

Genetische Strukturen von endlichen Pflanzenpopulationen konstanter Größe

I. Verteilung der Genotypen über die Population dargestellt als Markoff-Kette

Von H.-R. GREGORIUS¹⁾

(Eingegangen im August 1972)

Einleitung

Eine weitgehend erschöpfende Darstellung der genetischen Architektur einer Population, deren räumliche Verteilung zu einem gewissen Zeitpunkt vorgegeben sei, mußte explizit die Wahrscheinlichkeiten beschreiben, mit denen ein beliebiges Mitglied der Population bestimmte Genotypen trägt. Aus der Kenntnis dieser ‚Genotyp-Wahrscheinlichkeiten‘ können dann alle relevanten populationsgenetischen Parameter berechnet werden. Leider findet man jedoch kaum Arbeiten, welche diese Reihenfolge einhalten; vielmehr wird meistens davon ausgegangen, daß ein spezifischer populationsgenetischer Parameter zu bestimmen sei, und daher die Paarungsverhältnisse in der Population derart anzunehmen sind, daß unter Berücksichtigung der durch die Natur vorgeschriebenen Gesetzmäßigkeiten mit Hilfe der zur Verfügung stehenden Mittel ein ‚vernünftiges‘ Ergebnis entsteht. Hierbei würde in der Anwendung dieses ‚vernünftig‘ als ‚rechnerisch durchführbar‘ interpretiert werden müssen. So effektiv diese eng auf das eigentliche Vorhaben zugeschnittene Vorgehensweise auch sein mag, versäumt sie jedoch darauf hinzuweisen, welche Schwierigkeiten die Wahl gerade eines bestimmten Ansatzes erzwingen. Denn häufig ist es die Frage ‚warum gerade so und nicht anders‘, die viele Forscher veranlaßt, aus gutem Grunde von ihren Vorgängern verworfene und nicht publizierte Gedanken zu duplizieren. Andererseits kann die ausschließliche Verfolgung nur eines einzigen möglichen Ansatzes zur baldigen Stagnation führen. Ein Beispiel hierfür sind etwa die Arbeiten von WRIGHT (1943, 1946) über Distanzisolierung, die unter diesem Gesichtspunkt zu hervorragenden Ergebnissen kommen und daher auch zu Recht den meisten Autoren als Ausgangsbasis für weiterführende Betrachtungen dienen. Man sieht sich jedoch bereits bei Ausdehnung des WRIGHT'schen Modells auf nicht uniform über die Fläche verteilte Individuen gezwungen, nach neuen Ansätzen zu suchen und scheidet bald an der Komplexität eines generellen Modelles.

Nichtsdestoweniger soll hier versucht werden, ein möglichst generelles Modell der genetischen Entwicklung gewisser diploider Pflanzenpopulationen endlicher konstanter Größe zu diskutieren, deren Reproduktionsweise die Bedingungen einer MARKOFF-Kette erfüllt. Diese Forderung ist durchaus sinnvoll insofern, als die genetische Entwicklung fast aller Populationen exakt durch einen MARKOFF-Prozeß beschrieben werden kann. Eine Übersicht über Anwendungsmöglichkeiten von MARKOFF-Prozessen auf spezifische Fragestellungen der Populationsgenetik gibt KARLIN (1970). Das generelle Modell, das — wie eingangs erwähnt — eine Beschreibung der Genotyp-Wahrscheinlichkeiten ermöglichen muß, soll dann anschließend auf seine Praktikabilität untersucht und gegebenenfalls durch weitere Forderungen eingegrenzt werden.

