American Journal of Laboratory Medicine
2023; 8(1): 4-12
http://www.sciencepublishinggroup.com/j/ajlm
doi: 10.11648/j.ajlm.20230801.12
ISSN: 2575-3878 (Print); ISSN: 2575-386X (Online)

On the Basics of Pedigree Visualization and Feature Extraction for the Autosomal Recessive Inheritance Pattern

Ummay Tania Akter, Sajjad Bin Sogir, Tapati Basak*

Department of Statistics, Jahangirnagar University, Dhaka, Bangladesh

Email address:
tapati.basak@juniv.edu (Tapati Basak)

*Corresponding author

To cite this article:
Ummay Tania Akter, Sajjad Bin Sogir, Tapati Basak. On the Basics of Pedigree Visualization and Feature Extraction for the Autosomal Recessive Inheritance Pattern. *American Journal of Laboratory Medicine.* Vol. 8, No. 1, 2023, pp. 4-12. doi: 10.11648/j.ajlm.20230801.12

Received: May 11, 2023; **Accepted**: May 31, 2023; **Published**: June 9, 2023

Abstract: Pedigree analysis is carried out to understand the mode of inheritance of a particular disease, which includes recessive, dominant, partial dominant, autosomal, mitochondrial or sex-linked etc. It also determines the individual's probability of affecting in a given cross. Genetic disorders are transmitted from one generation to the next by following a particular inheritance pattern. Correct investigation of the transmission patterns of a trait under different circumstances is the crucial in genetic research. That is, identification of ancestry patterns for twins, full penetrance and reduced penetrance cases etc. This paper considers autosomal recessive case. Two simulated single nucleotide polymorphisms (SNPs) based genotype data sets for 14 and 47 individuals with three and four generations, respectively, were applied for this investigation. This evaluation looks for the probable features of ancestry patterns of a genetic disorder from one generation to the next based on the specified genetic conditions. Proper visualization of the pedigree charts for autosomal recessive case having different characteristics were demonstrated here. Since, sequencing of deoxyribonucleic acid (DNA), and handling of such massive amount of data depends on the availability of funding, dedicated software, high throughput data storage capacity etc. Hence, effective simulation for data generation would be the choice to cope with this situation for realizing the pipeline of such genetic research. The main objective of this paper is to provide a useful guideline for the introductory genetic researchers to whom real data sets are not available, and once available, dealing with this massive amount of sequencing data is a big challenge due to some limitations. This guideline will help to have an idea about such research. If opportunity is given, this idea could be applied for the real data sets, and the results would be similar.

Keywords: Pedigree, Inheritance, Autosomal Recessive, SNP, DNA

1. Introduction

The pattern of inheritance of a particular trait throughout a family can be determined by the ancestry analysis for which the pedigree chart is the visual method. These charts show the presence or absence of a particular trait as it relates to the relationship among parents, offspring, and siblings. Man has long noticed that families often have similar characteristics. Before Mendel's theory, the understanding of genetics was quite limited. Although, humans had been selectively breeding plants and animals for thousands of years, the scientific basis for this process was poorly understood. One of the earliest theories of heredity was proposed by the Greek philosopher Aristotle in the 4th century BC. Aristotle believed that the traits of offspring were determined by a blending of the traits of their parents. For example, he thought that if a tall and a short person had a child, the child would be of average height [1, 2].

This idea of blending inheritance persisted for centuries, and it was not until the 19th century that scientists began to question it. The development of the theories underlying evolution is where genetics got its start. Few biologists (not the biometricians) during this time, advanced blending from a description of breeding results to a blending process. According to Robert Olby, the most regrettable underlying premise of Darwin's theory of evolution was the mixture of heredity and environment. That is, bisexual reproduction blends parental differences, resulting in offspring with less variance over time [3]. After that, Mendel's work (1822-1884)

laid the foundation for the understanding of inheritance patterns and the principles of genetics [4].

Mendel's idea of heredity is predicated on equality and consistency across the whole life cycle. Thus, the contemporary synthesis of evolutionary biology was created through the fusion of Mendelian genetics with Darwin's theory of natural selection [5-7]. Genetic association is a fundamental concept in genetics and genomics, as it allows researchers to identify genetic factors that contribute to the risk of developing certain diseases or traits. Such association studies typically involve comparing the frequency of genetic variations, such as single nucleotide polymorphisms (SNPs), between individuals with a particular disease or trait and those without. By identifying SNPs that are more common in individuals with the disease or trait of interest, researchers can determine which genetic variations are associated with the condition [8, 9].

In recent years, advancements in technology have allowed scientists to study the human genome in greater detail, leading to a better understanding of the role of genes and chromosomes in health and disease. Human chromosomes are the structures that carry genetic information in the form of deoxyribonucleic acid (DNA) in human cells [10-12].

Inheritance is the process by which an individual inherits traits or characteristics from their parents through the passing down of genes. Genes are segments of DNA that contain the instructions for the development and functioning of an organism. The pattern of inheritance of a particular trait can be determined by the type of gene involved. There are two main types of genes: dominant and recessive. There are several types of inheritance patterns, including autosomal dominant, autosomal recessive, X-linked dominant, X-linked recessive, and mitochondrial inheritance etc. [2, 13]. According to one of these inheritance patterns, genetic diseases like Cystic fibrosis [14], hemophilia [15], Pseudo-trisomy 13 syndrome [16], Emery-Dreifuss syndrome [17], Sickle cell disease [18], Huntington's disease [19], Fraser-Cryptophthalmos Syndrome [20] and so on have been discovered to be passed down from one generation to the next.

