

(51) Int.Cl. ⁷	識別記号	F I	テ-マコード ⁸ (参考)
G06F 17/18		G06F 15/36	Z 2G045
G01N 33/48		G01N 33/48	Z 4B024
// C12N 15/09		C12N 15/00	A 5B056

審査請求 有 請求項の数 4 O L (全 8 頁)

(21)出願番号	特願平11 - 84422	(71)出願人	593059887 農林水産省水産庁養殖研究所長 三重県度会郡南勢町中津浜浦422 - 1
(22)出願日	平成11年 3月26日(1999.3.26)	(72)発明者	大原 一郎 三重県度会郡南勢町五ヶ所浦190番地25号
		(72)発明者	小林 敬典 三重県度会郡玉城町佐田411番地 2号
		(72)発明者	中山 一郎 三重県伊勢市古市町355番地10号
		(74)代理人	100091096 弁理士 平木 祐輔 (外 2名)

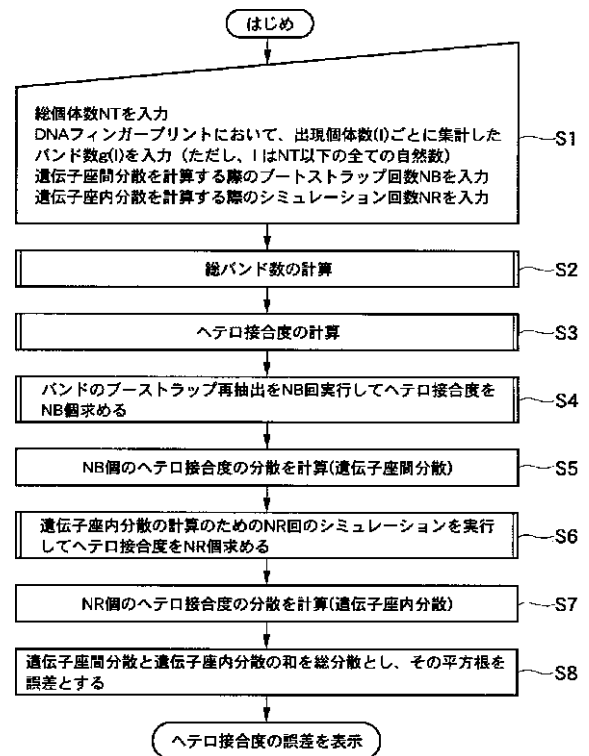
最終頁に続く

(54) 【発明の名称】ヘテロ接合度を推定するプログラムを記録した記録媒体

(57) 【要約】

【課題】 2倍体ゲノムを有する生物の集団の遺伝的多様性を検査して近親交配の程度を調べ、近交弱勢の危険を事前に防ぐなど、DNAフィンガープリントから対象生物の遺伝的特性を評価するプログラムを記録した記録媒体を提供すること。

【解決手段】 バンドのブートストラップ再抽出を実行して求めた(ステップS4)ヘテロ接合度の遺伝子座間分散(ステップS5)と、遺伝子座内分散の計算のためのシミュレーションを実行して求めた(ステップS6)ヘテロ接合度の遺伝子座内分散(ステップS7)との和を総分散としてその平方根を誤差とする(ステップS8)。



【特許請求の範囲】

【請求項 1】 コンピュータに、複数遺伝子座に対する優性遺伝子マーカーによる DNA フィンガープリントからヘテロ接合度の推定値を得るヘテロ接合度推定手順を実行させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体。

【請求項 2】 前記ヘテロ接合度推定手順によって推定されるヘテロ接合度に関して、DNA バンドのブートストラップ再抽出を複数回繰り返すことによりその遺伝子座間分散を推定する遺伝子座間分散推定手順を実行させるためのプログラムを記録したコンピュータ読み取り可能な請求項 1 記載の記録媒体。

【請求項 3】 前記ヘテロ接合度推定手順によって推定されるヘテロ接合度に関して、個体数 n のサンプリングで見かけ上の出現頻度が j / n であるバンドの数を f_o(j) と定義するとき、出現頻度が j / n であるバンドの真の数値 f_r(j) を、j < n のとき、

【数 1】

$$f_r(j) = f_o(j) \frac{1}{\left[1 - \left(1 - \frac{j}{n}\right)^n - \left(\frac{j}{n}\right)^n\right]}$$

