Asian Journal of Basic Science & Research
Volume 5, Issue 2, Pages 73-82, April-June 2023

Fundamentals of Pedigree Visualization and Feature Extraction for the Autosomal Dominant Inheritance Pattern

Sajjad Bin Sogir[1], Pranto Roy[1] & Tapati Basak[2*]

[1]Student, [2]Associate Professor, [1-2]Department of Statistics, Jahangirnagar University, Savar, Dhaka-1342, Bangladesh.
Corresponding Author (Tapati Basak) Email: tapati.basak@juniv.edu*

DOI: http://doi.org/10.38177/AJBSR.2023.5207

Copyright: © 2023 Sajjad Bin Sogir et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Article Received: 14 April 2023 *Article Accepted: 26 May 2023* *Article Published: 12 June 2023*

ABSTRACT

The purpose of this study was to examine the transmission of a trait, such as a disease, through an autosomal dominant inheritance pattern under the two distinct situations of full penetrance and reduced penetrance, as well as twins. To assess how a disease is passed down from one generation to the next based on the provided genetic conditions, two data sets were produced for 14 and 45 individuals with three and four generations, respectively. Then comparisons were drawn between the various situations. Due to certain restrictions, the original data was not available, yet the simulated data behaved exactly like the characteristics of autosomal inheritance. The inheritance patterns for several genetic disorders would match the findings if genuine data had been utilized in this analysis.

Keywords: Pedigree; Inheritance; Penetrance; Autosomal; Dominant; Trait; Transmission; Genotype.

1. Introduction

From the inception of the human race, some common questions arose in human mind such as why do we look like our ancestors? or why do we have some of the same diseases as our ancestors? or why do some diseases appear after a certain generation? etc. At that time, people thought that an offspring's traits were just a blend of those of its biological parents [1]. However, after Gregor Mendel's groundbreaking discovery in 1866 [2] based on experiments with pea plants, view shifted. On the basis of Mendel's findings, extensive research has been conducted to solidify our understanding of inheritance patterns [3-5]. Now, it is abundantly clear that this is not simply the result of environmental factors, but also for genetic association [6,7].

Modern molecular biology facilitates human understanding of the fact that the nucleus of every human cell contains 23 pairs of chromosomes, 22 of which are autosomes and one of which is a sex chromosome, and that these chromosomes are made up of genes carrying genetic information through the deoxyribonucleic acid (DNA) sequence [8-10]. A single gene or collection of genes regulate traits through providing instructions for protein synthesis. In the process of human reproduction, an equal number of chromosomes alone with genetic information of their ancestors are passed on to the offspring known as inheritance. This is the main cause for a family seeming identical or passing on a certain disease from generation to generation.

The particular inheritance of a disease in a family from generation to generation generally occurs by some specific patterns including autosomal dominant, autosomal recessive, X-linked dominant, X-linked recessive, Y-linked dominant [11]. Genetic disorders, such as Genetic Atypical Hemolytic-Uremic Syndrome [12], Congenital Ptosis [13], Emery-Dreifuss syndrome [14], Triphalangeal thumb [15], Pseudo-trisomy 13 syndrome [16], Fraser-cryptophthalmos syndrome [17] and Myotonia Congenital [18] etc. have found to transmitted from one generation to the next according to one of these inheritance patterns. The preliminary and most powerful method for discovering these inheritance patterns and study of inheritance of different traits can be conducted by pedigree

analysis. Genome sequencing data are used in pedigree analysis to show a visual manner that how a certain trait, such as a disease, is passed down from the past generation. In addition, genome sequencing information improves molecular diagnostic yield in comparison to current diagnostic methods [19]. Such data may be obtained from reputable websites or directly from a scientific laboratory. Due to the volume of these data, however, it is difficult to process them with normal technological assistance, and data from scientific laboratories is prohibitively expensive in a developing country like Bangladesh.

This type of issues discourage new researchers from making contributions to this domain. For overcoming this type of problem, we used simulated data, and with these data, some basic preliminary septs of genetic inheritance based on autosomal dominant mood were analyzed by pedigree chart, allowing us to observe the several conditions of penetrance, twins and to extract the characteristics of the pedigree chart.

