THE APPLICATION OF DISCRETE HIDDEN MARKOV MODEL ON CROSSES OF DIPLOID PLANT

Nahrul Hayati [1*]**, Berlian Setiawaty**[2]**, I Gusti Putu Purnaba**[3]

[1,2,3]*Mathematics Department, IPB University*
Raya Dramaga Street, Bogor, 16680, Indonesia

*Corresponding author's e-mail: * nahrul.ney@gmail.com*

ABSTRACT

The hidden Markov model consists of a pair of an unobserved Markov chain $\{X_k\}$ and an observation process $\{Y_k\}$. In this research, the crosses of diploid plant apply the model. The Markov chain $\{X_k\}$ represents genetic structure, which is genotype of the kth generation of an organism. The observation process $\{Y_k\}$ represents the appearance or the observed trait, which is the phenotype of the kth generation of an organism. Since it is unlikely to observe the genetic structure directly, the Hidden Markov model can be used to model pairs of events and unobservable their causes. Forming the model requires the use of the theory of heredity from Mendel. This model can be used to explain the characteristic of true breeding on crosses of diploid plants. The more traits crossed, the smaller probability of plants having a dominant phenotype in that period. Monohybrid, dihybrid, and trihybrid crosses have a dominant phenotype probability of 99% in the seventh, eighth, and ninth generations, with the condition of previous generations having a dominant phenotype. But in seventh generation, monohybrid crosses only have the probability of an optimal genotype of 50%, dihybrid crosses have a probability of an optimal genotype of 25% in the eighth generation, and trihybrid crosses have a probability of an optimal genotype of 12.5% in the ninth generation.

1. INTRODUCTION

Heredity is a trait inheritance or forwarding from one generation to the next. For thousand of years, the principles of heredity and variation have been exploited by farmers so they can produce plants and animals with desired traits. Gregor Mendel, who found the first acceptable theory of the trait inheritance system, proposed several hypotheses about the inheritance of genetic material from the parent to the offspring. According to Mendel, for each character, the organism inherits two alleles, one from each parent. He makes this deduction without knowing about the roles or even the existence of chromosomes. Any cell with two chromosome sets is called a diploid cell, where one set is inherited from each parent [1].

Although this law of Mendel's can provide hypotheses about the inheritance of genetic material from the parent to the offspring, searching for the trait is only done based on the observed appearance in the organism. The characteristics of an organism do not always reveal its genetic composition. Therefore, the appearance of the observed trait in the organism, namely phenotype, can be distinguished from its genetic structure, namely genotype. This genetic structure is unlikely to be observed. The hidden Markov model can help model pairs of events and their causes that cannot be observed directly [2]. This models have also been used in bioinformatics and biochemistry, some of them are [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13].

In the state space of the finite observation process and the time index is countable, the model is known as the discrete version of the hidden Markov model [2]. Thorvaldsen [14] in his research, revealed interesting ideas about the role of hidden Markov models in bioinformatics. In his research, he found that the hidden Markov model can be applied to Mendel's experiments. In this research the model is applied to crosses of diploid plant and used to explain the characteristic of true-breeding in diploid plant crosses. The true-breeding generation of a plant can help to find seeds that have both feasible appearances and genetics.

2. RESEARCH METHODS

2.1 Discrete Hidden Markov Models

The hidden Markov model (X, Y) is a model generated by a Markov chain $X = \{X_k\}_{k \in N}$ that is unobserved and an observation process $Y = \{Y_k\}_{k \in N}$. The observation is influenced by X_k, while X_k is influenced by X_{k-1}. This relationship is illustrated in **Figure 1**.
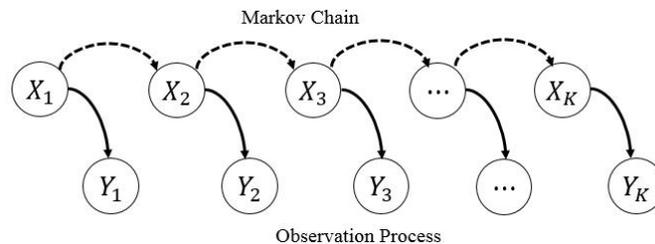

Markov Chain

Observation Process

Figure 1. Hidden Markov Model (X, Y)

Characteristics of hidden Markov model:

1) The sequence $\{X_k\}_{k \in N}$ forms a Markov chain with finite state space $S_X = \{1, 2, \cdots, N\}$ and $\boldsymbol{A} = \left(a_{ij}\right)_{N \times N}$ is a matrix of transition probability with
$$a_{ij} = P(X_{k+1} = j | X_k = i),$$
for $i, j = 1, 2, \cdots, N$; $a_{ij} \geq 0$ and $\sum_{j=1}^{N} a_{ij} = 1$. Furthemore, $\boldsymbol{\pi} = (\pi_i)_{N \times 1}$ is a probability matrix of the initial state with
$$\pi_i = P(X_1 = i),$$
for $i = 1, 2, \cdots, N$ and $\sum_{i=1}^{N} \pi_i = 1$.

2) The sequence $\{Y_k\}_{k \in N}$ is the observation process with state space $S_Y = \{1, 2, \cdots, M\}$. The relationship of $\{X_k, Y_k\}_{k \in N}$ process is represented by emission probability matrix $\boldsymbol{B} = \left(b_i(j)\right)_{N \times M}$ with
$$b_i(j) = P(Y_k = j | X_k = i),$$
for $i = 1, 2, \cdots, N$; $j = 1, 2, \cdots, M$; $b_i(j) \geq 0$ and $\sum_{j=1}^{M} b_i(j) = 1$.

3) It is assumed that $\{Y_k | X_k\}_{k \in N}$ are independent, this means
$P(Y_1 = y_1, \cdots, Y_K = y_K | X_1 = x_1, \cdots, X_K = x_K) = \prod_{k=1}^{K} P(Y_k = y_k | X_k = x_k)$

Therefore, based on those characteristics, we obtain a set of parameters as the characteristic of the hidden Markov model, namely

$$\lambda = (A, B, \pi).$$

Given observation sequence $y_1, y_2, \cdots, y_K$; $y_k \in S_y$ for $k = 1, 2, \cdots, K$ and $\lambda = (A, B, \pi)$. There are three main problems in the hidden Markov model. The first is calculating the probability of the emergence of an observation sequence $y_1, y_2, \cdots, y_K$ which is solved by forward and backward algorithms. The second problem is determining the sequence of a hidden state $x_1, x_2, \cdots, x_K$ that is optimally solved by the Veterbi algorithm. The third problem is estimating the parameters of the hidden Markov model $\lambda = (A, B, \pi)$, so that $P(Y_1 = y_1, \cdots, Y_K = y_K | \lambda)$ is maximum which is done with the Baum-Welch algorithm [15].

2.2 Forward and Backward Algorithms

The probability of the appearance of an observation sequence $y_1, y_2, \cdots, y_K$ given $\lambda = (A, B, \pi)$ i.e $P(Y_1 = y_1, \cdots, Y_K = y_K | \lambda)$ is calculated using forward and backward algorithms. Defined variables

$$\alpha_k(j) = P(Y_1 = y_1, \cdots, Y_K = y_K, X_k = j | \lambda), \tag{1}$$

for $k = 1, 2, \cdots, K$. The forward algorithm consists of three stages:

1) Initialization stage

$$\alpha_1(j) = P(Y_1 = y_1, X_1 = j | \lambda)$$
$$= P(X_1 = j | \lambda) P(Y_1 = y_1 | X_1 = j, \lambda)$$
$$= \pi_j b_j(y_1), \text{ for } j = 1, 2, \cdots, N.$$

2) Induction stage

$$\alpha_{k+1}(j) = P(Y_1 = y_1, \cdots, Y_{k+1} = y_{k+1}, X_{k+1} = j | \lambda)$$
$$= \left(\sum_{i=1}^{N} \alpha_k(i) a_{ij} \right) b_j(y_{k+1}), \quad \text{for } j = 1, 2, \cdots, N \text{ and } k = 1, 2, \cdots, K - 1.$$

3) Termination stage

$$P(Y_1 = y_1, \cdots, Y_K = y_K | \lambda) = \sum_{j=1}^{N} \alpha_K(j).$$

Defined variables

$$\beta_k(j) = P(Y_{k+1} = y_{k+1}, \cdots, Y_K = y_K | X_k = j, \lambda), \tag{2}$$

for $k = 1, 2, \cdots, K$. The backward algorithm consists of three stages:

