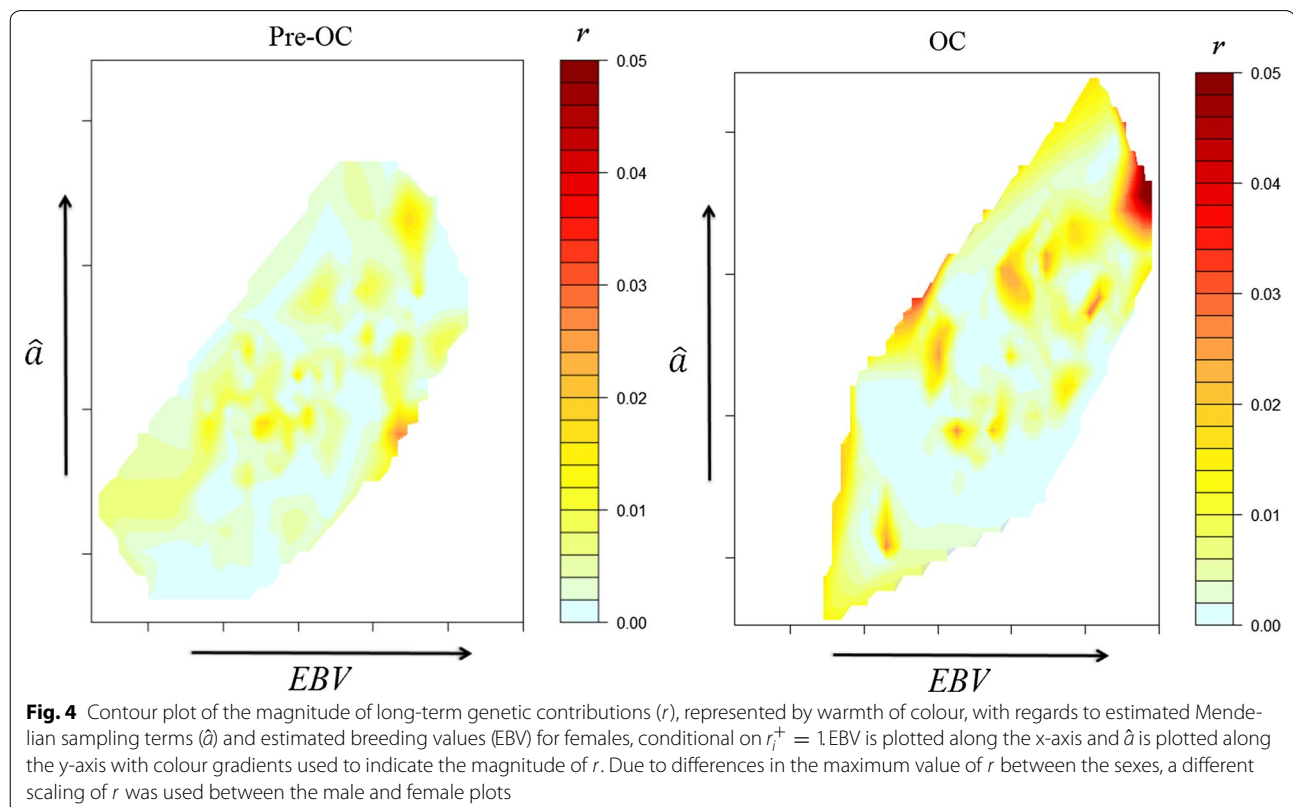
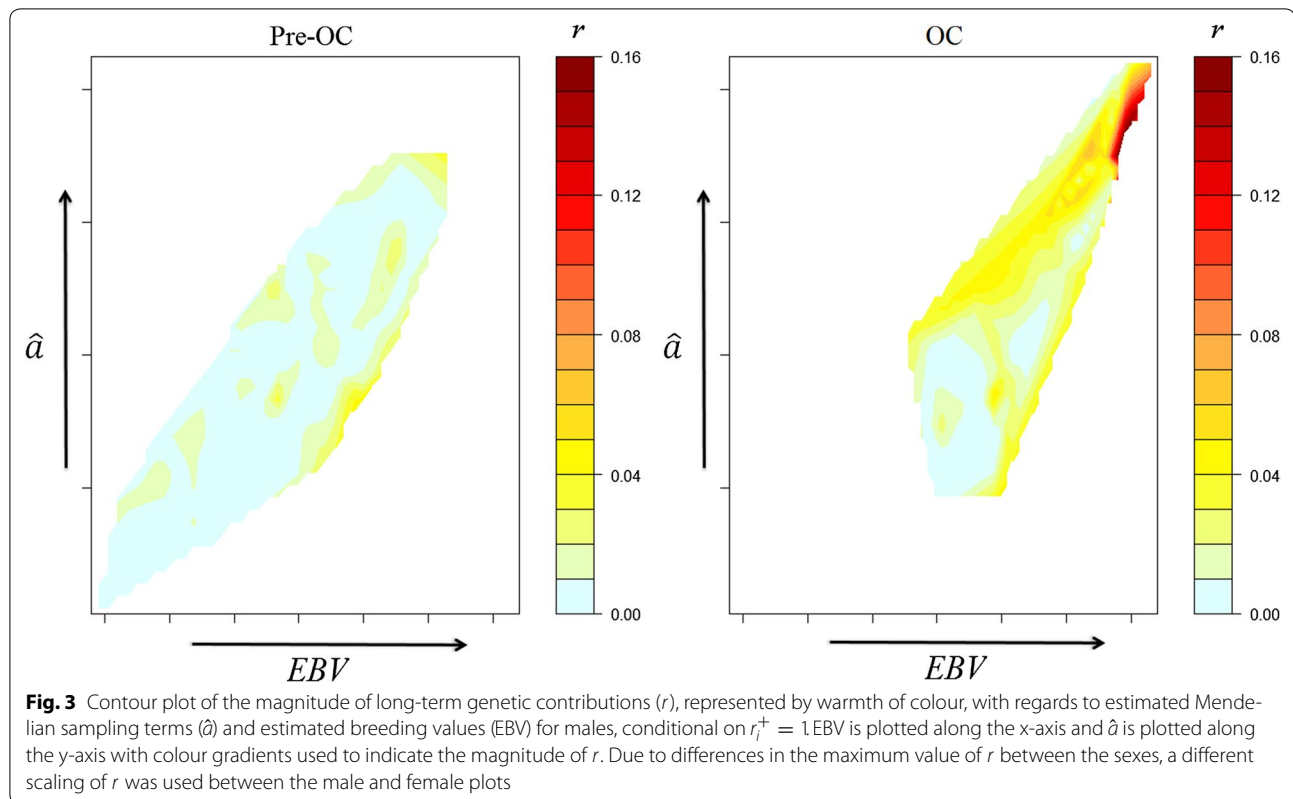
$x_i = 1$ ),  $EBV_i$  had a significant effect on  $r_i$  during both Pre-OCS ( $P < 0.05$ ) and OCS ( $P < 0.01$ ). In females (conditional on  $x_i = 1$ ), neither  $\hat{a}_i$  nor  $EBV_i$  had a significant effect on  $r_i$  Pre-OCS, but in the female OCS group,  $\hat{a}_i$  was a significant determinant of  $r_i$  ( $P < 0.001$ ). Estimates of univariate regression coefficients, conditional only on  $x_i = 1$  and conditional on  $r_i^+ = 1$  are in Additional file 2: Tables S3 and S4, respectively.

