Jurnal Matematika UNAND
Vol. **13** No. **4** Hal. 244 – 256
Edisi Oktober 2024
ISSN : 2303–291X
e-ISSN : 2721–9410

https://doi.org/10.25077/jmua.13.4.244-256.2024

UTILIZING DISCRETE HIDDEN MARKOV MODELS TO ANALYZE TETRAPLOID PLANT BREEDING

NAHRUL HAYATI*, EKO SULISTYONO, VITRI APRILLA HANDAYANI
Mathematics Department, Batam Institute of Technology
The Vitka City Complex, Batam, 29425, Indonesia
email : nahrul@iteba.ac.id, eko@iteba.ac.id, vitri@iteba.ac.id

Accepted May 15, 2024 Revised July 02, 2024 Published October 31, 2024

Abstract. *In plant heredity, the phenotype is the result of observation that can be directly observed, while the genotype is the underlying hidden factor that underlies the expression of the phenotype. The genotype is an important aspect that needs to be understood to explain the pattern of trait inheritance and predict trait inheritance in subsequent generations. The discrete hidden Markov model consists of two key components: an unseen Markov chain and an observable process. This model's structure makes it suitable for application in the study of tetraploid plant crosses by modeling genotypes as hidden state and phenotypes as the observation process. The probability of expressing the dominant phenotype in crosses involving monohybrid, dihybrid and trihybrid traits with the condition that the previous generation also has a dominant phenotype is 99% occurring in the sixteenth, twentieth, and twenty second generations consecutively. Furthermore, as more traits are crossed, the probability of dominant phenotype appearing decreases. When the dominant phenotype occurs, the same genotype can be obtained in crosses involving monohybrid, dihybrid and trihybrid traits, which is heterozygous in the first and second generations, while from the third to the next generation it is homozygous dominant.*

Keywords: Discrete hidden Markov model, Tetraploid

1. Introduction

The heredity of plants or the inheritance of traits from parents to offspring is a fundamental aspect in the field of genetics and plant breeding. Since the discovery of the laws of inheritance formulated by Gregor Mendel in 1865, the science of genetics has developed rapidly and become an important foundation in efforts to increase productivity and quality of cultivated plants [1].

In the study of plant heredity, the phenotype (physical appearance) is the result of observation that can be directly observed, while the genotype (genetic composition) is the underlying hidden factor that underlies the expression of the phenotype.

*Corresponding author

Nevertheless, the genotype is an important aspect that needs to be understood to explain the pattern of trait inheritance and predict trait inheritance in subsequent generations [2].

The discrete hidden Markov model is a statistical approach that can be used to model stochastic processes with hidden states that cannot be directly observed, but can be inferred from observations on observable variables. In plant heredity, this model can be applied by modeling genotypes as hidden state and phenotypes as the observation process. In addition, this model is also widely used in other field of biomathematics, including [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13].

Hayati et al. [2] in her research was interested in applying this discrete hidden Markov model to diploid plant crossings. In her research, it can be known that the probability of dominant phenotype in crosses involving monohybrid, dihybrid and trihybrid traits with the condition that the previous generation also has a dominant phenotype is 99% occurring in the 7^{th}, 8^{th}, and 9^{th} generations consecutively. When the dominant phenotype occurs, the same genotype can be obtained in crosses involving monohybrid, dihybrid and trihybrid traits, which is heterozygous in the first generation, while from the second to the next generation it is homozygous dominant.

In this study, the discrete hidden Markov model is utilized for tetraploid plant crossings. This is because tetraploid plants often have advantages such as larger organ size, higher plant vigor, and greater genetic diversity compared to diploid plants [14]. Furthermore, it is known how the probability of dominant phenotype and its genotype occurs in tetraploid plant crosses.

2. Discrete Hidden Markov Model

The discrete hidden Markov model (DHMM) consists of two components: an unobserved Markov chain $X = \{X_k\}_{k\in N}$ and an observed process $Y = \{X_k\}_{k\in N}$. The Markov chain X_k is influenced by X_{k+1}, and the observation Y_k is determined by X_k. The characteristics of DHMM are as follows.

(1) The transition probability matrix $\boldsymbol{A} = (a_{ij})_{N\times N}$ for $i, j = 1, 2, \cdots, N$; is defined as:

$$a_{ij} = P(X_{k+1} = j|X_k = i),\ a_{ij} \geq 0 \text{ and } \textstyle\sum_{j=1}^{N} a_{ij} = 1.$$

The initial state probability matrix is represented by $\boldsymbol{\pi} = (\pi_i)_{N\times 1}$ for $i = 1, 2, \cdots, N$; where:

$$\pi_i = P(X_1 = i), \text{ and } \textstyle\sum_{i=1}^{N} \pi_i = 1.$$

(2) The emission probability matrix $\boldsymbol{B} = (b_i(j))_{N\times M}$ for $i = 1, 2, \cdots, N$; $j = 1, 2, \cdots, M$; with:

$$b_i(j) = P(Y_k = j|X_k = i),\ b_i(j) \geq 0 \text{ and } \textstyle\sum_{j=1}^{M} b_i(j) = 1.$$

(3) It is assumed that the variables of Y_k and X_k are conditionally independent.

