# AlphaPart - R implementation of the method for partitioning genetic trends

Jana Obšteter1, Justin Holl2, John M. Hickey3,and Gregor Gorjanc3

1Department of Animal Science, Agricultural Institute of Slovenia, Hacquetova ulica 17, 1000 Ljubljana, Slovenia

*2*The Pig Improvement Company, Genus plc, 100 Bluegrass Commons Blvd., Ste 2200, Hendersonville, TN 37075, USA

*3*The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Midlothian, EH259RG, United Kingdom

*\**Jana Obšteter, Agricultural Institute of Slovenia, Department of Animal Science, Hacquetova ulica 17, 1000 Ljubljana, Slovenia

E-mail addresses:

J. O.: jana.obsteter@kis.si

J. H: justin.holl@genusplc.com

J. M. H.: [john.hickey@roslin.ed.ac.uk](mailto:john.hickey@roslin.ed.ac.uk)

G. G.: [gregor.gorjanc@roslin.ed.ac.uk](mailto:gregor.gorjanc@roslin.ed.ac.uk)

# Abstract

# **Background:** In this paper we present the AlphaPart R package, an open-source software that implements a method for partitioning genetic trends to identify sources of genetic gain. The package includes the main partitioning function AlphaPart and a set of functions for handling data and results. We demonstrate the practical use of the package in a simulated pig breeding example. We simulated a two-tier pig breeding programme and partitioned the genetic trends with AlphaPart.

# **Results: T**he multiplier tier achieved higher breeding values than the nucleus tier for traits measured and selected in the multiplier. The partition of genetic trends revealed that this increase depended on the accuracy and intensity of selection in the multiplier and the extent of gene flow from the nucleus. For traits measured only in the nucleus, the multiplier achieved comparable or smaller genetic gain than the nucleus depending on the amount of gene flow.

# **Conclusions:** AlphaPart is freely from CRAN repository at <http://CRAN.R-project.org/package=AlphaPart>. The AlphaPart R package implements a method for partitioning genetic trends. The method and package provide a useful tool for quantifying the sources of genetic gain in breeding programmes. Utilization of AlphaPart software will help breeders better understand sources of genetic gain in their breeding programmes.

Keywords: genetic trend, partition, Mendelian sampling term, R package, pig breeding

# Implications

1. The AlphaPart R package implements a method for partitioning genetic trends. The method and package provide a useful tool for quantifying the sources of genetic gain in breeding programmes. We demonstrate the package and apply it to identify sources of genetic trend in a pig breeding example. When traits measured and selected for in the multiplier are a subset of those in the nucleus, the multiplier surpasses the nucleus for these traits even though an index trend is similar between both tiers. Utilization of AlphaPart software will help breeders better understand sources of genetic gain in their breeding programmes.

# Background

In this paper we present the AlphaPart R package that implements a method for partitioning genetic trends and demonstrate it with a pig breeding example. Breeding programmes improve populations for a set of traits by selecting and intermating genetically superior individuals. Population improvement can be measured with a genetic trend calculated from averaged year of birth estimated breeding values of selection candidates (Blair and Pollak. 1984; Sorensen et al., 1994).

While sources of the overall genetic gain are generally known, their realised contributions are hard to quantify in complex breeding programmes. García-Cortés *et al.* (2008) proposed a method for such analysis. First, the method partitions breeding values into parent average and Mendelian sampling terms (Thompson, 1979), and allocates the terms to analyst-defined “paths” (males, females, tested sires, etc.). Next, it summarizes path specific terms to quantify path contributions to the overall genetic trend.

The partitioning method has been used in a number of cases. Gorjanc *et al.* (2011) and Gorjanc *et al.* (2012) estimated contributions of national breeding programmes to Brows-Swiss and Holstein country-specific and global genetic trends. Špehar *et al.* (2011) estimated contributions of national selection and importation in Croatian Simmental cattle. Škorput *et al.* (2015) estimated the contribution of national selection and importation in two pig breeds in Croatia, and extended the analysis with the quantification of uncertainty (Sorensen et al., 1994). However, these studies used dedicated software implementations of the partitioning method, for which no open-source software exists.

The aim of this paper is to: i) present the AlphaPart R package; and ii) demonstrate it with a simulated pig breeding example that quantifies nucleus-multiplier gene flow and the contribution of nucleus and multiplier selection on genetic gain in the two tiers.

# Implementation

# We first demonstrate the AlphaPart package and its functions on an example dataset. Next, we describe the simulation of a pig breeding example to demonstrate the use of AlphaPart.

## AlphaPart

AlphaPart is an R package available from CRAN repository at <https://CRAN.R-project.org/package=AlphaPart>. It consists of the main function AlphaPart for partitioning breeding values and auxiliary functions for manipulating data and summarizing, visualizing, and saving results. The package includes an example dataset AlphaPart.ped, which includes a four-generation pedigree and information about the generation, country, gender, and breeding values. Below we describe and demonstrate the functions with the dataset.