2. Materials and Methods

2.1. Autosomal dominant inheritance pattern

The traits, particularly diseases, are transmitted from one generation to the next through a variety of inheritance patterns, including autosomal dominant inheritance, autosomal recessive inheritance, X-chromosomal dominant inheritance, etc. Some common criteria distinguish these inheritance patterns. This study includes the ancestry patterns for the autosomal dominant inheritance [11].

One allele of a gene situated on one of the 22 autosomes is enough to cause an autosomal dominantly transmitted disease. Certain features can be observed during the studies with an autosomal dominant case. For example, the incidence rate is approximately similar for both genders and hence visible in every generation. If an individual is affected, at least one of his or her parents, either his or her father or his or her mother is also affected. About 50% of the offspring will be affected with a particular disease in case of a mating, where, one of them is affected [11].

The basic concept of an autosomal dominant inheritance pattern can be expressed for a Single Nucleotide Polymorphisms (SNP) data of genotype counts. For example, for a SNP with disease allele "A" and normal allele "a", the genotype counts are: homozygous (A/A), heterozygous (A/a) and opposite homozygous (a/a). Here, the dominant model (for A allele) assumes that having one or more copies of the A allele increases risk compared to a. Hence, the genotypes A/A or A/a have the higher risk [20].

2.2. Pedigree data preparation

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

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

In order to observe the features under several conditions (full penetrance, reduced penetrance, and twins), the two different PED files were generated under the model of autosomal dominant inheritance pattern. Among the two simulated PED files, the first one contains the ancestry information for 14 individuals with three generations of a certain family. For 45 individuals with four generations of another family the information are contained in another PED file. Each file includes the six column information (Family ID, Individual ID, Paternal ID and so on). The first PED file includes 14 individuals, where, 7 are male and 7 are female; 6 persons are afflicted and 8 persons were unaffected. The second file is comprised of 45 individuals, where, 23 are male and 22 are female; 13 are affected and the rest are unaffected. All analyses were performed with R programming language.

3. Ancestry Analysis Using the Pedigree Charts

Pedigree analysis is the first stages of an in-depth examination into how a trait (e.g., disease) is inherited from one generation to the next. Additionally, it enables the visual observation and identification of passing characteristics using PED data file. Understanding the PED file is the key to analyze the ancestry pattern. A brief description of the first PED file (Table 1) for the 14 individuals, obtained in Section 2 are as below.

Table 1. The PED file for 14 individuals with full penetrance

Family ID	Individual ID	Paternal ID	Maternal ID	Sex	Phenotype
1	101	0	0	1	1
1	102	0	0	2	0
1	103	0	0	1	0
1	104	101	102	1	0
1	105	101	102	2	1
1	106	101	102	1	1
1	107	101	102	2	0
1	108	0	0	2	0
1	109	103	105	1	1
1	110	103	105	2	0
1	111	103	105	1	0
1	112	103	105	2	1
1	113	106	108	1	1
1	114	106	108	2	0

The first column (Family ID) contains the number 1, as each individual is a member of the same family. The identification number of a person is listed in the second column, labeled as "Individual ID". The third and fourth

columns contain the father's and mother's IDs, respectively, who has already listed as an individual person in the second column. The fourth column, "Sex", indicates the genders, where, 1 represents men and 2 represents women. The affected status is listed in the fifth column (1: affected and 0: not affected). Similar description is sufficient for the second PED file of 45 individuals for the four generations. The feature of full penetrance is a genetic condition in which all individuals with the disease causing mutation exhibit clinical symptoms of the disease. The pedigree chart of the first family (Table 1) with complete penetrance is shown in Figure 1.

In Figure 1, the first-generation starts with two individuals having Individual IDs 101 and 102, where, 101 is male (1) and affected (1) and 102 is female (2) and unaffected (0). The father and mother IDs of the two persons with Individual ID numbers 101 and 102 are filled with zero because they are unknown(Table 1).The second generation comprises of the persons whose Individual IDs are 104, 105, 106, and 107 and they are siblings with father and mother's IDs 101 and 102, respectively. From the second generation, it was observed that, 50 percent of the children (105 and 106) are affected and the two individuals with Individual ID numbers 105 and 106 mate with 103 and 108 respectively. Finally, in the third generation, four individuals (109, 110, 111, 112) have the same parents, and fifty percent of them are affected since their mother (105) is affected. On the other hand, individuals 113 and 114 are siblings, and one of them is affected as their father (106) is affected.