と推定し、j = n のとき、

【数 2】

$$f_r(n) = f_o(n) - \sum_{i=1}^{n-1} \left(\frac{i}{n}\right)^n f_r(i)$$

と推定して、f_r(j) を用いたシミュレーションにより遺伝子座内分散を推定する遺伝子座内分散推定手順を実行させるためのプログラムを記録したコンピュータ読み取り可能な請求項 1 記載の記録媒体。

【請求項 4】 請求項 2 記載の遺伝子座間分散推定手順によって推定される遺伝子座間分散の推定値と請求項 3 記載の遺伝子座内分散推定手順によって推定される遺伝子座内分散の推定値の和の平方根を、ヘテロ接合度の誤差として推定する誤差推定手順を実行させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、2 倍体ゲノムを有する生物の集団の遺伝的多様性を検査して近親交配の程度を調べ、近交弱勢の危険を事前に防ぐなど、DNA フィンガープリントから対象生物の遺伝的特性を評価するプログラムを記録した記録媒体に関する。

【0002】

【従来の技術】従来、対象生物の遺伝的特性を評価する方法の一つに、RAPD や AFLP に代表される、複数遺伝子座に対する優性遺伝子マーカーによる DNA フィンガープリント法がある。

【0003】

【発明が解決しようとする課題】しかし、そのデータ解

析に関して重要な事は、複数集団間の遺伝的多様度の適切な指標を用い、その値の複数集団間での差を示すとともに、差の有意性を統計的に検定することである。本発明では、2 倍体ゲノムを有する生物の集団の遺伝的多様度を比較するために上述のフィンガープリント法を用いる。しかし、そのデータ解析に関して重要な事は、複数集団間の遺伝的多様度の適切な指標を用い、同じ生物種から任意に選んだ複数集団間での値の差を示すとともに、差の有意性を統計的に検定することである。そこで、遺伝的多様度の指標としてヘテロ接合度を用いることとし、ヘテロ接合度の誤差を集団ごとに推定することとする。しかしながら、これまでのところ、そのような方法は報告されていなかった。そこで、本発明は「複数遺伝子座 DNA フィンガープリント法からヘテロ接合度を求める方法」(Stephens et al. 1992) を参考にしつつ、優性遺伝子マーカーによるヘテロ接合度とその誤差に関する知見を得るプログラムを記録した記録媒体を提供することを目的とする。

【0004】

【課題を解決するための手段】本発明は、コンピュータに、複数遺伝子座に対する優性遺伝子マーカーによる DNA フィンガープリントからヘテロ接合度の推定値を得るヘテロ接合度推定手順を実行させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体である。また、本発明は、前記ヘテロ接合度推定手順によって推定されるヘテロ接合度に関して、DNA バンドのブートストラップ再抽出を複数回繰り返すことによりその遺伝子座間分散を推定する遺伝子座間分散推定手順を実行させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体である。

【0005】さらに、前記ヘテロ接合度推定手順によって推定されるヘテロ接合度に関して、個体数 n のサンプリングで見かけ上の出現頻度が j / n であるバンドの数を f_o(j) と定義するとき、出現頻度が j / n であるバンドの真の数値 f_r(j) を、j < n のとき、

【0006】

【数 3】

$$f_r(j) = f_o(j) \frac{1}{\left[1 - \left(1 - \frac{j}{n}\right)^n - \left(\frac{j}{n}\right)^n\right]}$$

【0007】と推定し、j = n のとき、

【0008】

【数 4】

$$f_r(n) = f_o(n) - \sum_{i=1}^{n-1} \left(\frac{i}{n}\right)^n f_r(i)$$

【0009】と推定して、f_r(j) を用いたシミュレーションにより遺伝子座内分散を推定する遺伝子座内分散推定手順を実行させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体である。また、前記遺伝子座間分散推定手順によって推定される遺伝子座間分

散の推定値と前記遺伝子座内分散推定手順によって推定される遺伝子座内分散の推定値の和の平方根を、ヘテロ接合度の誤差として推定する誤差推定手順を実行させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体である。

【 0 0 1 0 】

【発明の実施の形態】以下添付図面を参照しながら本発明の好適な実施の形態について詳細に説明する。本実施の形態では、2倍体ゲノムを有する生物の、同一種又は同一亜種に属する各集団の遺伝的多様度を比較するために上述のフィンガープリント法を用いる。そして、遺伝的多様度の指標としてヘテロ接合度を用いることとし、ヘテロ接合度の誤差を集団ごとに推定する。