Based on those characteristics, parameters are obtained as the characteristics of HMM, namely:

$$\lambda = (\boldsymbol{A},\boldsymbol{B},\boldsymbol{\pi}).$$

There are three main problems in DHMM, namely calculating the probability of the emergence of an observation sequence using the forward and backward algorithms, determining the optimal hidden state sequence using Veterbi algorithm, and re-estimating the DHMM parameters using the Baum-Welch algorithm so that the probability $P(Y_1 = y_1, \cdots, Y_K = y_K|\lambda)$ is maximum [15].

3. Crosses of Tetraploid Plant

Gregor Mendel, in 1865, established the theory of inheritance based on his work with garden peas. He asserted that parents pass discrete genes of their offspring, which preserve their identity through generations. Genes have various forms called alleles. An organism with a pair of identical alleles for a trait is homozygous, whereas one with two different alleles is heterozygous. The dominant allele masks the effect of the recessive one, meaning an organism's traits do not always show its genetic composition. Thus, a distinction is made between the appearance or observable trait of an organism, called the phenotype, and its genetic makeup, called genotype [1].

To obtain new varieties with superior traits, plant breeding or crossing of plants is performed. A cross involving a single pair of contrasting traits is called a monohybrid cross. A cross involving two pairs of contrasting traits is called a dihybrid cross, and one involving three pairs of contrasting traits is called a trihybrid cross. These crosses begin with a cross between a pair of true-breeding (homozygous) parental line (P_1) to produce the first filial generation (F_1). Then, a cross is made among the F_1 individuals (P_2) to produce the second filial generation (F_2). To determine all possible genotype of phenotype combination in such crosses, a Punnett square can be used [1].

In tetraploid plant, the crosses that occur result in offspring that are also tetraploid with four complete sets of chromosomes in their somatic cells. Tetraploid crosses can cause genetic variation and can produce new adventageous traits in plant breeding. This is because tetraploid plants often have advantages such as larger organ size, higher plant vigor, and greater genetic diversity compared to diploid plants [14].

3.1. *Monohybrid Crosses*

P_1 : $FFFF \times ffff$
F_1 : $FFff$
P_2 : $FFff \times FFff$
Gamete : FF, Ff, ff
F_2 :

The k^{th} descendant genotype denoted by X_k random variable and state space of a Markov chain $\{X_k\}$ is:

$$S_X = \{FFFF, FFFf, FFff, Ffff, ffff\}.$$

	FF	Ff	ff
FF	$FFFF$	$FFFf$	$FFff$
Ff	$FFFf$	$FFff$	$Ffff$
ff	$FFff$	$Ffff$	$ffff$

The k^{th} descendant phenotype denoted by Y_k random variable and state space of an observation process $\{Y_k\}$ is:

$$S_Y = \{F, f\}.$$

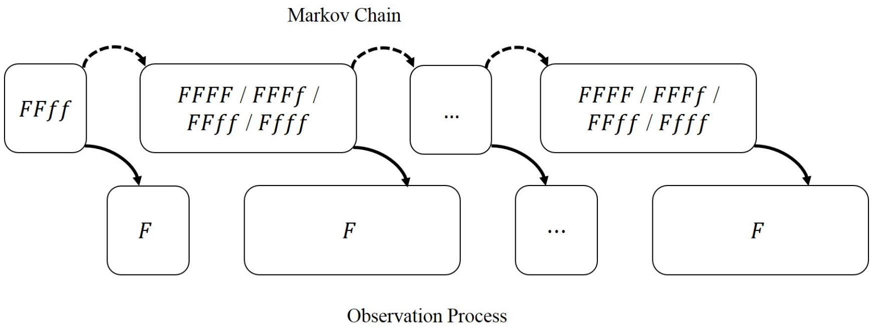

Figure. 1. Monohybrid Crosses of Tetraploid Plant

Based on Figure 1, it is known that the first generation has a heterozygote genotype $FFff$, so the first generation has a dominant phenotype F. The dominant phenotype F on the second generation is influenced by unknown genotype whether it is heterozygote $FFFf$, $FFff$, $Ffff$, or homozygote $FFFF$.

The characteristics of HMM for monohybrid crosses are transition probability matrix $\boldsymbol{A_1}$, emission probability matrix $\boldsymbol{B_1}$, and probability matrix of the initial state $\boldsymbol{\pi_1}$. The first is determining transition probability matrix $\boldsymbol{A_1}$. The dimension of matrix $\boldsymbol{A_1}$ is 5×5, that is determined by number of state space S_X. The entries of matrix $\boldsymbol{A_1}$ are obtained by punnett square of state space S_X.