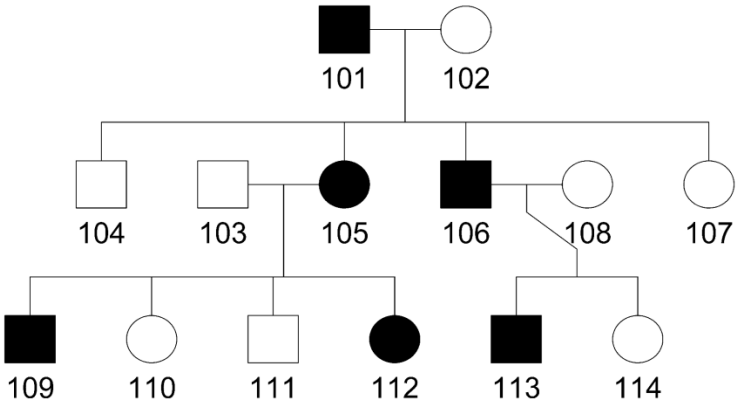

Figure 1. Pedigree for the autosomal dominant ancestry pattern with full penetrance for the first PED file (PED1)

The characteristics are available in both gender at similar frequency; disease is visible in every generation; at least one parents of an affected individual is affected; the square shape is indicating for male, whereas the circle is for the female, and the black is the indicator of affected status of an individual.

The condition of reduced penetrance is a condition where all carriers of the disease-causing mutation lack clinical manifestations of the disease. To demonstrate this condition with the previous first PED file a simple modification has been done in the given Table 1, and the modified part is shown in Table 2.

Table 2. The modified part of the first PED file for 14 individuals with full penetrance

Family ID	Individual ID	Paternal ID	Maternal ID	Sex	Phenotype
1	109	103	105	1	0
1	112	103	105	2	0

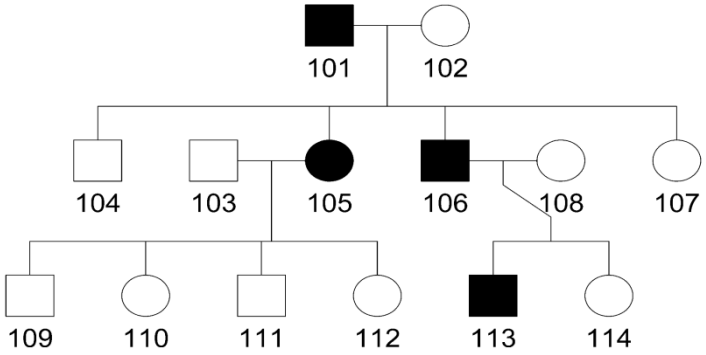

Figure 2. The autosomal dominant ancestry pattern with reduced penetrance for the first PED file (PED1) of 14 individuals from a family. The exposure of the trait is not 100% visible in the third generation

Due to the condition of reduced penetrance, the individuals with IDs 109 and 112 have not developed the disease (violates the characteristics of autosomal dominant pattern) though their mother was affected (Figure 2).

The inheritance mechanism by autosomal dominant inheritance with reduced penetrance depicted in Figure 2 that is similar as in the preceding chart (Figure 1), except the third generation. In the third generation, four individuals with IDs 109, 110, 111, 112 are all unaffected despite of their mother's phenotype status.

From Figure 1 and 2, it is observed that sometimes a phenotype may not present despite of the presence of a dominant allele. Though, the penetrance describes the probability of expressing a particular phenotype, but, the full expression of phenotype may not be found in some particular cases due the reduction of the penetrance. Hence, the exact ancestry pattern is observed for the complete or full penetrance; otherwise, it is referred to as reduced or incomplete penetrance.

The pedigree pattern was also examined for the genetic characteristics for 45 individuals of a second family spanning for the four generations. Initially, the analysis was conducted on the assumption of full penetrance, and afterwards, it was conducted under the assumption of reduced penetrance and twins.