【 0 0 1 1 】 DNA 多様度を集団間で比較することを目的とし、各集団の個体から RAPD 分析や AFLP 分析により DNA フィンガープリントを得る。データの解析に際し、ノンパラメトリックな統計的手法を用いて集団のヘテロ接合度の誤差を推定することで、集団間にヘテロ接合度の有意差があるかどうかを検討することができる。優性遺伝子マーカーによる DNA フィンガープリントから求められるヘテロ接合度と、その誤差に関する知見を次のようにして得る。

【 0 0 1 2 】 1) ヘテロ接合度の計算について、RAPD (random amplified polymorphic DNA) や AFLP (amplified fragment length polymorphism) などの優性遺伝子マーカー (Welsh and McClelland 1990; Williams et al. 1990; 島野・矢野 1997) からヘテロ接合度を計算する方法の報告は、現在のところ存在しない。しかしながら、ほぼ共優性マーカーと考えられるミニサテライト DNA フィンガープリントなどの複数遺伝子座 DNA フィンガープリントに関しては Stephens et al. (1992) による報告があり、さらに Jin and Chakraborty (1993) による改良案も出されている。そこで、優性遺伝子マーカーとしての個々のバンドを、複数遺伝子座における共優性マーカーからのランダム抽出であるとみなすことにより、Stephens et al. (1992) および Jin and Chakraborty (1993) によって与えられたものと同じ式で、ヘテロ接合度を計算する。

【 0 0 1 3 】 2) ヘテロ接合度における遺伝子座間分散について、まず、n 個体から DNA フィンガープリントを得たとき、あるバンドが n 個体中何個体で出現するかを求めてヒストグラムを作成すると、一般にその出現頻度分布は正規分布から大きく外れるため、誤差を推定するのにパラメトリックな方法を用いることは好ましくな

い。そこで本実施の形態ではノンパラメトリックな推定法として、「ブートストラップ再抽出」を使用し、各再抽出ごとにヘテロ接合度を求め、その分布からヘテロ接合度の遺伝子座間分散を推定することとする。

【 0 0 1 4 】 この際問題となるのは、ブートストラップの単位は独立な遺伝子座であるべきであるにも拘わらず、Stephens et al. (1992) で例示されているような複数遺伝子座 DNA フィンガープリントでは、1 つの遺伝子座に対応するバンドがどれとどれであるかがわからないので、遺伝子座ごとのブートストラップ再抽出が不可能であることである。しかしながら、幸いなことに、RAPD や AFLP では、出現するバンドのほとんどが独立であって、多くの遺伝子座からの対立遺伝子のランダムな抽出であるとみなしても大きな問題はないので、ミニサテライト DNA フィンガープリント (Stephens et al. 1992) が直面するバンド間の相関の問題を回避することができると考えられる。従って、RAPD や AFLP の DNA フィンガープリントでは各バンドを独立なサンプルと解釈しても、大きな問題は生じないと考えられる。

【 0 0 1 5 】 3) ヘテロ接合度における遺伝子座内分散について、Stephens et al. (1992) にコンピューターシミュレーションの方法が示されているが、この方法はヘテロ接合度が小さいとき、その値を過小評価するという欠陥があることがわかった。そこで「OKN 補正」により、シミュレーションの方法を改良した。

【 0 0 1 6 】 4) ヘテロ接合度の誤差について、遺伝子座間分散と遺伝子座内分散の和の平方根をとることにより、ヘテロ接合度の誤差が得られる。図 1 は、DNA フィンガープリントからヘテロ接合度の誤差を計算する全体動作を示すフローチャートである。ここで、ステップ S 2 とステップ S 3 とは入れ換え可能である。また、「ステップ S 4 ステップ S 5」と「ステップ S 6 ステップ S 7」も入れ換え可能である。

【 0 0 1 7 】 図 2 は、ステップ S 2 における総バンド数を計算する動作を示すフローチャートである。図 3 は、ステップ S 3 におけるヘテロ接合度を計算する動作を示すフローチャートである。ここで、計算に用いられている式は Stephens et al. (1992) からとられているが、これを Jin and Chakraborty (1993) による式で置き換えてもよい。その場合は、