A technique for examining how genetic characteristics are passed down through families is pedigree analysis. It entails creating a chart that depicts the connections between family members and their genetic traits [21]. To find genetic variations linked to particular characteristics or diseases, pedigree analysis can be used with data from genome sequencing. Researchers can find variations that are passed down through families in a particular manner by comparing the genome sequences of different family members [22].

DNA sequencing SNP genotype data are used in the analysis of pedigree as sequencing information from the human genome improves molecular diagnostic yield as compared to current diagnostic checking [23]. This kind of sequencing generally produces massive amount of the genotype with the relevant information about family history. Handling and analyzing of these DNA sequencing data become a great challenge, where the researcher has no funding.

Despite of these limitations, the researchers can still move with the simulated data that can be generated using some open source software, and the volume of the data set can be easily kept in that size, where the availability of a researcher can go. These simulated data preserves the different features of the real genome. This paper provides a practical guideline and some basic preliminary septs of analyzing genetic inheritance pattern for autosomal recessive case by considering several conditions like full and reduced penetrance, twins etc.

2. Materials and Methods

2.1. Autosomal Recessive Inheritance Pattern

There are basic two step analysis to define the ancestry pattern of an individual. One is to determine whether the trait is dominant or recessive, and the other is whether the chart shows an autosomal or sex-linked trait. When the trait is recessive, neither parent is required to have the trait since they can be heterozygous. In autosomal case, both males and females are equally likely to be affected (usually in equal proportions). This paper considers the patterns for the autosomal recessive inheritance [2].

A mutation in a gene on one of the first 22 non sex chromosomes can lead to an autosomal disorder. Genes come in pairs. One gene in each pair comes from the mother, and the other gene comes from the father. Recessive inheritance means both genes in a pair must be abnormal to cause disease. People with only one defective gene in the pair are called carriers. These people are most often not affected with the condition. However, they can pass the abnormal gene to their children [24]. For a child born to a couple who both carry the gene (but do not have signs of disease), the expected outcome for each pregnancy is:

A 25% possibility exists that the child will be born with two healthy genes. The likelihood that the kid will be born with one normal gene and one abnormal gene is 50% (carrier without disease). The likelihood of the kid being born with two abnormal genes is 25% (at risk for the disease) [2].

Specifically, an autosomal recessive genetic condition occurs when one variant is present on both alleles (copies) of a given gene. The important characteristics of this pattern include: disease can skip generations of the family, the parents of a person with the disease usually don’t carry it [2].

In this pattern, variants occur in both copies in each cell. Each of the parents of an individual with an autosomal recessive disorder carries one copy of the altered gene, but they typically don’t show signs and symptoms of the condition. If both copies of the gene have the same deleterious mutation, the defect is termed homozygous, and the defect is termed compound heterozygous if each of them has a different deleterious mutation. Each parent of an affected patient is typically a heterozygous carrier, and has one normal and one abnormal copy of the gene. In most cases, a normal copy of the gene can compensate for the defective copy; thus, heterozygous carriers are generally asymptomatic. When two carrier parents have offspring, one in four

offspring should have the disease from a statistical view point. That is, two should be carriers, and one should be the normal. Autosomal recessive disorders occur with increased frequency in offspring of consanguineous marriages or in isolated populations, where an original founder mutation that occurs in one individual at some point in history is subsequently propagated throughout the population.

The autosomal recessive pattern can be modelled mathematically for genetic association analysis of the SNP genotype data in a case-control setting. Every model has different predefined assumptions about the genetic effect in the data. For example, there are three genotype counts, A/A, A/B and B/B for a single SNP having two alleles, "A" (disease allele) and "B" (normal allele), respectively. Then the recessive model (for "A" allele) assumes that two copies of the "A" allele are required to alter the risk. So, the individuals having A/A genotype are compared to individuals having genotypes B/B, and A/B [25-28]. Also, the penetrance functions can be deployed to represent the modeling for the relationship between SNPs and risk of disease [28-31].

2.2. Preparation of Pedigree (PED) Data File

The PLINK format (https://zzz.bwh.harvard.edu/plink/data.shtml#ped) PED file was generated by simulation in the R programming language. The PED file consists of several columns, where the first six columns are mandatory. It is a white-space (space or tab) delimited file, whose columns have the information for the Family ID, Individual ID, Paternal ID, Maternal ID, Sex (1 = male; 2 = female; other = unknown) and Phenotype.

The combination of family ID and individual ID should uniquely identify an individual, where all the IDs are alphanumeric. The sixth column of a PED file must have only one phenotype, which can be either a quantitative trait or a column having affection status. Some dedicated software like PLINK will automatically identify which type based on the predefined coding for Genome-Wide Association Study (GWAS).

Two different PED files were generated under the autosomal recessive model of inheritance to compare the characteristics of recessive disorders under different conditions. These conditions include full and reduced penetrance and twins etc. The first PED file (PED1) contains the ancestry information of three generations of a certain family for the 14 individuals. The family history of another family over the four generations are contained in the second PED file (PED2) having 47 individuals. Among the 14 persons, there are 6 male and 8 female (for PED1), whereas, 23 male and 22 female are there among 47 persons (for PED2), respectively. In PED1, 2 persons are affected, 6 are the carriers, and the rest of them are unaffected individuals. There are 13 affected persons, 22 carriers, and the remaining individuals are unaffected for the PED2. All the computations were carried out using R programming language.