We install and load the package with:

> install.packages(pkg = “AlphaPart”)  
> library(package = "AlphaPart")

We use the AlphaPart function to partition breeding values (bv1) in the AlphaPart.ped by the country variable into domestic and import contributions:

> data(AlphaPart.ped)  
> part <- AlphaPart(x = AlphaPart.ped,  
 colPath = “country”,  
 colBV = “bv1”)

The partitioning function AlphaPart requires a data frame holding pedigree with animal/sire/dam or animal/sire/maternal-grandsire, a time-ordering variable such as year of birth, partition variable (path), and breeding values. Following the method described in García-Cortés *et al.* (2008), we recurse the pedigree from the oldest to the youngest individuals, for each individual calculate parent average and Mendelian sampling terms for any number of traits and assign terms to paths. We partition multiple traits by specifying a vector of variables, say colBV = c(“bv1”, “bv2”). The multiple trait option can also serve to partition samples from a posterior distribution to quantify uncertainty (Sorensen et al., 1994**;** Škorput et al.**,** 2015). To speed-up calculations we use C++ and trait-vectorised partitioning. The function can also directly partition and summarize path contributions “on-the-fly”, which is a useful computational speed-up for huge pedigrees. The output object of the function is either AlphaPart or summaryAlphaPart class.

We use the generic summary.AlphaPart function to summarize an AlphaPart object by a grouping variable, say generation (gen):

> sumPartByGen <- summary(part, by = “gen”)  
> print(sumPartByGen)

The summary function summarizes breeding values and their path partitions by levels of grouping variable. By default, we summarize with a mean, but the user can specify any R function via the FUN argument. The summary function can also summarize only a subset of the object via the subset argument.

We use the generic plot.summaryAlphaPart function to plot summarized partitions:

> plot(sumPartByGen)

We provide a number of utility functions that ease partitioning analysis. With the pedFixBirthYear function we impute missing or fix erroneous years of birth. With the pedSetBase function we set the base population by specifying founders and removing older pedigree records. With the AlphaPartSubset function we keep partitions for specified paths in the AlphaPart or summaryAlphaPart objects. With the AlphaPartSum function we sum the partitions of several paths in a summaryAlphaPart object. The AlphaPartSubset and AlphaPartSum functions simplify the presentation of partitioning analysis.

*Pig breeding example*

We applied the AlphaPart R package to a simulated pig breeding example to examine the nucleus-multiplier gene flow and the contribution of nucleus and multiplier selection on genetic gain in both tiers. Pig breeders select in the nucleus and multiply this improvement in the multiplier to supply large number of commercial animals. The multiplier generally has lower genetic mean than the nucleus due to time-lag. However, animals with higher breeding values are often observed in the multiplier for some traits and we aimed to use AlphaPart to explain the source of this observation. To this end we have first simulated a stylised pig breeding programme that exposes the drivers of real observations. We have next partitioned the genetic trend by a tier-gender variable to quantify sources of genetic gain in the nucleus and the multiplier.

We used the AlphaSimR package (Gaynor et al., 2019) to simulate a pig breeding programme for a single breed with 40 years of selection on two uncorrelated traits. Trait 1 had heritability 0.25 and trait 2 had heritability 0.10. We measured both traits in the nucleus and only trait 1 in the multiplier. We selected on the index of the two traits with equal emphasis. We split the simulation into initial 20 years of a “burn‑in” and 20 years of evaluation.

In the burn-in we simulated only the nucleus and selected animals based on the index of phenotype values for both traits. We selected 25 males and 500 females each year and randomly crossed them to produce a new generation of 6,000 progeny (12 per cross). At the end of the burn-in we generated 5,000 females to seed the multiplier.

In the evaluation we simulated both the nucleus and the multiplier and selected animals based on the index of estimated breeding values for both traits. In the nucleus, we selected 25 males and 500 females each year and randomly crossed them to produce a new generation of 6,000 progeny (12 per cross). In the multiplier, we selected 750 females each year and randomly crossed them to a set of males to produce a new generation of 9,000 progeny (12 per cross). To quantify the effect of selection in the multiplier on genetic gain we defined the set of males as either 1) the 25 best nucleus males (MaleFlow100 scenario) or 2) the 25 best nucleus males and 100 best multiplier males (MaleFlow20 scenario).

We estimated the breeding values for each trait independently before each nucleus or multiplier selection decision. We ran pedigree‑based model implemented in blupf90 (Misztal et al., 2002) and used all available data from evaluation years. The model included the mean as a fixed effect and animal breeding values as a random effect modelled hierarchically with pedigree.