1) Initialization stage

$$\beta_K(j) = P(Y_{K+1} = y_{K+1} | X_K = j, \lambda)$$
$$= 1, \quad \text{for } j = 1, 2, \cdots, N.$$

2) Induction stage

$$\beta_k(j) = P(Y_{k+1} = y_{k+1}, \cdots, Y_K = y_K | X_k = j, \lambda)$$
$$= \sum_{i=1}^{N} b_i(y_{k+1}) \beta_{k+1}(j) a_{ji}, \quad \text{for } j = 1, 2, \cdots, N \text{ and } k = K - 1, K - 2, \cdots, 2, 1.$$

3) Termination stage

$$P(Y_1 = y_1, \cdots, Y_K = y_K | \lambda) = \sum_{j=1}^{N} \pi_j b_j(y_1) \beta_1(j).$$

Forward and backward algorithms produce probability of observation sequence as follows

$$L_K = P(Y_1 = y_1, \cdots, Y_K = y_K | \lambda)$$
$$= \sum_{j=1}^{N} P(Y_1 = y_1, \cdots, Y_K = y_K, X_k = j | \lambda)$$
$$= \sum_{j=1}^{N} \alpha_k(j) \beta_k(j), \quad \text{for } k = 1, 2, \cdots, K. \tag{3}$$

2.3 Veterbi Algorithm

The selection of hidden state sequence $x_1, x_2, \cdots, x_K$ in order to optimize the probability of hidden state is done by using Veterbi algorithm. Defined variables

$$\delta_k(j) = \max_{x_1, \cdots, x_{k-1}} P(Y_1 = y_1, \cdots, Y_k = y_k, X_1 = x_1, \cdots, X_k = j | \lambda)$$

$$= \left(\max_{i=1,2,\cdots,N} \delta_{k-1}(i) a_{ij} \right) b_j(y_k), \quad \text{for } k = 1,2, \cdots, K.$$

The Veterbi algorithm consists of four stages:

1) Initialization stage

$$\delta_1(j) = P(Y_1 = y_1, X_1 = j)$$
$$= P(Y_1 = y_1 | X_1 = j) P(X_1 = j)$$
$$= \pi_j b_j(y_1),$$

$$\psi_1(j) = \arg \max_{i=1,2,\cdots,N} \delta_0(i) \, a_{ij}$$
$$= \emptyset.$$

2) Recursion stage

$$\delta_k(j) = b_j(y_k) \max_{1 \le i \le N} \{ a_{ij} \delta_{k-1}(i) \},$$

$$\psi_k(j) = \arg \max_{1 \le i \le N} \{ a_{ij} \delta_{k-1}(i) \}, \text{for } k = 2,3, \cdots, K-1.$$

3) Termination stage

$$P^* = \max_{1 \le i \le N} \{ \delta_K(i) \},$$

$${x_K}^* = \arg \max_{1 \le i \le N} \{ \delta_K(i) \}.$$

4) Backtracking stage

$${x_k}^* = \psi_{k+1}({x_{k+1}}^*), \text{for } k = K-1, K-2, \cdots 2,1.$$

The backtracking stage allows the optimal hidden state sequence to be found from the last point that is stored at the recursion stage.

2.4 Baum-Welch Algorithm

The problem that must be solved on the third problem is to obtain the best model that can explain an observation sequence $y_1, y_2, \cdots, y_K$. To solve this last problem, the Baum-Welch algorithm is used. Estimation problems are solved by the Maximum Likelihood (ML) method.

Defined likelihood function

$$\mathcal{L}(\lambda | Y_1 = y_1, Y_2 = y_2, \cdots, Y_K = y_K) = P(Y_1 = y_1, Y_2 = y_2, \cdots, Y_K = y_K | \lambda)$$

and

$$\hat{\lambda}_{ML} = \arg \max_{\lambda} P(Y_1 = y_1, Y_2 = y_2, \cdots, Y_K = y_K | \lambda).$$

Finding $\hat{\lambda}_{ML}$ cannot be done analytically. Hence, the numerical recursive method of Expectation Maximization (EM) is used. An initial value

$$\lambda = (A, B, \pi)$$

is given as the characteristic of the hidden Markov model by using Mendel's law. Then, we find the new estimation model parameter namely $\hat{\lambda}$. Furthermore, the $\hat{\lambda}$ model can be said to be a better model then the λ model if it fulfills

$$P(Y_1 = y_1, Y_2 = y_2, \cdots, Y_K = y_K | \hat{\lambda}) \ge P(Y_1 = y_1, Y_2 = y_2, \cdots, Y_K = y_K | \lambda).$$

If this process is done repeatedly using $\hat{\lambda}$ as λ for the new process, then the better probability value $Y_1 = y_1, Y_2 = y_2, \cdots, Y_K = y_K$ can be determined. The process is carried out until $\hat{\lambda}$, which maximizes probabilities

$Y_1 = y_1, Y_2 = y_2, \cdots, Y_K = y_K$, is obtained. The final result of re-estimating the parameters of the hidden Markov model is called the ML estimator.

The formula to re-estimate the parameters of the model is obtained by maximizing Baum-Welch's function

$$Q(\lambda, \hat{\lambda}) = \sum_x P(X_1 = x_1, \cdots, X_K = x_K | Y_1 = y_1, \cdots, Y_K = y_K, \lambda)$$
$$\log[P(Y_1 = y_1, \cdots, Y_K = y_K, X_1 = x_1, \cdots, X_K = x_K | \hat{\lambda})].$$

Maximizing Baum-Welch's function can cause an increase in the likelihood function, namely

$$\max_{\hat{\lambda}}[Q(\lambda, \hat{\lambda})] \rightarrow P(Y_1 = y_1, \cdots, Y_K = y_K | \hat{\lambda}) \geq P(Y_1 = y_1, \cdots, Y_K = y_K | \lambda).$$

The sequence $\{P(Y_1 = y_1, \cdots, Y_K = y_K | \hat{\lambda})\}$ is obtained so that it converges to a local maximum point. The formula to re-estimate the next parameters model can be expressed as an implementation of the EM algorithm.

Start   : given an estimator $\lambda = (\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\pi})$,

Step E : calculate the value

$$Q(\lambda, \hat{\lambda}) = \sum_x P(X_1 = x_1, \cdots, X_K = x_K | Y_1 = y_1, \cdots, Y_K = y_K, \lambda)$$
$$\log[P(Y_1 = y_1, \cdots, Y_K = y_K, X_1 = x_1, \cdots, X_K = x_K | \hat{\lambda})],$$

Step M : determine

$$\hat{\lambda}_{ML} = \arg \max_{\hat{\lambda}}[Q(\lambda, \hat{\lambda})] \rightarrow P(Y_1 = y_1, \cdots, Y_K = y_K | \hat{\lambda}) \geq P(Y_1 = y_1, \cdots, Y_K = y_K | \lambda).$$

If the sequence $\{P(Y_1 = y_1, \cdots, Y_K = y_K | \hat{\lambda})\}$ converges to a local maximum point, then it is done. Otherwise, $\hat{\lambda}$ is replaced with λ, then step E is repeated, and so on. Beside that, it is worth to note that in each iteration, there are conditions that must be fulfilled by the parameters of the new hidden Markov model, namely

$$\sum_{j=1}^{N} \hat{a}_{ij} = 1, \sum_{i=1}^{N} \hat{\pi}_i = 1, \text{ and } \sum_{j=1}^{N} \hat{b}_i(j) = 1 \quad \text{for } i = 1, 2, \cdots, N. \tag{4}$$

Looking at the estimation problem as an optimization constraint function of $P(Y_1 = y_1, \cdots, Y_K = y_K | \lambda)$, the Lagrange multiplication technique can be a solution to find the value of $\lambda = (\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\pi})$ that maximizes $P(Y_1 = y_1, \cdots, Y_K = y_K | \lambda)$. Based on the standard form of Lagrange optimization, Equation (4) is a constraint function and Equation $Q(\lambda, \hat{\lambda})$ is a maximized objective function. Defined variables

$$\xi_k(i, j) = P(X_k = i, X_{k+1} = j | Y_1 = y_1, \cdots, Y_K = y_K, \lambda) \tag{5}$$

for $i, j = 1, 2, \cdots, N$ and $k = 1, 2, \cdots, K$, is probability of i state at k time and j state at $k + 1$ time. If the λ model and observation sequence is given, by using **Equation (1)**, **Equation (2)**, **Equation (3)**, and **Equation (5)**, can be obtained

$$\xi_k(i, j) = \frac{\alpha_k(i) a_{ij} b_j(y_{k+1}) \beta_{k+1}(j)}{\sum_{j=1}^{N} \alpha_k(j) \beta_k(j)}.$$