$FFFF$	$\times$	$FFFF$
(FF)		(FF)
$FFFF$ 100%		

$$a_{FFFF,j} = [1 \quad 0 \quad 0 \quad 0 \quad 0]$$

$FFFf$	$\times$	$FFFf$
$(3FF, 3Ff)$		$(3FF, 3Ff)$

	$3FF$	$3Ff$
$3FF$	$9FFFF$	$9FFFf$
$3Ff$	$9FFFf$	$9FFff$
$FFFF(\frac{1}{4}), FFFf(\frac{1}{2}), FFff(\frac{1}{4}), Ffff(0), ffff(0)$		

$$a_{FFFf,j} = [\frac{1}{4} \quad \frac{1}{2} \quad \frac{1}{4} \quad 0 \quad 0]$$

$FFff$	$\times$	$FFff$
$(1FF, 4Ff, 1ff)$		$(1FF, 4Ff, 1ff)$

	$1FF$	$4Ff$	$1ff$
$1FF$	$1FFFF$	$4FFFf$	$1FFff$
$4Ff$	$4FFFf$	$16FFff$	$4Ffff$
$1ff$	$1FFff$	$4Ffff$	$1ffff$
$FFFF(\frac{1}{36}), FFFf(\frac{2}{9}), FFff(\frac{1}{2}), Ffff(\frac{2}{9}), ffff(\frac{1}{36})$			

$$a_{FFff,j} = [\frac{1}{36} \quad \frac{2}{9} \quad \frac{1}{2} \quad \frac{2}{9} \quad \frac{1}{36}]$$

$Ffff$	$\times$	$Ffff$
$(3Ff, 3ff)$		$(3Ff, 3ff)$

	$3Ff$	$3ff$
$3Ff$	$9FFff$	$9Ffff$
$3ff$	$9Ffff$	$9ffff$
$FFFF(0), FFFf(0), FFff(\frac{1}{4}), Ffff(\frac{1}{2}), ffff(\frac{1}{4})$		

$$a_{Ffff,j} = [0 \quad 0 \quad \frac{1}{4} \quad \frac{1}{2} \quad \frac{1}{4}]$$

$ffff$	$\times$	$ffff$
(ff)		(ff)
$ffff$ 100%		

$$a_{ffff,j} = [0 \quad 0 \quad 0 \quad 0 \quad 1]$$

Therefore, the transition probability matrix $\boldsymbol{A_1}$ is:

$$\boldsymbol{A_1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & 0 \\ \frac{1}{36} & \frac{2}{9} & \frac{1}{2} & \frac{2}{9} & \frac{1}{36} \\ 0 & 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

The second is determining emission probability matrix $\boldsymbol{B_1}$. The dimension of matrix $\boldsymbol{B_1}$ is 5×2, that is determined by number of state space S_X and S_Y. The first column of matrix $\boldsymbol{B_1}$ tells that type $FFFF, FFFf, FFff$ and $Ffff$ of genotype will be observed as dominant phenotype F. The second column of matrix $\boldsymbol{B_1}$ tells that type $ffff$ of genotype will be observed as recessive phenotype f.

$$\boldsymbol{B_1} = \begin{bmatrix} b_{FFFF}(F) & b_{FFFF}(f) \\ b_{FFFf}(F) & b_{FFFf}(f) \\ b_{FFff}(F) & b_{FFff}(f) \\ b_{Ffff}(F) & b_{Ffff}(f) \\ b_{ffff}(F) & b_{ffff}(f) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

The third is determining probability matrix of the initial state $\boldsymbol{\pi_1}$. The dimension of matrix $\boldsymbol{\pi_1}$ is 5×1, that is determined by nuber of state space S_X. Since Mendel start the parental generation with heterozygote plants and P_2 has only one heterozygote i.e. $FFff$, therefore the initial value of π_i is one for type $FFff$ of genotype and zero to the other.

$$\boldsymbol{\pi_1} = \begin{bmatrix} \pi_{FFFF} \\ \pi_{FFFf} \\ \pi_{FFff} \\ \pi_{Ffff} \\ \pi_{ffff} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}.$$

3.2. *Dihybrid Crosses*

P_1 : $FFFFGGGG \times ffffgggg$
F_1 : $FFffGGgg$
P_2 : $FFffGGgg \times FFffGGgg$
Gamete : $FFGG, FFGg, FFgg, FfGG, FfGg, Ffgg, ffGG, ffGg, ffgg$