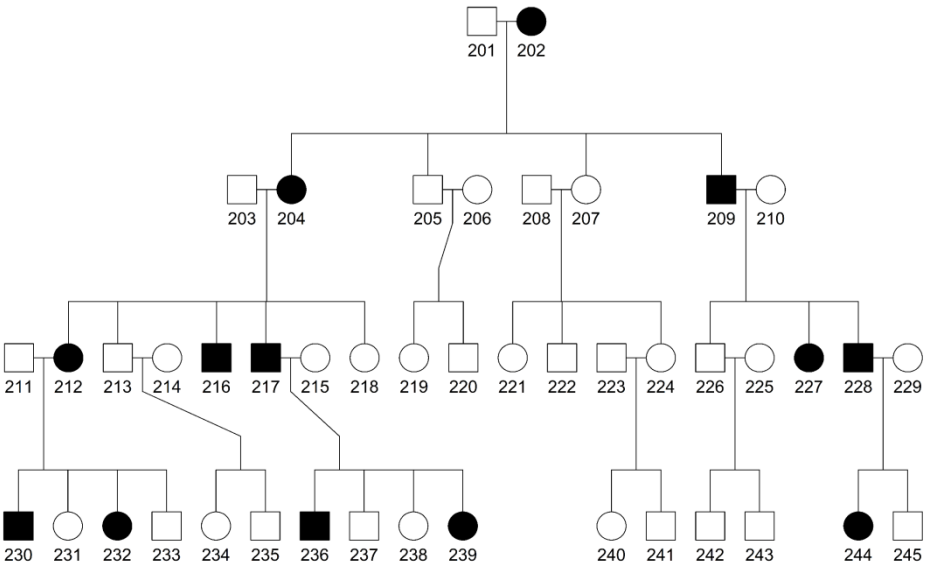

Figure 3. Ancestry chart with full penetrance of 45 individuals from a family over the four generations for the autosomal dominant pattern

Figure 3 illustrates another example of how a disease passes down from first generation to fourth generation. The first generation consists of two individuals with an affected female partner (202). Four individuals in the second generation (204, 205, 207, 209) are siblings, and two of them (204, 209) are affected. The four siblings have an unaffected father (201) and a mother (202) who is affected, and they meet four unaffected persons. In the generation that follows, there are four partial families: in the first family, there are eight individuals with IDs ranging from 211 to 218; five of them (212, 213, 215, 217, 218) are siblings, and three (212, 216, 217) are affected due to their affected mother (204). The second and third families are comprised of individuals with IDs ranging from 219 to 224, and no one in these families is affected because their parents are unaffected. In the fourth family, two individuals are affected out of three siblings (226, 228, 228), where, two unaffected individuals with IDs 225 and 229 meet two of the siblings with IDs 226 and 228, respectively.

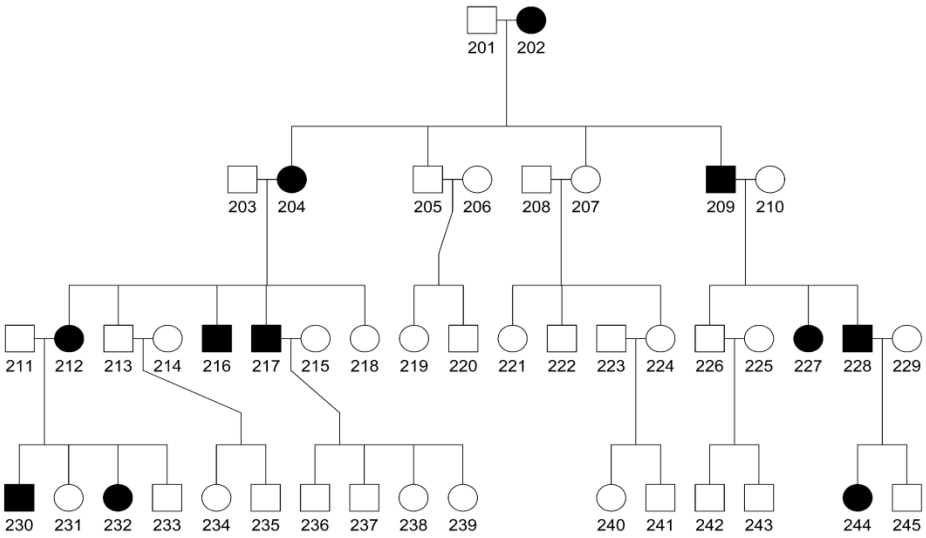

Figure 4. The autosomal dominant ancestry pattern with reduced penetrance for the second PED file (PED2) of 45 individuals from a family. The exposure of the trait is not 100% visible in the fourth generation

Finally in last generation, there are six partial families: the first family consists of two sons (230, 233) and two daughters (231, 232), of which one son (230) and one daughter (232) is affected due to an affected mother(212); the second family consists of one son (225) and one daughter (234), both of whom are unaffected; the third family has the same number of sons and daughters as the first family, and the result is identical due to an affected father (217). Due to the disease-free status of their parents, the fourth and fifth families (240 to 244) comprise unaffected individuals. The sixth family has one daughter (244) and one son (245), and the daughter (244) is affected due to her father's disease status (228).

