

## RESEARCH

## Open Access

# Optimising multistage dairy cattle breeding schemes including genomic selection using decorrelated or optimum selection indices

Vinzent Börner and Norbert Reinsch\*

**Abstract**

**Background:** The prediction of the outcomes from multistage breeding schemes is especially important for the introduction of genomic selection in dairy cattle. Decorrelated selection indices can be used for the optimisation of such breeding schemes. However, they decrease the accuracy of estimated breeding values and, therefore, the genetic gain to an unforeseeable extent and have not been applied to breeding schemes with different generation intervals and selection intensities in each selection path.

**Methods:** A grid search was applied in order to identify optimum breeding plans to maximise the genetic gain per year in a multistage, multipath dairy cattle breeding program. In this program, different values of the accuracy of estimated genomic breeding values and of their costs per individual were applied, whereby the total breeding costs were restricted. Both decorrelated indices and optimum selection indices were used together with fast multidimensional integration algorithms to produce results.

**Results:** In comparison to optimum indices, the genetic gain with decorrelated indices was up to 40% less and the proportion of individuals undergoing genomic selection was different. Additionally, the interaction between selection paths was counter-intuitive and difficult to interpret. Independent of using decorrelated or optimum selection indices, genomic selection replaced traditional progeny testing when maximising the genetic gain per year, as long as the accuracy of estimated genomic breeding values was  $\geq 0.45$ . Overall breeding costs were mainly generated in the path "dam-sire". Selecting males was still the main source of genetic gain per year.

**Conclusion:** Decorrelated selection indices should not be used because of misleading results and the availability of accurate and fast algorithms for exact multidimensional integration. Genomic selection is the method of choice when maximising the genetic gain per year but genotyping females may not allow for a reduction in overall breeding costs. Furthermore, the economic justification of genotyping females remains questionable.

**Introduction**

Genomic selection (GS) offers breeders the opportunity to reduce costs, decrease the generation interval [1] and possibly avoid inbreeding [2]. GS is based on the prediction of breeding values from individual genotypes (estimated genomic breeding values, GEBV). These genotypes consist of a large number of DNA markers in the form of single nucleotide polymorphisms (SNP), which are in linkage disequilibrium with quantitative trait loci coding for economically important traits.

In dairy cattle, accuracies ( $r$ ) of GEBV ( $r_{QEBV}$ ) for milk production traits can be as high as 0.75 [3], but are below those from progeny testing. However, the tremendously decreased generation interval ( $L$ ) may lead to a much higher genetic gain per year ( $\Delta G_a$ ) [1,4]. Schaeffer [1] summarised the potential effects of GS on dairy cattle breeding schemes assuming an  $r_{GEBV}$  of 0.75 and a cost for GS per genotyped individual ( $C_{QEBV}$ ) of 500 Canadian Dollars in a one-stage selection approach. As  $r_{GEBV}$  and  $C_{QEBV}$  may change, research work concerning optimum breeding schemes for different combinations of these parameters and possible multistage selection approaches are of interest.

\* Correspondence: [reinsch@fbn-dummerstorf.de](mailto:reinsch@fbn-dummerstorf.de)  
Leibniz Institute for Farm Animal Biology, Wilhelm-Stahl-Allee 2, 18196  
Dummerstorf, Germany

The availability of a variety of SNP chips at different prices and yielding breeding values of different accuracies makes it possible to use specific SNP chips in each selection path or more than one chip in a multistage preselection approach. Additionally, the advantage of GS might also be combined with traditional progeny performance testing (PPT) schemes as currently applied in certain dairy cattle breeding programs [5]. Thus, apart from the already answered questions concerning the applicability of GS, the answer to the question of how a breeding scheme should be structured has become more complex. This complexity results from the possibility of combining different information sources according to their costs and correlations with the aggregate genotype, allowing for a variety of one-, two- or multistage breeding schemes in each selection path to choose from.

It is a question of economic optimisation to select the breeding scheme which maximises a defined utility function. In multistage breeding schemes, the information about the selection candidates collected at all previous stages is combined with the information collected at the current stage. Therefore, the estimated breeding values (EBV) of successive stages are correlated and the EBV of selected individuals after the first selection stage are non-normally distributed. One of the major challenges of research on optimising multistage breeding schemes is the necessity of using computationally sophisticated multiple integration techniques to derive the selection intensities. When Ducrocq and Colleau [6] applied such methods on multistage dairy cattle breeding schemes they were faced with the problem that the convergence of such algorithms can be difficult to achieve if the EBV of successive stages are highly correlated. Furthermore, the computational time was seen as unacceptable if the number of selection stages became too high [7].

A possible solution to circumvent these problems is the decorrelation of the EBV of successive stages, as proposed by Xu and Muir [8,9]. Then, a normal distribution can be assumed for EBV at each stage, making it easy to calculate selection intensities. Xu and Muir [9] estimated the loss in the predicted genetic gain due to the decorrelation in a two-stage breeding scheme to be up to 10% compared to the exact calculation solving the integral. This loss was justified by the possibility of implementing an unlimited number of stages, which was otherwise not applicable.

Decorrelated indices have been used for model calculations of breeding schemes for poultry [9,10], beef cattle [11] and marker-assisted selection [12]. In all these applications accuracies of breeding values, selection intensities and generation intervals across selection paths were assumed identical. Thus, the applicability of decorrelated indices for complex breeding schemes has not been investigated in detail regarding a) different

selection paths, b) their interaction due to the effects of the selection strategy in one path on the accuracy of EBV of the other paths, c) different selection intensities and EBV accuracies in each path, d) the interaction between the generation interval and the number and intensities of successive selection stages and, finally, e) the opportunity to split financial investments between selection paths.

Numerical integration techniques developed more recently by Genz [13], in conjunction with the maximisation algorithm of Brent [14], allow for a fast and stable calculation of the exact selection intensities, even when using many stages. The aim of our work was to compare the results of breeding scheme optimisations when the approach of Xu and Muir [9] (decorrelated indices) or numerical integration (optimum indices) was used to derive selection intensities, breeding value accuracies, and  $\Delta G_a$ . Both methods were applied to a complex dairy cattle breeding scheme as mentioned above. The possibility of using GS as an optional selection stage in a way that it can be used in addition to or instead of PPT was allowed for. Optimisation was over a semi-continuous parameter range of  $r_{GEBV}$  and  $C_{GEBV}$  and financial resources were restricted. Therefore, the results provide insights into the sensitivity of dairy cattle breeding plans to variation in these parameters.