The k^{th} descendant genotype denoted by X_k random variable and state space of a Markov chain $\{X_k\}$ is:

$$\begin{aligned} S_X = \{ & FFFFGGGG,\ FFFFGGGg,\ FFFFGGgg,\ FFFFGggg,\ FFFFgggg, \\ & FFFfGGGG,\ FFFfGGGg,\ FFFfGGgg,\ FFFfGggg,\ FFFfgggg, \\ & FFffGGGG,\ FFffGGGg,\ FFffGGgg,\ FFffGggg,\ FFffgggg, \\ & FfffGGGG,\ FfffGGGg,\ FfffGGgg,\ FfffGggg,\ Ffffgggg, \\ & ffffGGGG,\ ffffGGGg,\ ffffGGgg,\ ffffGggg,\ ffffgggg\}. \end{aligned}$$

The k^{th} descendant phenotype denoted by Y_k random variable and state space of an observation process $\{Y_k\}$ is:

$$S_Y = \{FG, Fg, fG, fg\}.$$

The characteristics of DHMM for dihybrid crosses can be determined using the same method with the previous crosses.

$$\boldsymbol{A_2} = \begin{bmatrix} \mathbf{1}\boldsymbol{A_1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \frac{\mathbf{1}}{\mathbf{4}}\boldsymbol{A_1} & \frac{\mathbf{1}}{\mathbf{2}}\boldsymbol{A_1} & \frac{\mathbf{1}}{\mathbf{4}}\boldsymbol{A_1} & \mathbf{0} & \mathbf{0} \\ \frac{\mathbf{1}}{\mathbf{36}}\boldsymbol{A_1} & \frac{\mathbf{2}}{\mathbf{9}}\boldsymbol{A_1} & \frac{\mathbf{1}}{\mathbf{2}}\boldsymbol{A_1} & \frac{\mathbf{2}}{\mathbf{9}}\boldsymbol{A_1} & \frac{\mathbf{1}}{\mathbf{36}}\boldsymbol{A_1} \\ \mathbf{0} & \mathbf{0} & \frac{\mathbf{1}}{\mathbf{4}}\boldsymbol{A_1} & \frac{\mathbf{1}}{\mathbf{2}}\boldsymbol{A_1} & \frac{\mathbf{1}}{\mathbf{4}}\boldsymbol{A_1} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1}\boldsymbol{A_1} \end{bmatrix}, \quad \boldsymbol{B_2} = \begin{bmatrix} \boldsymbol{B_1} & \mathbf{0} \\ \boldsymbol{B_1} & \mathbf{0} \\ \boldsymbol{B_1} & \mathbf{0} \\ \boldsymbol{B_1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{B_1} \end{bmatrix}, \quad \boldsymbol{\pi_2} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \boldsymbol{\pi_1} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}.$$

3.3. *Trihybrid Crosses*

P_1 : $FFFFGGGGHHHH \times ffffgggghhhh$
F_1 : $FFffGGggHHhh$
P_2 : $FFffGGggHHhh \times FFffGGggHHhh$
Gamete : $FFGGHH, FFGGHh, FFGGhh, FFGgHH, FFGgHh, FFGghh,$
$FFggHH, FFggHh, FFgghh, FfGGHH, FfGGHh, FfGGhh,$
$FfGgHH, FfGgHh, FfGghh, FfggHH, FfggHh, Ffgghh,$
$ffGGHH, ffGGHh, ffGGhh, ffGgHH, ffGgHh, ffGghh,$
$ffggHH, ffggHh, ffgghh$

The k^{th} descendant genotype denoted by X_k random variable and state space of a Markov chain $\{X_k\}$ is:

$S_X = \{FFFFGGGGHHHH$, $FFFFGGGGHHHh$, $FFFFGGGGHHhh$,
$FFFFGGGGHhhh$, $FFFFGGGGhhhh$, $FFFFGGGgHHHH$,
$FFFFGGGgHHHh$, $FFFFGGGgHHhh$, $FFFFGGGgHhhh$,
$FFFFGGGghhhh$, $FFFFGGggHHHH$, $FFFFGGggHHHh$,
$FFFFGGggHHhh$, $FFFFGGggHhhh$, $FFFFGGgghhhh$,
$FFFFGgggHHHH$, $FFFFGgggHHHh$, $FFFFGgggHHhh$,
$FFFFGgggHhhh$, $FFFFGggghhhh$, $FFFFggggHHHH$,
$FFFFggggHHHh$, $FFFFggggHHhh$, $FFFFggggHhhh$,
$FFFFgggghhhh$, $FFFfGGGGHHHH$, $FFFfGGGGHHHh$,
$FFFfGGGGHHhh$, $FFFfGGGGHhhh$, $FFFfGGGGhhhh$,
$FFFfGGGgHHHH$, $FFFfGGGgHHHh$, $FFFfGGGgHHhh$,
$FFFfGGGgHhhh$, $FFFfGGGghhhh$, $FFFfGGggHHHH$,
$FFFfGGggHHHh$, $FFFfGGggHHhh$, $FFFfGGggHhhh$,
$FFFfGGgghhhh$, $FFFfGgggHHHH$, $FFFfGgggHHHh$,
$FFFfGgggHHhh$, $FFFfGgggHhhh$, $FFFfGggghhhh$,
$FFFfggggHHHH$, $FFFfggggHHHh$, $FFFfggggHHhh$,
$FFFfggggHhhh$, $FFFfgggghhhh$, $FFffGGGGHHHH$,
$FFffGGGGHHHh$, $FFffGGGGHHhh$, $FFffGGGGHhhh$,
$FFffGGGGhhhh$, $FFffGGGgHHHH$, $FFffGGGgHHHh$,
$FFffGGGgHHhh$, $FFffGGGgHhhh$, $FFffGGGghhhh$,
$FFffGGggHHHH$, $FFffGGggHHHh$, $FFffGGggHHhh$,
$FFffGGggHhhh$, $FFffGGgghhhh$, $FFffGgggHHHH$,
$FFffGgggHHHh$, $FFffGgggHHhh$, $FFffGgggHhhh$,
$FFffGggghhhh$, $FFffggggHHHH$, $FFffggggHHHh$,
$FFffggggHHhh$, $FFffggggHhhh$, $FFffgggghhhh$,