3. Ancestry Analysis Using the Pedigree Charts

3.1. Description of Pedigree (PED) Files

In pedigree analysis, researchers look for any clues that will allow them to decide if the trait is dominant or recessive, and whether it is linked to an autosomal chromosome, or other. This type of analysis is interesting as the findings from here can be used to do some detective work and are often used to study the genetics of inherited diseases. For example, pedigrees can be analyzed to determine the mode of transmission of a genetic disease. From graphical observation and proof of identity of passing features are done with the PED file. For a simulated data generation as well as for analyzing the family history by pedigree even through the dedicated software, the understanding of the PED file is necessary. Because, it is a representation of a family tree showing how individuals within a family are related to each other. Hence, understanding of PED file makes clear realization of such relations (links) of individuals within a family, also, the symbols used here in the PED file to represent the history. An overview of the PED file (PED1) with 14 individuals as outlined in Section 2 are as below (Table 1).

This complete data of Table 1 for small sample (14 individuals) was used for constructing the pedigree chart. The "Family ID" in first column presents the identification number of a family or number of family to be considered. For example, "1" indicates this data contains the information a single family. In second column, contains a list of each person's identification number labelled by the "Individual ID". The father and mother's IDs are located in the third and fourth columns under the labels of "Paternal ID" and "Maternal ID", respectively. Gender is indicated in the fifth column by the label of "Sex", where, "1" for males and "2" for females. The affected status of an individual is listed in sixth column, where, "1" and "0" are presenting the affected (Case) and unaffected (Control) status of an individual, respectively. Two additional or optional information are added to the seventh and eighth column, respectively, to the data of Table 1. The columns named "Carrier" and "Avail" are presenting the information regarding the availability and having one defective gene in the pair of each individual ("1" for carrier, "0" for not carrier). Similar description is sufficient for the second PED file (PED2) of 45 individuals for the four generations.

Table 1. *The PED File for 14 Individuals with Full Penetrance.*

Family ID	Individual ID	Paternal ID	Maternal ID	Sex	Phenotype	Carrier	Avail
1	101	0	0	1	1	1	0
1	102	0	0	2	1	1	0
1	103	0	0	1	0	0	0
1	104	101	102	1	1	0	0

Family ID	Individual ID	Paternal ID	Maternal ID	Sex	Phenotype	Carrier	Avail
1	105	101	102	2	0	0	0
1	106	101	102	2	1	0	1
1	107	101	102	1	1	1	0
1	108	0	0	2	0	0	1
1	109	101	102	2	1	1	1
1	110	103	106	1	1	1	0
1	111	103	106	2	0	0	0
1	112	107	108	2	0	0	0
1	113	107	108	1	0	0	1
1	114	107	108	2	1	1	0

3.2. Visualization of Pedigree

3.2.1. For Full and Reduced Penetrance

There are standard ways to draw pedigrees so that one can look at a pedigree and understand it. For example, squares are used to represent males and circles to represent females. The generations can be represented by roman numerals, so the top generation would be generation one, or roman numerals (i). Marriages can be indicated between individuals with a horizontal line connecting the two individuals. If an individual has a genetic trait, those individuals should be blackened in or shaded, so that it would be understood that they have a particular trait etc.

The genetic condition penetrance is one feature that should be considered for the visualization of a pedigree. It is a measurement of the relationship between a genotype and phenotype that refers to the likelihood that a clinical condition will occur when a particular genotype is present. That is, it measures the proportion of individuals in a population who carry a specific gene and express the related trait.

The penetrance of expression may also change in different age groups of a population. The condition can be in the form of full or reduced under different unknown circumstances viz. a combination of genetic, environmental, and lifestyle factors. A form of penetrance where a deleterious or disease-causing gene is expressed in all individuals carrying the gene is the condition of full or complete penetrance.

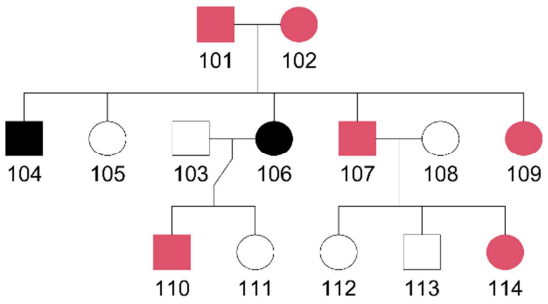


Figure 1. *The Autosomal Recessive Inheritance Pattern with Full Penetrance of PED1. Both genders are affected at similar frequency; disease is not visible in every generation; both the parents of an affected individual are the carriers; the square shape symbol presents a male whereas the circle is for a female; the black color and red color in any of the square or circle indicates the affected and the carrier status of an individual.*

Under this condition, a disease-causing gene shows 100% or complete penetrance if all individuals who have this gene develop the associated trait. For example, Huntington's disease is a dementia that is genetically inherited as an autosomal-dominant trait with a complete lifetime penetrance.

Figure 1 is presenting the pedigree chart with complete penetrance obtained from the first PED file (PED1) having 14 individuals (Table 1).