## Methods

### Construction of selection indices and the implementation of GEBV

Deterministic methods were used to optimise breeding plans. Accuracies of EBV ( $r_{EBV}$ ) were calculated based on the selection index theory and coefficients for regressing the aggregated genotype of the selection candidate on phenotypic measurements of informants were derived using two different methods.

The GEBV was included in the selection index as a trait with a heritability of one and a genetic correlation determined by its accuracy [15,16].

For the optimum selection indices (OSI), standard selection index methodology was used at each selection stage  $i = 1, \dots, m_j$  of each selection path  $j = 1, \dots, n$ .

For decorrelated selection indices (DSI) the regression coefficients  $b_{ij}$  were constructed according to Xu and Muir [9]:

$$\begin{aligned} b_{ij} &= P_{ij}^{-1} G_{ij} w \\ 0 &= b'_{ij} P_{i(i-1)j} B_{(i-1)j} (i > 1) \\ 1 &= b'_{ij} P_{ij} b_{ij} \end{aligned} \quad (1)$$

where  $w$  is the vector of economic weights of the traits in the aggregated genotype,  $b_{ij}$  is the vector of regression coefficients on all available information

sources,  $G_{ij}$  is the genetic-phenotypic covariance matrix,  $P_{ij}$  is the phenotypic covariance matrix of all information sources up to stage  $i$ ,  $P_{i(i-1)j}$  is the phenotypic covariance matrix of all information sources up to stage  $i - 1$  and  $B_{(i-1)j}$  is a matrix of regression coefficients for all stages previous to  $i$ . Note that second and third equations are used as constraints when solving the first equation, and are incorporated into it via Lagrange multipliers. The first constraint guarantees that the covariance between EBV of stage  $i$  and EBV of all other stages is equal to zero and the second that a solution exists.

The pedigree and associated phenotypic and genomic information for candidates and relatives available to construct case specific selection indices are given in Figure 1.

### Genetic gain

The genetic gain of each path ( $\Delta G_j$ ) was calculated differently depending on whether DSI or OSI was used. When using OSI, the following formula was applied:

$$\Delta G_j = z_{ij} \sigma_a \quad (2)$$

where  $z_{ij}$  is the expectation of the aggregated genotype (selection intensity) of the selected individuals after selecting at stage  $i$  (the last stage in path  $j$ ) and  $\sigma_a$  is the standard deviation of the aggregated genotype. Selection intensities were derived using the moment generating function of a truncated multivariate normal distribution including all used selection stages and the aggregated genotype [17], where the truncation points were calculated following the approach of Mi and Utz [18],

merging the integration algorithm of Genz [13] and the maximisation techniques of Brent [14].

Using the decorrelated index approach of Xu and Muir [9],  $\Delta G_j$  was calculated avoiding multiple integration according to

$$\Delta G_j = \sum_{i=1}^{m_j} z_{ij} \sigma_{ij} \quad (3)$$

where  $z_{ij}$  is the selection intensity and  $\sigma_{ij}$  is the standard deviation of the decorrelated EBV. Due to the zero covariance between EBV of successive stages,  $z_{ij}$  could be calculated by one-dimensional integration.

$\Delta G_a$  for the whole breeding scheme was calculated according to Rendel and Robertson [19]:

$$\Delta G_a = \frac{\sum_{j=1}^n \Delta G_j}{\sum_{j=1}^n L_j}, \quad (4)$$

where  $L_j$  is the generation interval in path  $j$ .

### Breeding program

A cooperative Holstein dairy cattle breeding program with a cow population of 100 000 was used. Bull dams were selected as heifers from all available heifers and were assumed to be used as bull dams only once. Male selection candidates were produced by contract mating to bull dams. For the purpose of comparing methods, only one milk trait with a heritability of 0.25 was set in the breeding goal. Selection took place on EBV combining performance data and GEBV.

The biological, technical and economical parameters of the program are given in Table 1. All parameters of the breeding program that are expressible in terms of

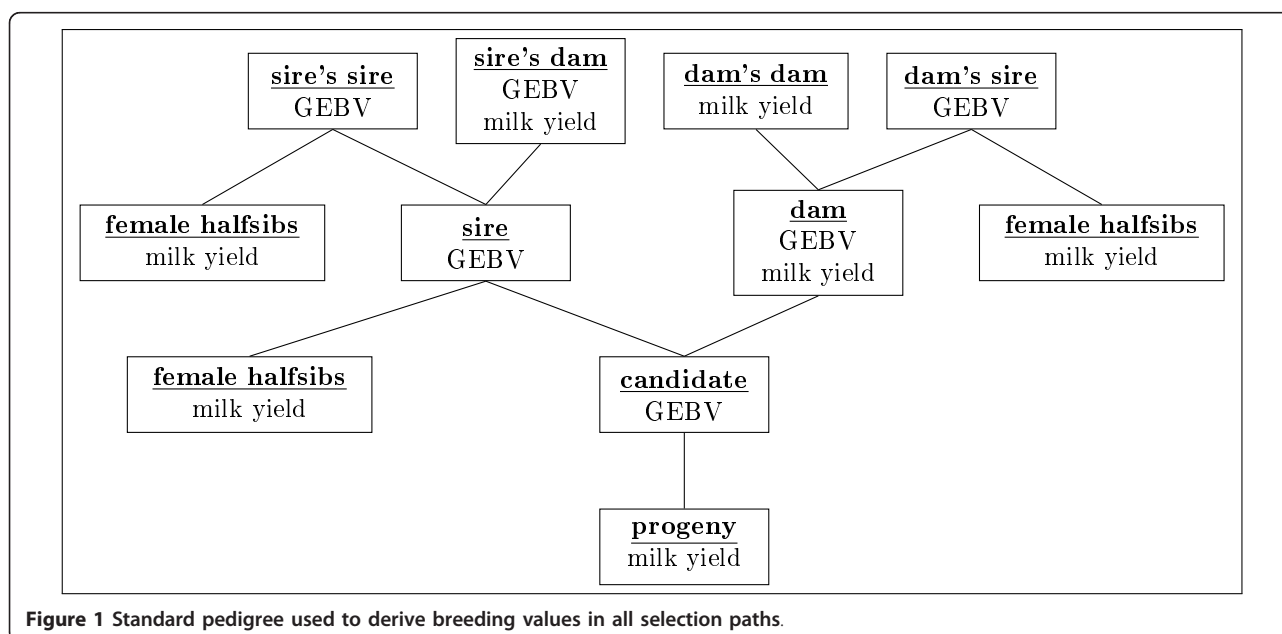


Figure 1 Standard pedigree used to derive breeding values in all selection paths.