$FfffGGGGHHHH$, $FfffGGGGHHHh$, $FfffGGGGHHhh$, $FfffGGGGHhhh$, $FfffGGGGhhhh$, $FfffGGGgHHHH$, $FfffGGGgHHHh$, $FfffGGGgHHhh$, $FfffGGGgHhhh$, $FfffGGGghhhh$, $FfffGGggHHHH$, $FfffGGggHHHh$, $FfffGGggHHhh$, $FfffGGggHhhh$, $FfffGGgghhhh$, $FfffGgggHHHH$, $FfffGgggHHHh$, $FfffGgggHHhh$, $FfffGgggHhhh$, $FfffGggghhhh$, $FfffggggHHHH$, $FfffggggHHHh$, $FfffggggHHhh$, $FfffggggHhhh$, $Ffffgggghhhh$, $ffffGGGGHHHH$, $ffffGGGGHHHh$, $ffffGGGGHHhh$, $ffffGGGGHhhh$, $ffffGGGGhhhh$, $ffffGGGgHHHH$, $ffffGGGgHHHh$, $ffffGGGgHHhh$, $ffffGGGgHhhh$, $ffffGGGghhhh$, $ffffGGggHHHH$, $ffffGGggHHHh$, $ffffGGggHHhh$, $ffffGGggHhhh$, $ffffGGgghhhh$, $ffffGgggHHHH$, $ffffGgggHHHh$, $ffffGgggHHhh$, $ffffGgggHhhh$, $ffffGggghhhh$, $ffffggggHHHH$, $ffffggggHHHh$, $ffffggggHHhh$, $ffffggggHhhh$, $ffffgggghhhh\}$.

The k^{th} descendant phenotype denoted by Y_k random variable and state space of an observation process $\{Y_k\}$ is:

$$S_Y = \{FGH, FGh, FgH, Fgh, fGH, fGh, fgH, fgh\}.$$

The characteristics of DHMM for trihybrid crosses can be determined using the same method with the previous crosses.

$$\boldsymbol{A_3} = \begin{bmatrix} \mathbf{1}\boldsymbol{A_2} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \frac{1}{4}\boldsymbol{A_2} & \frac{1}{2}\boldsymbol{A_2} & \frac{1}{4}\boldsymbol{A_2} & \mathbf{0} & \mathbf{0} \\ \frac{1}{36}\boldsymbol{A_2} & \frac{2}{9}\boldsymbol{A_2} & \frac{1}{2}\boldsymbol{A_2} & \frac{2}{9}\boldsymbol{A_2} & \frac{1}{36}\boldsymbol{A_2} \\ \mathbf{0} & \mathbf{0} & \frac{1}{4}\boldsymbol{A_2} & \frac{1}{2}\boldsymbol{A_2} & \frac{1}{4}\boldsymbol{A_2} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1}\boldsymbol{A_2} \end{bmatrix}, \ \boldsymbol{B_3} = \begin{bmatrix} \boldsymbol{B_2} & \mathbf{0} \\ \boldsymbol{B_2} & \mathbf{0} \\ \boldsymbol{B_2} & \mathbf{0} \\ \boldsymbol{B_2} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{B_2} \end{bmatrix}, \ \boldsymbol{\pi_3} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \boldsymbol{\pi_2} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}.$$

Based on Subsection 3.1, 3.2, and 3.3, the general form of the characteristics of HMM on crosses of tetraploid plant is as follows.