Figure 4 illustrates the another genetic condition that is reduced penetrance for the second family of 45 individuals spanning four generation. From this figure, it is evident that the disease is inherited in a regular autosomal pattern from the first to the third generation, but in the fourth generation, the individuals with IDs 236, 237, 238, 239 are unaffected, despite the fact that half of them should be affected due to the presence of an affected father (217) in the third generation.

Figures 3 and 4 demonstrate how a disease is transmitted from the first to the fourth generation maintaining autosomal dominance with full penetrance (Figure 3) and for the reduced penetrance (Figure 4). In fourth

generation, observing the offspring from the six families it was found that the two sons (236, 237) and two daughters (238, 239) of an affected father (217) are not affected. But, according to the transmission pattern of an autosomal dominant inheritance, it was predicted that around half of the children of an affected father would be affected. So, the pattern observed in Figure 4 is due to the reduced penetrance.

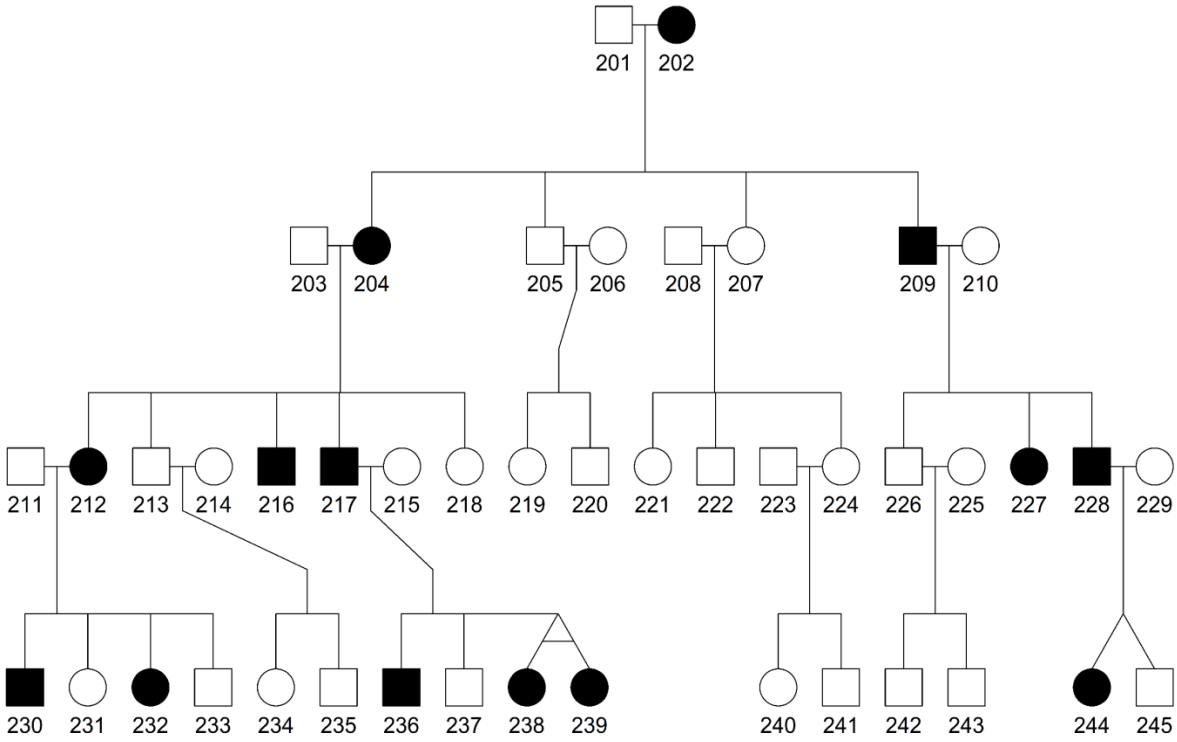

Figure 5. Inheritance pattern for twins (special relation) in a family of 45 individuals (PED2) for autosomal dominant case

In the case of monozygotic or dizygotic twins, a special scenario arises. Figure 5 depicts the transmission of a disease in the second family of 45 members within the fourth generation of twins. As previously explained, the transmission process is similar for the first three generations.