**Table 1 Biological, economical and technical parameters of the breeding program**

parameter	unit	value
$h^2$ milk trait		0.25
phenotypic standard deviation	kg	700
accuracy GEBV <sup>1</sup>		0.3 - 0.9
age at first calving	month	26
time between calving	month	12
length of lactation	month	10
maturity of test bulls	month	14
number of daughters per sire	head	99 <sup>3</sup>
insemination with test bulls	%	20
average age of bull calves at purchase	month	6
price of bull calves	€	4 000
keeping costs of bull calves until maturity	€/day	5
keeping costs of test bulls	€/day	4.5
cost for genomic selection <sup>2</sup>	€	25 - 400
population of cows	head	100 000
demand for cow sires	head/year	10
demand for bull sires	head/year	5
initial male selection candidates	head/year	500
demand for bull dams (contract matings)	head/year	1000
compensation payments for test bull matings	€/test bull	3 000
compensation payments to breeders for keeping genotyped selection candidates	€/candidate	150
maximum breeding program costs	€/year	719 050
$L_{SS}^4$ (PED <sup>5</sup> , GS <sup>6</sup> , PPT <sup>7</sup> )	months	23, 23, 71
$L_{SD}^8$ (PED, GS, PPT)	months	23, 23, 61
$L_{DS}^9$ (PED, GS)	months	26, 26

1: estimated genomic breeding values, 2: per genotyped individual, includes DNA isolation, genotyping and calculation of breeding values, 3: minimum number, recalculated according to the number of bulls entering the progeny testing stage, 4: generation interval in the path "sire-sire" as a function of the final selection stage, 5: selection stage selecting on ancestor and sib performance and GEBV, 6: selection stage selection on candidate's GEBV, 7: selection stage selecting on candidate's progeny performance, 8: generation interval in the path "sire-dam" as a function of the final selection stage, 9: generation interval in the path "dam-sire" as a function of the final selection stage

probabilities and not given in Table 1, such as success of first insemination, survival rate of calves, etc., were assumed to be zero or one.

The selection paths were "sire to sire"(SS), "sire to dam"(SD), "dam to sire"(DS) and "dam to dam" (DD). Since in practise, almost no selection takes place within the path DD, a selection intensity of zero was assumed for this path. Each selection path was structured in stages. The selection stages available in paths SS and SD were a) selection on performance data and GEBV of ancestors and half-sibs (PED stage), b) selection on the candidate's GEBV (GS stage) and c) selection on

progeny performance data (PPT stage). The selection stages available in path DS were a) selection on performance data and GEBV of ancestors and half-sibs (PED stage), and b) selection on the candidate's GEBV (GS stage).

To genotype and select male calves, a tissue sample was taken by a veterinarian on farm at birth and had to be paid for by the farmer. DNA-isolation and SNP-genotyping was carried out by a central laboratory, and was added to the expenses of the breeding organisation. Selection candidates were kept on farm until the age of six months. Farmers keeping male candidates selected at the PED stage and slaughtered after GS received a compensation payment from the breeding organisation. When genotyping female calves as potential bull dams, costs were similar, but since the bull dams were not bought by the breeding organisation, no compensation was paid.

#### Breeding cost

Costs not related to selection strategies (e.g. performance recording of females, calculating EBV, marketing, semen processing) were not considered. To allow for changes in costs for labour and infrastructure due to different breeding schemes, all expenses were derived from some invariant basic cost given in Table 1 via the cost function of NamKoong [20]:

$$C_{bj} = C_n n_j \prod_{k=1}^{i-1} p_{jk}, \quad (5)$$

where  $C_{bj}$  is the total cost of expense factor  $b$  in path  $j$ ,  $C_b$  is the cost of expense factor  $b$  per individual,  $n_j$  is the number of initial selection candidates,  $p_{jk}$  is the proportion of the available individuals that is selected at stage  $k$  and  $i$  is the stage within path  $j$  at which the selection stops concerning  $C_b$ .  $C_b$  can be costs for genotyping, purchasing, keeping the animals until maturity, keeping them from maturity until their breeding value is estimated from PPT, compensations for test bull insemination, or compensations for keeping the animals during genotyping stages.

Note that in the case of compensation for keeping the animals during genotyping, the cost formula changes to

$$C_{bj} = C_b n_j \prod_{k=1}^{i-1} p_{jk} (1 - p_{ji}). \quad (6)$$

The breeding costs of path  $j$  were the sum over all expense factors in the path, and the total cost for a certain breeding program was the sum over all paths.

The maximum breeding cost of 719 050 € was imposed as a constraint during maximisation and was derived assuming progeny performance testing of 50 young bulls per year:

$$(4000\text{€} + 213.5d \times 5\text{€/d} + 1403d \times 4.5\text{€/d} + 3000\text{€}) \times 50. \quad (7)$$

This cost included the purchase of 50 male calves from contract matings, their maintenance until maturity and from maturity until breeding value estimation from PPT, and compensation payments for test bull insemination.

#### Parameter variation

Various values for  $C_{QEBV}$  and  $r_{QEBV}$  were used during the calculation process,  $r_{GEBV}$  varied between 0.3 and 0.9 in steps of 0.025 and  $C_{QEBV}$  between 20 € and 400 € in steps of 10 €, resulting in 975 combinations of  $r_{GEBV}$  and  $C_{QEBV}$ .

#### Maximisation

Breeding schemes maximising  $\Delta G_a$  for each combination of  $C_{QEBV}$  and  $r_{GEBV}$  were obtained using a grid search in which the proportion of selected individuals varied at each selection stage in every path. Within the maximisation process, trait measurements available for ancestors, and, therefore the accuracy of the candidates' EBV were adjusted according to the selection strategy in the path from which the ancestor had been derived. Furthermore, to select bull sires from cow sires, the possibility of additional selection stages was allowed for, which requires additional information (e.g. PPT stage) instead of just increased selection intensity for the same group of males.

For each path, the initial number of selection candidates and the final number of selected individuals for reproduction were fixed. The product of the selected proportions at each stage  $i$  within path  $j$  ( $p_{ij}$ ) had to fulfil the equation:

$$\frac{s_j}{n_j} = \prod_{i=1}^{m_j} p_{ij}, \quad (8)$$

where  $s_j$  is the number of selected individuals actually used for reproduction and  $m_j$  is the number of selection stages in path  $j$ .