【 0 0 1 8 】

【数 5】

$$\left[\frac{\sum_{i=1}^{NT} g(i) \frac{1}{NT}}{g(NT) + \sum_{j=1}^{NT-1} g(j) \left(1 - \sqrt{1 - \frac{j}{NT}} \right) - \frac{1}{8NT} \sum_{j=1}^{NT-1} \frac{j}{NT \sqrt{1 - \frac{j}{NT}}} \right] - 1$$

【 0 0 1 9 】 となる。図 4 は、ステップ S 4 におけるバンドのブートストラップ再抽出を NB 回実行してヘテロ

接合度を NB 個求める動作を示すフローチャートである。図 5 及び図 6 は、ステップ S 6 における遺伝子座内

10

20

30

40

50

分散の計算のためのNR回のシミュレーションを実行してヘテロ接合度をNR個求める動作を示すフローチャートである。

【0020】なお、本発明は上記実施の形態に限定されるものではない。記録媒体は、例えば、磁気テープ、CD-ROM、ICカード、RAMカード等のいかなるタイプの記録媒体であってもよい。また、「ブートストラップ再抽出」とは、N個のサンプルから重複を許してN個のサンプルをランダムに抽出することである。

【0021】

【遺伝子座内分散における「OKN補正」の意義と方法】Stephens et al. (1992) による遺伝子座内分散のシミュレーションの方法の概略は、同論文のp740に記述されている。それによれば、真の出現頻度が j/n であるようなバンドが、n個体のサンプリングで全く検出されない確率は、 $(1 - j/n)^n$ であるから、見かけ上の出現頻度が j/n であるバンドの数を $f_o(j)$ とすると、出現頻度が j/n であるバンドの真の数値は、

【0022】

【数6】

$$f_T(j) = f_o(j) \frac{1}{[1 - (1 - \frac{j}{n})^n]}$$

【0023】であるとされている。この補正によって、出現頻度の低い所での真のバンド数は正しく見積もられており、特にヘテロ接合度が高い時はこの補正に基づくシミュレーションは有効である。

【0024】しかしながら、この補正法では、 $j = n$ のとき、

$$f_T(n) = f_o(n)$$

となり、全く補正が効いてない。ところが実際には、例えば $f_T(n-1)$ のバンドが $f_o(n)$ に貢献することもありうるのであるから、正しくは $f_T(n) < f_o(n)$ なくてはならない。そしてこのような傾向はヘテロ接合度の低い時に大きく効いてくるので、Stephens et al. (1992) の補正のみでは、シミュレーションによってヘテロ接合度の値そのものを過小評価してしまうことになる。そこで、上述の難点を克服するために新たに「OKN補正」を加えた。

【0025】<方法> 真の出現頻度が j/n であるようなバンドが、n個体のサンプリングでn個体すべてに検出されてしまう確率は、 $(j/n)^n$ であるから、見かけ上の出現頻度が j/n であるバンドの数を $f_o(j)$ とすると、出現頻度が j/n であるバンドの真の数値は、 $j < n$ として、

【0026】

【数7】

$$f_T(j) = f_o(j) \frac{1}{[1 - (1 - \frac{j}{n})^n - (\frac{j}{n})^n]}$$

【0027】で近似される。ただし、 j は1から $n-1$ までの数値をとる。右辺の分母の最後の項 $-(j/n)^n$ が「OKN補正」である。一方、 $j = n$ については、

【0028】

【数8】

$$f_T(n) = f_o(n) - \sum_{i=1}^{n-1} \left(\frac{i}{n}\right)^n f_T(i)$$

【0029】であり、右辺の最終項

【0030】

10 【数9】

$$\sum_{i=1}^{n-1} \left(\frac{i}{n}\right)^n f_T(i)$$

【0031】が「OKN補正」である。このようにして、ヘテロ接合度の小さい場合にも偏りのないヘテロ接合度をシミュレーションできるようになり、その結果、ヘテロ接合度に伴っている遺伝子座内分散についても、より正確な値が求められるようになったと考えられる。

【0032】ただし、この補正によって、ヘテロ接合度そのものの推定はかなり改善されるものの、遺伝子座内分散の値についてはあまり大きな変更が生じない。したがって、遺伝子座間分散が遺伝子座内分散よりもはるかに大きいケースにおいては、この「OKN補正」はあまり大きな貢献をする訳ではないことに注意する必要がある。あくまでも遺伝子座間分散の見積もりの方がより重要である。