Voraussetzungen des Modells:

1. Standplätze (Stellen) der Pflanzen und Ausmaß der Produktion männlicher (Pollen) bzw. weiblicher (Eizellen) Gameten an jeder Stelle bleiben über die Generationen unverändert.
2. Die Pollen- und Eizellenproduktion ist unendlich groß.
3. Diskrete Generationen.
4. Betrachtung eines diploiden Locus. Jeder Genotyp produziert genau zur Hälfte Gameten, die das eine bzw. das andere Allel tragen.
5. Die Pollenverteilung einer Pflanze ist unabhängig von ihrem Genotyp und der Generation.
6. Bildung der Nachkommengeneration: Die Menge der auf einer Pflanze gelandeten Pollen paart zufallsmäßig (i.e. unabhängig) mit den dort vorhandenen Eizellen. Aus dem gebildeten Samen wird zufällig einer ausgewählt und an derselben Stelle zu einer neuen Pflanze herangezogen. Die Auswahl der Samen geschieht unabhängig voneinander.
7. Die Selbstinkompatibilitätsrate ist unabhängig von der Stelle und dem an ihr vertretenen Genotyp.
8. Es werden nur Pflanzen gleichen Blühtermins zu einem Bestand zusammengefaßt, da zwischen Pflanzen mit verschiedenen, nicht überlappenden Blühterminen keine Befruchtung stattfinden kann.
9. Sowohl Immigration von Pollen als auch Mutation und Selektion seien ausgeschlossen.

Definitionen:

Um unter anderem auch die Wanderung der von den einzelnen Stellen stammenden Allele im Pflanzenbestand verfolgen zu können, sollen die bei N Stellen existierenden $2N$ Allele derart numeriert werden, daß $(1; 2)$ denjenigen Genotyp bezeichnet, der in der Ausgangsgeneration 0 an der Stelle 1 sitzt, und in dem die erste Komponente das von einem männlichen, die zweite Komponente das von einem weiblichen Gameten stammende Allel darstellt. Daher besitzt aufgrund der Voraussetzungen jeder mögliche Genotyp in späteren Generationen immer die Form $(m; n; u; v)$, wobei $m, u = 1, 2$ und $n, v = 1, \dots, N$ sein kann. Befinden sich nun an den Stellen l ($l = 1, \dots, N$) des Bestandes die Genotypen $(m_l; n_l; u_l; v_l)$, so wollen wir als ‚genotypischen Zustand‘ oder auch einfach ‚Zustand‘ des Bestandes die Anordnung dieser Genotypen zu einem Vektor der Länge N definieren. Folglich gibt es $4N^2$ mögliche Genotypen und daher $(4N^2)^N$ mögliche Zustände des Bestandes, die wir auf eine beliebige Weise indizieren und mit H_k ($k = 1, \dots, (4N^2)^N$) bezeichnen. Insbesondere sei H , immer gerade jener Vektor, dessen Komponenten l ($l = 1, \dots, N$) durch die Genotypen $(1; 2)$ beschrieben werden (H , ist also der Zustand der Ausgangsgeneration 0), während allgemein die Komponenten des Vektors H_k gleich $(m_k; n_k; u_k; v_k)$ gesetzt werden sollen, i. e.

$$H_k = ((m_k; n_k; u_k; v_k), \dots, (m_{kN}; n_{kN}; u_{kN}; v_{kN}))$$

Bestimmung der Übergangswahrscheinlichkeiten:

Mit den oben erläuterten Notationen können wir nun die

¹⁾ Lehrstuhl für Forstgenetik und Forstpflanzenzüchtung, Universität Göttingen, 34 Göttingen-Weende, Büsingenweg 2.

in den Voraussetzungen getroffenen Annahmen wie folgt zusammenfassen: Die genetische Entwicklung des Bestandes vollzieht sich in einer zufälligen zeitlichen Aufeinanderfolge von Zuständen in der Weise, daß der Zustand in einer jeden Generation lediglich vom Zustand der jeweilig vorangehenden Generation abhängt, und daß mit den Paarungsverhältnissen auch die Übergangswahrscheinlichkeiten von einem in einen weiteren folgenden Zustand zu jedem Zeitpunkt gleich sind. Beschreiben wir eine solche zufällige Aufeinanderfolge von Zuständen durch eine diskrete Folge von Zufallsvariablen X_t ($t = 0, 1, 2, 3, \dots$), welche Werte aus der Menge der möglichen genotypischen Zustände des Bestandes annehmen, so bilden diese Variablen eine MARKOFF-Kette mit stationären Übergangswahrscheinlichkeiten $P(X_t = H_i | X_{t-1} = H_k)$ und der Anfangsverteilung $P(X_0 = H_i) = 1$ nach Annahme. $P(\dots)$ bezeichnet hier das zugehörige Wahrscheinlichkeitsmaß.