The proportion  $p_{ij}$  varied between 0.01 and 1 in steps of 0.025. Stages with  $p_{ij} = 1$  were treated as not used. For the stages used ( $p_{ij} < 1$ ), the constraint of equation 8 was fulfilled by calculating  $p_{ij}$  of the last used stage as a dependent variable. Only if this value was  $\leq 1$ , was the stage combination considered as valid. The valid stage combinations of all paths were completely cross-classified to obtain all possible breeding schemes. Breeding costs were derived for each of these schemes but  $\Delta G_a$  was only calculated if the cost constraint was fulfilled. The breeding scheme with the highest  $\Delta G_a$  was seen as the optimum for the given combination of  $C_{QEBV}$  and  $r_{QEBV}$ .

For each combination of  $r_{QEBV}$  and  $C_{QEBV}$ , a grid of about 60 000 breeding schemes was searched for optimisation. The total amount of evaluated breeding plans was 58 million for each calculation method, OSI and DSL

All calculations were carried out with a FORTRAN 90 program written by the first author. The calculation of the selection intensity of an optimum index in a multi-stage breeding program used the FORTRAN routines of Genz [21] and Brent [22].

## Results

### Comparison of methods to calculate genetic gain

Parameters and results of the breeding schemes that maximise  $\Delta G_a$  and fulfill the cost constraint were compared between OSI and DSL Table 2 summarises the frequency of certain selection strategies of optimum breeding schemes in paths SS, SD and DS as a function of  $r_{GEBV}$  and the application of either OSI or DSL For selection in path SD with  $r_{QEBV}$  ranging from 0.3 to 0.4, both methods found that combined selection at PED and GS stage or a three-stage selection approach maximised  $\Delta G_a$ . The same holds for path SS for an

**Table 2 Number of different selection strategies in each selection path**

path	$r_{GEBV}^1$	$n^2$	$n_{Pd}^3$		$n_{Pd+PT}^4$		$n_{Pd+GS}^5$		$n_{Pd+GS+PT}^6$		$n_{GS}^7$	
			DSI <sup>8</sup>	OSI <sup>9</sup>	DSI	OSI	DSI	OSI	DSI	OSI	DSI	DSI
SD <sup>10</sup>	0.3-0.4	195	0	0	120	0	23	22	52	173	0	0
SS <sup>11</sup>	0.3-0.45	273	0	0	120	0	30	25	123	248	0	0
SD	0.45-0.9	741	0	0	0	0	61	741	0	0	680	0
DS <sup>12</sup>	0.425-0.5	156	0	0	0	0	156	156	0	0	0	0
DS	0.3-0.9	975	221	121	0	0	754	854	0	0	0	0

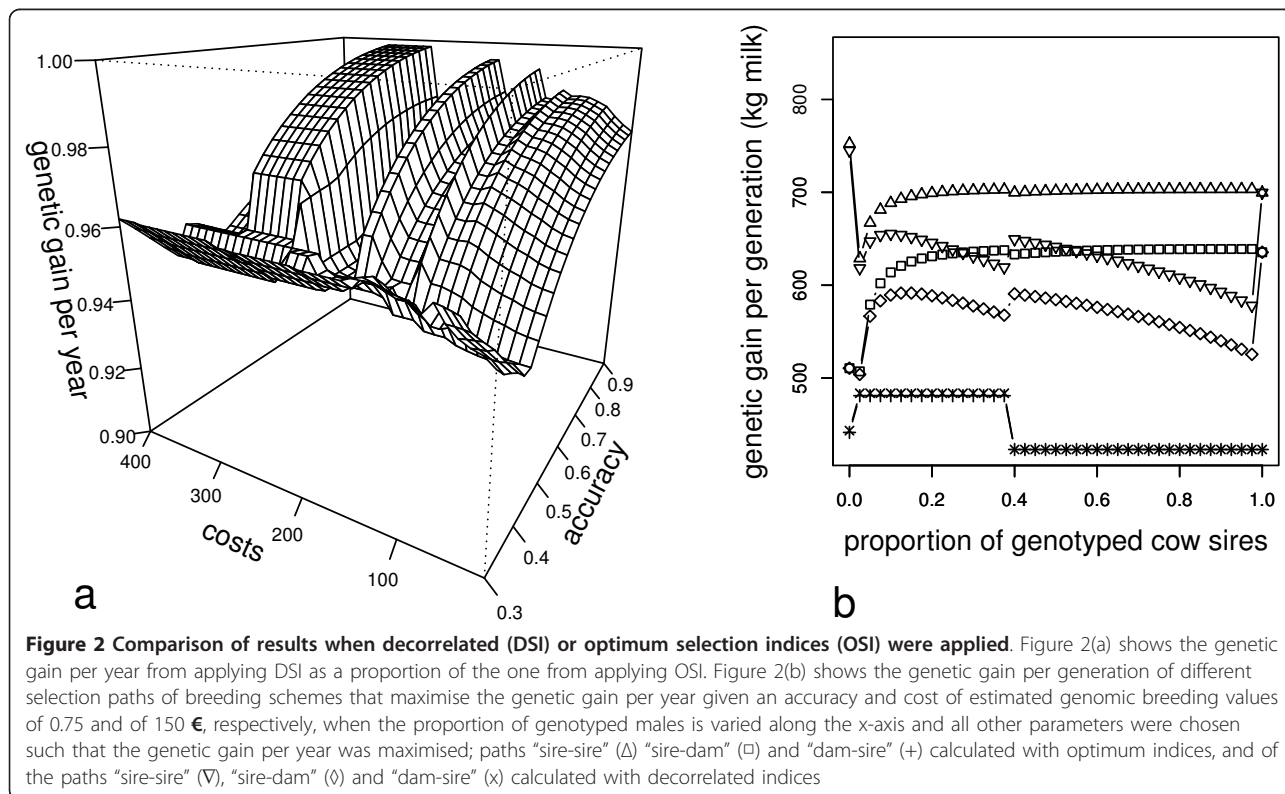
Number of different selection strategies in each selection path of breeding schemes which maximised  $\Delta G_a$  and the accuracy of estimated genomic breeding values. 1: parameter space for the accuracy of estimated genomic breeding values, 2: number of possible breeding schemes within the given parameter space, 3: number of schemes with selection only on pedigree data, 4: number of schemes with selection on pedigree data and progeny performance, 5: number of schemes with selection on pedigree data and estimated genomic breeding values, 6: number of schemes with selection on pedigree data, estimated genomic breeding values and progeny performance, 7: number of schemes with selection only on estimated genomic breeding values, 8: decorrelated indices are used, 9: optimum indices are used, 10: selection path "sire-dam", 11: selection path "sire-sire", 12: selection path "dam-sire", the parameter space for the cost of genomic selection per genotyped individual is 20-400 €



