

機関番号：14301

研究種目：挑戦的萌芽研究

研究期間：2012～2013

課題番号：24658235

研究課題名(和文) 絶滅危惧動物の飼育集団に関するDNA多型と血統の情報による集団構造解析手法の開発

研究課題名(英文) Analyzing the genetic population structure using DNA marker and pedigree information in the artificially-reared population of an endangered animal species

研究代表者

祝前 博明(IWASAKI, Hiroaki)

京都大学・(連合)農学研究科(研究院)・教授

研究者番号：00109042

交付決定額(研究期間全体)：(直接経費) 3,100,000円、(間接経費) 930,000円

研究成果の概要(和文)：国内コウノトリ集団に多くの子孫を残している基礎個体5羽を対象として、一度に百億塩基以上のDNA配列データの取得が可能な次世代シーケンサーを利用し、遺伝的多様性を評価するために必要なDNA多型マーカーの大規模な開発を実施した。1億8千万本を超える101塩基のDNA配列データを取得し、バイオインフォマティクスによる解析の結果、7万以上の多型マーカー候補を検出した。これらのうち、5羽の基礎個体のすべてについてマーカー型データが得られた約3万8千の多型マーカーを利用して様々な解析を実施し、個体間の遺伝的関係を明らかにした。

研究成果の概要(英文)：Since the current Japanese population of oriental stork originates from a small number of founders, understanding the genetic diversity between them is critical for the long-term protection and management of the population. The polymorphic DNA markers are essential tools for evaluating the genetic relatedness between the founders. To discover genome-wide DNA markers among five dominant founder genomes, we used a next-generation sequencer which produced sequences of over 10 Gb. We generated 185 million 101-bp reads and detected 74,552 putative single nucleotide polymorphisms (SNPs) by using bioinformatics. Out of them, 38,928 putative SNPs with the genotyping data in every founder were applied to the various analyses and the genetic relatedness between the founders were evaluated.

研究分野：農学

科研費の分科・細目：畜産学・獣医学・応用動物科学

キーワード：絶滅危惧動物 コウノトリ 人工飼育集団 次世代シーケンサー 一塩基多型 血統情報 ゲノム関係行列 遺伝的多様性

1. 研究開始当初の背景

環境省のレッドデータブック(2007年)によれば、日本の哺乳類の24%、鳥類の13%などが絶滅の恐れのある動物種となっており、その数は年々増加している。このような状況のもとで、わが国においては、人工飼育集団の増殖と野生復帰に係る国家的なプロジェクトとして、特別天然記念物のコウノトリとトキの人工飼育および野外への(試験)放鳥が実施されている。しかし、コウノトリでは、飼育集団の遺伝的構造の総合的な評価に基づく保全戦略は十分には採用されておらず、トキにおいては、始祖個体を非近交・無血縁とみなした血統情報のみに基づく集団構造解析の戦略がとられているのが現状である(日本のトキ集団の遺伝的多様性の現状、祝前博明、トキ再発見30周年記念中日韓国際ワークショップ報告書, 2011)。したがって、いずれのケースについても、さらに科学的なアプローチの構築による集団構造解析の実施が課題であると考えられた。

2. 研究の目的

絶滅危惧動物の人工飼育集団では、自然界への再導入(野生復帰)を目指し、少数の始祖個体からの成功裡の増殖が必須である。その際、近親交配の回避、遺伝的多様性の維持および飼育環境への遺伝的適応の最小化が重要であり、そのためには、集団の遺伝的構造の的確な把握が不可欠である。しかし、現時点では、集団構造の評価は、先進事例の場合でも、血統記録の解析のみに依っているか、あるいはごく少数のDNA多型情報を用いた解析のみに依っているかのいずれかである場合が多い。

そこで、本研究課題では、コウノトリ人工飼育集団の複数の始祖個体について、数千~数万ヶ所に及ぶDNA多型を効率的に検出し、個体間の分子近縁関係を適切に評価するスキームを確立するとともに、当該分子近縁情報と血統情報とを合理的に統合して、人工飼育集団の遺伝的構造のより適切な評価法の確立を目指そうとした。

3. 研究の方法

(1)日本コウノトリ集団の主要な始祖5個体(A:J047F, B:J048M, C:J273F, D:J305M, E:J366M)からゲノムDNAを精製し、制限酵素HaeIIIまたはMboIで完全消化した後、1.5%アガロースゲルを用いて分離した。250-450bpのサイズを持つDNA断片をゲルから精製した。

(2)精製したDNA断片にアダプター配列を付加した後、再度アガロースゲルを用いて分離し、400-500bpのサイズを持つDNA断片(アダプター120bp+ゲノムDNA断片280-380bp)をゲルから精製し、Reduced Representation Library(RRL)を作製した。各個体からのRRLは異なるタグ配列を持つアダプターを付加することで識別できるようにした。