Gemäß Vorauss. 6) müssen wir davon ausgehen, daß die Samen zur Bildung der Nachkommengeneration an den verschiedenen Stellen unabhängig voneinander ausgewählt werden, so daß die Komponenten X_{lt} ($l = 1, \dots, N$) der X_t bedingt unabhängig sind, i. e. es gilt:

$$P_{ki} = P(X_t = H_i | X_{t-1} = H_k) = \prod_{l=1}^N P(X_{lt} = (m_l^n n_l; u_l^k v_l) | X_{t-1} = H_k)$$

Da durch Angabe der Übergangswahrscheinlichkeiten P_{ki} der Prozeß vollständig bestimmt ist, können wir uns in diesem Falle mit der Berechnung der Wahrscheinlichkeiten $P(X_{lt} = (ri; sj) | X_{t-1} = H_k)$ dafür, daß der Bestand in der Generation t einen Zustand annimmt, der an der Stelle l den Genotyp $(ri; sj)$ aufweist unter der Bedingung, daß der vorangehende Zustand gerade H_k war, begnügen; auch diese Wahrscheinlichkeiten sind natürlich unabhängig von der Zeit t .

Wir führen das folgende verallgemeinerte KRONECKER-Symbol ein:

$$\delta_{sj}^{ri} := \begin{cases} 1 & \text{falls } r = s \text{ und } i = j \\ 0 & \text{falls } r \neq s \text{ oder } i \neq j \end{cases}$$

Weiterhin seien y_{1l}, \dots, y_{Nl} ($\sum_{a=1}^N y_{al} = 1$) die Wahrscheinlichkeiten dafür, daß die von den N Stellen des Bestandes stammenden Pollen an der Stelle l zur Befruchtung gelangen (eine weitere Aufschlüsselung der y_{al} wird sich noch anschließen).

Wird nun der genotypische Zustand des Bestandes mit $H_k = ((m_{1N}^k n_{1N}^k; u_{1N}^k v_{1N}^k), \dots, (m_{NN}^k n_{NN}^k; u_{NN}^k v_{NN}^k))$ beschrieben, dann wird an der Stelle l eine Eizelle, welche das Allel (sj)

trägt mit der Wahrscheinlichkeit $\frac{1}{2} (\delta_{sj}^{m_{1l}^k n_{1l}^k} + \delta_{sj}^{u_{1l}^k v_{1l}^k})$ produziert, während ein von der Stelle a stammender Pol-

len mit der Wahrscheinlichkeit $\frac{1}{2} (\delta_{ri}^{m_{al}^k n_{al}^k} + \delta_{ri}^{u_{al}^k v_{al}^k})$ das Allel (ri) trägt und mit der Wahrscheinlichkeit y_{al} an der Stelle l zur Befruchtung gelangt. Folglich wird aus dem Beitrag der von allen Stellen a ($a = 1, \dots, N$) stammenden Pollen an der Stelle l ein Same, der den Genotyp $(ri; sj)$ trägt mit der Wahrscheinlichkeit

$$P(X_{lt} = (ri; sj) | X_{t-1} = H_k) = \frac{1}{4} (\delta_{sj}^{m_{1l}^k n_{1l}^k} + \delta_{sj}^{u_{1l}^k v_{1l}^k}) \cdot \sum_{a=1}^N y_{al} (\delta_{ri}^{m_{al}^k n_{al}^k} + \delta_{ri}^{u_{al}^k v_{al}^k}) \quad (1a)$$

gebildet.