rQEBV range between 0.3 and 0.45. Beside these similarities, applying DSI also led to strategies for male selection that did not include GS, which was never the case when using OSI. Furthermore, a three-stage selection approach in paths SS and SD was found to maximise  $\Delta G_a$  much more frequently when using OSI. The differences between the proposed selection strategies in path SD are even more obvious within  $r_{GEBV}$  ranging between 0.45 and 0.9, including 741 possible parameter combinations. When using OSI, in all these cases the breeding schemes that maximised  $\Delta G_a$  used two-stage selection of sires, whereas when using DSI in 680 cases one-stage selection of sires was found to be optimum. Similar results were obtained for path DS, where two-stage selection was the most frequent strategy that maximised  $\Delta G_a$  when using OSI compared to DSI. Except for one-stage selection procedures, DSI always suggested breeding schemes yielding less  $\Delta G_a$  than OSI, with a maximum reduction of 5.5% and a mean of 2% (see Figure 2 (a)). If all optimization results obtained by applying DSI were recalculated using OSI, which means using the selection intensities obtained from DSI and using these with OSI, the loss in the predicted  $\Delta G_a$  due to DSI was up to 5.5%, and up to 7% for the predicted  $\Delta G_{SS}$  (results not shown). For the reverse recalculation, i.e. using selection intensities of optimum breeding schemes obtained from OSI in DSI, the loss in the predicted  $\Delta$

$G_a$  increased up to 29%, and the loss in the predicted  $\Delta G_{SS}$  up to 40% (results not shown).

A key to understand the different results when using DSI or OSI is given in Figure 2(b) and Table 3. Decorrelating EBV of subsequent selection stages is equal to decomposing e.g. the variance at stage 2,  $\sigma_2^2$ , into  $\sigma_1^2 + 2\sigma_{1,2} + \sigma_e^2$  (the  $\sigma_1^2$  is the variance at the previous stage and  $\sigma_{1,2}$  is the respective covariance) and reducing the variance of the information source at stage 2 to  $\sigma_e^2$ . Thus, an increase in the covariance between stages leads to a further reduction of the variance of EBV in the last selection stage. This is exemplified in the last two rows of Table 3. Including a GS stage in path DS increased the accuracy of the PED stage EBV and the covariance between PED stage and GS stage in path SD, but led to a reduction of the accuracy of the GS stage EBV in this path to a value below that of the PED stage EBV. The effect of this interaction on  $\Delta G$  is given in Figure 2(b), which reflects  $\Delta G_{SD}$ ,  $\Delta G_{SS}$  and  $\Delta G_{DS}$  of breeding schemes that maximized  $\Delta G_a$  as a function of DSI, OSI and the proportion of genotyped initial male selection candidates (PGsd) given an rQEBV of 0.75 and Cqebv of 150 €. All other parameters, e.g. the number of genotyped females were chosen such that  $\Delta G_a$  was maximised. If PGsd was set to zero, a breeding scheme selecting males at PED and PPT stages and females only at the PED stage was found to maximise  $\Delta G_a$  (GS of



**Table 3 The accuracy of estimated breeding values of successive selection stages,  $r_{GEBV^1} = 0.75$**

Selection path	Selection stage		Accuracy of estimated breeding values by stage			
			Optimum indices		decorrelated indices	
	PED <sup>2</sup>	GS <sup>3</sup>	PED	PED + GS	PED	GS
SD <sup>4</sup>	yes	no	0.177	-	0.177	-
DS <sup>5</sup>	yes	no	0.347	-	0.347	-
SD	yes	no	0.427	-	0.427	-
DS	yes	yes	0.379	0.766	0.379	0.606
SD	yes	yes	0.441	0.754	0.441	0.612
DS	yes	no	0.499	-	0.499	-
SD	yes	yes	0.537	0.754	0.537	0.530
DS	yes	yes	0.50	0.773	0.50	0.589

The accuracy of estimated breeding values of successive selection stages as a function of the implementation of genomic selection and the used selection indices. 1: accuracy estimated genomic breeding value, 2: selection on performance and GEBV of ancestors and sibs, 3: selection on candidates own GEBV, 4: selection path "sire to dam", 5: selection path "dam to sire"

females was not possible due to cost limitations). In this case,  $\Delta G_{DS}$  of both methods were equal because of the one-stage selection approach. The difference in  $\Delta G_{SD}$  and  $\Delta G_{SS}$  was small because of the high selection intensity at the PED stage and a DSI induced reduction of the accuracy of the PPT stage EBV from 0.993 to 0.814. If genotyping was allowed, as indicated by  $PG_{SD} > 0$ , breeding schemes that maximised  $\Delta G_a$  excluded the PPT stage in paths SD and SS in favour of inclusion of a GS stage in all paths. GS of females had to be abandoned due to cost constraints if  $PG_{SD}$  exceeded 0.375. The DSI induced reduction of the accuracy of GS stage EBV below that of PED stage EBV led to a strong peak of  $\Delta G_{SS}$  and  $\Delta G_{SD}$  at an already low  $PG_{SD}$  value. Furthermore, abandoning pre-selection at the PED stage ( $PG_{SD}$  equal to one) led to a sharp increase in  $\Delta G_{SS}$  and  $\Delta G_{SD}$  because one-stage selection made decorrelation unnecessary. The counter-intuitive interaction between paths due to DSI became even more obvious if GS of females was abandoned, which slightly decreased  $\Delta G_{SD}$  and  $\Delta G_{SS}$  when OSI was applied but increased these genetic gains when DSI was applied. This resulted from an inversion of the process well described in the last two rows of Table 3. Thus, the reason behind the given interaction is the characteristic of DSI to exploit only the residual variance. This can reduce the accuracy of EBV of a certain stage below those of the preceding stages, which leads to a loss of  $\Delta G_a$  if two- or multi-stage selection is applied. Additionally, more accurate EBV in one path can have negative side effects on EBV accuracies and  $\Delta G$  in other paths. Such results are impossible when using OSI.