Finally, we partitioned the true breeding values with the AlphaPart as demonstrated above. We used AlphaPart function to partition true breeding values from the 20 evaluation years by the tier‑gender variable and summary.AlphaPart function to summarize the partitions by year to quantify the contribution of each tier-gender level to genetic trend in the nucleus and the multiplier.

We repeated the simulation 10 times. We present mean standardized true breeding values and their partitions with mean set to zero and genetic standard deviation set to one in the year 20. We chose to present true (instead of estimated) breeding values to assess the true sources of genetic gain. We plot path partitions by year and report final values, but note that partitions can change over time. Code for the simulation is available at <https://git.ecdf.ed.ac.uk/HighlanderLab_public/jobsteter_alphapart>.

# Results

The results show true breeding values and their partitions obtained with the AlphaPart for the two simulated pig breeding scenarios. Partitioning showed that the superiority depends on the extent of nucleus-multiplier gene flow as well as accuracy and intensity of multiplier selection.

## Distribution of breeding values in the last evaluation year

The simulation created a scenario where some multiplier animals had higher breeding values than the nucleus animals for some traits. This is shown in Figure 1 that presents distribution of true breeding values by trait, scenario, and tier in the last evaluation year of one simulation replicate. The difference between the nucleus and the multiplier was particularly large in MaleFlow20 scenario. Below we quantify the sources of these distribution differences with the partitioning of distributions means separately for each scenario over replicates.

## Partitioning the genetic trend of MaleFlow100 scenario

In MaleFlow100 scenario the multiplier achieved a higher genetic gain than the nucleus for trait 1 due to selection of multiplier females. This is shown in Figure 2 that presents the partitioning of genetic trends for trait 1 and 2 and their index in the nucleus and the multiplier for MaleFlow100 scenario. The partitioning expectedly showed that selection in the nucleus was the main source of genetic gain for both traits in both tiers, with selection of nucleus males contributing the most. In the nucleus the genetic gain for trait 1 (9.75) and trait 2 (8.34) stemmed from selection of nucleus females (4.10 for trait 1 and 3.42 for trait 2) and nucleus males (5.65 for trait 1 and 4.92 for trait 2). In the multiplier the genetic gain for trait 1 was higher (10.00) than in the nucleus. This increase was partly due to increased contribution of gene flow from the nucleus males (5.75) and partly due to selection of multiplier females (0.14). The contribution of multiplier selection changed across years. Genetic gain and path partitions for trait 2 in the multiplier were comparable to the nucleus.

## Partitioning the genetic trend of MaleFlow20 scenario

In the MaleFlow20 scenario selection of multiplier males further increased the genetic gain for trait 1 in the multiplier compared to the nucleus, but decreased the genetic gain for trait 2. This is shown in Figure 3 which presents partitioning of genetic trends for trait 1 and 2 and their index in the nucleus and the multiplier for MaleFlow20 scenario. As in MaleFlow100 scenario, the nucleus genetic gain for trait 1 (10.09) and trait 2 (8.39) stemmed from selection of nucleus females (4.40 for trait 1 and 3.22 for trait 2) and nucleus males (5.69 for trait 1 and 5.17 for trait 2). In the multiplier the final genetic gain for trait 1 was again higher (10.36) than in the nucleus. This increase was a result of selection of multiplier females (0.30) and multiplier males (0.15), and reduced contribution of gene flow from the nucleus females (via reduced use of nucleus males) (4.21). As observed in Figure 2, contribution of multiplier selection changed across years. Genetic gain for trait 2 was lower in the multiplier (8.14) than in the nucleus due to a small negative contribution of multiplier females (-0.05) and multiplier males (‑0.03), and reduced contribution of gene flow from the nucleus females (3.13) and nucleus males (5.09).

# Discussion

In this paper we present the AlphaPart R package that implements a method for partitioning genetic trends. The method and the package are valuable for deciphering and quantifying the sources of genetic gain. The package streamlines such analysis into a few lines of R code, while enabling advanced handling of data and results, and plotting. Here we demonstrated the partitioning method in a simulated pig breeding example with a higher genetic trend for some traits in the multiplier compared to the nucleus. This example showed the investigative power of the partitioning method and raised two discussion topics: i) what were the sources of genetic gain in the two tiers of a pig breeding programme; and ii) implications for pig breeding programmes. We conclude with plans for future development of AlphaPart.

By partitioning the genetic trend in a simulated pig breeding programme, we disentangled the observation of some multiplier animals having higher breeding values for some traits compared to the nucleus animals. While larger number of recombinations in the multiplier can reveal more variation, we expected lower breeding values in the multiplier due to time-lag between the nucleus and multiplier. The partitioning revealed that the gene flow from the nucleus into the multiplier was the main source of genetic gain in the multiplier, with the nucleus males contributing the most. This was expected due to nucleus-multiplier gene flow and higher intensity of selection in males.

