

# XDesign 1.0: a tool for generation of complex mating designs with optimal contributions

Verktøy for generering av komplekse korsningsplaner med optimalisert utvalg



PHOTO: CURT ALMQVIST/SKOGFORSK

## Summary

Tree breeders working with advanced-generation material have the difficult task of preparing mating plans that represent optimum unequal contributions from selected parents, applying positive assortment of parents, while avoiding excessive relatedness between parents that might give rise to inbreeding depression in the test progeny. To generate such plans for more than a few parents using pencil and paper is hugely difficult and prone to error. XDesign was developed as a generalized tool to prepare suitable mating designs for such difficult situations.

This report describes the use of XDesign on Windows-based computers. The main input file is a list of selected candidates, their pedigree, estimated breeding values (EBVs) and the desired frequency of use in the mating plan. The user can specify if mating is to be positive-assortative or random, and can specify a threshold on the maximum coancestry between parents in a given cross. The program prepares a crossing list, giving the parents to be crossed, their mid-parent EBV and the coancestry between parents in each cross.

## Sammanfattning

Skogsträdsförädlare som arbetar med andra till tredje generationens förädlingsmaterial står inför den svåra uppgiften att ta fram korsningsplaner som inbegriper ett obalanserat bidrag från valda föräldrar och att föräldrarna skall rangordnas i fallande ordning efter deras avelsvärde, samtidigt som det eftersträvas att minimera släktskapet i varje enskild parkorsning för att undvika inavelsdepression i efterföljande generation. Att generera sådana korsningsplaner med mer än ett fåtal föräldrar för hand är en mycket svår uppgift där det är lätt att göra misstag. XDesign utvecklades för att förenkla arbetet vid tillverkning av komplexa korsningsplaner.

Denna rapport beskriver användningen av XDesign på Windows-baserade datorer. Indata till programmet består av en textfil med utvalda kandidater, fullständig stamtavla, skattade avelsvärden samt önskat antal korsningar för varje kandidat. Användaren kan specificera om föräldrarna skall rangordnas i fallande ordning och det finns även möjlighet att ange ett tröskelvärde för maximalt tillåtet släktskapet i varje enskild parkorsning. Programmet presenterar en korsningslista med valda föräldrar för varje parkorsning, deras genomsnittliga avelsvärde och släktskap.

## **Preface**

This research has received funding from (1) Föreningen Skogsträdsförädling (The Swedish Tree Breeding Association), (2) the European Union Horizon 2020 Research and Innovation Program under grant agreement no. 676876 (Project GenTree), and (3) Skogforsk (The Swedish Forestry Research Institute)

2017-09-05

Tim J. Mullin

## Contents

Summary.....	2
Sammanfattning.....	2
Preface .....	3
Introduction.....	5
How XDesign goes about the task.....	6
The design algorithm steps.....	7
Hardware/software requirements for XDesign.....	8
Becoming a registered user of XDesign .....	8
A note on copyright, licensing and warranty .....	8
Distribution of XDesign .....	8
Installing XDesign .....	9
Running the examples provided with XDesign .....	9
The candidate file .....	9
Interactive dialogue or command line .....	10
Running example 1 from the online dialogue .....	10
Reviewing the design .....	12
Running example 2 from the command line.....	13
When things go wrong! .....	14
Some possibilities for future upgrades .....	14
Acknowledgements .....	14
References .....	15

## Introduction

Most tree breeding programs advance through repeated cycles of selection, breeding and testing. In the first cycle, the breeding generally uses a mating design where selected parents are crossed an equal number of times according to a standard template representing a partial diallel, factorial, single- or double-pair mating design (White et al. 2007; Zobel and Talbert 1984). Such designs are straightforward to plan and implement, and can provide valuable insight on genetic parameters.