$$\boldsymbol{A_{k+1}} = \begin{bmatrix} \mathbf{1}\boldsymbol{A_k} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \frac{1}{4}\boldsymbol{A_k} & \frac{1}{2}\boldsymbol{A_k} & \frac{1}{4}\boldsymbol{A_k} & \mathbf{0} & \mathbf{0} \\ \frac{1}{36}\boldsymbol{A_k} & \frac{2}{9}\boldsymbol{A_k} & \frac{1}{2}\boldsymbol{A_k} & \frac{2}{9}\boldsymbol{A_k} & \frac{1}{36}\boldsymbol{A_k} \\ \mathbf{0} & \mathbf{0} & \frac{1}{4}\boldsymbol{A_k} & \frac{1}{2}\boldsymbol{A_k} & \frac{1}{4}\boldsymbol{A_k} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1}\boldsymbol{A_k} \end{bmatrix}, \boldsymbol{B_{k+1}} = \begin{bmatrix} \boldsymbol{B_k} & \mathbf{0} \\ \boldsymbol{B_k} & \mathbf{0} \\ \boldsymbol{B_k} & \mathbf{0} \\ \boldsymbol{B_k} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{B_k} \end{bmatrix}, \boldsymbol{\pi_{k+1}} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \boldsymbol{\pi_k} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}.$$

4. Utilizing Discrete Hidden Markov Models

The initial problem in DHMM is determining the probability of an observation sequence occurring using the forward and backward algorithms. The following is the probability that model $\boldsymbol{\lambda} = (\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\pi})$ produces the observation sequence $O = \{F, F, F, F, F, F, F, F, F, F\}$ in monohybrid crosses of tetraploid plants.

Table 1. The Forward and Backward Algorithm of Monohybrid Crosses

t	1	2	3	4	$\cdots$	8	9	10
$\alpha_t(1)$	0	0,02778	0,09722	0,16281	$\cdots$	0,33720	0,36434	0,38695
$\alpha_t(2)$	0	0,22222	0,22222	0,19136	$\cdots$	0,09303	0,07752	0,06460
$\alpha_t(3)$	1	0,5	0,36111	0,29167	$\cdots$	0,13954	0,11628	0,09690
$\alpha_t(4)$	0	0,22222	0,22222	0,19136	$\cdots$	0,09303	0,07752	0,06460
$\alpha_t(5)$	0	0	0	0	$\cdots$	0	0	0
$P(O\|\lambda)$								0,61305

t	10	9	8	7	$\cdots$	3	2	1
$\beta_t(1)$	1	1	1	1	$\cdots$	1	1	1
$\beta_t(2)$	1	1	0,99306	0,97222	$\cdots$	0,87015	0,85077	0,83430
$\beta_t(3)$	1	0,97222	0,90278	0,83719	$\cdots$	0,66280	0,63566	0,61305
$\beta_t(4)$	1	0,75	0,61806	0,53472	$\cdots$	0,37405	0,35273	0,33528
$\beta_t(5)$	1	0	0	0	$\cdots$	0	0	0
$P(O\|\lambda)$								0,61305

The calculation in the forward and backward algorithms yields the same probability value. Therefore, the probability of the dominant phenotype F occurring over ten generations during that period is 0,61305 or 61,305%.

After knowing that the probability of observation is 0,61305, the next problem is determining the optimal hidden state sequence in this case, which is genotype $\{FFFF, FFFf, FFff, Ffff, ffff\}$. This second problem can be solved using the Veterbi algorithm.

Based on the calculation from the Veterbi algorithm, it can be known that when the dominant phenotype F over ten generations occurs, the most optimal sequence of hidden state is:

$$x^* = \{FFff, FFFf, FFFF, FFFF, FFFF, \cdots, FFFF, FFFF\}.$$

The last problem is re-estimating the HMM parameters using the Baum-Welch algorithm. The process of re-estimating the HMM parameters is carried out to obtain a new parameter $\hat{\boldsymbol{\lambda}} = (\hat{\boldsymbol{A}}, \hat{\boldsymbol{B}}, \hat{\boldsymbol{\pi}})$. These new parameters are what result in $P(O|\hat{\lambda}) \geq P(O|\lambda)$.

Table 2. The Veterbi Algorithm of Monohybrid Crosses

t	1	2	3	4	⋯	8	9	10
$\delta_t(1)$	0	0,02778	0,05556	0,05556	⋯	0,05556	0,05556	0,05556
$\delta_t(2)$	0	0,22222	0,11111	0,05556	⋯	0,00347	0,00174	0,00087
$\delta_t(3)$	1	0,5	0,25	0,125	⋯	0,00781	0,00391	0,00195
$\delta_t(4)$	0	0,22222	0,11111	0,05556	⋯	0,00347	0,00174	0,00087
$\delta_t(5)$	0	0	0	0	⋯	0	0	0
P^*								0,05556
${x_5}^*$								1 ($FFFF$)