【0033】

【発明の効果】以上のように、本発明によれば、複数遺伝子座DNAフィンガープリントから得られる平均ヘテロ接合度の誤差の範囲を推定するという、従来存在しなかった推定を可能とするものである。

30

【図面の簡単な説明】

【図1】DNAフィンガープリントからヘテロ接合度の誤差を計算する全体動作を示すフローチャートである。

【図2】総バンド数を計算する動作を示すフローチャートである。

【図3】ヘテロ接合度を計算する動作を示すフローチャートである。

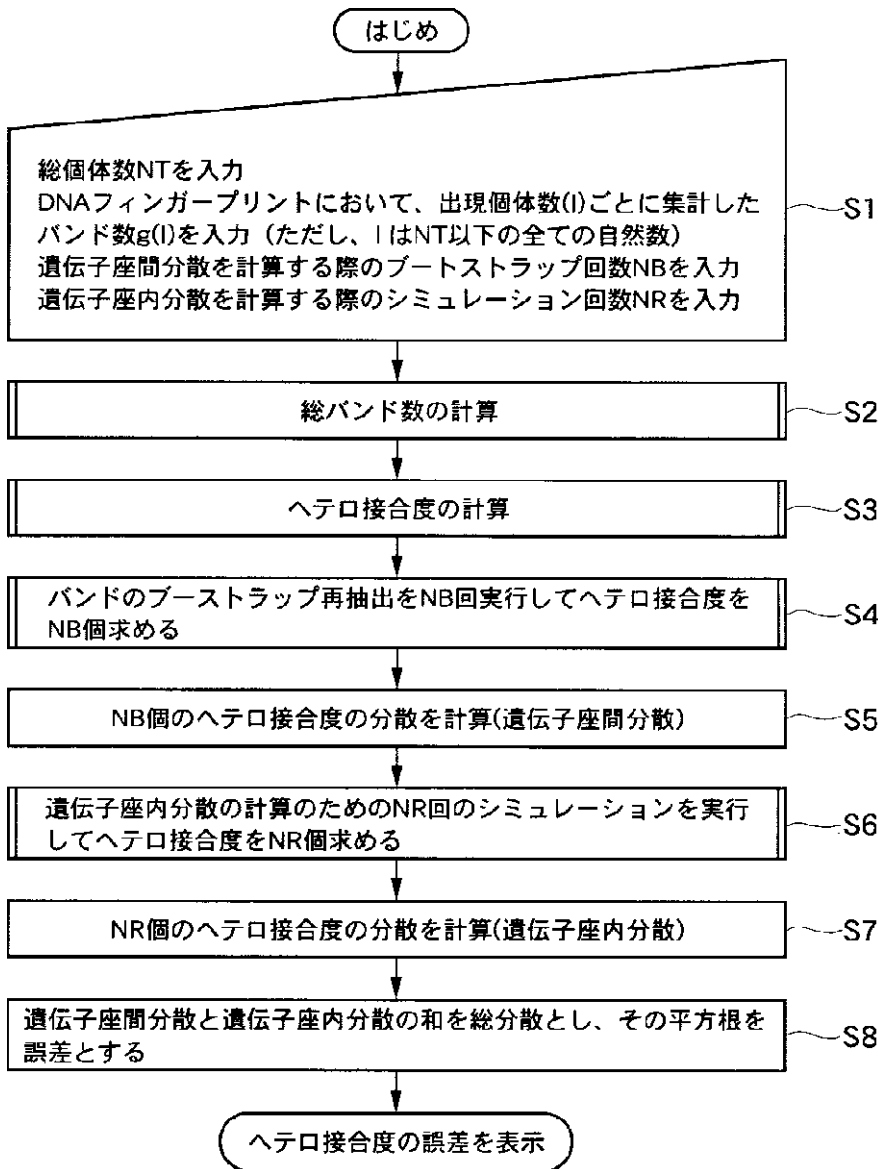
【図4】バンドのブートストラップ再抽出をNB回実行してヘテロ接合度をNB個求める動作を示すフローチャートである。

