

Pedigree Analysis Programme GTree 1.0

Daisuke Ogino ¹

ogino@matsys.org

Masato Nose ³

masanose@m.ehime-u.ac.jp

Shiro Mori ²

siromori@mail.cc.tohoku.ac.jp

Hideki Sawada ¹

sawada@matsys.org

¹ Department of Mathematical Sciences, Faculty of Science, Yamagata University, Yamagata 990-8560, Japan

² Tohoku University, Graduate School of Dentistry, Sendai 980-8575, Japan

³ Department of Pathology, School of Medicine, Ehime University, Ehime 791-0295, Japan

Keywords: pedigree analysis, algorithm, GTree

1 Introduction

GTree 1.0 is a C-language programme on LINUX which analyze pedigrees. In this version we have developed 18 commands for finding out characteristics of a given strain data by going up to forefathers from a descendant.

We explain the algorithm of the programme in Section 2, then we discuss its computational complexity in Section 3. In Section 4 you will find the list of the commands.

2 Algorithm

In this section we should like to review the outline of the algorithm of our programme GTree according to [2]. The input data is a matrix. The column of matrix is fixed but the row of the matrix is allowed to add a new data. That is, we represent the each mouse's or object's informations as a row. A column of matrix consists of ten elements (identity number, breeding diary + or −, their father's number, their mother's number, sex, generation, disease + or −, statistic of sample1, statistic of sample2 and date of birth). Table 1 is a row of the input data. Then the formats of output data are pedigrees or statistical diagrams. Genealogical Analysis Programme, GTree is written by C and is able to search pedigree (Fig. 1) and contribute to analyze genetic statistics.

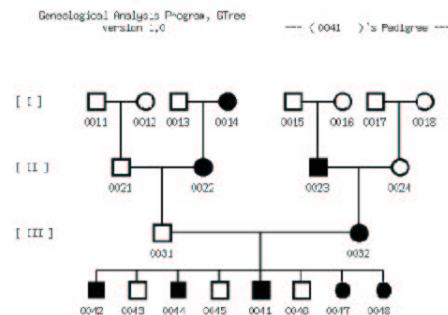


Figure 1: Pedigree.

3 Computational Complexity

Now we should like to explain the computational complexity on the analysis of pedigrees. For a general reference of the terminologies in population genetics see [1, 3].

Table 1: Data form of GTree.

identity number	breeding diary + or -	father's number	mother's number	sex	genera- tion	disease + or -	statistic of sample1	statistic of sample2	date of birth
--------------------	--------------------------	--------------------	--------------------	-----	-----------------	-------------------	-------------------------	-------------------------	------------------

Let n be the number of total generations of a given pedigree. Without loss of generality we can assume that $n \geq 2$. First we show upper bounds of the total number of forefathers, according to the following 3 cases.

- Case 1. Random mating: In this case the i -th generation has 2^{n-i} forefathers, where $1 \leq i \leq n - 1$. Then the sum of forefathers from the first to $n - 1$ th generation is $2(2^{n-1} - 1)$.
- Case 2. Brother-sister mating: Since any i -th generation ($1 \leq i \leq n - 1$) has 2 forefathers, the sum of all forefather from the first to $n - 1$ th generation is $2(n - 1)$.
- Case 3. Selfing: This happens to plants. Since each generation has just one forefather, we have $n - 1$ forefathers at the n -th generation.

From this classification we can conclude that the number of forefathers of the n -th generation is between $2(n - 1)$ and $2(2^{n-1} - 1)$, and the complexity analyzing the pedigree is from $O(n)$ to $O(2^n)$. Since congenic mouse strains are made up through inbreeding, the number of forefathers is always less than $2(2^{n-1} - 1)$ in practice.

4 Commands of GTree 1.0

In this section we should like to list the commands of GTree 1.0. We use the standard symbols of pedigree (see, e.g., [1, 3]).

Command	Classification	Output
all, id	CUI	The data matrix, Individual data
fchld, mchld	CUI	Children of father, Children of mother
fmly, pts	CUI	The pedigree, Parents
fr, mr	CUI	Father's side, Mother's side
gwl, gws	CUI	Statistic of sample1 or sample2 for each generation
wl, ws	CUI	Statistic of sample1 or sample2 for the total generations
xchld, xfmlly	GUI	children, pedigree
xgwls	GUI	Scatter diagram statistics of sample1, 2 for each generation
xwl, xws	GUI	Bar graph statistic of sample1 or sample2 for the total generations
xwls	GUI	Scatter diagram statistics of samples for each generation

In our recent paper [2] sample1 is weight of axillary lymphnds and sample2 is weight of spleen.

References

- [1] Crow, J.F., *Genetic Notes 8th edition*, Burgess Publishing Company, 1983.
- [2] Ogino, D., Mori, S., Nose, M., and Sawada, H., Complexity and application of pedigree analysis programme GTree, *Proc. IEEE Computer Society Bioinformatics Conference*, 333–335, 2002.
- [3] Ooba, S., *Population genetios (in Japanese)*, Tokyo University Press, 1997.