(3)5つのRRLを混合した後、次世代シーケンサー(イルミナ社製HiSeq2000)の1レーンを使用し、各DNA断片の両端101baseの塩基配列を決定した。

(4)得られた配列データ(101baseのRead)に対し、アダプタートリミングおよびフィルター処理(QV10未満の塩基を含むReadを削除)した。

(5)フィルター済みの全Readを用いてSEEDプログラムによるクラスタリングを実施した。

(6)クラスタリングされた101baseの配列に対し、DNA断片の両端の配列を連結する処理を実施し、コンセンサス配列(両端の101baseの配列間に98個のNを加え、300baseの配列とした)を作製した。このクラスタリングでのdepthが10以上であったコンセンサス配列を以下のマッピングに使用した。

(7)bowtieプログラムにより、各個体からのRead-pair(DNA断片の両端101baseの配列を連結)を作製したコンセンサス配列に対しマッピングした。

(8)SNP部位の検出では、始祖5個体全体でのマッピングのdepthが100以上であったSNP候補に対し、さらにフィルタリング処理を実施した。

(9)マイクロサテライトマーカの検出では、2塩基の繰り返しを8回以上、3塩基の繰り返しを5回以上あるいは4塩基の繰り返しを4回以上含むコンセンサス配列を抽出し、繰り返し配列の数のみが異なるコンセンサス配列をbowtieプログラムによりグループ化した。さらに(7)のマッピングデータを利用し、各アレルの検出数を集計した。

(10)次に、これらのゲノムの全域にわたる多数のDNA多型の情報をもとに、基礎集団内のゲノム配列の類似性を推定し、始祖個体間のゲノム関係行列を導出した。

(11)さらに、血統情報に基づく相加的血縁行列における始祖個体群に対応する部分行列を、このゲノム関係行列で置き換え、始祖個体の次世代以降の全後代個体について、このゲノム関係行列を考慮に入れたより合理的な遺伝的関係行列の導出を図ろうとした。また、この種の情報に基づいて、人工飼育集団の遺伝的構造に係る諸パラメータ値の算出を図ろうとした。

4. 研究成果

(1)シーケンスの結果、約1.85億本の101baseのReadデータが得られた(表1)。アダプタートリミングおよびフィルター処理を通過したReadは約1.27億本であった(表1)。各始祖個体でのフィルター済みRead数は、2千2百万-2千9百万であった(表2)。

(2)クラスタリングにより作製された総コンセンサス配列数は10,497,263であり、これらのうちdepthが10以上であったものは772,830であった(表3)。マッピングの結果、74,552のSNP候補が検出された(表3)。SNP

候補の検出頻度からは、始祖5個体の集団におけるゲノム全体でのSNP数は約4.8百万箇所と推定された(表3)。全SNP候補74,552箇所のうち、始祖5個体のいずれにおいてもマッピングでのdepthが20以上であったものは38,928であった(表3)。

表1. HiSeq2000によるシーケンスデータの概要

| 項目 | 全体 |
|--------------|-------------|
| 総Read数 | 185,675,654 |
| 総塩基数(Mb) | 18,753 |
| トリミング後総Read数 | 176,008,198 |
| フィルター済総Read数 | 127,299,750 |

表2. HiSeq2000によるシーケンスデータの個別別概要

| 項目 | 個体A | 個体B |
|-----------------------|------------|------------|
| 総Read数 | 43,236,738 | 39,936,468 |
| 総塩基数(Mb) | 4,367 | 4,034 |
| % of \geq Q30 Bases | 93.66 | 93.74 |
| トリミング後総Read数 | 40,805,300 | 37,959,648 |
| フィルター済総Read数 | 29,668,094 | 27,683,522 |
| 項目 | 個体C | 個体D |
| 総Read数 | 33,009,492 | 35,024,658 |
| 総塩基数(Mb) | 3,334 | 3,537 |
| % of \geq Q30 Bases | 93.46 | 93.21 |
| トリミング後総Read数 | 31,305,812 | 33,237,826 |
| フィルター済総Read数 | 22,524,654 | 23,670,450 |
| 項目 | 個体E | |
| 総Read数 | 34,468,298 | |
| 総塩基数(Mb) | 3,481 | |
| % of \geq Q30 Bases | 93.61 | |
| トリミング後総Read数 | 32,699,612 | |
| フィルター済総Read数 | 23,753,030 | |

% of \geq Q30 Bases: フィルター済の総塩基数に占めるQ30以上のHigh Quality Baseの割合

表3. 始祖5個体におけるSNP候補の検出

| 項目 | 全体 | | |
|--|-------------------|-------------------|-------------------|
| Read-pair数 | 63,649,875 | | |
| 総Consensus配列数 | 10,497,263 | | |
| Consensus配列数 (clustering depth \geq 10) | 772,830 | | |
| Consensus配列数 (mapping depth \geq 100 In total of 5 founders) | 115,815 | | |
| 総SNP候補数 | 74,552 | | |
| SNP候補数 (mapping depth \geq 20 In each of 5 founders) | 38,928 | | |
| ゲノム当たりの推定SNP数 | 4.8×10^6 | | |
| 項目 | グループ | | |
| | Both-end HaeIII | Both-end Mbol | Others |
| Read-pair数 | | | |
| 総Consensus配列数 | 575,190 | 551,170 | 9,370,903 |
| Consensus配列数 (clustering depth \geq 10) | 332,224 | 269,737 | 170,869 |
| Consensus配列数 (mapping depth \geq 100 In total of 5 founders) | 112,683 | 926 | 2,206 |
| 総SNP候補数 | 72,729 | 546 | 1,277 |
| SNP候補数 (mapping depth \geq 20 In each of 5 founders) | 38,218 | 210 | 500 |
| ゲノム当たりの推定SNP数 | 4.8×10^6 | 4.4×10^6 | 4.3×10^6 |