In later cycles, however, the mating design may become more complicated as it attempts to apply positive assortment of mates (Rosvall 1999; Rosvall et al. 2011), and possibly unequal optimal contributions from selected genotypes (Mullin 2014, 2017; Yamashita et al. 2017), such that selection satisfies some constraint on “group coancestry” (Cockerham 1967), “status number” (Lindgren et al. 1996), “proportional gene diversity” (Lacy 1995), or some other measure of diversity. It is also often desired to avoid mating parents that are strongly related, so as to avoid inbreeding depression in the test progeny. All of these requirements greatly complicate the planning and generation of the next recruitment population. Pencil and paper were quite adequate to plan the mating of founders, but such approaches for advanced cycles and complex pedigrees are hugely difficult and prone to error.

XDesign was developed to simplify the task of generating mating designs that accommodate many of the complications experienced in advanced generations. Selected genotypes need not be used equally, and one can impose positive assortment on the crossing. Should the relatedness (coancestry) between parents of a prescribed cross exceed a threshold specified by the user, XDesign will systematically “swap” parents with other crosses, in order to avoid excessive relatedness. The output from XDesign is a simple list of prescribed crosses to be carried out by the technical crew in the field. The crosses are listed with the estimated mid-parent breeding value, as well as the coancestry between the parents.

## How XDesign goes about the task

The program is supplied with a list of all selected genotypes and their desired frequency of use in the mating design, as well as all ancestors of these individuals. The list is organized in a conventional pedigree format, with columns for Individual, Female and Male, and ordered such that parents always precede their progeny. Additional columns in the file describe the estimated breeding value (EBV) and frequency of use (Table 1.).

The program is also supplied with a small number of parameters describing the constraints on the design. These include whether the selected genotypes are to be used in positive-assortative or random order (by EBV) and what threshold level of coancestry can be tolerated between crossed parents.

Table 1. Small example of a selection candidate file for designing the mating of unequal parent contributions. The Frequency column will always contain positive integer values, indicating how many times each individual is to appear in the mating design. Ancestors to the selected individuals must be given, even if their contribution frequency is zero.

Individual	Female	Male	EBV	Frequency
1	0	0	90.4	1
2	0	0	0.0	0
3	0	0	75.5	0
4	1	2	100.4	3
5	1	3	130.5	6
6	2	4	150.3	5
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.

## THE DESIGN ALGORITHM STEPS

1. XDesign begins by examining the order and coding of the pedigree. If the genotype IDs are sequential from 1 to  $N$ , this check can be skipped. In most cases, however, genotypes are identified with unique alphanumeric labels and the check is required.
2. XDesign extracts the information for those genotypes with contribution frequency of 1 or more, and determines the total number of crosses in the design as this sum divided by two (there being two genotypes required for each cross); obviously, the total number of contributions must be divisible by two.
3. These genotypes are transferred to a list, ranked either by EBV (in the case of PAM) or randomly.
4. A coancestry matrix is calculated for all possible crosses between individuals in the ranked list.
5. XDesign starts at the top of the list to assign the next available genotypes to crosses. These crosses must not be selfs, nor can they be reciprocals of crosses already assigned. At this point in the design, the coancestry between parents is not considered, so that the crossing is in strict rank order.
6. Xdesign now reviews the ranked list of crosses, again starting at the top, transferring the coancestry between parents from the matrix prepared in step 4 and evaluating if the coancestry is below the threshold specified by the user. If not, XDesign sets aside the parent with the lower rank, and scans down the list in rank order to find the next possible parent that satisfies the coancestry requirement. The program then considers “swapping” this new parent for the one set aside earlier. Provided the cross is not a self or a reciprocal of an earlier cross, the new cross is judged acceptable and the “swap” is confirmed. This step is repeated until all crosses have been verified and swaps made as required.
7. The completed list of crosses is prepared as an Excel-compatible file, giving the identities and EBVs of the parents, as well as the mid-parent EBV and coancestry of the crossed parents.

In the event that the swapping routine reaches the end of the ranked list without finding a parent that satisfies the coancestry constraint, XDesign will simply accept the cross and move to the next. This is most likely to occur when there is a great deal of relatedness among low-ranking parental candidates, an overly ambitious coancestry constraint or the program is trying to swap out a parent from the lowest-ranking cross. In any event, the user should review the design prepared by the program, including the coancestry calculated for each cross.