Defined variables

$$\gamma_k(i) = P(X_k = i | Y_1 = y_1, \cdots, Y_K = y_K, \lambda)$$

for $k = 1, 2, \cdots, K$, is probability of i state at k time if given an observation sequence $Y_1 = y_1, \cdots, Y_K = y_K$ and λ model, so can be obtained

$$\gamma_k(i) = P(X_k = i | Y_1 = y_1, \cdots, Y_K = y_K, \lambda)$$
$$= \frac{\alpha_k(i) \beta_k(i)}{\sum_{i=1}^{N} \alpha_k(i) \beta_k(i)}.$$

Then, the relationship between $\xi_k(i, j)$ and $\gamma_k(i)$ is obtained by summing up $\xi_k(i, j)$ for each j as follows:

$$\gamma_k(i) = P(X_k = i | Y_1 = y_1, \cdots, Y_K = y_K, \lambda)$$
$$= \sum_{j=1}^{N} \xi_k(i, j),$$

for $i = 1,2, \cdots, N$ and $k = 1,2, \cdots, K$. Summing up $\xi_k(i,j)$ on $1 \leq k \leq K - 1$ produces the probability value of moving i state to j state until $K - 1$ time, as follows:

$$\sum_{k=1}^{K-1} \xi_k(i,j) = \sum_{k=1}^{K-1} P(X_k = i, X_{k+1} = j | Y_1 = y_1, \cdots, Y_K = y_K, \lambda).$$

Summing up $\gamma_k(i)$ on $1 \leq k \leq K - 1$ produces the probability value at i state until $K - 1$ time, as follows:

$$\sum_{k=1}^{K-1} \gamma_k(i) = \sum_{k=1}^{K-1} P(X_k = i | Y_1 = y_1, \cdots, Y_K = y_K, \lambda).$$

The formula to re-estimate the parameters of the hidden Markov model is as follows:

$\hat{\pi}_i$ = the probability value in i state at $k = 1$ time

$\quad = P(X_1 = i | Y_1 = y_1, \cdots, Y_K = y_K, \lambda)$

$\quad = \gamma_1(i)$, for $i = 1,2, \cdots, N$.

$$\hat{a}_{ij} = \frac{\text{the transition probability of } i \text{ state to } j \text{ state until } K - 1 \text{ time}}{\text{the probability value in } i \text{ state until } K - 1 \text{ time}}$$

$$= \frac{\sum_{k=1}^{K-1} \xi_k(i,j)}{\sum_{k=1}^{K-1} \gamma_k(i)}, \quad \text{for } i = 1,2, \cdots, N \text{ and } j = 1,2, \cdots, K.$$

$$\hat{b}_i(j) = \frac{\text{the probability of appearance } X_k = i \text{ and } Y_k = j \text{ until } K \text{ time}}{\text{the probability value in } i \text{ state until } K \text{ time}}$$

$$= \frac{\sum_{k=1, s.t \, y_k=j}^{K} \gamma_k(i)}{\sum_{k=1}^{K} \gamma_k(i)}.$$

3. RESULTS AND DISCUSSION

3.1 Mendel's Law

Gregor Mendel developed the theory of inheritance by experimenting with various varieties, which for several generations carried out self-fertilization and only produced the same variety as its parent plant. Plants that have these properties are called true breeding. In his experiments, Mendel crossed two contradicting variations of true breeding in peas. The crossing of two true-breeding varieties is called hybridization. The true-breeding parent is called a parental generation (P generation), hybrid offspring is called the first filial generation (F_1 generation), and the offspring produced from self-fertilization F_1 hybrids are called the second filial generation (F_2 generation). Mendel concluded the result of quantitative analysis of F_2, which has become the fundamental principle of heredity known as segregation law and free sorting law.

He describes four concepts used to develop a model that explains the pattern of trait inheritance. First, alternative versions of genes (alleles) cause variations in inherited characters. For example, at the color locus of the peas flower, there are alleles of purple flowers and alleles of white flowers. Second, for each character, the organism inherits two alleles, one from each parent. Every somatic cell in a diploid organism has two set of chromosomes, one set inherited from each parent. Third, if the two alleles at a locus are different, then one of them is the dominant allele that determines the organism's appearance, and the other is the recessive allele that has no visible effect on the appearance of the organism. Fourth, at the same time, the law of segregation from Mendel state that two alleles for an inherited character segregate during gametes formation, and eventually, it is at the different gametes.

Organisms that have a pair of identical alleles for a character are called homozygotes. Meanwhile, organisms with two different alleles for a character are called heterozygotes. The characteristics of an organism do not always reveal its genetic composition. Therefore, the appearance of the observed trait in the organism, namely the phenotype, can be distinguished from its genetic structure, namely the genotype. Individuals that are heterozygotes only for one character are called monohybrids, and crosses between heterozygotes are called monohybrid crosses. Likewise, individuals that heterozygotes for two characters are dihybrids, while three characters are called trihybrids. The results of Mendel's dihybrid experiments are the basis of the free sorting law which states that each pair of alleles segregates freely against other alleles during gamete formation [1].

3.1.1 Monohybrid Crosses

P_1 : $AA \times aa$
F_1 : Aa
P_2 : $Aa \times Aa$
Gamete : (A, a) (A, a)

Table 1. The Second Filial Generation of Monohybrid Crosses

	A	a
A	AA	Aa
a	Aa	aa

The random variable X_k is the kth progenic genotype and $\{X_k\}$ is a Markov chain with state space $S_X = \{AA, Aa, aa\}$.

The random variable Y_k is the kth progenic phenotype and $\{Y_k\}$ is an observation process with state space $S_Y = \{A, a\}$.

The observation Y_k is influenced by X_k, while X_k is influenced by X_{k-1}. **Figure 2** shows this relationship for a case when the observation (phenotype) sequence is $A, A, \cdots, A$.
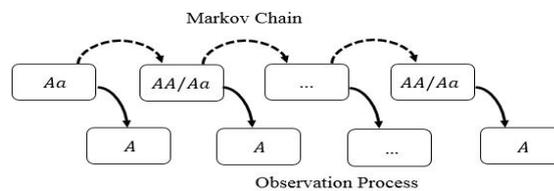

Figure 2. Hidden Markov Model of Monohybrid Crosses

Based on **Figure 2** it is known that in the first generation, there is dominant phenotype A and heterozygote genotype Aa. The second generation has dominant phenotype A, but its genotype is unknown whether it is heterozygote Aa or homozygote AA. This happens until k^{th} generation.

The characteristic of hidden Markov model for monohybrid crosses are as follows:

1) The determination of matrix A

The first is determining matrix $A = \left(a_{ij}\right)_{N \times N}$. The dimensions of matrix A is determined by the number of state space S_X, that is $N = 3$. Then, a_{ij} represents a conditional probability of genotype j with genotype i given. Index i represents a genotype of an organism at $k + 1$ time, while index j represents genotype of an organism at k time. Matrix A is obtained by using Mendel's law (**Table 2**).