40

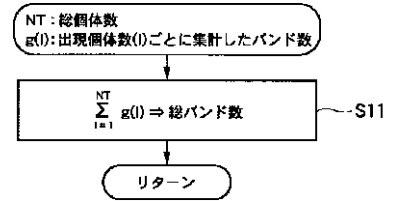
【図5】遺伝子座内分散の計算のためのNR回のシミュレーションを実行してヘテロ接合度をNR個求める動作を示すフローチャート(その1)である。

【図6】遺伝子座内分散の計算のためのNR回のシミュレーションを実行してヘテロ接合度をNR個求める動作を示すフローチャート(その2)である。

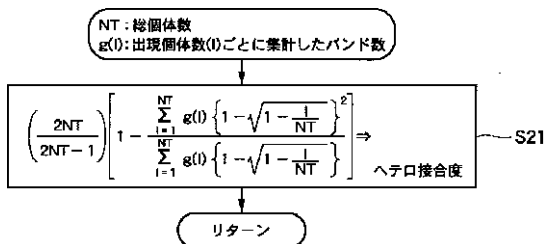
【図 1】



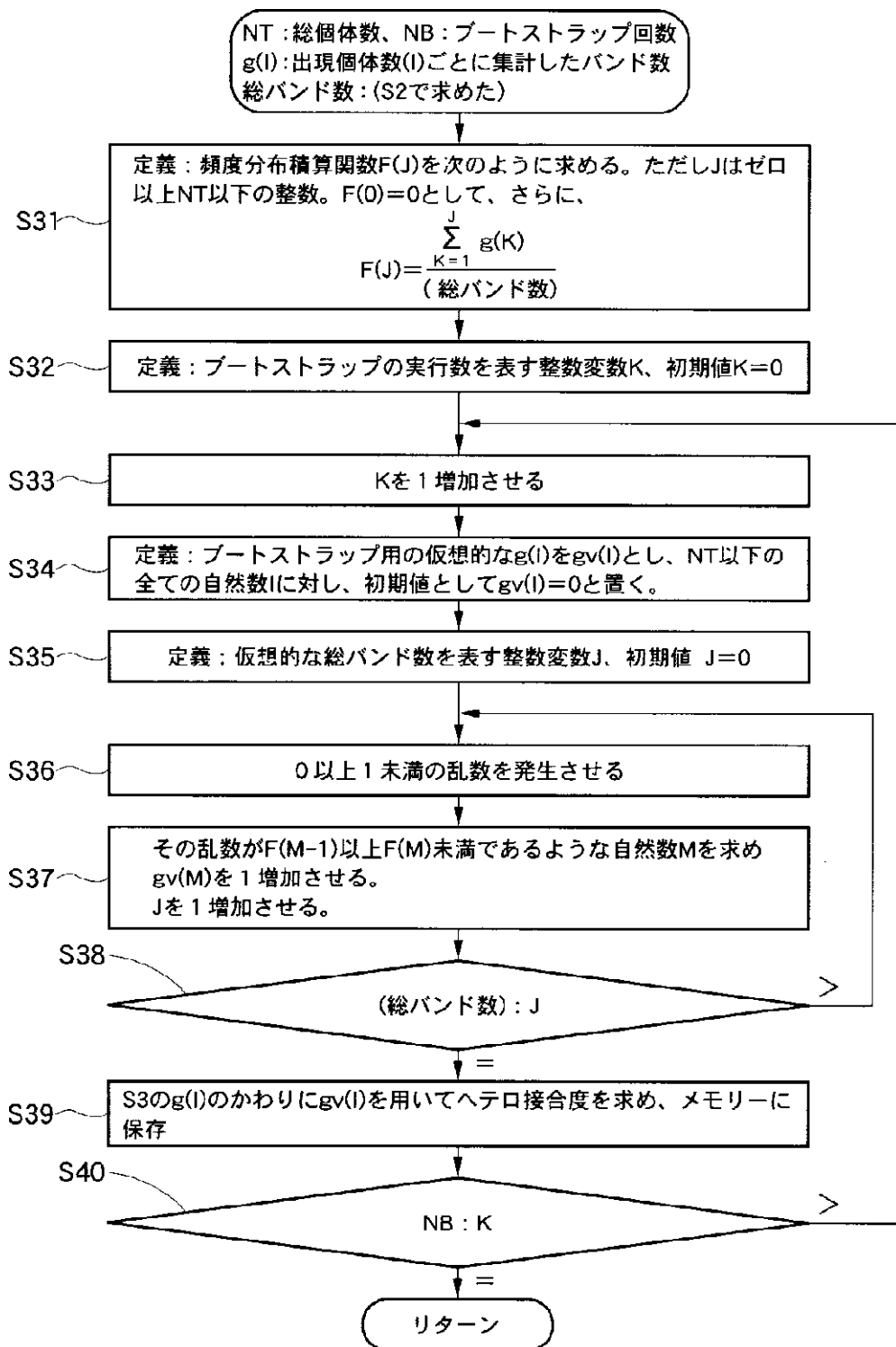
【図 2】



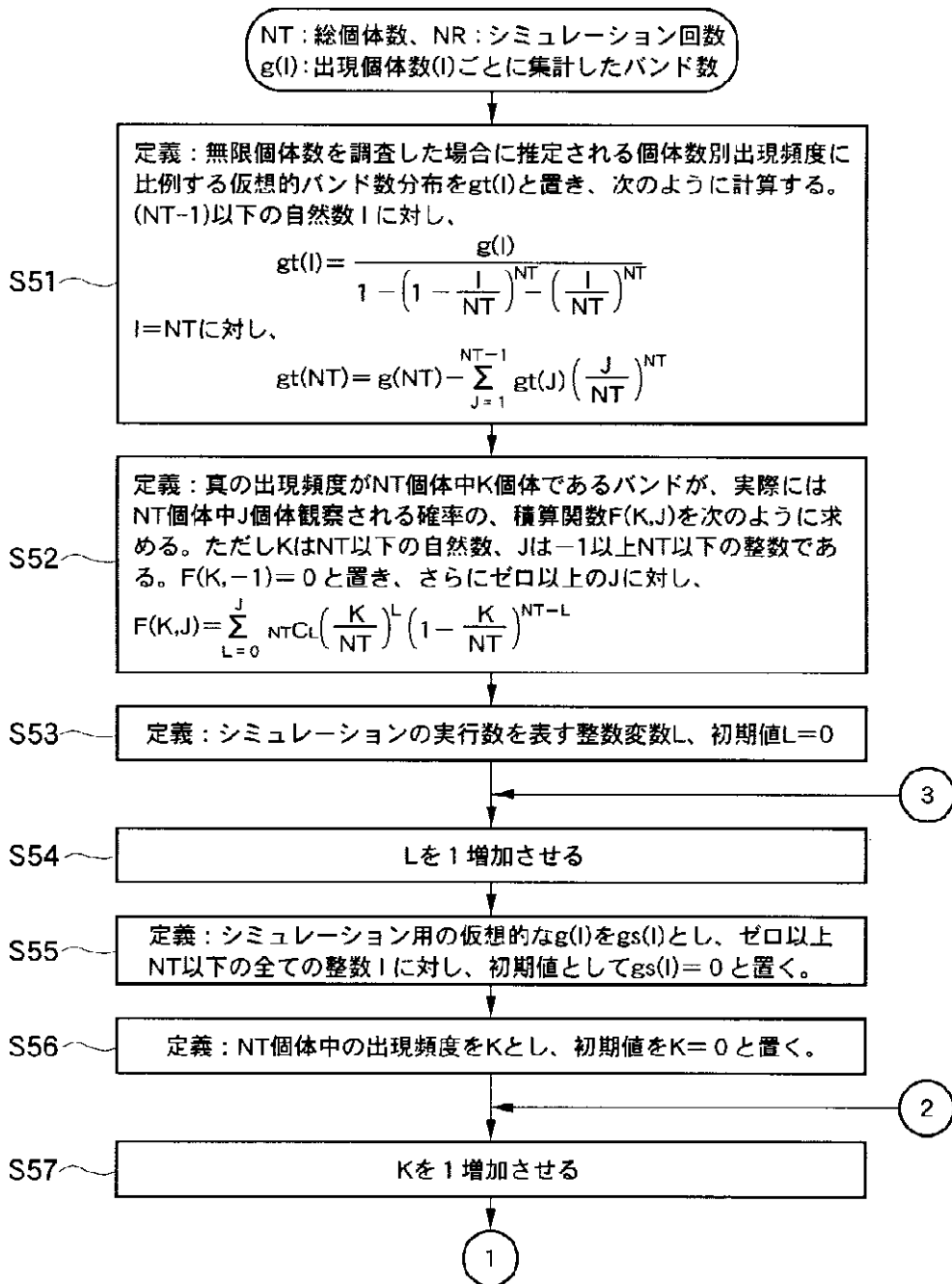
【図 3】