Somit stellen sich die Übergangswahrscheinlichkeiten als

$$P_{ki} = \frac{1}{4N} \prod_{l=1}^N \left[(\delta_{i_1 v_1}^{m_{1l}^k n_{1l}^k} + \delta_{i_1 v_1}^{u_{1l}^k v_{1l}^k}) \cdot \sum_{a=1}^N y_{al} (\delta_{i_1 n_1}^{m_{al}^k n_{al}^k} + \delta_{i_1 n_1}^{u_{al}^k v_{al}^k}) \right] \quad (1)$$

dar.

Bevor wir aus (1) weitere Schlüsse ziehen, soll die bereits angekündigte Aufschlüsselung der y_{al} vorgenommen werden.

In vielen Fällen ist es praktisch möglich, folgende Größen zu bestimmen:

Ausmaß der Pollenproduktion der einzelnen Stelle relativ zu der des gesamten Bestandes (hier mit f_l für die Stelle l bezeichnet, $\sum_{l=1}^N f_l = 1$).

Pollenverteilung der einzelnen Stelle über die Fläche (hier als zweidimensionale Dichte $h_l(x, y)$ bzgl. der Stelle l dargestellt). Die Selbstinkompatibilitätsrate der Population, definiert als die Wahrscheinlichkeit dafür, daß von dersel-

ben Pflanze stammende Pollen und Eizellen Samen bilden (hier mit R bezeichnet).

Ordnet man jeder Stelle k eine Fläche B_k zu (B_k seien disjunkte Borelmengen der zweidimensionalen Ebene), auf der all jene Pollen landen, die an dieser Stelle zur Befruchtung gelangen können, dann beschreibt $p_{lk} = \int_{B_k} h_l(x, y) dx dy$ multipliziert mit f_l den Anteil der auf der Pflanze k gelandeten Pollen, der von der Pflanze l stammt. Normierung dieses Anteils führt zu

$$q_{lk} := \frac{p_{lk} f_l}{\sum_{i=1}^N p_{ik} f_i}$$

Weiterhin bedingt Selbstinkompatibilität eine Reduktion des Anteils der von der Stelle k stammenden Pollen, die an dieser selben Stelle zur Befruchtung gelangen um einen Bruchteil R , so daß wir endlich nach nochmaliger Normierung erhalten:

$$y_{lk} = \frac{q_{lk}}{1 - (1 - R)q_{kk}} \quad \text{für } l \neq k \quad \text{und} \quad y_{kk} = \frac{Rq_{kk}}{1 - (1 - R)q_{kk}}$$

Kehren wir nun zu unserer mit Hilfe der Übergangswahrscheinlichkeiten (1) und der Anfangsverteilung $P(X_0 = H_i) = 1$ vollständig bestimmten MARKOFF-Kette zurück und gehen auf die Besonderheiten einiger genotypischer Zustände etwas ausführlicher ein.

Aus der Gleichung (1) ersieht man, daß der Übergang von einem Zustand H_k in einen solchen H_i mit positiver Wahrscheinlichkeit (i. e. $P_{ki} \neq 0$) genau dann möglich ist, wenn für $l = 1, \dots, N$ gilt: Das von einer Eizelle stammende Allel an der Stelle l des Zustandes H_i kommt an derselben Stelle im Genotyp des Zustandes H_k vor, während das von einem Pollen stammende Allel im Genotyp einer Stelle a des Zustandes H_k vorkommen muß, für welche $y_{al} \neq 0$ ist, i. e. von der noch Pollen an der Stelle l landet. Diese Vorstellung soll nun benutzt werden, um die Bedingungen zusammenzustellen, unter denen ein Allel im Laufe der Zeit von einer Stelle zu einer anderen Stelle 'springen' kann. Enthält nämlich die Folge $y_{1l_1}, y_{1l_1 l_2}, y_{1l_1 l_2 l_3}, \dots, y_{1l_1 l_2 \dots l_k}$ nur Glieder $\neq 0$, dann wollen wir sagen, daß von der Stelle l