## Hardware/software requirements for XDesign

XDesign is written as a 64-bit program for PC-compatible computers, but will run on any computer under the Microsoft Windows Operating System from version “XP” to present, provided a 64-bit version is installed. A text editor is required to prepare input files and interpret some files created by XDesign. While not absolutely necessary, most users will also find it helpful to have access to an Excel-compatible spreadsheet program for editing of pedigree and breeding value data, converting these to text or comma-separated files for input to XDesign, and manipulation of files created when XDesign is run.

## Becoming a registered user of XDesign

XDesign and its supporting tools are available from Skogforsk at [www.skogforsk.se/xdesign](http://www.skogforsk.se/xdesign). A download link is provided to supply users with a zipped package of components and examples. While distribution of XDesign is not restricted, registration is strongly encouraged. Periodic updates, bug fixes and lists of known issues will be circulated to all registered users as they are released.

## A note on copyright, licensing and warranty

While copyright for XDesign is held by Skogforsk, distribution is without fee for research and non-commercial use under GNU General Public License Version 2.0. A copy of the GPLv2 license must accompany distribution of XDesign and work derived from it. In addition, there is ABSOLUTELY NO WARRANTY expressed or implied, and the user bears all the risk when using XDesign. In accepting the conditions of use on the opening screen, the user agrees to these conditions of use – please read these!

## Distribution of XDesign

XDesign will usually be distributed as a “zip” file. The following files are provided in the package:

- [XDesign.exe](#) – the executable binary for running XDesign.
- [GNUv2\\_license.txt](#) – GNU General Public License, Version 2, applies to XDesign.
- [Example1.csv](#) and [Example2.csv](#) – the example datasets discussed in this document, formatted as comma-separated values.
- [Example1.ctl](#) and [Example2.ctl](#) – control parameter files for running the examples from the command line in Command Prompt window.
- [XDesign\\_Userguide\\_v1.pdf](#) – this document.
- Release [Notes xx-xx-xxxx.pdf](#) – fixes and upgrades since publication of the Userguide.

## Installing XDesign

The following procedure is recommended for installing the program, although file locations could be changed, once successful operation of the program has been verified.

1. Extract the zip file to a convenient folder on your machine, such as `C:\XDesign\`. The “`xdesign.exe`” file may have had the “.exe” extensions stripped, to enable the sending of the zip file through institution firewalls; if so, you will need to restore the names by adding “.exe” to distributed file name. We recommend that all files in the package be placed in the folder where crossing design is to be carried out, i.e., the same file as the user-prepared input files.
2. Test your installation by working through the example data provided with the package and detailed in the next sections.

## Running the examples provided with XDesign

### THE CANDIDATE FILE

XDesign is distributed with example candidate files, formatted as “comma-separated values” (.csv). These describe the pedigree of selection candidates and their ancestors, using the “Me-Mum-Dad” standard that will be familiar to most users of pedigree analysis tools and databases, and appends extra fields to describe the genetic values and contributions to the selected mating population. For users of OPSEL (Mullin 2014, 2017), a compatible solution file is provided when optimizing “unequal” contributions, and this file can be used directly with XDesign without modification.

Open this file in a text editor or in Excel and note the following structure:

1. The first line of the file will be skipped, so can contain column headers to document the file.
2. Each of the subsequent lines describes an individual with its ID, those of its Female and Male parents, its estimated genetic value, and finally the desired frequency of contributions to the mating design, i.e., the number of times the individual is to be crossed.

The pedigree must respect the usual rule that parents precede their progeny in the file sequence, and that all ancestors must be listed. Unknown Female and Male IDs are indicated with a 0 (zero). All other pedigree IDs are alphanumeric labels up to 256 characters in length. As spaces, commas and tab characters are used to delimit fields, these characters are not permitted in any field. A small example of a candidate file is illustrated in Table 1.

The 4<sup>th</sup> field for each record contains the genetic value (EBV) of the individual as a real number. The values in this column are ignored when the frequency constraint specified in field 5 is zero.