Table 2. The Mendel's Law of Monohybrid Crosses

AA (A) $\times$ AA (A)	Aa (A, a) $\times$ Aa (A, a)	aa (a) $\times$ aa (a)
AA 100%	$AA \left(\frac{1}{4}\right), Aa \left(\frac{1}{2}\right), aa \left(\frac{1}{4}\right)$	aa 100%
$a_{AA,j} = [1 \quad 0 \quad 0]$	$a_{Aa,j} = \left[\frac{1}{4} \quad \frac{1}{2} \quad \frac{1}{4}\right]$	$a_{aa,j} = [0 \quad 0 \quad 1]$

So, we have the transition probability matrix as follows:

$$A = \begin{bmatrix} a_{AA,AA} & a_{AA,Aa} & a_{AA,aa} \\ a_{Aa,AA} & a_{Aa,Aa} & a_{Aa,aa} \\ a_{aa,AA} & a_{aa,Aa} & a_{aa,aa} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 1 \end{bmatrix}.$$

2) The determination of matrix B

The second characteristic is determining matrix $B = \left(b_i(j)\right)_{N \times M}$. The dimensions of matrix B is determined by the number of state space S_X and S_Y, that is $N = 3$ and $M = 2$. Then, $b_i(j)$ represents a conditional probability of phenotype j with genotype i given. Index i represents a genotype of an organism at k time while index j represents phenotype of an organism at the same time. The first column in matrix B tells that type AA and Aa of genotype will be observed as phenotype type A, whereas the second column

tells that type aa of genotype will be observed as phenotype type a. Thus, we have obtained the emission probability matrix as follows:

$$\boldsymbol{B} = \begin{bmatrix} b_{AA}(A) & b_{AA}(a) \\ b_{Aa}(A) & b_{Aa}(a) \\ b_{aa}(A) & b_{aa}(a) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

3) The determination of matrix $\boldsymbol{\pi}$

The third is determining matrix $\boldsymbol{\pi} = (\pi_i)_{N \times 1}$. The dimensions of matrix π is determined by the number of state space S_X, that is $N = 3$. Then, π_i represents an initial probability of genotype i. Index i represents a genotype of an organism at an initial time. Since Mendel started with a founder population of heterozygote plants, the initial value of π_i is zero to homozygote plants and one to heterozygote plants. Hence, we have obtained the initializing parameter of matrix $\boldsymbol{\pi}$ for monohybrid crosses, which is as follows:

$$\pi = \begin{bmatrix} \pi_{AA} \\ \pi_{Aa} \\ \pi_{aa} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}.$$

3.1.2 Dihybrid Crosses

P_1　　:　　$AABB$　　$\times$　　$aabb$
F_1　　:　　　　　$AaBb$
P_2　　:　　$AaBb$　　$\times$　　$AaBb$
Gamete : (AB, Ab, aB, ab)　(AB, Ab, aB, ab)

Table 3. The Second Filial Generation of Dihybrid Crosses

	AB	Ab	aB	ab
AB	$AABB$	$AABb$	$AaBB$	$AaBb$
Ab	$AABb$	$AAbb$	$AaBb$	$Aabb$
aB	$AaBB$	$AaBb$	$aaBB$	$aaBb$
ab	$AaBb$	$Aabb$	$aaBb$	$aabb$

The random variable X_k is the kth progenic genotype and $\{X_k\}$ is a Markov chain with state space
$$S_X = \{AABB, AaBB, aaBB, AABb, AaBb, aaBb, AAbb, Aabb, aabb\}.$$

The random variable Y_k is the kth progenic phenotype and $\{Y_k\}$ is an observation process with state space
$$S_Y = \{AB, Ab, aB, ab\}.$$

The observation Y_k is influenced by X_k, while X_k is influenced by X_{k-1}. **Figure 3** shows this relationship for a case when the observation (phenotype) sequence is $AB, AB, \cdots, AB$.

Figure 3. Hidden Markov Model of Dihybrid Crosses

Based on **Figure 3** it is known that in the first generation, there is dominant phenotype AB and heterozygote genotype $AaBb$. The second generation has dominant phenotype AB, but its genotype is unknown whether it is heterozygote $AaBB, AABb, AaBb$ or homozygote $AABB$. This happens until k^{th} generation. The initializing parameters of the hidden Markov model on dihybrid crosses can be obtained in the same way with monohybrid crosses. So, we have the initializing parameters as follows:

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & 0 & 0 & \frac{1}{2} & 0 & 0 & \frac{1}{4} & 0 & 0 \\ \frac{1}{16} & \frac{1}{8} & \frac{1}{16} & \frac{1}{8} & \frac{1}{4} & \frac{1}{8} & \frac{1}{16} & \frac{1}{8} & \frac{1}{16} \\ 0 & 0 & \frac{1}{4} & 0 & 0 & \frac{1}{2} & 0 & 0 & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, B = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ and } \pi = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

3.1.3 Trihybrid Crosses

P_1 : $AABBCC$ × $aabbcc$
F_1 : $AaBbCc$
P_2 : $AaBbCc$ × $AaBbCc$
Gamete : $(ABC, ABc, AbC, Abc, aBC, aBc, abC, abc)$ $(ABC, ABc, AbC, Abc, aBC, aBc, abC, abc)$

Table 4. The Second Filial Generation of Trihybrid Crosses

	ABC	ABc	AbC	Abc	aBC	aBc	abC	abc
ABC	AABBCC	AABBCc	AABbCC	AABbCc	AaBBCC	AaBBCc	AaBbCC	AaBbCc
ABc	AABBCc	AABBcc	AABbCc	AABbcc	AaBBCc	AaBBcc	AaBbCc	AaBbcc
AbC	AABbCC	AABbCc	AAbbCC	AAbbCc	AaBbCC	AaBbCc	AabbCC	AabbCc
Abc	AABbCc	AABbcc	AAbbCc	AAbbcc	AaBbCc	AaBbcc	AabbCc	Aabbcc
aBC	AaBBCC	AaBBCc	AaBbCC	AaBbCc	aaBBCC	aaBBCc	aaBbCC	aaBbCc
aBc	AaBBCc	AaBBcc	AaBbCc	AaBbcc	aaBBCc	aaBBcc	aaBbCc	aaBbcc
abC	AaBbCC	AaBbCc	AabbCC	AabbCc	aaBbCC	aaBbCc	aabbCC	aabbCc
abc	AaBbCc	AaBbcc	AabbCc	Aabbcc	aaBbCc	aaBbcc	aabbCc	aabbcc

The random variable X_k is the kth progenic genotype and $\{X_k\}$ is a Markov chain with state space
$$S_X = \{AABBCC, AABBCc, AABBcc, AABbCC, AABbCc, AABbcc, AAbbCC, AAbbCc, AAbbcc,$$

$$AaBBCC, AaBBCc, AaBBcc, AaBbCC, AaBbCc, AaBbcc, AabbCC, AabbCc, Aabbcc,$$

$$aaBBCC, aaBBCc, aaBBcc, aaBbCC, aaBbCc, aaBbcc, aabbCC, aabbCc, aabbcc\}.$$

The random variable Y_k is the kth progenic phenotype and $\{Y_k\}$ is an observation process with state space
$$S_Y = \{ABC, ABc, AbC, Abc, aBC, aBc, abC, abc\}.$$

The observation Y_k is influenced by X_k, while X_k is influenced by X_{k-1}. Figure 4 shows this relationship for a case when the observation (phenotype) sequence is $ABC, ABC, \cdots, ABC$.
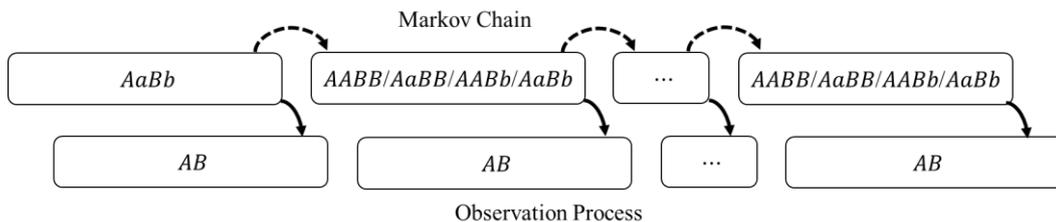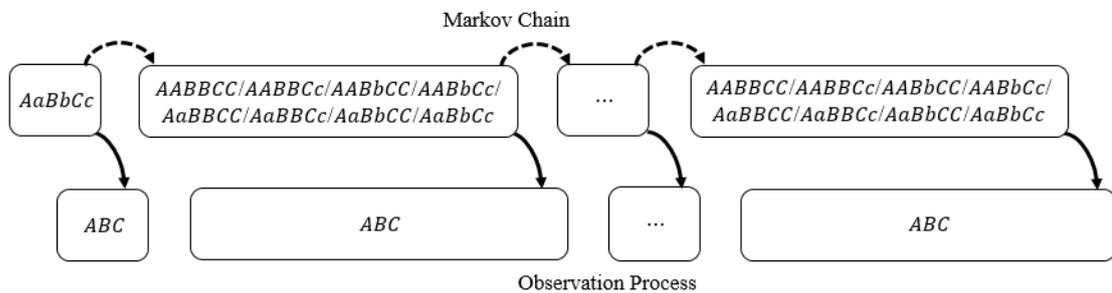