A contour map illustrating the relationship of  $r_i$  with  $EBV_i$  and with  $\hat{a}_i$  is in Fig. 3 for selected males with  $r_i^+ = 1$  and in Fig. 4 for selected females with  $r_i^+ = 1$ . More active control of  $r_i$  was generally observed in the OCS selection period compared to the Pre-OCS period. There was a shift in the coloured area of the plot both horizontally to the right and vertically upwards following the introduction of OCS, which shows that animals with  $r_i^+ = 1$  had higher  $EBV_i$  and  $\hat{a}_i$  in the OCS period than animals with  $r_i^+ = 1$  in the Pre-OCS period. Examining the colour dispersal of the Pre-OCS contour plot revealed that  $EBV_i$ , along the x-axis, was the more important determinant for the magnitude of  $r_i$  in the Pre-OCS period. In the OCS period, there was a change in the distribution of  $r_i$ , with warmer colours found predominantly in the upper-most half of the contour plot. This demonstrates that after the introduction of OCS, the  $\hat{a}_i$ , on the y-axis, was more important for determining the magnitude of  $r_i$ . For males, the largest concentration of warmer colour in the OCS plot was in the upper right hand corner of the plot (Fig. 3), highlighting that both  $EBV_i$  and  $\hat{a}_i$  were important for males with the very highest  $r_i$ .