The 5<sup>th</sup> field for each record gives the number of times the individual is to be used as a parent in the mating design, as a positive integer. The number of crosses that will appear in the design is the sum of this column, divided by two.

When preparing the input file, the user will normally use a spreadsheet program to organize and create the various fields. Simply save the spreadsheet as a comma-separated value (*csv*) or tab-delimited txt file before submitting to XDesign. If the solution file from an OPSEL run is used, the columns will already be in the required order and format. When submitted to XDesign, any additional columns will be ignored and not transferred to the output file.

## INTERACTIVE DIALOGUE OR COMMAND LINE

XDesign is a Windows program. It can be started by double-clicking on the program icon (*xdesign.exe*), which will take the user through an online dialogue to define the constraints, declare the path to the candidate file and initiate the design process. Alternatively, the user can run the program from the command-line in the Command Prompt window, declaring a control file as a parameter that defines the constraints and declares the location of the candidate file. The following sections describe the running of each example through the online dialogue (Example1) and via the command line (Example2).

## RUNNING EXAMPLE 1 FROM THE ONLINE DIALOGUE

In this example, we consider an optimum selection solution generated by OPSEL. The solution file is in *.csv* format and specifies a total of 200 contributions from candidates in the list, or 100 crosses. The candidate list presented to OPSEL contained more than 48,000 identities, from which the optimization selected 50 genotypes to contribute optimally between 1 and 13 times to the progeny of the 100 crosses.

The candidate file must conform to the format and sort order described earlier, and must contain the following 5 columns (other columns in this file are ignored and have been deleted from the OPSEL solution):

<i>Indiv</i>	<i>Female</i>	<i>Male</i>	<i>EBV</i>	<i>Frequency</i>
--------------	---------------	-------------	------------	------------------

The estimated breeding values (EBVs) would normally be extracted, together with the appropriate pedigree entries, from a third-party database, such as DATAPLAN<sup>®</sup>, prior to performing selection. The Frequency values must be positive integers.

The following is a detailed description of an XDesign job to prepare a positive-assortative mating design for this candidate list, while avoiding excessive coancestry between mates, illustrated by processing [Example1.csv](#):

1. Start XDesign by double-clicking on [xdesign.exe](#)
2. Accept the conditions of use on the splash screen by typing “Y”, then pressing Enter.
3. The description on the next screen is a summary of the information in this user guide on how to prepare the pedigree file. Note that the “root” name of the input file, in this case “[Example1](#)”, is used by XDesign when referencing all other files, using different extensions appended to the root.
4. Enter the name of the candidate file containing the pedigree and breeding-value data: [Example1.csv](#).
5. Pedigree verification and recoding is sometimes a lengthy process, but can be skipped IF you’re absolutely sure that the pedigree is valid AND the pedigree identities are sequential integers starting at 1. This is a rather special case, so normally we answer the question “Skip recoding and verification?” with “N”.
6. XDesign reports the current working directory.
7. You must now specify if the parents are to be sorted by EBV (positive assortative mating, answer “Y”) or randomly. We will answer “Y” for this example. (If you choose random assortment, you will be asked to provide a “seed” number for the random-number generator – the sort order is repeatable on subsequent runs, provided you use the same seed number.)
8. XDesign also needs to know if you wish to avoid a level of coancestry between parents. If you answer “Y”, you will be asked to enter a threshold level of coancestry; this is a real number between 0 and 1. For this example, you can enter 0, meaning that XDesign should to avoid related mates completely.
9. By default, XDesign will store the mating design as a tab-delimited text file, but you can specify comma-separated values ([.csv](#)) here, if you wish. Doing so will make it easier to read the solution with Excel, as [.csv](#) files will open directly.
10. XDesign proceeds with processing the candidate parents and their pedigree, and carries out the design. If successful, it will report back the number of selected genotypes and the name of the mating design file, in this case [Example1\\_design.csv](#).
11. Xdesign will ask you if you want to store your responses to the online dialogue so that you can run the problem again later from the command line; provide a filename, such as [Example1.ctl](#).