Figure 4. Hidden Markov Model of Trihybrid Crosses

Based on **Figure 4** it is known that in the first generation, there is dominant phenotype ABC and heterozygote genotype $AaBbCc$. The second generation has dominant phenotype ABC, but its genotype is unknown whether it is heterozygote $AABBCc, AABbCC, AABbCc, AaBBCC, AaBBCc, AaBbCC, AaBbCc$ or homozygote $AABBCC$. This happens until k^{th} generation. The initializing parameters of the hidden Markov model on trihybrid crosses can be obtained in the same way with monohybrid crosses. So, we have the initializing parameters as follows:

$$A = \begin{bmatrix} A_1 & 0 & 0 \\ A_2 & A_3 & A_2 \\ 0 & 0 & A_1 \end{bmatrix}, B = \begin{bmatrix} B_1 \\ B_1 \\ B_2 \end{bmatrix} \text{ and } \pi = \begin{bmatrix} 0 \\ \pi_1 \\ 0 \end{bmatrix} \text{ with}$$

$$A_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & 0 & 0 & \frac{1}{2} & 0 & 0 & \frac{1}{4} & 0 & 0 \\ \frac{1}{16} & \frac{1}{8} & \frac{1}{16} & \frac{1}{8} & \frac{1}{4} & \frac{1}{8} & \frac{1}{16} & \frac{1}{8} & \frac{1}{16} \\ 0 & 0 & \frac{1}{4} & 0 & 0 & \frac{1}{2} & 0 & 0 & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, A_2 = \begin{bmatrix} \frac{1}{4} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{16} & \frac{1}{8} & \frac{1}{16} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{16} & 0 & 0 & \frac{1}{8} & 0 & 0 & \frac{1}{16} & 0 & 0 \\ \frac{1}{64} & \frac{1}{32} & \frac{1}{64} & \frac{1}{32} & \frac{1}{16} & \frac{1}{32} & \frac{1}{64} & \frac{1}{32} & \frac{1}{64} \\ 0 & 0 & \frac{1}{16} & 0 & 0 & \frac{1}{8} & 0 & 0 & \frac{1}{16} \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{16} & \frac{1}{8} & \frac{1}{16} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{4} \end{bmatrix},$$

$$A_3 = \begin{bmatrix} \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{8} & \frac{1}{4} & \frac{1}{8} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{8} & 0 & 0 & \frac{1}{4} & 0 & 0 & \frac{1}{8} & 0 & 0 \\ \frac{1}{32} & \frac{1}{16} & \frac{1}{32} & \frac{1}{16} & \frac{1}{8} & \frac{1}{16} & \frac{1}{32} & \frac{1}{16} & \frac{1}{32} \\ 0 & 0 & \frac{1}{8} & 0 & 0 & \frac{1}{4} & 0 & 0 & \frac{1}{8} \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{8} & \frac{1}{4} & \frac{1}{8} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} \end{bmatrix}, B_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$B_2 = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \text{ and } \pi_1 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Searching for the genotype, from generation to generation, is done to obtain good quality seeds (dominant homogenous of genotype). Although the phenotype that is planted in k^{th} generation is dominant, it is uncertain whether it is good quality seeds. The fewer generations that are necessary to be planted to obtain quality seeds, the better.

3.2 True Breeding Generation in Diploid Plant Crosses

Suppose the discrete hidden Markov model is applied to monohybrid, dihybrid and trihybrid crosses. With forward, backward, Veterbi and Baum-Welch algorithms, and A, B, and π in Subsection 2.2, 2.3, and 2.4 in Section 2 as initial values which are obtained by Mendel's law, the conditional probability of $P = (Y_{k+1} = y_{k+1}|Y_1 = y_1, \cdots, Y_k = y_k)$ and $P = (X_1 = x_1, \cdots, X_k = x_k|Y_1 = y_1, \cdots, Y_k = y_k)$ can be calculated as follows:

3.2.1 Monohybrid Crosses

We have obtained the optimal value of $\hat{\lambda} = (\hat{A}, \hat{B}, \hat{\pi})$ for monohybrid crosses over 2734 plants during twenty generations by using package Mathematica 11 software, as follows:

$$\hat{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0.254892 & 0.49355 & 0.251558 \\ 0 & 0 & 1 \end{bmatrix}, \hat{B} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \text{ and } \hat{\pi} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}.$$

Here, we have obtained the conditional probability $P = (Y_{k+1} = y_{k+1}|Y_1 = y_1, \cdots, Y_k = y_k)$ and $P = (X_1 = x_1, \cdots, X_k = x_k|Y_1 = y_1, \cdots, Y_k = y_k)$ on monohybrid crosses (**Table 5**). To simplify the notation, it is defined that $Y_1^n = A$ represents $Y_1 = A, Y_2 = A, \cdots, Y_n = A$ and $X_2^n = AA$ represents $X_2 = AA, X_3 = AA, \cdots, X_n = AA$.

Table 5. The Conditional Probability of Monohybrid Crosses

Probability	Value	Probability	Value
$P(Y_2 = A\|Y_1 = A)$	0.748442	$P(X_1 = Aa, X_2 = AA\|Y_1{}^2 = A)$	0.340563
$P(Y_3 = A\|Y_1{}^2 = A)$	0.834113	$P(X_1 = Aa, X_2{}^3 = AA\|Y_1{}^3 = A)$	0.408294
$P(Y_4 = A\|Y_1{}^3 = A)$	0.901844	$P(X_1 = Aa, X_2{}^4 = AA\|Y_1{}^4 = A)$	0.452732
$P(Y_5 = A\|Y_1{}^4 = A)$	0.946282	$P(X_1 = Aa, X_2{}^5 = AA\|Y_1{}^5 = A)$	0.478433
$P(Y_6 = A\|Y_1{}^5 = A)$	0.971983	$P(X_1 = Aa, X_2{}^6 = AA\|Y_1{}^6 = A)$	0.492223
$P(Y_7 = A\|Y_1{}^6 = A)$	**0.985773**	$P(X_1 = Aa, X_2{}^7 = AA\|Y_1{}^7 = A)$	**0.499327**
$P(Y_8 = A\|Y_1{}^7 = A)$	0.992877	$P(X_1 = Aa, X_2{}^8 = AA\|Y_1{}^8 = A)$	0.502909
$P(Y_9 = A\|Y_1{}^8 = A)$	0.996459	$P(X_1 = Aa, X_2{}^9 = AA\|Y_1{}^9 = A)$	0.504696
$P(Y_{10} = A\|Y_1{}^9 = A)$	0.998246	$P(X_1 = Aa, X_2{}^{10} = AA\|Y_1{}^{10} = A)$	0.505583
$P(Y_{11} = A\|Y_1{}^{10} = A)$	0.999133	$P(X_1 = Aa, X_2{}^{11} = AA\|Y_1{}^{11} = A)$	0.506022
$P(Y_{12} = A\|Y_1{}^{11} = A)$	0.999672	$P(X_1 = Aa, X_2{}^{12} = AA\|Y_1{}^{12} = A)$	0.506239
$P(Y_{13} = A\|Y_1{}^{12} = A)$	0.999789	$P(X_1 = Aa, X_2{}^{13} = AA\|Y_1{}^{13} = A)$	0.506346
$P(Y_{14} = A\|Y_1{}^{13} = A)$	0.999896	$P(X_1 = Aa, X_2{}^{14} = AA\|Y_1{}^{14} = A)$	0.506398
$P(Y_{15} = A\|Y_1{}^{14} = A)$	0.999948	$P(X_1 = Aa, X_2{}^{15} = AA\|Y_1{}^{15} = A)$	0.506425
$P(Y_{16} = A\|Y_1{}^{15} = A)$	0.999975	$P(X_1 = Aa, X_2{}^{16} = AA\|Y_1{}^{16} = A)$	0.506437
$P(Y_{17} = A\|Y_1{}^{16} = A)$	0.999987	$P(X_1 = Aa, X_2{}^{17} = AA\|Y_1{}^{17} = A)$	0.506444
$P(Y_{18} = A\|Y_1{}^{17} = A)$	0.999994	$P(X_1 = Aa, X_2{}^{18} = AA\|Y_1{}^{18} = A)$	0.506447
$P(Y_{19} = A\|Y_1{}^{18} = A)$	0.999997	$P(X_1 = Aa, X_2{}^{19} = AA\|Y_1{}^{19} = A)$	0.506448
$P(Y_{20} = A\|Y_1{}^{19} = A)$	0.999998	$P(X_1 = Aa, X_2{}^{20} = AA\|Y_1{}^{20} = A)$	0.506449

From **Table 5**, it is known that the seventh generation monohybrid crosses has the probability of a dominant phenotype of 99% but the probability of an optimal genotype is only 50%. With more generations of plants, the probability of the plants having a dominant phenotype and an optimal genotype is bigger. However, the increase is very slow.