t	1	2	3	4	⋯	8	9	10
$\psi_t(1)$	0	3	2	1	⋯	1	1	1
$\psi_t(2)$	0	3	2	2	⋯	2	2	2
$\psi_t(3)$	1	3	3	3	⋯	3	3	3
$\psi_t(4)$	0	3	3	3	⋯	3	3	3
$\psi_t(5)$	0	3	4	4	⋯	4	4	4

$$\hat{\boldsymbol{A}}_{\mathbf{1}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0,27415 & 0,51136 & 0,21449 & 0 & 0 \\ 0,04018 & 0,28986 & 0,52535 & 0,14461 & 0 \\ 0 & 0 & 0,43363 & 0,56637 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \hat{\boldsymbol{B}}_{\mathbf{1}} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \hat{\boldsymbol{\pi}}_{\mathbf{1}} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}.$$

The problems of HMM for dihybrid and trihybrid crosses can be solved in the same way with monohybrid crosses. In the dihybrid crosses, it is known that the probability of dominant phenotype FG occurring over ten generations during that period is 0,37583 or 37,583%. Next, when the dominant phenotype FG over ten generations occurs, the most optimal sequence of hidden state is:

$$x^* = \{FFffGGgg, FFFfGGGg, FFFFGGGG, \cdots, FFFFGGGG\}.$$

In the trihybrid crosses, it is known that the probability of dominant phenotype FGH occurring over ten generations during that period is 0,23041 or 23,041%. Next, when the dominant phenotype FGH over ten generations occurs, the most optimal sequence of hidden state is:

$$x^* = \{FFffGGggHHhh, FFFfGGGgHHHh, FFFFGGGGHHHH, \cdots\}.$$

In this context, we have derived the conditional probability $P(Y_k = y_k | Y_1 = y_1, \cdots, Y_{k-1} = y_{k-1})$ in monohybrid, dihybrid and trihybrid crosses (Table 3). For ease of notation, it is defined as:

(1) $P(Y_n = dominant|{Y_1}^{n-1} = dominant)$ represents $P(Y_n = F|{Y_1}^{n-1} = F)$ in monohybrid crosses.
(2) $P(Y_n = dominant|{Y_1}^{n-1} = dominant)$ represents $P(Y_n = FG|{Y_1}^{n-1} = FG)$ in dihybrid crosses.
(3) $P(Y_n = dominant|{Y_1}^{n-1} = dominant)$ represents $P(Y_n = FGH|{Y_1}^{n-1} = FGH)$ in trihybrid crosses.
(4) ${Y_1}^n = F$ represents $Y_1 = F, Y_2 = F, \cdots, Y_n = F$ in monohybrid crosses.
(5) ${Y_1}^n = FG$ represents $Y_1 = FG, Y_2 = FG, \cdots, Y_n = FG$ in dihybrid crosses.
(6) ${Y_1}^n = FGH$ represents $Y_1 = FGH, Y_2 = FGH, \cdots, Y_n = FGH$ in trihybrid crosses.

Table 3. Conditional Probability of Tetraploid Plant

Conditional Probability	Monohybrid	Dihybrid	Trihybrid
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$P(Y_{11} = dominant\|{Y_1}^{10} = dominant)$	0,969265	0,939474	0,910599
$P(Y_{12} = dominant\|{Y_1}^{11} = dominant)$	0,973575	0,947849	0,922802
$P(Y_{13} = dominant\|{Y_1}^{12} = dominant)$	0,977382	0,955275	0,933668
$P(Y_{14} = dominant\|{Y_1}^{13} = dominant)$	0,980715	0,961802	0,943254
$P(Y_{15} = dominant\|{Y_1}^{14} = dominant)$	0,983614	0,967495	0,951641
$P(Y_{16} = dominant\|{Y_1}^{15} = dominant)$	**0,986117**	0,972427	0,958926
$P(Y_{17} = dominant\|{Y_1}^{16} = dominant)$	0,988268	0,976673	0,965215
$P(Y_{18} = dominant\|{Y_1}^{17} = dominant)$	0,990107	0,980312	0,970614
$P(Y_{19} = dominant\|{Y_1}^{18} = dominant)$	0,991674	0,983417	0,975228
$P(Y_{20} = dominant\|{Y_1}^{19} = dominant)$	0,993003	**0,986055**	0,979156
$P(Y_{21} = dominant\|{Y_1}^{20} = dominant)$	0,994128	0,988291	0,982488
$P(Y_{22} = dominant\|{Y_1}^{21} = dominant)$	0,995078	0,990180	**0,985306**
$P(Y_{23} = dominant\|{Y_1}^{22} = dominant)$	0,995878	0,991773	0,987685
$P(Y_{24} = dominant\|{Y_1}^{23} = dominant)$	0,996551	0,993113	0,989688
$P(Y_{25} = dominant\|{Y_1}^{24} = dominant)$	0,997116	0,994240	0,991372

From Table 3, it is known that the probability of dominant phenotype in crosses involving monohybrid, dihybrid and trihybrid traits with the condition that the previous generation also has a dominant phenotype is 99% occurring in the 16^{th}, 20^{th}, and 22^{nd} generations consecutively. As the number of plant generation grows, the probability of plants showing a dominant phenotype becomes higher. Nevertheless, the increase is quite slow. Moreover, the probability of a dominant phenotype decreases as more traits are crossed.