zur Stelle k ($i \neq k$) ein Pfad in $t + 1$ Schritten führt. Gleichbedeutend hiermit ist offenbar nach (1) die Existenz einer Folge von Zuständen mit positiven Übergangswahrscheinlichkeiten, in der sich die Wanderung eines Allels von der Stelle i zur Stelle k ausdrückt. Befindet sich insbesondere ein Allel einmal an einer Stelle, so verbleibt es dort mit positiver Wahrscheinlichkeit, denn ist z. B. H_k ein Zustand, in welchem etwa das Allel (s_j) an der Stelle l sitzt, i. e. $(m_{kj}^k n_{kj}) = (s_j)$ oder $(u_{kj}^k v_{kj}) = (s_j)$, dann ist es nach (1) immer möglich, ein H_i so zu wählen, daß $(u_{ij}^k v_{ij}) = (s_j)$ und $P_{ki} \neq 0$ erfüllt wird.

Existiert nun eine Stelle im Bestand, von der aus Pfade zu allen weiteren Stellen des Bestandes führen, so kann ein ursprünglich dort vorhandenes Allel mit positiver Wahrscheinlichkeit im Laufe der Zeit an allen Stellen auftauchen und zugleich dort verbleiben. Daher endet die Entwicklung mit wiederum positiver Wahrscheinlichkeit in einem Zustand, der nur noch ein einziges Allel enthält. Dieser Zustand ist absorbierend in dem Sinne, daß er im nächsten Schritt mit Wahrscheinlichkeit 1 in sich selbst übergeht; es gibt also keine Rückkehr in einen anderen vorangehenden Zustand. In der Populationsgenetik spricht man hier von Fixierung der Population auf ein Allel. Folglich sind in diesem Falle alle Zustände außer den fixierten nicht rekurrent (transient), und es strebt $P^{(t)}_{ki}$ (die wohlbekannte Übergangswahrscheinlichkeit von einem Zustand H_k zu einem solchen H_i in t Schritten) mit wachsendem t gegen 0 für beliebige H_k und nicht fixierte — also nicht rekurrente — H_i (siehe z. B. K.-L. CHUNG (1960)). Weiterhin wächst natürlich die Folge der Wahrscheinlichkeiten dafür, daß sich der Bestand in der Generation t auf ein bestimmtes Allel fixiert monoton mit t , so daß der Bestand mit Wahrscheinlichkeit 1 auf ein Allel fixiert wird, das im Ausgangszustand H_l an einer Stelle saß, von der aus Pfade zu allen weiteren Stellen des Bestandes führen.

Existiert andererseits keine Stelle im Bestand, von der aus Pfade zu allen weiteren Stellen führen, gibt es also keinen Zustand, in welchem an allen Stellen nach einiger Zeit ein gewisses Allel sitzt, dann kann sich der Bestand selbstverständlich niemals auf ein Allel fixieren; die Wahrscheinlichkeit für Fixierung auf ein Allel ist daher in diesem Falle gleich 0.

Die Genotyp-Wahrscheinlichkeiten:

Neben der Verteilung der genotypischen Zustände in jeder Generation besitzt vor allem die Kenntnis der Genotyp-Wahrscheinlichkeiten für die genetische Entwicklung des Bestandes große Bedeutung. Letztere sind definiert als die Wahrscheinlichkeit dafür, in der Generation t einen Zustand zu erhalten, der etwa den Genotyp $(ri; sj)$ an der Stelle l aufweist, i. e.

$$\begin{aligned} P(X_{1t} = (ri; sj) \mid X_0 = H_l) &= G_t^l(ri; sj) = \\ &= \sum_k P(X_{1t} = (ri; sj) \mid X_{t-1} = H_k) \cdot P^{(t-1)}_{lk}. \end{aligned} \quad (2)$$

Für $t = 1$ erhält man nach (1a):

$$G_1^l(ri; sj) = \frac{1}{4} (\delta_{sj}^{11} + \delta_{sj}^{21}) \cdot \sum_{a=1}^N y_{al} (\delta_{ri}^{1a} + \delta_{ri}^{2a}) \text{ und daher}$$

$G_1^l(ri; sj) = 0$ für $j \neq 1$ und $G_1^l(ri; sl) = \frac{1}{2} y_{l1}$ durchaus in Übereinstimmung mit der Anschauung.