3.2.2 Dihybrid Crosses

We have obtained the optimal value of $\hat{\lambda} = (\widehat{A}, \widehat{B}, \widehat{\pi})$ for dihybrid crosses over 2734 plants during twenty generations by using package Mathematica 11 software, as follows:

$$\widehat{A} = \begin{bmatrix} \widehat{A}_1 & 0 & 0 \\ \widehat{A}_2 & \widehat{A}_3 & \widehat{A}_4 \\ 0 & 0 & \widehat{A}_5 \end{bmatrix}, \widehat{B} = \begin{bmatrix} \widehat{B}_1 \\ \widehat{B}_1 \\ \widehat{B}_2 \end{bmatrix} \text{ and } \widehat{\pi} = \begin{bmatrix} 0 \\ \widehat{\pi}_1 \\ 0 \end{bmatrix} \text{ with }$$

$$\widehat{A}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0.274512 & 0.515042 & 0.210446 \\ 0 & 0 & 1 \end{bmatrix}, \widehat{A}_2 = \begin{bmatrix} 0.257295 & 0 & 0 \\ 0.066717 & 0.117863 & 0.065342 \\ 0 & 0 & 0.266882 \end{bmatrix},$$

$$\widehat{A}_3 = \begin{bmatrix} 0.507058 & 0 & 0 \\ 0.12782 & 0.241037 & 0.125859 \\ 0 & 0 & 0.48594 \end{bmatrix}, \widehat{A}_4 = \begin{bmatrix} 0.235647 & 0 & 0 \\ 0.064249 & 0.132817 & 0.058296 \\ 0 & 0 & 0.247178 \end{bmatrix},$$

$$\widehat{A}_5 = \begin{bmatrix} 1 & 0 & 0 \\ 0.241372 & 0.501032 & 0.257596 \\ 0 & 0 & 1 \end{bmatrix}, \widehat{B}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \widehat{B}_2 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ and } \widehat{\pi}_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}.$$

Here, we have obtained the conditional probability $P = (Y_{k+1} = y_{k+1}|Y_1 = y_1, \cdots, Y_k = y_k)$ and $P = (X_1 = x_1, \cdots, X_k = x_k|Y_1 = y_1, \cdots, Y_k = y_k)$ on dihybrid crosses (**Table 6**). To simplify the notation, it is defined that $Y_1{}^n = AB$ represents $Y_1 = AB, Y_2 = AB, \cdots, Y_n = AB$ and $X_2{}^n = AABB$ represents $X_2 = AABB, X_3 = AABB, \cdots, X_n = AABB$.

Table 6. The Conditional Probability of Dihybrid Crosses

Probability	Value	Probability	Value
$P(Y_2 = AB\|Y_1 = AB)$	0.553437	$P(X_1 = AaBb, X_2 = AABB\|Y_1{}^2 = AB)$	0.120550
$P(Y_3 = AB\|Y_1{}^2 = AB)$	0.706268	$P(X_1 = AaBb, X_2{}^3 = AABB\|Y_1{}^3 = AB)$	0.170686
$P(Y_4 = AB\|Y_1{}^3 = AB)$	0.827998	$P(X_1 = AaBb, X_2{}^4 = AABB\|Y_1{}^4 = AB)$	0.206144
$P(Y_5 = AB\|Y_1{}^4 = AB)$	0.905671	$P(X_1 = AaBb, X_2{}^5 = AABB\|Y_1{}^5 = AB)$	0.227614
$P(Y_6 = AB\|Y_1{}^5 = AB)$	0.949937	$P(X_1 = AaBb, X_2{}^6 = AABB\|Y_1{}^6 = AB)$	0.239610
$P(Y_7 = AB\|Y_1{}^6 = AB)$	0.973878	$P(X_1 = AaBb, X_2{}^7 = AABB\|Y_1{}^7 = AB)$	0.246037

$P(Y_8 = AB\|Y_1^7 = AB)$	**0.986496**	$P(X_1 = AaBb, X_2^8 = AABB\|Y_1^8 = AB)$	**0.249405**
$P(Y_9 = AB\|Y_1^8 = AB)$	0.993056	$P(X_1 = AaBb, X_2^9 = AABB\|Y_1^9 = AB)$	0.251149
$P(Y_{10} = AB\|Y_1^9 = AB)$	0.996440	$P(X_1 = AaBb, X_2^{10} = AABB\|Y_1^{10} = AB)$	0.252046
$P(Y_{11} = AB\|Y_1^{10} = AB)$	0.998178	$P(X_1 = AaBb, X_2^{11} = AABB\|Y_1^{11} = AB)$	0.252506
$P(Y_{12} = AB\|Y_1^{11} = AB)$	0.999068	$P(X_1 = AaBb, X_2^{12} = AABB\|Y_1^{12} = AB)$	0.252742
$P(Y_{13} = AB\|Y_1^{12} = AB)$	0.999524	$P(X_1 = AaBb, X_2^{13} = AABB\|Y_1^{13} = AB)$	0.252862
$P(Y_{14} = AB\|Y_1^{13} = AB)$	0.999756	$P(X_1 = AaBb, X_2^{14} = AABB\|Y_1^{14} = AB)$	0.252924
$P(Y_{15} = AB\|Y_1^{14} = AB)$	0.999876	$P(X_1 = AaBb, X_2^{15} = AABB\|Y_1^{15} = AB)$	0.252956
$P(Y_{16} = AB\|Y_1^{15} = AB)$	0.999936	$P(X_1 = AaBb, X_2^{16} = AABB\|Y_1^{16} = AB)$	0.252972
$P(Y_{17} = AB\|Y_1^{16} = AB)$	0.999967	$P(X_1 = AaBb, X_2^{17} = AABB\|Y_1^{17} = AB)$	0.252980
$P(Y_{18} = AB\|Y_1^{17} = AB)$	0.999983	$P(X_1 = AaBb, X_2^{18} = AABB\|Y_1^{18} = AB)$	0.252984
$P(Y_{19} = AB\|Y_1^{18} = AB)$	0.999992	$P(X_1 = AaBb, X_2^{19} = AABB\|Y_1^{19} = AB)$	0.252986
$P(Y_{20} = AB\|Y_1^{19} = AB)$	0.999996	$P(X_1 = AaBb, X_2^{20} = AABB\|Y_1^{20} = AB)$	0.252987

From **Table 6**, it is known that the eighth generation dihybrid crosses has the probability of a dominant phenotype of 99% but the probability of an optimal genotype is only 25%. With more generations of plants, the probability of the plants having a dominant phenotype and an optimal genotype is rising. However, the growth is slow-moving.

3.2.3 Trihybrid Crosses

We have obtained the optimal value of $\hat{\lambda} = (\hat{A}, \hat{B}, \hat{\pi})$ for trihybrid crosses over 2734 plants during twenty generations by using package Mathematica 11 software, as follows:

$$\hat{A} = \begin{bmatrix} \hat{A}_1 & 0 & 0 \\ \hat{A}_2 & \hat{A}_3 & \hat{A}_4 \\ 0 & 0 & \hat{A}_5 \end{bmatrix}, \hat{B} = \begin{bmatrix} \hat{B}_1 \\ \hat{B}_1 \\ \hat{B}_2 \end{bmatrix} \text{ and } \hat{\pi} = \begin{bmatrix} 0 \\ \hat{\pi}_1 \\ 0 \end{bmatrix} \text{ with } \hat{A}_1 = \begin{bmatrix} \hat{A}_{11} & 0 & 0 \\ \hat{A}_{12} & \hat{A}_{13} & \hat{A}_{14} \\ 0 & 0 & \hat{A}_{15} \end{bmatrix}, \hat{A}_2 = \begin{bmatrix} \hat{A}_{21} & 0 & 0 \\ \hat{A}_{22} & \hat{A}_{23} & \hat{A}_{24} \\ 0 & 0 & \hat{A}_{25} \end{bmatrix},$$