#### Genetic gain of optimum indices

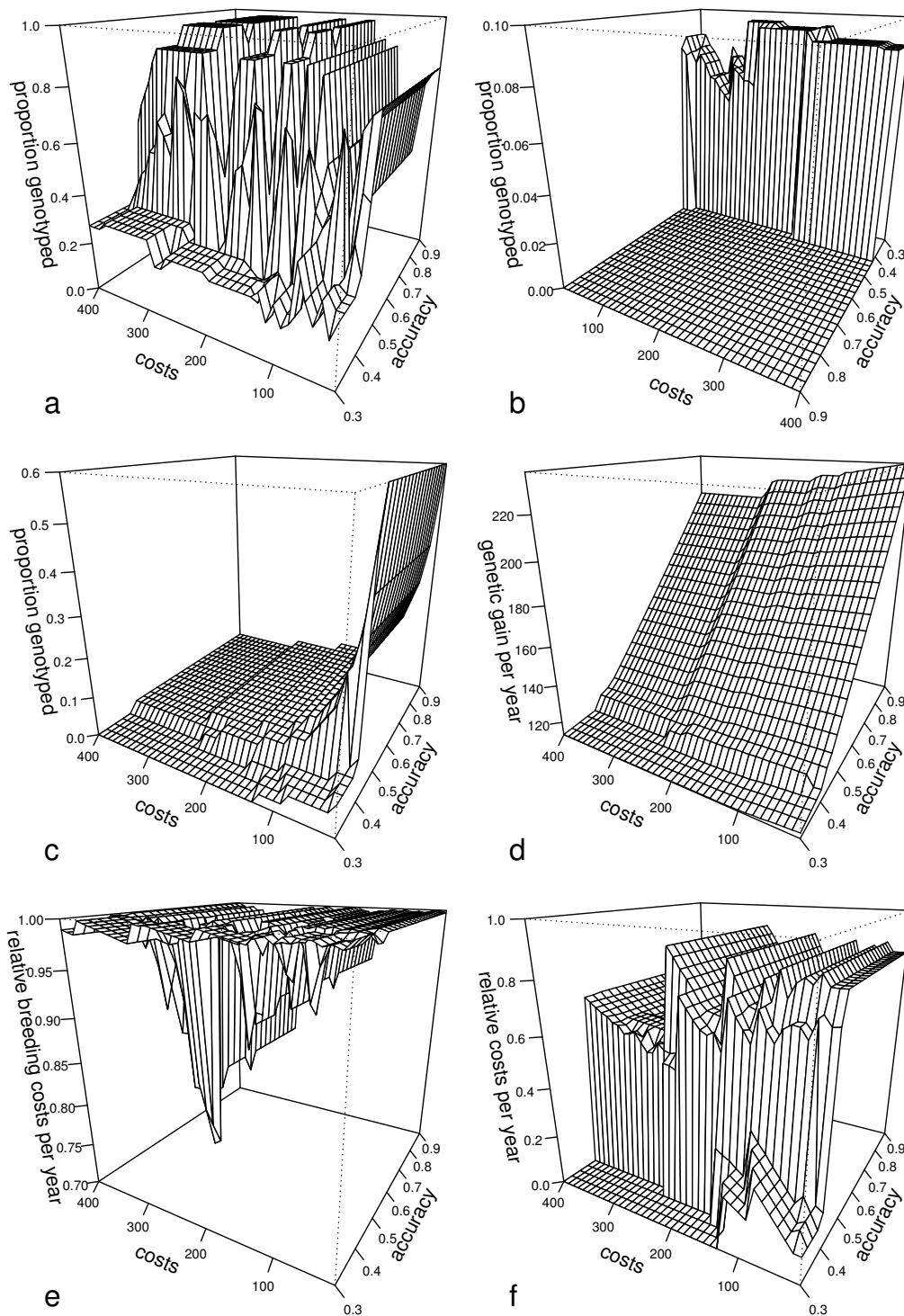
Results when using OSI are given in Figure 3.  $\Delta G_a$  increased with increasing  $r_{GEBV}$  but was almost independent of  $C_{GEBV}$ . The highest  $\Delta G_a$  of 236.94 kg or 0.67

genetic standard deviations was achieved at a parameter combination of  $r_{QEBV} = 0.9$  and  $C_{GEBV} = 20$  €, whereas the lowest  $\Delta G_a$  of 113.09 or 0.32 genetic standard deviations was achieved at a parameter combination of  $r_{QEBV} = 0.3$  and  $C_{GEBV}$  between 380 and 400 €.

The relative contribution of the different selection paths to the total genetic gain of breeding schemes was between 0.46 and 0.34 for path SS, between 0.36 and 0.29 for path SD, and between 0.35 and 0.19 for path DS. Only in 21 of 975 parameter combinations did the relative contribution of path DS exceed that of path SS, but by no more than 0.01. The relative contribution of path SD was exceeded by that of path DS in 202 of 975 cases; wherein the maximum excess was 0.04. Thus, in the vast majority of parameter combinations, the main contribution to genetic gain came from path SS (results not shown).

The total cost of breeding schemes that maximised  $\Delta G_a$  ranged from 544 685 to 718 973 € but only in 218 of the 975 parameter combinations was it below 700 000 €. Breeding costs incurred by genotyping females ranged from zero (no genotyping) to > 90% of the total breeding costs but was greater than 50% in 802 of the 975 parameter combinations. As given in Figure 3(f), a prerequisite for such a high proportion of total cost from genotyping females was an  $r_{QEBV} > 0.4$ . Below this value, the marginal benefit of a reduced selection intensity at the PED stage and an increased selection intensity at the GS stage in path DS was not found to maximise  $\Delta G_a$ .

The highest proportion of genotyped bull dams i.e. 0.6, which equals 30 000 heifers, was found when genotyping costs were the lowest and  $r_{GEBV}$  was  $\geq 0.45$ . This proportion decreased with increasing  $C_{GEBV}$ , but GS was applied to select females in almost all parameter combinations except when  $C_{GEBV}$  exceeded 120 € in conjunction with an  $r_{GEBV}$  range between 0.3 and 0.4. As



**Figure 3** Characteristics of breeding schemes that maximise the genetic gain per year when optimum selection indices were used. 3(a) Proportion of genotyped initial selection candidates in the sire-dam path 3(b) Proportion of progeny tested initial selection candidates in the sire-dam path 3(c) Proportion of genotyped initial selection candidates in the dam-sire path 3(d) Genetic gain per year 3(e) Total breeding cost as proportion of the maximum cost 3(f) Breeding cost of the dam-sire path as proportion of total breeding cost



$PG_{SD}$  was always below one, selection in paths SS and SD included a PED and a GS stage if  $r_{GEBV}$  ranged between 0.45 and 0.9 and between 0.4 and 0.9 for bull sires and cow sires, respectively. Below an  $r_{GEBV}$  of 0.4 (SD) and 0.45 (SS) three-stage selection strategies with a PPT stage were also found to maximise  $\Delta G_a$ . Furthermore, for  $r_{GEBV}$  between 0.425 and 0.45, bull sires were selected from cow sires on the basis of an additional PPT stage. Nevertheless, a pure two-stage selection strategy with only a PED and a PPT stage was not found to maximise  $\Delta G_a$  across all 975 parameter combinations.