Die Definition (2) der Genotyp-Wahrscheinlichkeiten gestattet eine sehr einfache Darstellung des Inzuchtcoeffizienten I_1^l einer Pflanze, die sich in der Generation t an der Stelle l befindet, wenn man MALECOR'S Erklärung des Inzuchtcoeffizienten als die Wahrscheinlichkeit dafür, daß ein Individuum zwei durch Abstammung identische Allele trägt zugrundelegt:

$I_1^l = \sum_{sj} G_1^l(sj; sj)$, speziell in der ersten Generation ist

$I_1^l = \frac{1}{2} y_{l1}$, also nur vom Selbstungsanteil bestimmt.

Eine weitere aus (2) leicht zu erhaltende Information über die genetische Entwicklung des Bestandes geht aus der Verteilung der Allele in der Population hervor. Diese Verteilung wird beschrieben durch die Allel-Wahrscheinlichkeiten, i. e. die Wahrscheinlichkeit dafür, etwa in der Generation t einen Zustand zu erhalten, der an der Stelle l das Allel (ri) aufweist:

$$\begin{aligned} G_1^l(ri) &:= \frac{1}{2} \cdot \sum_{sj} (G_1^l(ri; sj) + G_1^l(sj; ri)) = \text{nach (1) und (2)} \\ &= \frac{1}{2} \cdot \sum_k P_{lk}^{(t-1)} \cdot \frac{1}{4} \cdot \left[2 (\delta_{ri}^{m_{lk}^k n_{lk}^k} + \delta_{ri}^{u_{lk}^k v_{lk}^k}) + \right. \\ &\quad \left. + 2 \cdot \sum_{a=1}^N y_{al} (\delta_{ri}^{m_{lk}^k n_{lk}^k} + \delta_{ri}^{u_{lk}^k v_{lk}^k}) \right] = \frac{1}{4} \cdot \sum_{mnuv} (\delta_{ri}^{mn} + \\ &\quad + \delta_{ri}^{uv}) \cdot G_1^{t-1}(mn; uv) + \frac{1}{4} \cdot \sum_a y_{al} \cdot \sum_{mnuv} (\delta_{ri}^{mn} + \\ &\quad + \delta_{ri}^{uv}) \cdot G_a^{t-1}(mn; uv) = \frac{1}{2} \cdot G_1^{t-1}(ri) + \\ &\quad + \frac{1}{2} \cdot \sum_a y_{al} G_a^{t-1}(ri) = \sum_a x_{al} G_a^{t-1}(ri), \end{aligned} \quad (3)$$

wobei

$$x_{al} = \frac{1}{2} y_{al} \text{ für } a \neq 1 \text{ und } x_{l1} = \frac{1}{2} (1 + y_{l1}) \text{ ist } \left(\sum_{a=1}^N x_{al} = 1 \right)$$

Durch Induktion nach t zeigt man leicht, daß für alle l , t und i gilt: $G_t^l(1i) = G_t^l(2i)$. (3) läßt sich offenbar in kompakterer Form als Matrizenprodukt schreiben, wenn man die Matrix X der x_{al} und die Zeilenvektoren $g_t^l := (G_t^l(1i), G_t^l(2i), \dots, G_t^l(Ni))$ einführt: $g_t^l = g^{t-1} \cdot X$, und nach Ausführung dieser Rekursion: $g_t^l = g^0 \cdot X^t$, wobei also g^0 ein Zeilenvektor ist, dessen i -te ($3a$) Komponente den Wert $\frac{1}{2}$ und ansonsten nur Werte 0 annimmt.