## Discussion

The research reported here was made possible by the availability of data from a single, large ( $n=107,895$ ) nucleus breeding line from Genus-PIC. This was a novel opportunity to use hypothesis-driven analyses to assess the source of the selective advantage under different selection strategies using longitudinal data. The particular hypotheses investigated were that adoption of OCS, to maximise gain conditional on the target rate of inbreeding, shifts the emphasis of selective advantages from  $EBV$  towards estimates of Mendelian sampling terms,  $\hat{a}_i$ , based on both theory and previously published simulations [10]. The outcome of the analyses provides some qualified support for these hypotheses in females.

To maximise accuracy and minimise bias in an assessment of OCS, a number of conditions are required: (i) adequate periods of convergence for each method of selection; (ii)  $EBV$  that are relevant to decision-making; (iii) stability in breeding objectives, selection pressures and rates of inbreeding; and, (iv) adequate population size over time. Not all these conditions were met in this commercial operation. In relation to (i), analysis of



convergence of contributions for selected individuals (Fig. 1) suggested that eight generations were required to reach convergence, rather than the five generations found in simulations [1], and this was achieved only for the Pre-OCS group. However, the process of convergence within the OCS group was well advanced after five generations (Fig. 2) and this is supported by the evidence from the Pre-OCS group where the correlation between contributions at 5 and 8 generations after birth was 0.86. In relation to (ii), an individual's  $EBV_i$  and  $\hat{a}_i$  are dynamic and increase in accuracy as more information becomes available from descendants. Selection decisions were made using  $EBV_i$  and  $\hat{a}_i$  that were available at the time of selection, but for the current study these values were only available at a single time point (16th August 2012). Therefore, the accuracies of  $EBV_i$  and  $\hat{a}_i$  used in the analyses varied among individuals. Differences in accuracy are greatest between animals selected as parents, i.e. those that have offspring, and those that were not selected. Therefore, analyses conditional on initial selection as a parent (steps 2 and 3) are likely to be less affected by variation in accuracies of  $EBV_i$  and  $\hat{a}_i$  than analyses for step 1. In relation to (iii), it is expected that changes in breeding objectives occur over time within a commercial population, and the approximate correlation between the selection index applied in 2000 and that applied in 2005 (calculated from the (co)variance matrix of the EBV in the indices and the two sets of weights applied) was only moderate at 0.39. It is clear from Table 1 that selection intensity increased after initial establishment and expansion of the line, which affects quantitative comparisons of Pre-OCS and OCS groups. In relation to (iv), empirically, this population of more than 100,000 individuals born over the course of 15 years was sufficient to detect sources of selective advantages. The standard errors of the partial regression coefficients for  $EBV_i$  and  $\hat{a}_i$  for selected males were relatively large compared to those for females, as a result of the higher selection intensity applied to males. This higher selection intensity in males increased the magnitude of the potential effects [1], but also greatly reduced the sample size.

