



RESEARCH  
PROGRAM ON  
Fish  
Led by WorldFish



July, 2021

# AllocateMate: An R package for mate allocation

*Matthew G Hamilton*

## Citation

This publication should be cited as: Hamilton, MG. 2021. AllocateMate: An R package for mate allocation. Penang, Malaysia: CGIAR Research Program on Fish Agri-Food Systems. Program Report.

## About FISH

The [CGIAR Research Program on Fish Agri-Food Systems \(FISH\)](#) is a multidisciplinary research program. Designed in collaboration with research partners, beneficiaries and stakeholders, FISH develops and implements research innovations that optimize the individual and joint contributions of aquaculture and small-scale fisheries to reducing poverty, improving food and nutrition security and sustaining the underlying natural resources and ecosystems services upon which both depend. The program is led by [WorldFish](#), a member of the CGIAR Consortium. [CGIAR](#) is a global research partnership for a food secure future.

## Acknowledgments

The author acknowledges Trịnh Quốc Trọng for clarifying mating practices adopted in tilapia genetic improvement programs. This work was undertaken as part of the CGIAR Research Program on Fish Agri-Food Systems (FISH) led by WorldFish. The program is supported by contributors to the CGIAR Trust Fund.

## Contact

WorldFish Communications and Marketing Department, Jalan Batu Maung, Batu Maung, 11960 Bayan Lepas, Penang, Malaysia. Email: [fish@cgiar.org](mailto:fish@cgiar.org)

## Creative Commons License



Content in this publication is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License ([CC BY-NC 4.0](#)), which permits non-commercial use, including reproduction, adaptation and distribution of the publication provided the original work is properly cited.

© 2021 CGIAR Research Program on Fish Agri-Food Systems.

## Photo credits

Front cover, Matthew Hamilton.

# Table of contents

|   |   |
|---|---|
| 1. Introduction .....                           | 3 |
| 2. AllocateMate package.....                    | 3 |
| 3. Installing AllocateMate.....                 | 4 |
| References .....                                | 5 |
| Annex 1. allocate.mate.ped function R help..... | 6 |
| Annex 2. allocate.mate.H function R help .....  | 8 |

# 1. Introduction

WorldFish manages multiple family-based genetic improvement programs for tilapia and carp species (Charo-Karisa et al., 2020; Hamilton et al., 2019a; Hamilton et al., 2019b; Hamilton et al., 2021; Hamzah et al., 2014). In family-based genetic improvement programs, mate allocation is a process whereby selected parents (or in some cases families) are allocated to mating pairs prior to a spawning event. Mating pairs are then crossed to produce next-generation families of known pedigree.

## 2. AllocateMate package

AllocateMate is a R package (R Core Team, 2020) comprised of two primary functions: *allocate.mate.ped* and *allocate.mate.H*. These functions allocate parents to mating pairs based on genetic relationships provided by the user in the form a pedigree file (*allocate.mate.ped*) or a relationship matrix (*allocate.mate.H*). Mating lists are generated:

- that minimise the average Wright's inbreeding coefficient ( $F$ ; Wright, 1922) of offspring, by minimising the average coefficient of coancestry between parents in mating pairs; or
- according to assortative mating principles (Saura et al., 2017), by maximising the standard deviation of the parental mean EBVs of mating pairs.

To achieve this, all possible parental combinations are identified for a given set of parents. The optimal combination of these mating pairs is then found, using linear programming (Berkelaar, 2020), while constraining (Annex 1 and Annex 2):

- the number of full-sibling families each parent must contribute to (i.e. `N_AS_PARENT`);
- the maximum inbreeding coefficient allowed in the offspring (i.e. `max_F`); and
- the maximum number of siblings from one family allowed to be crossed with siblings of another family (i.e. `n_fam_crosses`) – this represents a constraint on the generation of double first cousins and is only applicable when using the *allocate.mate.ped* function.

Inputs, outputs and examples of use for the *allocate.mate.ped* and *allocate.mate.H* functions are detailed in Annex 1 and Annex 2.