## Discussion

### Comparison of the results for the decorrelated and optimum index

The results of this study quantify the respective loss in predicted  $\Delta G_a$  from using DSI instead of OSI to be up to 5.5% and in  $\Delta G$  per generation to be up to 6%. This is within the range given by Xu and Muir [9]. Nevertheless, DSI changes the functional dependencies of selection intensity and  $\Delta G$  as well as  $\Delta G_a$ . Although the modelled breeding scheme was less complex, tracing the interaction among selection stages and paths was quite difficult. This might become impossible if DSI is applied to more complex breeding schemes with numerous selection stages in a variety of paths with different selection intensities and generation intervals. Thus, DSI has its mathematical intrinsic logic, but for breeders its results are difficult to interpret, counter-intuitive or suboptimal.

The genetic variance is known to be reduced by the "Bulmer" effect, selection-induced gametic disequilibrium [23], leading to an overestimation of asymptotic rates of genetic response [24-26] if this effect is not accounted for as was the case in our study. The ranking and relative differences between alternative breeding programs have, however, been found to be little affected by ignoring this effect [24,25]. Comparisons between OSI and DSI-types of selection indices are not affected, because ignoring the Bulmer effect is equivalent to a comparison in terms of one-generation responses.

One of the advantages of decorrelated indices mentioned by Xu and Muir [9] is the ability to use maximisation methods that use first and second derivatives of the goal function. However, this option is limited to breeding schemes with equal generation intervals because otherwise the goal function might change in a non-continuous manner with the number of selection stages. Furthermore, the cost function of NamKoong [20] is not continuous if the selection intensities of certain selection stages converge to zero in the maximisation process. Thus, grid searches or heuristic approaches are still the methods of choice when goal functions are

difficult or impossible to differentiate or non-continuous. As limitations due to central processing unit time have almost been overcome because of developments in efficient hardware and fast algorithms [13], the exact calculation of optimum indices in combination with the above mentioned methods is the better alternative.

### Results using the optimum index

In this study, we found for a given  $r_{GEBV}$  of 0.75, a  $\Delta G_a$  between 0.53 and 0.57 genetic standard deviation, which is higher than found in other published results [1,27]. This may result from differences in the underlying selection intensities in paths SS and SD, which were higher in our calculations. As already mentioned, the effect of selection on genetic variance in later generations ("Bulmer" effect) was not accounted for. However, the results are still comparable to other studies that did not consider this effect either [1,28]. Furthermore, since ranking the breeding schemes might not be affected, the main conclusion remains the same. Ignoring reduction in variances in later stages also affects optimal breeding schemes in terms of the proportion selected in each stage and whether or not a given stage is utilised. So including this effect is important, which our methods do.

The strong fluctuation of  $PG_{SD}$  among programs that maximise genetic gain (see Figure 3(a)) can be explained by the non-linearity of  $\Delta G$  as a function of selection intensity (see the OSI curves in Figure 2(b)). In optimum breeding schemes,  $PG_{SD}$  was mostly at its upper limit, where the marginal benefit or loss in  $\Delta G_a$  from a small increase or decrease of selection intensity was very small. On the contrary, due to the imposed cost constraint, the proportion of genotyped females was small, leading to a high sensitivity of  $\Delta G_{DS}$  to any change in selection intensity. In the case of increasing  $C_{GEBV}$  three scenarios were possible. A: Cost was increased as long as the number of genotyped females could be maintained by reducing  $PG_{SD}$  with negligible effect on  $\Delta G_a$ . B: Cost was increased until a reduction of  $PG_{SD}$  led to a loss in  $\Delta G_{SD}$  being higher than a loss in  $\Delta G_{DS}$  due to reduced genotyping of females. C: Cost was increased till even not genotyping males could not save enough money to maintain the number of genotyped females. In the latter two scenarios the number of genotyped females had to be decreased and the available funds could be reinvested to increase  $PG_{SD}$  up to the maximum achievable  $\Delta G_{SD}$ .

Schaeffer [1] found that path DS became the main source of genetic gain but he assumed that for each possible bull dam a highly accurate GEBV was estimated. However, when calculating the breeding costs for such a GS scheme, genotyping costs for only 2000 bull dams were regarded. In our breeding programs genotyping

large proportions of the population of potential bull dams was not possible due to cost limitations. Thus, paths SS and SD were generally the main sources of  $\Delta G_a$ . Furthermore, genotyping bull dams was the major source of breeding costs, leading to values higher than 90% of the maximum possible breeding costs in the majority of parameter combinations. This exemplifies a trade-off between decreased overall breeding costs and the importance of path DS for  $\Delta G$ .

Since implementing GS allows to gather information on selection candidates relatively cheaply compared to PPT, the selection intensity can be increased because of a higher number of selection candidates. When doing so, breeding organisations should take into consideration that the additional  $\Delta G_a$  from an extended selection basis is approaching zero. Additionally, optimising breeding schemes should also include the cost of the invested capital and  $\Delta G_a$  per €. Such parameters may not allow for the inclusion of GS in path SD, as given in our results and for the application of GS in general in path DS, and may also question the utility of genotyping even whole sub-populations, as proposed by König and Swalve [28].

Some breeding organisations rely on using GS as a preselection stage followed by PPT [5]. The continuation of progeny testing in combination with GS was found to be economic only at an TGE<sub>EBV</sub>  $\leq 0.4$  in path SD and  $\leq 0.45$  in path SS. Since an  $r_{\text{QEBV}}$  of 0.7 can be achieved in practical breeding programs [29], there may be no alternative to replacing conventional progeny testing by GS in order to maximise the genetic gain per year.

## Conclusions

Applying decorrelated indices to multistage dairy cattle breeding schemes including genomic selection in an optimisation approach taking into consideration the strong interaction between selection paths led to results that were not only difficult to interpret but also counter-intuitive or suboptimal. This may result in improper advice to breeding organisations. Since fast and stable calculation of selection intensities in multistage breeding programs is possible even for highly correlated EBV of successive stages and small proportions of selected individuals, the optimum selection index is the method of choice for the deterministic optimisation of breeding schemes using the selection index methodology.