## REVIEWING THE DESIGN

After XDesign completes execution, the mating design can be found in the folder where you started XDesign: [Example1\\_design.csv](#) or [Example1\\_design.txt](#).

A partial example of a completed mating design for Example 1 is given in Table 2.

Table 2. A portion of the mating design for Example 1.

Cross	Parent1	Parent2	Parent1_R	Parent2_R	Parent1_EBV	Parent2_EBV	Avg_EBV	Coancestry
1	EHMGCD	EH0MEGB	1	6	440.54	353.71	397.125	0
2	EHMGCD	EH0HBDH	1	7	440.54	353.52	397.030	0
3	EHMGCD	EH0HACH	1	8	440.54	353.29	396.915	0
4	EHMGCD	EHEMDCM	1	9	440.54	348.31	394.425	0
5	EHMGCD	EH0MEGF	1	10	440.54	345.25	392.895	0
6	EHMGCD	EHFFFA	1	11	440.54	344.91	392.725	0
7	EHMGCD	EHFGMMA	1	13	440.54	341.30	390.920	0
8	EHMGCD	EH0HCBH	1	14	440.54	340.49	390.515	0
9	EHMGCD	EHHMEHB	1	15	440.54	337.81	389.175	0
10	EHMGCD	EHEHDCD	1	16	440.54	337.54	389.040	0
11	EHMGCD	FADMAAE	1	17	440.54	335.03	387.785	0
12	EHMGCD	EHMG0HC	1	18	440.54	334.13	387.335	0
13	EHMGCD	EMFOHBF	1	19	440.54	330.00	385.270	0
14	EHMAADH	EH0MEGB	2	6	416.48	353.71	385.095	0
15	EHMAADH	EH0HBDH	2	7	416.48	353.52	385.000	0
16	EHMAADH	EH0HACH	2	8	416.48	353.29	384.885	0
17	EHMAADH	EHEMDCM	2	9	416.48	348.31	382.395	0
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.

The columns in the file are:

**Cross** – a unique sequence number for each cross

**Parent1** – ID of one parent in the cross.

**Parent2** – ID of the second parent in the cross.

**Parent1\_R** – the sorting rank of Parent1.

**Parent2\_R** – the sorting rank of Parent2.

**Parent1\_EBV** – the EBV of Parent1.

**Parent2\_EBV** – the EBV of Parent2.

**Avg\_EBV** – the mid-parent EBV for the cross

**Coancestry** – the coancestry between the cross parents  
(expected inbreeding coefficient of the progeny)

Note that XDesign does not specify the gender of parents 1 and 2; the cross can be made in either direction.

It is important when using XDesign to review the design output. This is especially true when a threshold level of coancestry is specified, as there can be situations where XDesign cannot completely satisfy a coancestry threshold. There may be too many crosses required from certain genotypes, or there may simply be too much relatedness among the parents to be crossed.

The user should be aware that, in cases where XDesign cannot satisfy the threshold level of coancestry, no errors are generated, but the crosses with excess coancestry are normally found near the bottom of the list. It is even possible in extreme cases for some of these crosses to be selfs. The user can try a number of things to generate a more suitable crossing design:

1. If the sort order is random, try a different random seed number to change the parent sequence.
2. Raise the threshold level of coancestry between parents.
3. Make the contribution frequencies of parents more equal. If the contributions are optimized by OPSEL, this can be accomplished simply by setting a lower “maximum” contribution for each candidate.
4. Try some manual parent swaps between crosses. If only one or two crosses exceed the threshold coancestry, one can often scan the list and identify parent swaps to made manually.