In Figure 1, the persons having Individual ID numbers 101 and 102, respectively, are the two members of the first generation, where, 101 is male and 102 is female, and both of them are carriers. These two people with IDs 101 and 102, respectively, have zeros for their Paternal ID and Maternal ID columns in Table 1 as they are not known. The second generation consists of 5 persons with IDs 104, 105, 106, 107 and 109. They are siblings and their Paternal ID and Maternal IDs are 101 and 102, respectively. Observing the disease status of the second generation, the affected individuals are 104 and 106, and, the persons with IDs 107 and 109 are the carriers. The person having Individual ID 105 is the only unaffected person in the second generation. From the Figure 1, it was also observed that the disease were exposed in both genders with same frequencies (half of the offspring and half of them from each gender are affected). The two people with Individual IDs 106 and 107 mate with the individuals 103 and 108, respectively, in the second generation. In the third generation, there are also 5 individuals, where, two individuals (110, 111) are the children of the persons 103 (Not affected or Healthy) and 106 (Affected). As the mode of inheritance is autosomal recessive, both of the parents should be the carriers to expose the disease in the next generation. Hence, none of the children (110, 111) are affected in the third generation, but one is the carrier (Individual ID: 110). On the other hand, the persons with IDs 112, 113, 114 are siblings and their parents IDs are 107 and 108, respectively. As the father (107: Carrier) and the mother (108: Not affected), so, none of their children exposed with the disease as per the predefined characteristics of autosomal recessive inheritance pattern. But, one of their daughter (ID: 114) is the carrier of the disease (Figure 1). This chart also shows that the disease is not appearing in every generation. That is, among the three generations, the disease or phenotype is exposed in the second generation (Figure 1).

Reduced (or incomplete) penetrance occurs when some individuals with the variant do not exhibit symptoms of the disorder. Reduced penetrance is likely caused by a number of unknown genetic, environmental, and behavioral variables.

This condition were demonstrated with a simple modification of the first PED file (PED1) having 14 individuals. The modification of Table 1 is shown in Table 2.

The features of reduced penetrance are portrait in Figure 2. The features are similar as in Figure 1, having full penetrance except for the phenotype exposure of the person with ID 104 in the second generation. According to the defined characteristics autosomal recessive inheritance pattern, this person should be exposed with disease as both of his father and mother are carriers. Hence, this situation is considered as autosomal recessive ancestry pattern with reduced penetrance (Figure 2).

Table 2. *The Modification of the First PED File (PED1) Under the Condition of Reduced Penetrance.*

Family ID	Individual ID	Paternal ID	Maternal ID	Sex	Phenotype	Carrier	Avail
1	104	101	102	1	1	0	0
	Modified to the unaffected Phenotype status in the sixth column as below				⇩		
1	104	101	102	1	0	0	0

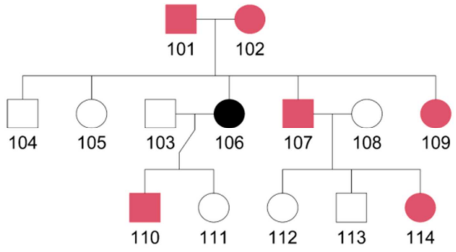


Figure 2. *Autosomal Recessive Inheritance Pattern with Reduced Penetrance of PED1. Due to Reduced Penetrance, the Characteristics are not Visible in both Genders at a Similar Frequency in the Second Generation.*

The difference between autosomal recessive inheritance patterns with full penetrance (Figure 1) and those with reduced penetrance (Figure 2) are as follows. The first pedigree chart (Figure 1) contains two affected individuals (104, 106). On the other hand, the second one (Figure 2) contains only one affected individual (106). Such reduction in penetrance may occur when all individuals carrying the mutation do not develop the associated phenotype. The reasons include the environmental factors, genetic modifiers, epigenetic factors and stochastic or random events etc. The environmental factors, such as exposure to toxins, may increase or decrease the likelihood of expressing a particular phenotype in individuals with a genetic mutation. The genetic modifiers are the genes that can interact with the gene carrying the mutation and modify its effects. A modifier gene may enhance or suppress the expression of the phenotype associated with the mutation. The epigenetic modifications, such as DNA methylation or histone modifications, can affect gene expression and may contribute to reduced penetrance. Even in genetically identical individuals, there can be differences in gene expression or other factors that can affect the likelihood of developing a particular phenotype that are the results of stochastic or random phenomena.

The features of autosomal recessive inheritance pattern were also extracted for the second PED file (PED2) in the similar ways. The PED2 is completely a different data set for four generations of another family with 47 individuals. For this family history, the characteristics were observed under the conditions of full and reduced penetrance.