One advantage of the study is that it was conducted on data collected on a real population rather than simulation, but this also means that day-to-day decision-making is inevitably driven by commercial requirements rather than strictly following selection recommendations. In summary, although the Pre-OCS group acted as a practical control for the OCS-group, the commercial setting of the data indicates that comparisons between the Pre-OCS and OCS groups should be considered qualitatively and with care. However, the comparison of outcomes from the OCS group with the theoretical expectations remains valid.

### Accumulation and convergence of contributions

The observed linear increase in the accumulation of  $r_i$  demonstrated a relatively uniform generation interval over time of approximately 468 days. Alternative generation intervals can be calculated for this population by either averaging the age of parents at birth of their first offspring (368 days), or after the birth of the first offspring that is selected (415 days). The latter is closer in concept to the generation interval in quantitative genetics since it includes the concept of replacement, and moves towards that provided by  $r_i$ , which is the length of time to renew the gene pool that is destined to maintain the population in the long term.

There was evidence that the generation interval calculated from parent ages increased over time (results not shown) and this was likely associated with operational decisions concerned with expansion of the line (see Table 1). The stochastic simulation work of Bijma and Woolliams [2] and Woolliams et al. [1] found that the generation interval calculated using  $r_i$  was shorter than that using the average age of parents of a cohort under simple mass selection with a pre-determined age structure. The primary reason for this qualitative difference between these two methods for calculating generation intervals is that there is a selective advantage to offspring born to younger parents in the mass selection scenario, since their parents have higher breeding values than the population average. With mixed-model evaluations, e.g. best linear unbiased prediction (BLUP), the genetic trend across cohorts is accounted for, which results in more effective selection among candidates of different ages.

### Sources of selective advantages of optimal contributions selection

The OCS algorithm guides the optimum  $r_i$  over generations until convergence, through consideration of both an individual's  $EBV_i$  as well as its degree of relatedness with the rest of the population. This approach is expected to reduce the  $r_i$  of individuals from over-represented families, allowing the selection of superior animals from families with lower EBV. As a result, individuals with greater  $\hat{a}_i$  are selected to provide comparatively larger  $r_i$  [7, 20]. This process results in the maximum gain ( $\Delta G$ ) conditional on the target rate of inbreeding ( $\Delta F$ ), since  $\Delta G \propto \Sigma r_i a_i$  (i.e. not  $\Sigma r_i EBV_i$ ) and  $\Delta F \propto \Sigma r_i^2$  (see Woolliams et al. [1]). In the case of perfect accuracy of EBV (e.g. with a heritability  $h^2 = 1$ ) and no interdependence of generations, this would take the form of a linear allocation above a cut-off, whereby selected individuals would be those with  $a_i$  above a threshold and  $r_i$  is expected to be linearly related to  $a_i$ , conditional on selection [12]. Where these conditions do not hold, the relationship between  $r_i$  and  $a_i$  will have additional random noise and the true



breeding value or  $EBV_i$  will also influence  $r_i$ . To examine these different aspects, the selection process was broken down into three consecutive steps, each more closely focused on the ultimate genetic contributors to the population, and each with its associated hypothesis.

The first hypothesis was concerned with whether  $EBV_i$  or  $\hat{a}_i$  was the principal factor that determined the selection of an individual to become a parent. In practice, the selection of an individual to become a parent, unlike the development of  $r_i$ , is only influenced by selection decisions within a narrow time-period. For females,  $\hat{a}_i$  was observed to be the most important determinant for selection of an individual to become a parent, although it was unclear if this had emerged as a result of the introduction of OCS, in spite of the influence being stronger in the OCS group.