## 3. Application in mouth-breeding tilapias

To ensure mouth-breeding tilapia successfully spawn in a timely fashion, it is common practice to place multiple females with one male in a mating hapa/tank (Trong, 2013). To use AllocateMate in this circumstance, the following guidelines can be adopted.

If females in each mating hapa/tank are to be from a single full-sibling family:

- the *allocate.mate.ped* function should be used (Annex 1);
- female parental identifiers (i.e. ID's present in the 'parents' data frame; Annex 1) should be modified so that there is only one female parental ID per mating hapa/tank, each representing multiple females from the same family; and
- `N_AS_PARENT` in the 'parents' data frame should be specified as 1.

If females in each mating hapa/tank are to be from multiple families and the average Wright's inbreeding coefficient is to be minimised (i.e. method = "min\_F"; Annex 1 and Annex 2):

- for males, N\_AS\_PARENT should be specified as the number of females to be placed in the hapa/tank with the male;
- for females, N\_AS\_PARENT should be specified as 1.

If females in each mating hapa/tank are to be from multiple families and assortative mating principles are to be applied (i.e. method = "assortative"; Annex 1 and Annex 2), to ensure that the female parents with the highest EBVs are not placed in the same mating hapa/tank:

- female parents should be sorted by their estimated breeding value (EBV; Annex 1 and Annex 2) and divided into groups – individuals with the highest EBVs should be grouped together, etc. The number of groups should equal the number of females in each mating hapa/tank and groups should be of equal size;
- N\_AS\_PARENT should be specified as 1; and
- *allocate.mate.ped* or *allocate.mate.H* should be run separately for each group of females with a common group of males; and
- the mating lists from all groups should be combined upon completion.

## 4. Installing AllocateMate

AllocateMate Version 1.0 is available at <https://github.com/mghamilton/AllocateMate> or for download as a .tar.gz 'source file' (<https://doi.org/10.7910/DVN/Z1PXJJ>). To install the package from the .tar.gz source file:

- use the *install.packages* function in R – `install.packages(path-to-source-file, repos = NULL, type = "source");` or
- within R studio, go to Tools -> Install Packages and then set 'Install From' to 'Package Archive File (.zip; .tar.gz)' before browsing to find the downloaded .tar.gz file.

## References

- Berkelaar, M., 2020. Package 'lpSolve'. Version 5.6.15. CRAN.  
<https://cran.uib.no/web/packages/lpSolve/lpSolve.pdf>.
- Charo-Karisa, H., Ali, S., Marijani, E., Ibrahim, N.A., Trinh, T.Q., Chadag, M.V., Benzie, J.A., 2020. Genetic parameters for black spot disease (diploptostomiasis) caused by *Uvulifer* sp. infection in Nile tilapia (*Oreochromis niloticus* L.). *Aquaculture*. 532, 736039.  
<https://doi.org/10.1016/j.aquaculture.2020.736039>
- Hamilton, M.G., Mekkawy, W., Benzie, J.A.H., 2019a. Sibship assignment to the founders of a Bangladeshi *Catla catla* breeding population. *Genetics Selection Evolution*. 51, 17.  
<https://doi.org/10.1186/s12711-019-0454-x>
- Hamilton, M.G., Mekkawy, W., Kilian, A., Benzie, J.A.H., 2019b. Single Nucleotide Polymorphisms (SNPs) reveal sibship among founders of a Bangladeshi rohu (*Labeo rohita*) breeding population. *Front Genet*. 10. <https://doi.org/10.3389/fgene.2019.00597>
- Hamilton, M.G., Mekkawy, W., Barman, B.K., Alam, M.B., Karim, M., Benzie, J.A.H., 2021. Genetic relationships among founders of a silver carp (*Hypophthalmichthys molitrix*) genetic improvement program in Bangladesh. *Aquaculture*, 736715.  
<https://doi.org/https://doi.org/10.1016/j.aquaculture.2021.736715>
- Hamzah, A., Ponzoni, R.W., Nguyen, N.H., Khaw, H.L., Yee, H.Y., Mohd Nor, S.A., 2014. Performance of the Genetically Improved Farmed Tilapia (GIFT) strain over ten generations of selection in Malaysia. *Pertanika J Trop Agric Sci*. 37, 411-429
- R Core Team, 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Saura, M., Villanueva, B., Fernández, J., Toro, M.A., 2017. Effect of assortative mating on genetic gain and inbreeding in aquaculture selective breeding programs. *Aquaculture*. 472, 30-37.  
<https://doi.org/10.1016/j.aquaculture.2016.05.013>
- Trong, T., 2013. Optimisation of selective breeding program for Nile tilapia (*Oreochromis niloticus*).
- Wright, S., 1922. Coefficients of inbreeding and relationship. *Am Nat*. 56, 330-338