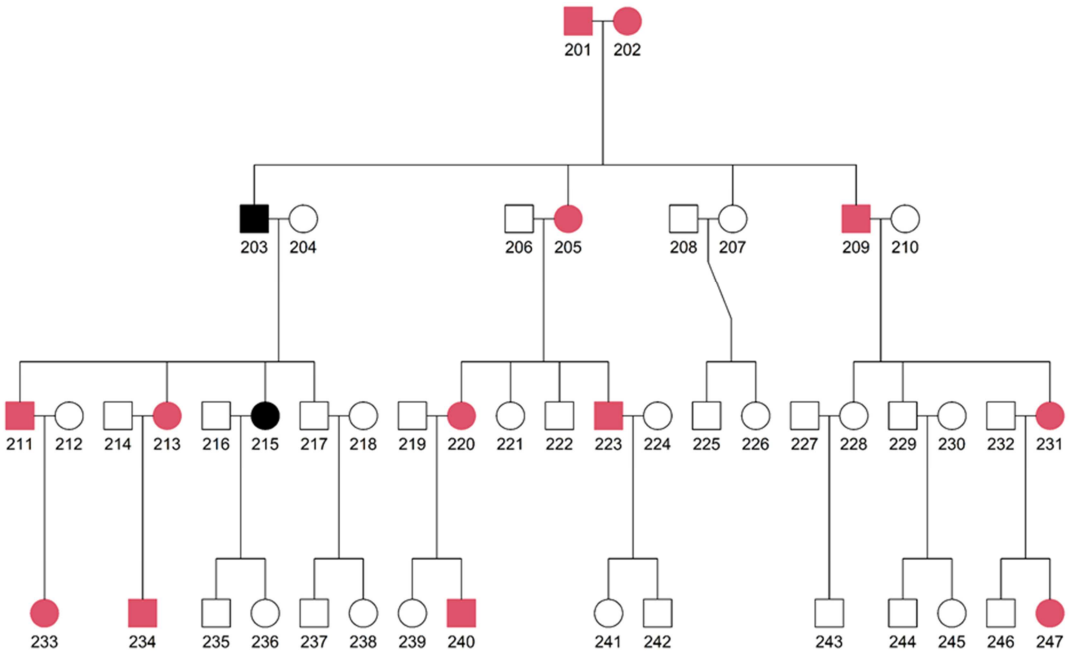


Figure 3. *Autosomal Recessive Inheritance Pattern with full Penetrance of PED2 for 47 Individuals.*

Figure 3 shows the transmission of a disease from the first to the fourth generations for the family history contained in PED2. The first generation starts with a disease carrier couple having Individual IDs 201 and 202, respectively. There are four siblings in the second generation (203, 205, 207, and 209), where one (203) of them is affected, two (205, 209) of them are carriers, one (207) is the only healthy person. These four siblings mate with four unaffected persons (204, 206, 208, and 210). Thus the second generation shows four individual families, where the first family (Father ID: 203, Mother ID: 204) consists of eight individuals, the second family (Father ID: 206, Mother ID: 205) consists of six individuals, the third family (Father ID: 208, Mother ID: 207) consists of two individuals and finally the fourth family (Father ID: 209, Mother ID: 210) consists of six individuals in their generation. In the first family, four individuals (211, 213, 215, and 217) are the siblings, and they mate with individuals 212, 214, 216 and 218, respectively, in third generation. The individuals having IDs: 211 and 213 are the carriers, 215 is affected and 218 is unaffected. In the second family, Individual IDs range from 219 to 224, two of them (220, 223) are carrier and the rest are the healthy persons. The third family contains two unaffected individuals (206, 207) and consequently their offspring (225, 226) are unaffected in the third generation. In the fourth family, there are six individuals (227, 228, 229, 230, 231, and 232), one of them (231) is carrier, and the rest are the unaffected individuals (Figure 3).

The final and fourth generation consists of offspring from the mating of third generation. The individuals 233 and 234 are the carriers as one of their parent are the carriers (211, 213). Individuals (235, 236) are the healthy persons despite of their mother (215) is affected. This is because, both of them (235, 236) are the heterozygous. The siblings 237 and 238 are unaffected as both of their parents are unaffected. Between two children (239 and 240), 240 is the carrier as his mother (220) is the carrier (Figure 3). Similarly, the features can be identified for other individuals of this generation.

For the second PED file (PED2), the genetic condition reduced penetrance is outlined in Figure 4. The ancestry pattern is similar to that of Figure 3, except for the affected status of the person having Individual ID 215 in the third generation (Figure 4).

Both the Figures 3 and 4 are showing the features of transmission of a particular disease from the first generation to the fourth by following autosomal recessive pattern. The features of the full penetrance are in Figure 3, and for the reduced penetrance are in Figure 4, respectively. For the condition of full penetrance (Figure 3), there were two affected individuals (203 and 215) in second and third generation, respectively. But, the disease skips the third generation because of reduced penetrance condition as shown in Figure 4.

3.2.2. For Special Relations: Twins

Twins are babies who are born at the same time, to the same mother. There are two main types of twins: monozygotic and dizygotic. The monozygotic twins share the same genetic material, and are always of the same sex. On the other hand, dizygotic twins can be of the same or opposite sex, and they share approximately 50% of their genetic material. This study also extracted the features for the family having twins. For the autosomal recessive case, this demonstration was carried out for the family history contained in the second PED file (PED2) by creating this special relations of twins. Figure 5 is presenting the ancestry chart of such special relation (twins). The fourth generation of this figure there are two pairs of twins (235, 236) and (246, 247), who are the dizygotic twins. As they are dizygotic, the disease transmission occurs as normal since they do not have similar DNA sequence.

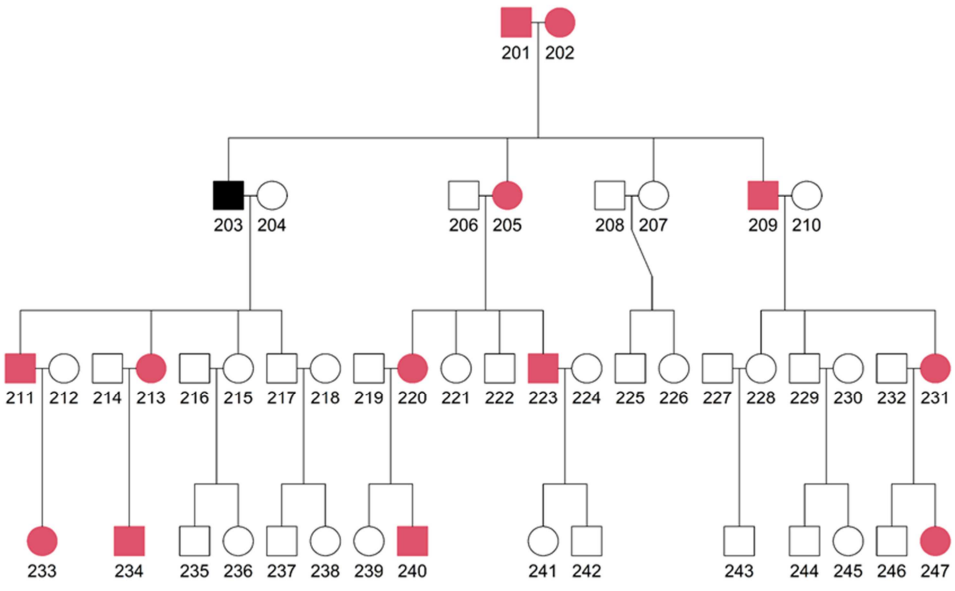


Figure 4. *Autosomal Recessive Inheritance Pattern with Reduced Penetrance in the Third Generation of PED2 for 47 Individuals.*