Um nun das Grenzverhalten der Allel-Wahrscheinlichkeiten für große t betrachten zu können, nehmen wir einen Bestand an, in welchem Pfade zwischen allen Stellen existieren. Aufgrund der Beziehungen zwischen den y_{al} und x_{al} muß es dann einen Exponenten t' geben, so daß die Matrix $X^{t'}$ nur noch Elemente $\neq 0$ enthält. Einem bekannten Theorem zufolge (siehe z. B. GIRAULT 1966) existiert in diesem Falle der $\lim_t X^t =: Y$, und die Spalten der Matrix

Y sind untereinander gleiche Wahrscheinlichkeitsvektoren, deren Komponenten wiederum alle $\neq 0$ sind. Diese Feststellung ist gleichbedeutend damit, daß alle Allel-Wahrscheinlichkeiten konvergieren und nach (3a) gilt: $\lim_t G_t^l$

$(ri) = \lim_t G_t^2(ri) = \dots = \lim_t G_t^N(ri)$, i. e. im Laufe der Zeit

gleichens sich die Wahrscheinlichkeiten eines Allels an allen Stellen mehr und mehr einander an.

Anwendbarkeit des Modelles:

Für den Praktiker werden die Grenzen der Anwendbarkeit eines Modells auf spezifische Fragestellungen fast ausschließlich von dem hierfür notwendigen rechnerischen Aufwand bestimmt, der im Zeitalter des Computers wiederum identisch mit der Anzahl der benötigten Speicherplätze ist. Würde man in unserem Falle einen populationsgenetischen Parameter berechnen wollen, dessen Definition die Genotyp-Wahrscheinlichkeiten verwendet (z. B. der Inzuchtcoeffizient), so brauchte man nach (2) die Information aller $(4N^2)^{2N}$ Übergangswahrscheinlichkeiten und würde damit bereits für kleine N die Speicherkapazität je-

des Rechners übersteigen. Ist es andererseits möglich, die Definition eines Parameters auf die Allel-Wahrscheinlichkeiten zurückzuführen, dann reduziert sich aufgrund der Beziehung (3a) und der dort eingehenden Matrix X die Speicherkapazität im wesentlichen auf nur $2N^2$ Plätze. In diesem Falle bliebe das Modell also auch für größere N praktikabel.

Als Ausweg aus den rechentechnischen Schwierigkeiten, welche sich bei einer exakten Ermittlung der Verteilung der Zustände in jeder Generation ergeben, bietet sich die Simulation von Ketten genotypischer Zustände an. Die Wahrscheinlichkeiten gewisser interessierender Zustände in den einzelnen Generationen lassen sich dann mit Hilfe ihrer relativen Häufigkeiten aus einer entsprechend großen Anzahl von unabhängigen Wiederholungen solcher Simulationen schätzen. Abschließend soll noch kurz die Vorgehensweise bei der Simulation einer Kette beschrieben werden: Man beginnt mit dem Ausgangszustand H_1 und erzeugt zufällig einen genotypischen Zustand der Generation 1, indem man z. B. an der Stelle l eines der beiden dort vorhandenen Allele (1l) bzw. (2l) jeweils mit der Wahrscheinlichkeit 1/2 als Eizelle greift. Weiterhin ist die Verteilung der von den einzelnen Stellen stammenden, befruchtungsfähigen Pollen, die sich an der Stelle l befinden mit $y_{1l}, y_{2l}, \dots, y_{Nl}$ gegeben. Man wählt also aus dieser Grundgesamtheit eine Stelle und von dieser Stelle wiederum mit der Wahrscheinlichkeit 1/2 eines der beiden Allele als Pollen, der mit der bereits gegriffenen Eizelle die Zygote bildet. Dies geschieht für alle Stellen $l = 1, \dots, N$. Alle folgenden Generationen entstehen immer wieder auf die gleiche Weise aus den vorangehenden.