$$\hat{A}_3 = \begin{bmatrix} \hat{A}_{31} & 0 & 0 \\ \hat{A}_{32} & \hat{A}_{33} & \hat{A}_{34} \\ 0 & 0 & \hat{A}_{35} \end{bmatrix}, \hat{A}_4 = \begin{bmatrix} \hat{A}_{41} & 0 & 0 \\ \hat{A}_{42} & \hat{A}_{43} & \hat{A}_{44} \\ 0 & 0 & \hat{A}_{45} \end{bmatrix}, \hat{A}_5 = \begin{bmatrix} \hat{A}_{51} & 0 & 0 \\ \hat{A}_{52} & \hat{A}_{53} & \hat{A}_{54} \\ 0 & 0 & \hat{A}_{55} \end{bmatrix}, \hat{B}_1 = \begin{bmatrix} \hat{B}_{11} \\ \hat{B}_{11} \\ \hat{B}_{12} \end{bmatrix}, \hat{B}_2 = \begin{bmatrix} \hat{B}_{21} \\ \hat{B}_{21} \\ \hat{B}_{22} \end{bmatrix},$$

$$\hat{\pi}_1 = \begin{bmatrix} 0 \\ \hat{\pi}_{11} \\ 0 \end{bmatrix}, \hat{A}_{11} = \begin{bmatrix} 1 & 0 & 0 \\ 0.241436 & 0.466738 & 0.291826 \\ 0 & 0 & 1 \end{bmatrix}, \hat{A}_{12} = \begin{bmatrix} 0.23068 & 0 & 0 \\ 0.0681 & 0.18086 & 0.048 \\ 0 & 0 & 0.265266 \end{bmatrix},$$

$$\hat{A}_{13} = \begin{bmatrix} 0.415959 & 0 & 0 \\ 0.2438 & 0.13 & 0.11354 \\ 0 & 0 & 0.42853 \end{bmatrix}, \qquad \hat{A}_{14} = \begin{bmatrix} 0.353361 & 0 & 0 \\ 0.0381 & 0.1046 & 0.073 \\ 0 & 0 & 0.306204 \end{bmatrix},$$

$$\hat{A}_{15} = \begin{bmatrix} 1 & 0 & 0 \\ 0.2602 & 0.42375 & 0.31605 \\ 0 & 0 & 1 \end{bmatrix}, \qquad \hat{A}_{21} = \begin{bmatrix} 0.276713 & 0 & 0 \\ 0.058241 & 0.143769 & 0.046177 \\ 0 & 0 & 0.26342 \end{bmatrix},$$

$$\hat{A}_{22} = \begin{bmatrix} 0.04922 & 0 & 0 \\ 0.0146097 & 0.0288199 & 0.0161306 \\ 0 & 0 & 0.07683 \end{bmatrix}, \hat{A}_{23} = \begin{bmatrix} 0.161175 & 0 & 0 \\ 0.0282563 & 0.0622467 & 0.0285505 \\ 0 & 0 & 0.24469 \end{bmatrix},$$

$$\hat{A}_{24} = \begin{bmatrix} 0.0353531 & 0 & 0 \\ 0.0166043 & 0.036169 & 0.0150533 \\ 0 & 0 & 0.03124 \end{bmatrix}, \hat{A}_{25} = \begin{bmatrix} 0.242718 & 0 & 0 \\ 0.06527 & 0.14009 & 0.0793 \\ 0 & 0 & 0.21098 \end{bmatrix},$$

$$\hat{A}_{31} = \begin{bmatrix} 0.35587 & 0 & 0 \\ 0.201831 & 0.208235 & 0.121611 \\ 0 & 0 & 0.44262 \end{bmatrix}, \hat{A}_{32} = \begin{bmatrix} 0.15848 & 0 & 0 \\ 0.0241246 & 0.0662044 & 0.0292756 \\ 0 & 0 & 0.199195 \end{bmatrix},$$

$$\hat{A}_{33} = \begin{bmatrix} 0.314799 & 0 & 0 \\ 0.0677946 & 0.115186 & 0.0623103 \\ 0 & 0 & 0.10961 \end{bmatrix}, \hat{A}_{34} = \begin{bmatrix} 0.081455 & 0 & 0 \\ 0.0352632 & 0.0668199 & 0.0308375 \\ 0 & 0 & 0.116435 \end{bmatrix},$$

$$\hat{A}_{35} = \begin{bmatrix} 0.498554 & 0 & 0 \\ 0.11702 & 0.212 & 0.11109 \\ 0 & 0 & 0.56514 \end{bmatrix}, \hat{A}_{41} = \begin{bmatrix} 0.367417 & 0 & 0 \\ 0.026431 & 0.130234 & 0.063471 \\ 0 & 0 & 0.29396 \end{bmatrix},$$

$$\hat{A}_{42} = \begin{bmatrix} 0.02725 & 0 & 0 \\ 0.0189525 & 0.0294926 & 0.0146145 \\ 0 & 0 & 0.04226 \end{bmatrix}, \hat{A}_{43} = \begin{bmatrix} 0.109018 & 0 & 0 \\ 0.0328038 & 0.0637264 & 0.0338015 \\ 0 & 0 & 0.10112 \end{bmatrix},$$

$$\widehat{A}_{44} = \begin{bmatrix} 0.0632499 & 0 & 0 \\ 0.0139362 & 0.029969 & 0.0184471 \\ 0 & 0 & 0.07862 \end{bmatrix}, \quad \widehat{A}_{45} = \begin{bmatrix} 0.258728 & 0 & 0 \\ 0.08733 & 0.12302 & 0.06488 \\ 0 & 0 & 0.22388 \end{bmatrix},$$

$$\widehat{A}_{51} = \begin{bmatrix} 1 & 0 & 0 \\ 0.216276 & 0.529313 & 0.254411 \\ 0 & 0 & 1 \end{bmatrix}, \quad \widehat{A}_{52} = \begin{bmatrix} 0.267289 & 0 & 0 \\ 0.080306 & 0.205324 & 0.0249849 \\ 0 & 0 & 0.217642 \end{bmatrix},$$

$$\widehat{A}_{53} = \begin{bmatrix} 0.471316 & 0 & 0 \\ 0.11792 & 0.189926 & 0.0967128 \\ 0 & 0 & 0.463864 \end{bmatrix}, \quad \widehat{A}_{54} = \begin{bmatrix} 0.261395 & 0 & 0 \\ 0.070001 & 0.15723 & 0.0575953 \\ 0 & 0 & 0.318494 \end{bmatrix},$$

$$\widehat{A}_{55} = \begin{bmatrix} 1 & 0 & 0 \\ 0.24405 & 0.51502 & 0.24093 \\ 0 & 0 & 1 \end{bmatrix}, \quad \widehat{B}_{11} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\widehat{B}_{12} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \widehat{B}_{21} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix},$$

$$\widehat{B}_{22} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \widehat{\pi}_{11} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}.$$

Here, we have obtained the conditional probability $P = (Y_{k+1} = y_{k+1}|Y_1 = y_1, \cdots, Y_k = y_k)$ and $P = (X_1 = x_1, \cdots, X_k = x_k|Y_1 = y_1, \cdots, Y_k = y_k)$ on trihybrid crosses (Table 7). To simplify the notation, it is defined that $Y_1^n = ABC$ represents $Y_1 = ABC, Y_2 = ABC, \cdots, Y_n = ABC$ and $X_2^n = AABBCC$ represents $X_2 = AABBCC, X_3 = AABBCC, \cdots, X_n = AABBCC$.