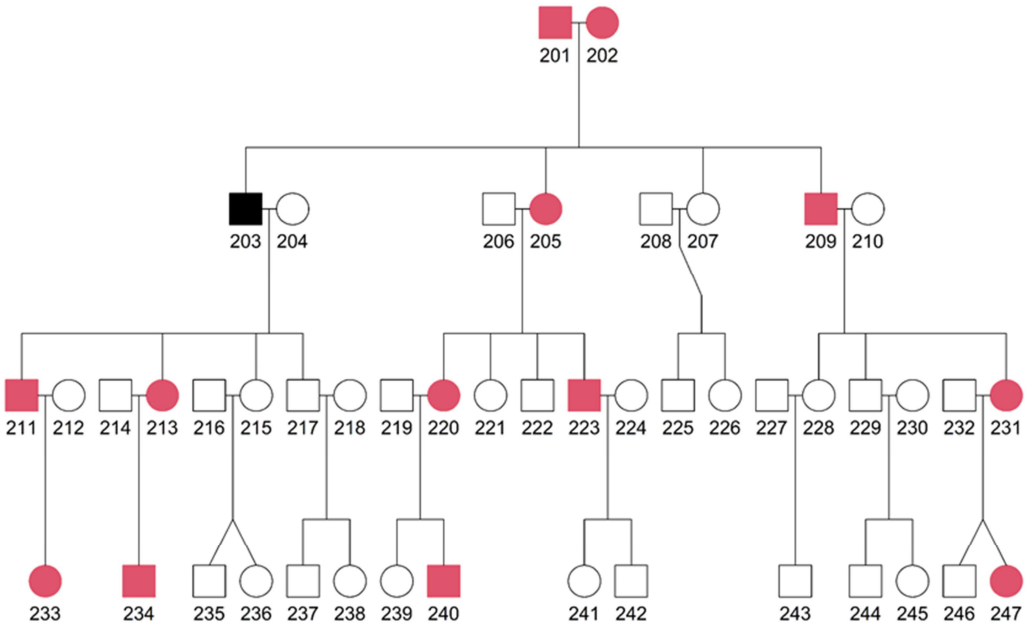
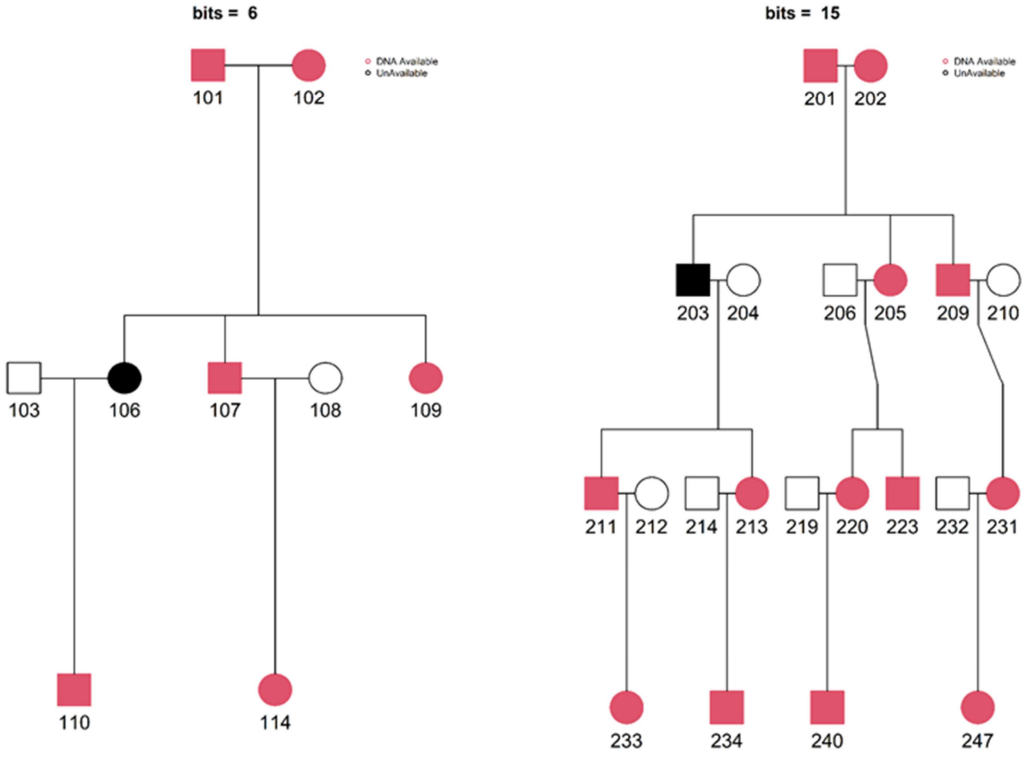