Genomic selection might meet its promises concerning the increase in genetic gain per year, although the effects on breeding costs are still unclear. However, the relatively low financial efforts to obtain estimated genomic breeding values compared to progeny performance testing make it possible to optimise breeding schemes in

a holistic across-path approach, which also includes the risk of losing money due to opportunity costs.

## Acknowledgements

This work was financed by the German Federal Ministry for Education and Research (BMBF), project "FUGATO-Plus Brain" (FKZ: 0315136D).

## Authors' contributions

VB worked out the conception and design of the simulation study, wrote computer program and analysed the results and wrote the article. NR worked out the conception and design of the simulation study and revised the article. **All authors have read and approved the final manuscript.**

## Competing interests

The authors declare that they have no competing interests.

Received: 21 February 2011 Accepted: 17 January 2012

Published: 17 January 2012

## References

- Schaeffer LR: **Strategy for applying genome-wide selection in dairy cattle.** *J Anim Breed Genet* 2006, **123**:218-223.
- Daetwyler HD, Villanueva B, Bijma P, Woolliams JA: **Inbreeding in genome-wide selection.** *J Anim Breed Genet* 2007, **124**:369-376.
- Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME: **Invited review: Genomic selection in dairy cattle: progress and challenges.** *J Dairy Sci* 2009, **92**:433-443.
- Meuwissen TH, Hayes BJ, Goddard ME: **Prediction of total genetic value using genome-wide dense marker maps.** *Genetics* 2001, **157**:1819-1829.
- Spelman R, Arias J, Keehan M, Obolonkin V, Winkelman A, Johnson D, Harris B: **Application of genomic selection in the New Zealand dairy cattle industry.** In *Proceedings of the 9th World Congress on Genetics Applied to Livestock Production: 1-6 August 2010; Leipzig* Edited by: Gesellschaft für Tierzuchtwissenschaften eV (German Society for Animal Science) 2010, Abstract 0311.
- Ducrocq V, Colleau J: **Optimum truncation points for independent culling level selection on a multivariate normal distribution, with an application to dairy cattle selection.** *Genet Sel Evol* 1989, **21**:185-191.
- Saxton A: **Indcull version 3.0: Independent culling for two or more traits.** *J Hered* 1989, **80**:166-167.
- Xu S, Muir WM: **Multistage selection for genetic gain by orthogonal transformation.** *Genetics* 1991, **129**:963-974.
- Xu S, Muir WM: **Selection index updating.** *Theor Appl Genet* 1992, **83**:451-458.
- Hicks C, Muir W, Stick D: **Selection index updating for maximizing rate of annual genetic gain in laying hens.** *Poult Sci* 1998, **77**:1-7.
- Xu S, Martin TG, Muir WM: **Multistage selection for maximum economic return with an application to beef cattle breeding.** *J Anim Sci* 1995, **73**:699-710.
- Xie C, Xu S: **Efficiency of multistage marker-assisted selection in the improvement of multiple quantitative traits.** *Heredity* 1998, **80**:489-498.
- Genz A: **Numerical computation of multivariate normal probabilities.** *J Comput Graph Stat* 1992, **1**:141-149.
- Brent RP: **Algorithms for Minimisation without Derivatives.** New Jersey: Prentice-Hall; 1973.
- Schrooten C, Bovenhuis H, van Arendonk JAM, Bijma P: **Genetic progress in multistage dairy cattle breeding schemes using genetic markers.** *J Dairy Sci* 2005, **88**:1569-1581.
- Dekkers JCM: **Prediction of response to marker-assisted and genomic selection using selection index theory.** *J Anim Breed Genet* 2007, **124**:331-341.
- Tallis GM: **The moment generating function of the truncated multinormal distribution.** *J R Statist Soc B* 1961, **23**:223-229.
- Mi X, Utz F: **R-Package selection gain.** [http://cran.r-project.org/web/packages/selectiongain/index.html].
- Rendel JM, Robertson A: **Estimation of genetic gain in milk yield by selection in a closed herd of dairy cattle.** *J Genet* 1950, **50**:1-8.
- Namkoong G: **Optimum allocation of selection intensity in two stages of truncation selection.** *Biometrics* 1970, **26**:465-476.

21. Genz A: **A subroutine for computing non-central multivariate t probabilities.**[<http://www.math.wsu.edu/faculty/genz/software/fort77/mvtdstpack.f>].
22. Brent R: **BRENT is a FORTRAN90 library which contains algorithms for finding zeros or minima of a scalar function of a scalar variable.**[[http://people.sc.fsu.edu/~jburkardt/f\\_src/brent/brent.f90](http://people.sc.fsu.edu/~jburkardt/f_src/brent/brent.f90)].
23. Bulmer M: **The effect of selection on genetic variability.** *Am Nat* 1971, **105**:201-211.
24. Wray NR, Hill WG: **Asymptotic rates of response from index selection.** *Anim Prod* 1989, **49**:217-227.
25. Dekkers J: **Asymptotic response to selection on best linear unbiased predictors of breeding values.** *Anim Prod* 1992, **54**:351-360.
26. Villanueva B, Wray N, Thompson R: **Prediction of asymptotic rates of response from selection on multiple traits using univariate and multivariate best linear unbiased predictors.** *Anim Prod* 1993, **57**:1-13.
27. Lillehammer M, Meuwissen T, Sonesson A: **Effects of alternative genomic selection breeding schemes on genetic gain in dairy cattle.** In *Proceedings of the 9th World Congress on Genetics Applied to Livestock Production: 1-6 August 2010; Leipzig* Edited by: Gesellschaft für Tierzuchtwissenschaften eV (German Society for Animal Science) 2010, Abstract 0130.
28. KSnig S, Swalve HH: **Application of selection index calculations to determine selection strategies in genomic breeding programs.** *J Dairy Sci* 2009, **92**:5292-5303.
29. Hayes BJ, Bowman PJ, Chamberlain AC, Verbyla K, Goddard ME: **Accuracy of genomic breeding values in multi-breed dairy cattle populations.** *Genet Sel Evol* 2009, **41**:51.

doi:10.1186/1297-9686-44-1

**Cite this article as:** Börner and Reinsch: **Optimising multistage dairy cattle breeding schemes including genomic selection using decorrelated or optimum selection indices.** *Genetics Selection Evolution* 2012 **44**:1.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