Zusammenfassung

Es wurde die Verteilung der Genotypen eines Locus über eine diploide Pflanzenpopulation endlicher konstanter Größe unter folgenden zusätzlichen Modellannahmen untersucht: Ortsgebundenheit der Pflanzen; diskrete Generationen; Konstanz der Eizellen- sowie Pollenproduktion und Pollenverteilungen der einzelnen Pflanzen in allen Generationen; Ausschluß von Pollenimmigration, Mutation und

Selektion. Die wahrscheinlichkeitstheoretische Formulierung dieser Annahmen führte zu einer Darstellung als MARKOFF-Kette, welche ihrerseits Bedingungen nahelegte, unter welchen sich die Population mit Wahrscheinlichkeit 1 auf ein Allel fixiert, bzw. unter welchen sich im Laufe der Zeit die Wahrscheinlichkeiten eines Allels an allen Stellen mehr und mehr einander angleichen.

Aufgrund des unverhältnismäßig hohen Rechenaufwandes, der sich bei der explizierten Bestimmung gewisser populationsgenetischer Parameter aus den hergeleiteten Beziehungen ergäbe, wurde ein auf diesen Beziehungen aufbauendes Simulationsverfahren beschrieben, das mit gleicher Allgemeinheit Schätzungen beliebiger Parameter bei wesentlich geringerem Rechenaufwand ermöglicht.

Schlagworte: Pflanzenpopulationen, MARKOFF-Kette.

Summary

We investigated the distribution of the genotypes on one locus over a diploid plant population of finite constant size using the following additional model assumptions: stationary plants; discrete generations; constancy of egg cell productions, pollen productions, and pollen distributions of the plants in all generations; exclusion of pollen immigration, mutation and selection. The probability-theoretical formulation of these assumptions led to a representation of a MARKOV chain. From this we found out the conditions under which the probability of fixation of the population on one allele is 1, and under which the probabilities of one allele at every place converge with time to the same value.

Because of the excessively high extense of computation needed by using the model formulas for determining some specific population genetic parameters we derived from the same model assumptions a simulation procedure with which any parameter can be estimated.

Literatur

CHUNG, KAI-LAI: Markov Chains with Stationary Transition Probabilities, Springer Verlag, 1960. — GIRAULT, M.: Stochastic Processes, Springer Verlag, 1966. — KARLIN, S.: Equilibrium Behavior of Population Genetic Models with non-Random Mating, Gordon a. Breach, 1970. — WRIGHT, S.: Isolation by Distance, *Genetics* 28, 114—138 (1943). — WRIGHT, S.: Isolation by Distance Under Diverse Systems of Mating, *Genetics* 31, 39—59 (1946).

Genetische Strukturen von endlichen Pflanzenpopulationen konstanter Größe

II. Verwandtschaftsgrad und Inzuchtgrad von Pflanzen an speziellen Stellen im Bestand

Von H.-R. GREGORIUS¹⁾

(Eingegangen September / Revision November 1972)

Gegenstand der folgenden Ausführungen ist die Elimination zweier populationsspezifischer Größen aus einem allgemeinen Modell zur Beschreibung einer als MARKOFF-Kette darstellbaren Verteilung der Genotypen über die Population (GREGORIUS 1973). Alle Termini und Bezeichnungen, die im Zusammenhang nicht ausdrücklich neu definiert sind, wurden aus dieser Arbeit übernommen.

Die Beschreibung der genetischen Entwicklung eines

Pflanzenbestandes in voller Allgemeinheit mittels der Verteilungen seiner möglichen genotypischen Zustände in jeder Generation, führt bei expliziter Bestimmung gewisser populationsgenetischer Parameter aus dieser Darstellung zu unüberwindlichen rechentechnischen Schwierigkeiten. Man sieht sich daher in solchen Fällen gezwungen, einfachere Ansätze, die selbstverständlich auf denselben grundsätzlichen Gedanken beruhen und dieselbe Allgemeinheit wahren, zu wählen.

In vielerlei Beziehungen, insbesondere für die Präzision des Begriffes der Nachbarschaft besitzen die Parameter In-

¹⁾Lehrstuhl für Forstgenetik und Forstpflanzenzüchtung, Universität Göttingen, 34 Göttingen-Weende, Büsgenweg 2.