Figure 5. *Autosomal Recessive Inheritance Pattern for the Special Relation (twins) Using PED2 for 47 Individuals.*



(a) (b)

Figure 6. *Autosomal Recessive Inheritance Pattern After Trimming and Shrinking. (a) for PED1 of 14 Individuals, (b) for PED2 of 47 Individuals.*

3.2.3. Extract Only Important Features: Trimming and Shrinking

Shrinking the pedigree to a useful size, where all the important information is retained is convenient to visualize, compare and interpret. That is, sorting out the unnecessary information and trimming them out by a scientific way can be the crucial of pedigree analysis. Now a days, software-based algorithms simplify this task by transforming the size of the chart to a specified bit size with priority placed on trimming less informative subjects. The existing methods remove the less informative subjects iteratively from the existing pedigree by random selection. This process is controlled by a seed argument. Here, less informative subjects means not available (not genotyped) with no available descendants. Hence, gradually the existing pedigree shrinks to an informative one in terms of size and

information. This provides a useful way for assessing the inheritance pattern of a family [32].

The trimming and shrining were applied to the Figure 1 and 4, respectively. The trimmed versions are shown in Figure 6. Here, the left one (Figure 6a) is for the first family (PED1) and the right one (Figure 6b) represents for the second family (PED2). Though, they are the trimmed version, the autosomal recessive pattern is very clearly visible in both charts.

4. Conclusions

High throughput technologies and methods are continuously developing in the field of genetics. The prior knowledge about the inheritance pattern is crucial in genetic association studies for setting an appropriate model. Sometimes, access to the real data repositories, and handling it with limited resources like funding, server space to locate these massive amount of data etc. is a very big challenge of the researchers. The feature based computer simulation could be an aid to cope with this challenges. Hence, the demonstration of the features of the autosomal recessive ancestry pattern using simulated data would be an initial pipeline for the introductory genetic researchers to help them to adopt with such kind of analysis. The methodologies will be the similar when the real data is applied. Further investigation can be carried out for calculating relative disease risk of a person through Genotypic Relative Risk (GRR) for any type of inheritance pattern. Using GRR, a comparative study among all modes of ancestry patterns (autosomal dominant and recessive, sex-linked etc.) would be an informative approach for extracting features of such conditions.

Limitations

Due to several limitations like insufficient funding, inadequate laboratory support for sequencing real DNA, limited space to handle massive amount of data etc., this study used simulated genotype data.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

[1] Vinci T, Robert JS (2005) Aristotle and modern genetics. Journal of the History of Ideas 66 (2): 201–221. https://doi.org/10.1353/jhi.2005.0041

[2] Ziegler A, König IR (2010) A Statistical Approach to Genetic Epidemiology: Concepts and Applications. Weinheim: Wiley-VCH, Germany. https://www.10.1002/9783527633654

[3] Porter TM (2014) The curious case of blending inheritance. Studies in History and Philosophy of Science Part C : Studies in History and Philosophy of Biological and Biomedical Sciences 46 (1): 125–132. https://doi.org/10.1016/j.shpsc.2014.02.003

[4] Abbott S, Fairbanks DJ (2016) Experiments on Plant Hybrids by Gregor Mendel. Genetics 204 (2): 4.7-422. https://doi.org/10.1534/genetics.116.195198

[5] Galla SJ, Brown L, Couch-Lewis Y, Cubrinovska I, Eason D, Gooley RM et al. (2021) The relevance of pedigrees in the conservation genomics era. Molecular Biology 31 (1): 41-54. https://doi.org/10.1111/mec.16192

[6] Wijsman EM (2012) The role of large pedigrees in an era of high-throughput sequencing Hum Genet. 131 (10): 1555–1563. https://doi.org/10.1007/s00439-012-1190-2

[7] Smýkal PK, Varshney RK, Singh V, Coyne CJ, Domoney C, Kejnovský E, Warkentin T (2016) From Mendel's discovery on pea to today's plant genetics and breeding. Theoretical and Applied Genetics 129 (12): 2267–2280. https://doi.org/10.1007/s00122-016-2803-2

[8] Timm J, Oberste N, Schmiemann P (2022) Which factors influence the success in pedigree analysis? International Journal of Science Education. https://doi.org/10.1080/09500693.2022.2155494

[9] Middeldorp CM, de Moor MHM, McGrath LM, Gordon SD, Blackwood DH, Costa PT et al. (2011) The genetic association between personality and major depression or bipolar disorder. A polygenic score analysis using genome-wide association data. Translational Psychiatry, 1 (e50). https://doi.org/10.1038/tp.2011.45

[10] Singh R, Sophiarani Y (2020) A report on DNA sequence determinants in gene expression. Bioinformation, 16 (5), 422. https://doi.org/10.6026/97320630016422

[11] Waller RG, Darlington TM, Wei X, Madsen MJ, Thomas A, Curtin K et al. (2018) Novel pedigree analysis implicates DNA repair and chromatin remodeling in multiple myeloma risk. PLoS Genet 14 (2): e1007111. https://doi.org/10.1371/journal.pgen.1007111

[12] Xia C, Amador C, Huffman J, Trochet H, Campbell A, Porteous, D et al. (2016) Pedigree- and SNP-Associated Genetics and Recent Environment are the Major Contributors to Anthropometric and Cardiometabolic Trait Variation PLoS Genet 12 (2): e1005804. https://doi.org/10.1371/journal.pgen.1005804

[13] Xu J (2005) The inheritance of organelle genes and genomes: Patterns and mechanisms. Genome 48 (6): 951–958. https://doi.org/10.1139/g05-082