The most convincing evidence for the importance of  $\hat{a}_i$  in determining  $r_i$  when using OCS selection came from analyses that addressed the third hypothesis, which related the magnitude of  $r_i$  to  $EBV_i$  and  $\hat{a}_i$ , conditional on  $i$  being selected initially. Of the analyses conducted in this study, these analyses most closely followed a previous simulation study [10] and directly examine the theoretical expectation that OCS will guide contributions in direct relation to  $\hat{a}_i$  among animals that are initially selected. The footprint of OCS that was expected from the simulation study was observed among females of the OCS group, with  $\hat{a}_i$  positively associated with  $r_i$  and dominating the influence of  $EBV_i$ . However, the reverse was observed for males, for which inferences are made more difficult due to the large standard errors of the bivariate coefficients. It is possible that short-term commercial needs that do not follow the outcome of the OCS algorithm are more likely to occur among males, compared to females, as male selection opportunities are greater and offer a more rapid response to the need.

## Conclusions

This work represents the first evaluation in commercial practice of the impact and validation of the theory underlying OCS, which predicts a closer alignment of  $r_i$  with  $\hat{a}_i$  than with  $EBV_i$ . There was some evidence of a re-weighting in the emphasis of selection away from the  $EBV_i$  towards  $\hat{a}_i$  in the final magnitude of  $r_i$ , with females providing stronger evidence for these conclusions. The evidence from males was equivocal, in part because of higher selection intensity among males, leading to smaller numbers of males with which to conduct the regression analyses. As a result, this study provides some support to the assertion that there is an advantage in combining OCS with genomic information, which can achieve more accurate prediction of  $\hat{a}_i$  at an earlier age,

to generate greater genetic progress over time, without negatively impacting the rate of inbreeding.

## Additional files

**Additional file 1: Figure S1.** Accumulation of long-term genetic contributions over time with a line of best fit.

**Additional file 2: Tables S1, S2, S3, S4.** Estimates of regression coefficients ( $\beta$ ) from univariate regressions of selection score ( $x_i$ ) on estimated breeding values (EBV) and estimated Mendelian sampling terms ( $\hat{a}$ ). Standard errors (s.e.) are given in parentheses. **Table S2.** Estimates of regression coefficients ( $\beta$ ) from univariate regressions of  $r_i^+$  (conditional on individual  $i$  was selected) on estimated breeding values (EBV) and estimated Mendelian sampling terms ( $\hat{a}$ ). Standard errors (s.e.) are given in parentheses. **Table S3.** Estimates of regression coefficients ( $\beta$ ) from simple linear regression of long-term genetic contribution (with all selected individuals, i.e. conditional on  $x_i = 1$ ) on estimated breeding values (EBV) and estimated Mendelian sampling terms ( $\hat{a}$ ). Standard errors (s.e.) are given in parentheses. **Table S4.** Estimates of regression coefficients ( $\beta$ ) from simple linear regression of long-term genetic contribution (restricted to those individuals with a non-zero long-term genetic contribution) on estimated breeding values (EBV) and estimated Mendelian sampling terms ( $\hat{a}$ ). Standard errors (s.e.) are given in parentheses.

## Authors' contributions

DMH contributed to experimental design, data analysis, interpretation and manuscript writing, RP-W to experimental design, data analysis, interpretation and manuscript writing, PWK and VDK to sample acquisition, experimental design and interpretation, JAW to experimental design, data analysis, interpretation and manuscript writing. All authors read and approved the final manuscript.

## Author details

<sup>1</sup>The Roslin Institute and R(D)SVS, The University of Edinburgh, Edinburgh, UK. <sup>2</sup>Genus-PIC, Ratsteich 31, 24837 Schleswig, Germany. <sup>3</sup>Genus PLC, 100 Bluegrass Commons Blvd, Suite 2200, Hendersonville, TN 37075, USA. <sup>4</sup>Genus-PIC, Edinburgh, UK.

## Acknowledgements

The authors acknowledge Genus-PIC for providing the data.

## Competing interests

The authors declare that they have no competing interests.

## Data availability

The data that support the findings of this study are available from Genus-PIC but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of Genus-PIC.

## Ethics approval and consent to participate

Not applicable.