However, the results also showed that selection in the multiplier can contribute genetic gain in addition of the gene flow from the nucleus. The multiplier outperformed the nucleus for trait 1, because with the 10,500 recorded multiplier animals there was substantial amount of information for accurate multiplier selection that generated additional genetic gain. The partitioning of genetic trend for trait 1 showed that when we used only the nucleus males in the multiplier (MaleFlow100), the multiplier generated additional gain from two sources. First, compared to the nucleus, the contribution of the nucleus males increased because they contributed through the gene flow and through the selection of multiplier females. Second, the selection of multiplier females contributed as well. When we used both the nucleus males and the multiplier males in the multiplier (MaleFlow20), the multiplier generated further gain through a combination of the sources. First was the contribution of the selection of multiplier females and males. In contrast, the contribution of nucleus decreased due to the reduced gene flow. This decrease was due to a smaller number of progeny per nucleus male compared to the MaleFlow100 scenario.

On the contrary, trait 2 was not measured in the multiplier and had comparable or smaller genetic trend in the multiplier than in the nucleus. For trait 2 the multiplier animals were selected only on estimated parent average, which resulted in low accuracy selection. In the MaleFlow100 scenario this low accuracy selection resulted in a null contribution of multiplier females to the genetic trend for trait 2 and comparable genetic trends between the nucleus and the multiplier. In the MaleFlow20 scenario with a reduced nucleus-multiplier gene flow this low accuracy selection resulted in the reduced genetic gain for trait 2.

Our future work on AlphaPart will include extending the partitioning method in three area. The first extension will utilise genomic information to inform which genome regions drive genetic change and what are sources of specific haplotypes or alleles. The second extension will use the partitioning method to analyse changes in genetic variance in addition to the genetic mean. The third extension will simplify handling of uncertainty of path contributions when working with samples from posterior distributions (Sorensen et al., 1994; Škorput et al., 2015).

# Acknowledgments

The authors acknowledge support from the BBSRC to The Roslin Institute (BBS/E/D/30002275) and The University of Edinburgh’s Data-Driven Innovation Chancellor’s fellowship.

# References

Blair HT and Pollak EJ 1984. Estimation of genetic trend in a selected populaton with and without the use of a control population. Journal of Animal Science 58, 878-886.

García-Cortés LA, Martínez-Ávila JC, Toro MA 2008. Partition of the genetic trend to validate multiple selection decisions. Animal 2, 821–824.

Gaynor RC, Gorjanc G, Wilson D, Money D, Hickey JM 2019. AlphaSimR Breeding Program Simulations. Version 0.11.0. <http://CRAN.R-project.org/package=AlphaSimR>

Gorjanc G, Potočnik K, García-Cortés LA, Jakobsen J and Dürr J 2011. Partitioning of international genetic trends by origin in Brown Swiss bulls. Interbull Bulletin 44, 81-86.

Gorjanc G, Hely FS, Amer PR 2012. Partitioning international genetic trends by origin in Holstein bulls. Paper presented at the ICAR conference, 28 May – 1 June 2012, Cork, Ireland.

Misztal I, Tsuruta S, Strabel T, Auvray B, Druet T and Lee DF 2002. BLUPF90 and related programs (BGF90). Proceedings from the 7th World Congress on Genetics Applied to Livestock Production, August 2002, Montpellier, France, 1-2.

Sorensen DA, Wang CS, Jansen J and Gianola D 1994. Bayesian analysis of genetic change due to selection using Gibbs sampling. Genetics Selection Evolution 26, 333-360.

Škorput D, Gorjanc G., Kasap A and Luković Z 2015. Partition of genetic trends by origin in Landrace and Large-White pigs. Animal 9, 1605–1609.

Špehar M, Ivkić Z, Bulić V, Barać Z, and Gorjanc G 2011. Partitioning of Genetic Trends by Origin in Croatian Simmental Cattle. Agriculturae Conspectus Scientificus 76, 301-304.

Thompson R 1979. Sire Evaluation. Biometrics 35, 339-353.

**Figure captions**

**Figure 1** Distribution of true breeding values by trait, scenario, and tier in the last evaluation year in one simulation replicate. In MaleFlow100 scenario the multiplier uses nucleus males, while in MaleFlow20 scenario the multiplier uses nucleus and multiplier males. Trait 1 is measured in the nucleus and the multiplier, while trait 2 is measured only in the nucleus.

**Figure 2** Partitioning of genetic trend by tier-gender in MaleFlow100 scenario that uses nucleus males in the multiplier. Trait 1 is measured in the nucleus and the multiplier, while trait 2 is measured only in the nucleus.

**Figure 3** Partitioning of the genetic trend by tier-gender in MaleFlow20 scenario that uses nucleus and multiplier males in the multiplier. Trait 1 is measured in the nucleus and the multiplier, while trait 2 is measured only in the nucleus.