[14] Rafeeq MM, Murad HAS (2017) Cystic fibrosis: Current therapeutic targets and future approaches. Journal of Translational Medicine 15 (84). https://doi.org/10.1186/s12967-017-1193-9

[15] Miller CH, Soucie MJ, Byams VR, Payne AB, Sidonio RF (Jr.), Buckner TW, Bean CJ (2021) Women and girls with haemophilia receiving care at specialized haemophilia treatment centres in the United States. Haemophilia 27 (6): 1037-1044. https://doi.org/10.1111/hae.14403

[16] Cohen MM, Gorlin RJ (1991) Pseudo-Trisomy 13 Syndrome. Am J Med Genet 39 (3): 332-5. https://doi.org/10.1002/ajmg.1320390316.

[17] Emery AEH (1989) Emery-Dreifuss syndrome. Med Genet 26 (10): 637-41. https://doi.org/10.1136/jmg.26.10.637

[18] Pace BS, Starlard-Davenport A, Kutlar A (2021) Sickle cell disease: progress towards combination drug therapy. British Journal of Haematology 194 (2): 240–251. https://doi.org/10.1111/bjh.17312

[19] McColgan P, Tabrizi SJ (2018) Huntington's disease: a clinical review. European Journal of Neurology 25 (1): 24–34. https://doi.org/10.1111/ene.13413

[20] Kabra M, Gulati S, Ghosh M, Menon PSN (2000) Fraser-Cryptophthalmos Syndrome. The Indian Journal of Pediatrics 67: 775–778. https://doi.org/10.1007/BF02723939

[21] Lacy RC (2012) Extending pedigree analysis for uncertain parentage and diverse breeding systems. Journal of Heredity 103 (2): 197–205. https://doi.org/10.1093/jhered/esr135

[22] Pareek CS, Smoczynski R, Tretyn A (2011) Sequencing technologies and genome sequencing. Journal of Applied Genetics 52 (4): 413–435. https://doi.org/10.1007/s13353-011-0057-x

[23] Ellingford JM, Barton S, Bhaskar S, Williams SG, Sergouniotis PI, O'Sullivan J et al. (2016) Whole Genome Sequencing Increases Molecular Diagnostic Yield Compared with Current Diagnostic Testing for Inherited Retinal Disease. Ophthalmology 123 (5): 1143–1150. https://doi.org/10.1016/j.ophtha.2016.01.009

[24] Alberts B, Bray D, Hopkin K, Johnson A, Lewis J, Raff M, Roberts K, Walter P (2013) Essential Cell Biology. Taylor & Francis Group: LLC, New York, USA. https://cdn.bc-pf.org/resources/biology/Molecular_biology/Alberts-Essential_Cell_Biology_4th_Edition.pdf

[25] Bush WS, Moore JH (2012) Chapter 11: Genome-Wide Association Studies. PLoS Computational Biology, 8, Article ID: e1002822. https://doi.org/10.1371/journal.pcbi.1002822

[26] Clarke GM, Anderson CA, Pettersson FH, Cardon LR, Morris AP, Zondervan KT (2011) Basic Statistical Analysis in Genetic Case-Control Studies. Nature Protocols, 6: 121-133. https://doi.org/10.1038/nprot.2010.182

[27] Horita N, Kanek T (2015) Genetic Model Selection for a Case-Control Study and a Meta-Analysis. Meta Gene, 5: 1-8. https://doi.org/10.1016/j.mgene.2015.04.003

[28] Setu TJ, Basak T (2021) An Introduction to Basic Statistical Models in Genetics. Open Journal of Statistics 11: 1017-1025. https://doi.org/10.4236/ojs.2021.116060

[29] Moore JH, Hahn LW, Ritchie MD, Thornton TA, White BC (2004) Routine Discovery of Complex Genetic Models using Genetic Algorithms. Applied Soft Computing, 4: 79-86. https://doi.org/10.1016/j.asoc.2003.08.003

[30] Cooper DN, Krawczak M, Polychronakos C, Tyler-Smith C, Kehrer-Sawatzk H (2013) Where Genotype Is Not Predictive of Phenotype: Towards an Understanding of the Molecular Basis of Reduced Penetrance in Human Inherited Disease. Human Genetics, 132: 1077-1130. https://doi.org/10.1007/s00439-013-1331-2

[31] Ford D, Easton DF, Stratton M, Narod S, Goldgar D, Devilee P, Bishop DT, Weber B, Lenoir G, Chang-Claude J, Sobol H, Teare MD, Struewing J, Arason A, Scherneck S, Peto J, Rebbeck TR, Tonin P, Neuhausen S, Barkardottir R, Eyfjord J, Lynch H, Ponder BAJ, Gayther SA, Birch JM, Lindblom A, Stoppa-Lyonnet D, Bignon Y, Borg A, Hamann U, Haites N, Scott RJ, Maugard CM, Vasen H, Seitz S, Cannon-Albright LA, Schofield A, Zelada-Hedman M, The Breast Cancer Linkage Consortium (1998) Genetic Heterogeneity and Penetrance Analysis of the BRCA1 and BRCA2 Genes in Breast Cancer Families. American Journal of Human Genetics, 62: 676-689. https://doi.org/10.1086/301749

[32] Pérez-Rodríguez P, Campos G (2022) Multitrait Bayesian shrinkage and variable selection models with the BGLR-R package. Genetics, 222 (1): iyac112. https://doi.org/10.1093/genetics/iyac112