In contrast, there are two sets of twins in the fourth generation. Among the six partial families in the fourth generation, the third family has monozygotic twins with the identifiers 238 and 239, whereas, the sixth family has dizygotic twins with the identifiers 244 and 245. In the third family, there are four siblings, two of whom are identical twins (monozygotic). About half of the four siblings should be affected at a similar frequency based on gender because of affected father (217), but due to the monozygotic chromosome, both identical twin sisters (238 and 239) are affected, whereas, only one would be affected according to the general inheritance patterns discussed earlier. The afflicted status of these monozygotic twins is due to the fact that their deoxyribonucleic acid (DNA) sequences are identical. And, the final family of the fourth generation likewise has twins (244 and 245), but they are dizygotic and disease transmission occurs as normal since they do not have a similar DNA sequence.

A comparison between two families of 14 individuals and 45 individuals is made by shrinking the pedigrees of the two families, which is represented in Figure 6. The left pedigree represents the first family, while the right represents the second. The pattern of autosomal dominant inheritance is very clearly visible in both pedigrees.

Asian Journal of Basic Science & Research
Volume 5, Issue 2, Pages 73-82, April-June 2023

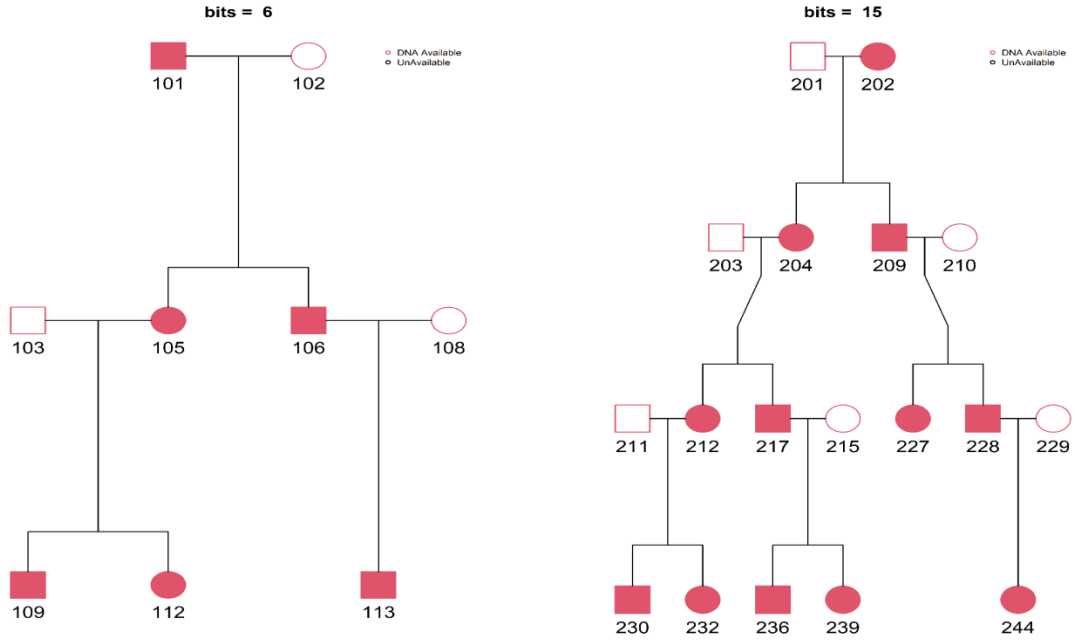

Figure 6. Trimming and shrinking for the pedigree charts with full penetrance. The left one and right one are the outcomes from shrinking of Figure 1 and 2, respectively

If the size of the pedigree chart is comparatively large, it is convenient to shrink the chart using a scientific way so that the main characteristics of the chart remain unchanged. Different software-based algorithm facilitate in this regard by converting the size of the chart to a specified bit size with priority placed on trimming uninformative subjects. Such algorithms are useful for generating a pedigree compressed to a minimally informative size. These methods randomly remove the uninformative subjects iteratively from the existing pedigree, which is controlled by a seed argument. Here, uninformative subjects means unavailable (not genotyped) with no available descendants. In this way, iteratively the pedigree shrinks to an informative one in terms of its size and information that in turns provides a useful way for examining the inheritance pattern of individuals [21].