## Funding

DMH gratefully acknowledges financial support from Biotechnology and Biological Sciences Research Council (BBSRC) Grant BB/J500847/1. RP-W and JAW gratefully acknowledge financial support from BBSRC Programme Grant BB/J004235/1.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 10 August 2017 Accepted: 23 April 2018

Published online: 10 May 2018

**References**

1. Woolliams JA, Bijma P, Villanueva B. Expected genetic contributions and their impact on gene flow and genetic gain. *Genetics*. 1999;153:1009–20.
2. Bijma P, Woolliams JA. Prediction of genetic contributions and generation intervals in populations with overlapping generations under selection. *Genetics*. 1999;151:1197–210.
3. Villanueva B, Woolliams JA. Optimization of breeding programmes under index selection and constrained inbreeding. *Genet Res*. 1997;69:145–58.
4. Grundy B, Villanueva B, Woolliams JA. Dynamic selection for maximizing response with constrained inbreeding in schemes with overlapping generations. *Anim Sci*. 2000;70:373–82.
5. Pong-Wong R, Woolliams JA. Optimisation of contribution of candidate parents to maximise genetic gain and restricting inbreeding using semidefinite programming. *Genet Sel Evol*. 2007;39:3–25.
6. Meuwissen THE. Maximizing the response of selection with a predefined rate of inbreeding. *J Anim Sci*. 1997;75:934–40.
7. Grundy B, Villanueva B, Woolliams JA. Dynamic selection procedures for constrained inbreeding and their consequences for pedigree development. *Genet Res*. 1998;72:159–68.
8. Dagnachew BS, Meuwissen THE. A fast Newton–Raphson based iterative algorithm for large scale optimal contribution selection. *Genet Sel Evol*. 2016;48:70.
9. Hinrichs D, Wetten M, Meuwissen THE. An algorithm to compute optimal genetic contributions in selection programs with large numbers of candidates. *J Anim Sci*. 2006;84:3212–8.
10. Avendaño S, Villanueva B, Woolliams JA. Expected increases in genetic merit from using optimized contributions in two livestock populations of beef cattle and sheep. *J Anim Sci*. 2003;81:2964–75.
11. Villanueva B, Avenano S, Woolliams JA. Prediction of genetic gain from quadratic optimisation with constrained rates of inbreeding. *Genet Sel Evol*. 2006;38:127–46.
12. Avendaño S, Woolliams JA, Villanueva B. Mendelian sampling terms as a selective advantage in optimum breeding schemes with restrictions on the rate of inbreeding. *Genet Res*. 2004;83:55–64.
13. Sorensen AC, Berg P, Woolliams JA. The advantage of factorial mating under selection is uncovered by deterministically predicted rates of inbreeding. *Genet Sel Evol*. 2005;37:57–81.
14. Newman S, Kremer VD, Kinghorn BP. Mate selection strategies. Balancing genetic gain and diversity in livestock improvement programs. In: *Proceedings of the 58th Annual National Breeders Roundtable: 7–8 May 2009; St Louis; 2009*.
15. Kinghorn BP. An algorithm for efficient constrained mate selection. *Genet Sel Evol*. 2011;43:4.
16. Wright S. Coefficients of inbreeding and relationship. *Am Nat*. 1922;56:330–8.
17. Akima H, Gebhardt A. Interpolation of irregular and regularly spaced data. R package version 0.6-2; 2016.
18. Wray NR, Thompson R. Prediction of rates of inbreeding in selected populations. *Genet Res*. 1990;55:41–54.
19. Woolliams JA, Wray NR, Thompson R. Prediction of long-term contributions and inbreeding in populations undergoing mass selection. *Genet Res (Camb)*. 1993;62:231–42.
20. Woolliams JA, Thompson R. A theory of genetic contributions. In: *Proceedings of the 5th World Congress on Genetics Applied to Livestock Production: 7–12 August 1994; Guelph; 1994*.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