Table 7. The Conditional Probability of Trihybrid Crosses

Probability	Value	Probability	Value
$P(Y_2 = ABC|Y_1 = ABC)$	0.407242	$P(X_1 = AaBbCc, X_2 = AABBCC|Y_1^2 = ABC)$	0.035875
$P(Y_3 = ABC|Y_1^2 = ABC)$	0.592022	$P(X_1 = AaBbCc, X_2^3 = AABBCC|Y_1^3 = ABC)$	0.060597
$P(Y_4 = ABC|Y_1^3 = ABC)$	0.711046	$P(X_1 = AaBbCc, X_2^4 = AABBCC|Y_1^4 = ABC)$	0.085222
$P(Y_5 = ABC|Y_1^4 = ABC)$	0.817162	$P(X_1 = AaBbCc, X_2^5 = AABBCC|Y_1^5 = ABC)$	0.104290
$P(Y_6 = ABC|Y_1^5 = ABC)$	0.900785	$P(X_1 = AaBbCc, X_2^6 = AABBCC|Y_1^6 = ABC)$	0.115777
$P(Y_7 = ABC|Y_1^6 = ABC)$	0.952064	$P(X_1 = AaBbCc, X_2^7 = AABBCC|Y_1^7 = ABC)$	0.121607
$P(Y_8 = ABC|Y_1^7 = ABC)$	0.978333	$P(X_1 = AaBbCc, X_2^8 = AABBCC|Y_1^8 = ABC)$	0.124300
$P(Y_9 = ABC|Y_1^8 = ABC)$	**0.990516**	$P(X_1 = AaBbCc, X_2^9 = AABBCC|Y_1^9 = ABC)$	**0.125490**
$P(Y_{10} = ABC|Y_1^9 = ABC)$	0.995902	$P(X_1 = AaBbCc, X_2^{10} = AABBCC|Y_1^{10} = ABC)$	0.126006
$P(Y_{11} = ABC|Y_1^{10} = ABC)$	0.998235	$P(X_1 = AaBbCc, X_2^{11} = AABBCC|Y_1^{11} = ABC)$	0.126229
$P(Y_{12} = ABC|Y_1^{11} = ABC)$	0.999239	$P(X_1 = AaBbCc, X_2^{12} = AABBCC|Y_1^{12} = ABC)$	0.126325
$P(Y_{13} = ABC|Y_1^{12} = ABC)$	0.999670	$P(X_1 = AaBbCc, X_2^{13} = AABBCC|Y_1^{13} = ABC)$	0.126367
$P(Y_{14} = ABC|Y_1^{13} = ABC)$	0.999857	$P(X_1 = AaBbCc, X_2^{14} = AABBCC|Y_1^{14} = ABC)$	0.126385
$P(Y_{15} = ABC|Y_1^{14} = ABC)$	0.999937	$P(X_1 = AaBbCc, X_2^{15} = AABBCC|Y_1^{15} = ABC)$	0.126393
$P(Y_{16} = ABC|Y_1^{15} = ABC)$	0.999972	$P(X_1 = AaBbCc, X_2^{16} = AABBCC|Y_1^{16} = ABC)$	0.126397
$P(Y_{17} = ABC|Y_1^{16} = ABC)$	0.999988	$P(X_1 = AaBbCc, X_2^{17} = AABBCC|Y_1^{17} = ABC)$	0.126398
$P(Y_{18} = ABC|Y_1^{17} = ABC)$	0.999995	$P(X_1 = AaBbCc, X_2^{18} = AABBCC|Y_1^{18} = ABC)$	0.126399
$P(Y_{19} = ABC|Y_1^{18} = ABC)$	0.999998	$P(X_1 = AaBbCc, X_2^{19} = AABBCC|Y_1^{19} = ABC)$	0.126399
$P(Y_{20} = ABC|Y_1^{19} = ABC)$	0.999999	$P(X_1 = AaBbCc, X_2^{20} = AABBCC|Y_1^{20} = ABC)$	0.126399

From Table 7, it is known that the ninth generation trihybrid crosses has the probability of a dominant phenotype of 99% but the probability of an optimal genotype is only 12.5%. With more generations of plants, the probability of the plants having a dominant phenotype and an optimal genotype is increasing. However, the rise is significantly slow.

The more traits that are crossed, the smaller probability of plants having a dominant phenotype in that period. Therefore, although the probability of diploid plants to obtain dominant phenotype reaches 99% on a generation, the probability of having an optimal genotype for those plants on that generation is tiny.

4. CONCLUSIONS

The discrete hidden Markov model can be applied to diploid plant crosses and can be used to explain the characteristic of true breeding. Initializing the parameters to estimate genotype from diploid plants with

the dominant phenotype produces the same genotype for monohybrid, dihybrid, and trihybrid crosses that are heterozygous in the first generation and homozygous dominant in the next generation. The more traits crossed, the smaller probability of plants having a dominant phenotype in that period. Monohybrid, dihybrid, and trihybrid crosses have a dominant phenotype probability of 99% in the seventh, eighth, and ninth generations, with the condition of previous generations having a dominant phenotype. But in seventh generation, monohybrid crosses only have the probability of an optimal genotype of 50%, dihybrid crosses have a probability of an optimal genotype of 25% in the eighth generation, and trihybrid crosses have a probability of an optimal genotype of 12.5% in the ninth generation. Although the probability for diploid plants has obtained a dominant phenotype by 99% on a generation, the probability of plants having an optimal genotype on that generation is tiny.

REFERENCES

[1] L. A. Urry, M. L. Cain, S. A. Wasserman, P. V. Minorsky and R. Orr, Campbell Biology 12th Edition, San Francisco: Pearson, 2021.

[2] Kastolan, B. Setiawaty and N. K. Ardana, "Point and Figure Portfolio Optimization using Hidden Markov Models and Its Application on the Bumi Resources Tbk Shares," *InPrime: Indonesian Journal of Pure and Applied Mathematics,* vol. 3, no. 1, pp. 44-52, 2021.

[3] L. D. Borchers, W. Zucchini, M. Heide-Jørgensen, A. Cañadas and R. Langrock, "Using hidden Markov models to deal with availability bias on line transect surveys," *Biometrics,* vol. 69, no. 3, pp. 703-713, 2013.

[4] P. D. Harris, A. Narducci, C. Gebhardt, T. Cordes, S. Weiss and E. Lerner, "Multi-parameter photon-by-photon hidden Markov modeling," *Nature Communications,* vol. 13, pp. 1-12, 2022.

[5] J. H. Hsiao, J. An, V. K. S. Hui, Y. Zheng and A. B. Chan, "Understanding the role of eye movement consistency in face recognition and autism through integrating deep neural networks and hidden Markov models," *npj Science of Learning,* vol. 7, pp. 1-13, 2022.

[6] N. Mousavi, M. Monemian, P. G. Daneshmand, M. Mirmohammadsadeghi, M. Zekri and H. Rabbani, "Cyst identification in retinal optical coherence tomography images using hidden Markov model," *Scientific Reports,* vol. 13, pp. 1-15, 2023.

[7] E. Marco, W. Meuleman, J. Huang, K. Glass, L. Pinello, J. Wang, M. Kellis and G.-C. Yuan, "Multi-scale chromatin state annotation using a hierarchical hidden Markov model," *Nature Communications,* vol. 8, pp. 1-9, 2017.

[8] F. Mohammadi, S. Visagan, S. M. Gross, L. Karginov, J. Lagarde, L. M. Heiser and A. S. Meyer, "A lineage tree-based hidden Markov model quantifies cellular heterogeneity and plasticity," *Communications Biology,* vol. 5, pp. 1-14, 2022.

[9] Y. Marchuk, R. Magrans, B. Sales, J. Montanya, J. López-Aguilar, C. d. Haro, G. Gomà, C. Subirà, R. Fernández, R. M. Kacmarek and L. Blanch, "Predicting Patient-ventilator Asynchronies with Hidden Markov Models," *Scientific Reports,* vol. 11, pp. 1-7, 2018.

[10] R. Ludwig, B. Pouymayou, P. Balermpas and J. Unkelbach, "A hidden Markov model for lymphatic tumor progression in the head and neck," *Scientific Reports,* vol. 11, pp. 1-17, 2021.

[11] T. Nguyen, A. Khosravi, D. Creighton and S. Nahavandi, "Hidden Markov models for cancer classification using gene expression profiles," *Information Sciences,* vol. 316, pp. 293-307, 2015.

[12] M. Momenzadeh, M. Sehhati and H. Rabbani, "Using hidden Markov model to predict recurrence of breast cancer based on sequential patterns in gene expression profiles," *Journal of Biomedical Informatics,* vol. 111, pp. 1-9, 2020.

[13] M. Momenzadeh, M. Sehhati and H. Rabbani, "A novel feature selection method for microarray data classification based on hidden Markov model," *Journal of biomedical informatics,* vol. 95, pp. 1-8, 2019.

[14] S. Thorvaldsen, "A tutorial on Markov models based on Mendel's classical experiments," *Journal of Bioinformatics and Computational Biology,* vol. 3, pp. 1441-1460, 2005.

[15] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," in *Proceedings of the IEEE*, Scottsdale, 1989.