(3)これら38,928のSNP候補のマーカー型データを用いて、始祖個体間の遺伝的な関係について検討した。遺伝子型がヘテロ型であるSNP候補の割合は、いずれの始祖個体においても約38%であった(表4)。特定の始祖個体に特異的なSNPアリルの数は1,824(始祖個体C)から3,434(始祖個体E)であった。pair-wiseでの遺伝子型の比較では、マーカー型が一致するSNPの割合は47.2-57.1%の範囲であり、始祖個体BとC間で57.1%と最も高く、次いで始祖個体CとD間の53.1%、それ以外の組み合わせでは、いずれも約47%であった(表5)。複数のSNP候補部位を含むコンセンサス配列は9,030検出され、ハプロタイプ数は2-8の範囲であり、ハプロタイプ数3のものが最も多かった。

(4)マイクロサテライトマーカーは、計127個が検出され、2、3、4および5アリルのマーカーはそれぞれ104、16、6および1個であった。

(5)上述のSNP候補の情報を用い、Yangら(2010)の方法によって求めた始祖5個体についてのゲノム関係行列は、表6のとおりであった。また、各SNP候補での遺伝子型の情報に基づいた主成分分析による始祖5個体の関係は、図1のように表された。

(6)現在、血統情報に基づく相加的血縁行列とゲノム関係行列とを考慮に入れたより合理的な遺伝的關係行列を導出中であり、当該行列の情報から、集団構造に関する各種のパラメータ値を算出する。

表4. 各始祖個体におけるSNP候補のマーカー型

| 項目 | 個体A | 個体B | 個体C |
|------------|--------|--------|--------|
| メジャーアリルホモ型 | 22,816 | 22,918 | 23,226 |
| マイナーアリルホモ型 | 1,473 | 1,168 | 822 |
| ヘテロ型 | 14,623 | 14,824 | 14,860 |
| ヘテロ型の割合(%) | 37.6 | 38.1 | 38.2 |
| 項目 | 個体D | 個体E | |
| メジャーアリルホモ型 | 22,814 | 22,918 | |
| マイナーアリルホモ型 | 1,173 | 1,391 | |
| ヘテロ型 | 14,922 | 14,600 | |
| ヘテロ型の割合(%) | 38.3 | 37.5 | |

表5. マーカー型が一致したSNPの数(上段)と割合(下段)

| | 始祖個体 | | | | |
|---|------|--------|--------|--------|--------|
| | A | B | C | D | E |
| A | - | 18,224 | 18,543 | 18,513 | 18,649 |
| B | 46.8 | - | 22,233 | 18,378 | 18,424 |
| C | 47.6 | 57.1 | - | 20,665 | 18,413 |
| D | 47.6 | 47.2 | 53.1 | - | 18,641 |
| E | 47.9 | 47.3 | 47.3 | 47.9 | - |

表6. ゲノム関係行列

| | 始祖個体 | | | | |
|---|-------|-------|-------|-------|------|
| | A | B | C | D | E |
| A | 0.87 | | | | |
| B | -0.25 | 0.83 | | | |
| C | -0.25 | -0.05 | 0.81 | | |
| D | -0.22 | -0.26 | -0.14 | 0.84 | |
| E | -0.20 | -0.24 | -0.26 | -0.22 | 0.87 |

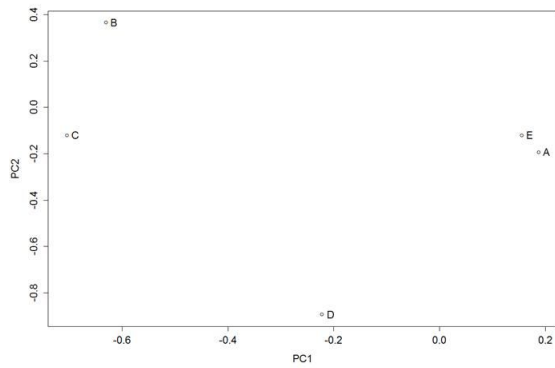


図1. 主成分分析の結果

5. 主な発表論文等
該当なし

6. 研究組織

(1) 研究代表者

祝前 博明 (IWAISAKI, Hiroaki)
京都大学・大学院農学研究科・教授
研究者番号：00109042

(2) 連携研究者

谷口 幸雄 (TANIGUCHI, Yukio)
京都大学・大学院農学研究科・助教
研究者番号：10252496