## **RUNNING EXAMPLE 2 FROM THE COMMAND LINE**

XDesign can also be run from the command line by specifying the various job parameters in a separate text file. This can be very convenient if you want to experiment with different threshold levels of coancestry, different random number seeds, or if changes are made in the parental contribution frequencies. The simplest way to do this is to run through online dialogue in the previous section and save the parameters in a text file (see step 11 above). A sample version of [Example2.ctl](#) is provided in the XDesign distribution package and can be modified with any text editor. The [Example2.ctl](#) is illustrated below:

Example2.csv

```
1 : assortment of parents 0 = random, 1 = PAM
54321 : seed number for random number generator
0.075 : threshold level of parent coancestry 0 <= MaxC <= 1
1 : complete pedigree check = 1, else = 0
1 : output as txt = 0, output as csv = 1
```

Open a Command Prompt window -- usually found in the Accessories folder. Navigate to the location of [xdesign.exe](#) and the parameter file. Launch XDesign by entering `xdesign`, followed by the path to the parameter file, for example:

```
C:\xdesign>xdesign example2.ctl.
```

## **When things go wrong!**

The user should be aware that XDesign is in a continuing state of development. While it has already seen operational application, the limitations have not been thoroughly explored, as that will happen naturally as users attempt to apply it. Unexpected things can happen! XDesign may provide messages that give some information on failed runs or problems with data input. If you cannot rectify a problem, please note the messages received and send your input file and problem parameters (sort type, threshold parent coancestry, etc.) to: [tim.mullin@skogforsk.se](mailto:tim.mullin@skogforsk.se)

## **Some possibilities for future upgrades**

In its initial release, XDesign is applicable to breeding of monoecious species, where an individual can be used, and is available for use, as both seed parent or pollen parent. In operational application, even for monoecious species, male or female reproductive structures may not always be available on an individual genotype for crossing. A cross can only be made when at least one of the genotypes identified in the cross has pollen and the other has receptive seed structures.

We hope to address these limitations in a future upgrade, where a user will be able to declare the availability of female and male structures for each selected genotype, and the mating design adjusted as required to ensure that crosses are both biologically and operationally possible.

As noted in an earlier section, the initial release of XDesign also has limited capability to avoid a small number of crosses with exceptionally high relatedness, including selfs, at the bottom of the crossing list, where swapping opportunities are exhausted. We hope to provide alternative swapping approaches in future versions that reduce the occurrence of such related crosses, while still respecting the rank order of the selected genotypes.

## **Acknowledgements**

The beta testing of earlier versions by Drs. Torgny Persson and Johan Kroon (Skogforsk, Sävar) greatly helped to improve the program. Funding was provided by Föreningen Skogsträdsförädling (The Swedish Tree Breeding Association), the European Union Horizon 2020 Research and Innovation Program under grant agreement no. 676876 (Project GenTree), and Skogforsk (The Swedish Forestry Research Institute). All trademark terms are the property of their respective.

## References

- Cockerham, C.C. 1967. Group inbreeding and coancestry. *Genetics*, 56: 89-104.
- Lacy, R.C. 1995. Clarification of genetic terms and their use in the management of captive populations. *Zoo Biology* 14: 565-578.
- Lindgren, D., Gea, L., and Jefferson, P. 1996. Loss of genetic diversity monitored by status number. *Silvae Genet.* 45(1): 52-59.
- Mullin, T.J. 2014. OPSEL 1.0: A Computer Program for Optimal Selection in Forest Tree Breeding. Arbetsrapport från Skogforsk Nr 814-2014.
- Mullin, T.J. 2017. OPSEL 2.0: a computer program for optimal selection in tree breeding. Arbetsrapport från Skogforsk Nr 954-2017.
- Rosvall, O. 1999. Enhancing gain from long-term forest tree breeding while conserving genetic diversity. Swedish University of Agricultural Sciences, Umeå, Sweden. p. 65.
- Rosvall, O., Almqvist, C., Lindgren, D., and Mullin, T. 2011. Breeding strategies. In *Review of the Swedish Tree Breeding Programme*. Edited by O. Rosvall. Skogforsk, Uppsala, Sweden. pp. 55-66.
- White, T.L., Adams, W.T., and Neale, D.B. 2007. *Forest genetics*. CABI Publishing, Wallingford, Oxfordshire, UK.
- Yamashita, M., Mullin, T.J., and Safarina, S. 2017. An efficient second-order cone programming approach for optimal selection in tree breeding. *Optimization Letters* (in press).
- Zobel, B., and Talbert, J. 1984. *Applied Forest Tree Improvement*. John Wiley & Sons, New York, USA.