5. Conclusions

The DHMM can be utilized for tetraploid plant crosses. The probability of dominant phenotype in crosses involving monohybrid, dihybrid and trihybrid traits with the condition that the previous generation also has a dominant phenotype is 99% occurring in the sixteenth, twentieth, and twenty second generations consecutively. As the number of plant generation grows, the probability of plants showing a dominant phenotype becomes higher. Nevertheless, the increase is quite slow. Moreover, the probability of a dominant phenotype decreases as more traits are crossed. When the dominant phenotype occurs, the same genotype can be obtained in crosses involving monohybrid, dihybrid and trihybrid traits, which is heterozygous in the first and second generations, while from the third to the next generation it is homozygous dominant.

Bibliography

[1] Reece, J.B., Urry, L.A., Cain, M.L., Wasserman, S.A., Minorsky, P.V., Jackson, R.B., 2011, *Campbell Biologi*, 9th Edition, Pearson Education Inc, USA

[2] Hayati, N., Setiawaty, B., Purnaba, I.G.P., 2023, The Application of Discrete Hidden Markov Model on Crosses of Diploid Plant, *Barekeng: Journal of Mathematics and Its Applications* Vol. **17** : 1449 – 1462

[3] Harris, P.D., Narducci, A., Gebhardt, C., Cordes, T., Weiss, S., Lerner, E., 2022, Multi-parameter Photon-by-Photon Hidden Markov Modeling, *Nature Communications* Vol. **13** : 1 – 12

[4] Hsiao, J.H., An, J., Hui, V.K.S., Zheng, Y., Chan, A.B., 2022, Understanding The Role of Eye Movement Consistency in Face Recognition and Autism Through Integrating Deep Neural Networks and Hidden Markov Models, *njp Science of Learning* Vol. **7** : 1 – 13

[5] Mousavi, N., Monemian , M., Daneshmand, P.G., Mirmohammadsadeghi, M., Zekri, M., Rabbani, H., 2023, Cyst Identification in Retinal Optical Coherence Tomography Images using Hidden Markov Model, *Scientific Reports* Vol. **13** : 1 – 15

[6] Mohammadi, F., Visagan, S., Gross, S.M., Karginov, L., Jagarde, J., Heiser, L.M., Meyer, A.S., 2022, A Lineage Tree-based Hidden Markov Model Quantifies Cellular Heterogeneity and Plasticity, *Communications Biology* Vol. **5** : 1 – 14

[7] Ludwig, R., Pouymayou, B., Balermpas, P., Unkelbach, J., 2021, A Hidden Markov Model for Lymphatic Tumor Progression in The Head and Neck, *Scientific Reports* Vol. **11** : 1 – 17

[8] Thorvaldsen, S., 2022, A Tutorial on Markov Models Based on Mendel's Classical Experiments, *Journal of Bioinformatics and Computational Biology* Vol. **3** : 1441 – 1460

[9] Paniello, I., 2021, In-evolution operators in genetic coalgebras, *Linear Algebra and Its Application* Vol. **614** : 197 – 207

[10] Azis, D., Syanur, M.N., Amanto, Zakaria, L., 2022, An Analysis Genotype Inheritance in a Trihybrid Cross by Applying a Diagonalization of Matrix Method, *Desimal: Jurnal Matematika* Vol. **5** : 305 – 314

[11] Zhai, Q., 2023, A Simple Markov Chain, *EDP Sciences*, **174** : 03001

[12] Stojkovic, N., Koceva, L., Limonka, Stojanova, A., 2023, Application of Markov Chains in the Biology,*Proceedings of The CODEMA 2022*: 83 – 93

[13] Pradana, Y.A., Azka, D.A., Aji, A.C., Fauzi, I.M., 2022, Analysis of Weather Changes for Estimation of Shallot Crops Fluctuation Using Hidden Markov, *Barekeng: Jurnal Ilmu Matematika dan Terapan* Vol. **16** : 333 – 342

[14] Sattler, M.C., Carvalho, C.R., Clarindo, W.R., 2015, The Polyploidy and Its Key Role in Plant Breeding, *Planta* Vol. **243** : 281 – 296

[15] Rabiner, L.R., 1989, A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, *Proceedings of The IEEE, Scottsdale* Vol. **77** : 257 – 285