# Annex 1. allocate.mate.ped function R help

allocate.mate.ped {AllocateMate}

R Documentation

## allocate.mate.ped

### Description

This function generates a mating list for a set of parents. The mating list can be generated i) to minimise the average inbreeding coefficient ( $F$ ) of families generated or ii) according to assortative mating principles. Inputs include a list of parents and a 3-column pedigree file specifying the ancestry of these candidates.

### Usage

```
allocate.mate.ped(ped, parents, max_F = 1, method = "min_F", n_fam_crosses = 1)
```

### Arguments

- `ped` is a 3-column dataframe with the following columns (class in parentheses):
- 'ID' is the individual identifier of parents and their ancestors (character).
  - 'DAM' is the identifier of the individual's dam (NA if unknown) (character).
  - 'SIRE' is the identifier of the individual's sire (NA if unknown) (character).
- `parents` data frame with the following columns (class in parentheses):
- 'ID' is the individual identifier (character).
  - 'SEX' is the sex of the individual - 'M' or 'F', for male and female respectively (character).
  - 'EBV' Estimated breeding value (EBV) of the individual - can not be NA if applying assortative mating as the method (numeric).
  - 'N\_AS\_PARENT' The number of families the individual is to contribute to (integer).
- `max_F` is the maximum inbreeding coefficient allowed in the offspring of parents (numeric between 0 and 1)
- `method` either 'min\_F' (to minimise the average inbreeding in offspring) or 'assortative' (to apply assortative mating) (character)
- `n_fam_crosses` is the maximum number of siblings from one family allowed to be crossed with siblings of another family. This represents a constraint on the generation of 'double first cousins'. (integer)

### Value

'summary' is a data frame containing a summary of all possible families generated from matings between parents:

- 'SELECTED' N represents unselected families, Y represents selected families (i.e. mate allocations to be generated) and All represents all possible families.
- 'COUNT\_FAMS' count of families.
- 'MEAN\_EBV' mean of family EBVs.
- 'SD\_EBV' standard deviation of family EBVs.
- 'MIN\_EBV' minimum of family EBVs.
- 'MAX\_EBV' maximum of family EBVs.
- 'MEAN\_F' mean of family inbreeding coefficients ( $F$ ).
- 'SD\_F' standard deviation of family inbreeding coefficients ( $F$ ).
- 'MIN\_F' minimum of family inbreeding coefficients ( $F$ ).
- 'MAX\_F' maximum of family inbreeding coefficients ( $F$ ).

'all\_families' is a data frame containing details of all possible families able to be generated from matings between parents:

- 'SIRE' male parent.
- 'DAM' female parent.
- 'F' inbreeding coefficient of family members (i.e. the 'family F').
- 'EBV' mean of parent EBVs (i.e. the 'family EBV').
- 'SELECTED' N represents an unselected family (i.e. family is not to be generated), Y represents a selected family (i.e. mate allocated - family to be generated).

'optimal\_families' is a data frame containing details of mate allocations (i.e. families to be generated from matings between parents):

- 'SIRE' male parent.
- 'DAM' female parent.
- 'F' inbreeding coefficient of family members (i.e. the 'family F').
- 'EBV' mean of parent EBVs (i.e. the 'family EBV').