4. Conclusions

Numerous developing countries have entered a new era of genetic data analysis. Despite their great interest in this topic, many researchers are unable to conduct in-depth analyses due to constraints (e.g., lack of financial or technological support).This investigation will serve as a guide for new scholars. The approach described for a particular inheritance pattern with different conditions, which was based on simulated data, will serve as a guide for future researchers. The procedure will be identical if real data are used. This study could be extended for calculating the disease risk of an individual using Genotypic Relative Risk (GRR) approach for inheritance pattern considered here in this paper (autosomal dominant). The GRR approach can further be applicable for investigating individual's relative risk of developing disease under the autosomal recessive condition. A comparison between autosomal dominant and recessive patterns can also be an informative study for identifying important features for these cases.

Limitations

The real genotype data was infeasible due to the lack of funding. In addition, there is insufficient laboratory support for handling huge amount of sequencing data that could be obtained from few accessible data repository website.

Declarations

Source of Funding

This study did not receive any grant from funding agencies in the public or not-for-profit sectors.

Conflict of Interest

The authors declare that they have no conflict of interest.

Consent for Publication

The authors declare that they consented to the publication of this study.

References

[1] Porter T.M. (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. doi: https://doi.org/10.1016/j.shpsc.2014.02.003.

[2] Abbott S. and Fairbanks D.J. (2016). Experiments on Plant Hybrids by Gregor Mendel. Genetics, 204(2): 417–422. doi: https://doi.org/10.1534/genetics.116.195198.

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

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

[5] Smýkal P.K., Varshney R.K., Singh V., Coyne C.J., 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. doi: https://doi.org/10.1007/s00122-016-2803-2.

[6] Timm J., Oberste N., Schmiemann P. (2022). Which factors influence the success in pedigree analysis? International Journal of Science Education, 45(3): 204–222. doi: https://doi.org/10.1080/09500693.2022.2155494.

[7] Middeldorp C.M., de Moor M.H.M., McGrath L.M., Gordon S.D., Blackwood D.H., Costa P.T. et al. (2011). The genetic association between personality and major depressi0on or bipolar disorder. A polygenic score analysis using genome-wide association data. Translational Psychiatry, 1(e50). doi: https://doi.org/10.1038/tp.2011.45.

[8] Singh R. & Sophiarani Y. (2020). A report on DNA sequence determinants in gene expression. Bioinformation, 16(5): 422–431. doi: https://doi.org/10.6026/97320630016422.

[9] Waller R.G., Darlington T.M., Wei X., Madsen M.J., 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. doi: https://doi.org/10.1371/journal.pgen.1007111.

[10] 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 Gen., 12(2): e1005804. doi: https://doi.org/10.1371/journal.pgen.1005804.

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

[12] Noris M., Bresin E., Mele C., Remuzzi G., Adam M.P., Mirzaa G.M. et al. (2007). Genetic Atypical Hemolytic-Uremic Syndrome. Gene Reviews, https://pubmed.ncbi.nlm.nih.gov/20301541/.

[13] Braathen G.J. (2012). Genetic Epidemiology of Charcot-Marie-Tooth Disease. Neurologica, 126(s193): doi: https://doi.org/10.1111/ane.12013.

[14] Emery A.E.H. (1989). Emery-Dreifuss syndrome. Med Genet., 26(10): 637–41. doi: https://doi.org/10.1136/jmg.26.10.637.

[15] Qazi Q., & Kassnert G. (1988). Triphalangeal thumb. J Med Genet, 25(8): 505–520. doi: https://doi.org/10.1136/jmg.25.8.505.

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

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

[18] Dunø M. and Vissing J. (2021). Myotonia Congenita. Gene Reviews. https://www.ncbi.nlm.nih.gov/books/NBK1355/.

[19] Ellingford J.M., Barton S., Bhaskar S., Williams S.G., Sergouniotis P.I., 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. doi: https://doi.org/10.1016/j.ophtha.2016.01.009.

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

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