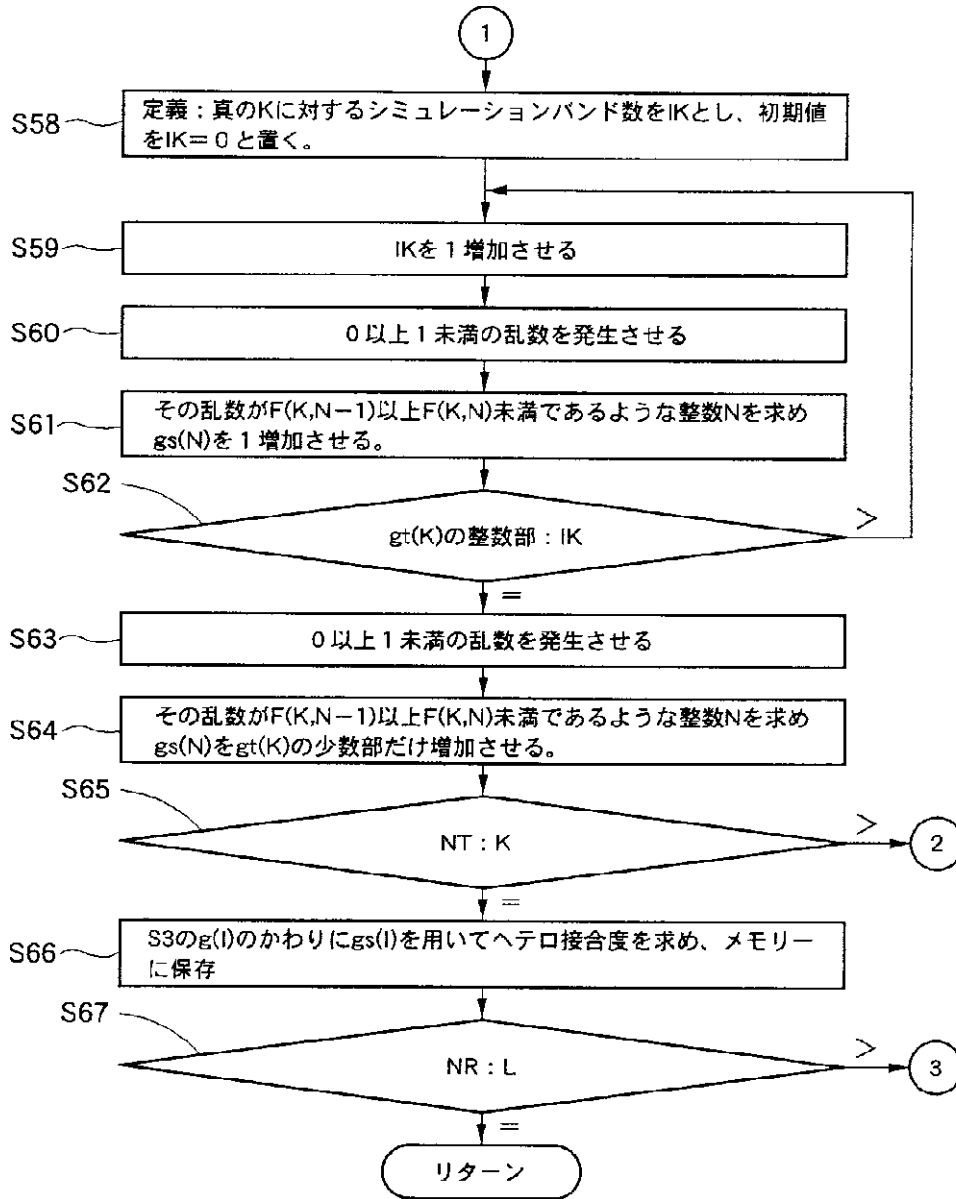
【図4】



【図 5】



【図6】



フロントページの続き

Fターム(参考) 2G045 AA35 AA40 DA13 JA01 JA02
4B024 AA11 AA19 AA20 CA01 HA11
HA19
5B056 BB64 HH00