## Examples

```
#Retrieve example data
data(ped)
data(parents)

output <- allocate.mate.ped(ped, parents, max_F = 0.08, method = "min_F", n_fam_crosses = 1)
output$summary
head(output$all_families)
head(output$optimal_families)
```

# Annex 2. allocate.mate.H function R help

allocate.mate.H {AllocateMate}

R Documentation

## allocate.mate.H

### Description

This function generates a mating list for a set of parents. The mating list can be generated i) to minimise the average inbreeding coefficient (F) of families generated or ii) according to assortative mating principles. Inputs include a list of parents and a relationship matrix (H) including these parents.

### Usage

```
allocate.mate.H(H, parents, max_F = 1, method = "min_F")
```

### Arguments

- H** is square additive genetic relationship matrix with individual identifiers as rownames and colnames.
- parents** data frame with the following columns (class in parentheses):
- 'ID' is the individual identifier (character).
  - 'SEX' is the sex of the individual - 'M' or 'F', for male and female respectively (character).
  - 'EBV' Estimated breeding value (EBV) of the individual - can not be NA if applying assortative mating as the method (numeric).
  - 'N\_AS\_PARENT' The number of families the individual is to contribute to (integer).
- max\_F** is the maximum inbreeding coefficient allowed in the offspring of parents (numeric between 0 and 1)
- method** either 'min\_F' (to minimise the average inbreeding in offspring) or 'assortative' (to apply assortative mating) (character)

### Value

'summary' is a data frame containing a summary of all possible families generated from matings between male parents:

- 'SELECTED' N represents unselected families, Y represents selected families (i.e. mate allocations to be generated) and All represents all possible families.
- 'COUNT\_FAMS' count of families.
- 'MEAN\_EBV' mean of family EBVs.
- 'SD\_EBV' standard deviation of family EBVs.
- 'MIN\_EBV' minimum of family EBVs.
- 'MAX\_EBV' maximum of family EBVs.
- 'MEAN\_F' mean of family inbreeding coefficients (F).
- 'SD\_F' standard deviation of family inbreeding coefficients (F).
- 'MIN\_F' minimum of family inbreeding coefficients (F).
- 'MAX\_F' maximum of family inbreeding coefficients (F).

'all\_families' is a data frame containing details of all possible families able to be generated from matings between parents:

- 'SIRE' male parent.
- 'DAM' female parent.
- 'F' inbreeding coefficient of family members (i.e. the 'family F').
- 'EBV' mean of parent EBVs (i.e. the 'family EBV').
- 'SELECTED' N represents an unselected family (i.e. family is not to be generated), Y represents a selected family (i.e. mate allocated - family to be generated).

'optimal\_families' is a data frame containing details of mate allocations (i.e. families to be generated from matings between male parents):

- 'SIRE' male parent.
- 'DAM' female parent.
- 'F' inbreeding coefficient of family members (i.e. the 'family F').
- 'EBV' mean of parent EBVs (i.e. the 'family EBV').

## Examples

```
#Retrieve example data
data(H)
data(parents)

output <- allocate.mate.H(H, parents, max_F = 0.08, method = "min_F")
output$summary
head(output$all_families)
head(output$optimal_families)
```



RESEARCH  
PROGRAM ON  
Fish

Led by WorldFish

### **About FISH**

The CGIAR Research Program on Fish Agri-Food Systems (FISH) is a multidisciplinary research program. Designed in collaboration with research partners, beneficiaries and stakeholders, FISH develops and implements research innovations that optimize the individual and joint contributions of aquaculture and small-scale fisheries to reducing poverty, improving food and nutrition security and sustaining the underlying natural resources and ecosystems services upon which both depend. The program is led by WorldFish, a member of the CGIAR Consortium. CGIAR is a global research partnership for a food secure future.

For more information, please visit [fish.cgiar.org](http://fish.cgiar.org)