

国立環境研究所研究プロジェクト報告 第124号  
NIES Research Project Report, No.124

SR - 124 - 2017

# 絶滅過程解明のための絶滅危惧種ゲノムデータベース構築 (所内公募型提案研究)

Establishing genome database to understand extinction process

平成25～27年度  
FY2013～2015

NIES



国立研究開発法人 国立環境研究所  
NATIONAL INSTITUTE FOR ENVIRONMENTAL STUDIES  
<http://www.nies.go.jp/>



国立環境研究所研究プロジェクト報告 第124号

NIES Research Project Report, No.124

SR - 124 - 2017

# 絶滅過程解明のための絶滅危惧種ゲノムデータベース構築 (所内公募型提案研究)

Establishing genome database to understand extinction process

平成25～27年度  
FY2013～2015

所内公募型提案研究「絶滅過程解明のための絶滅危惧種ゲノムデータベース構築」  
(期間：平成 25 ～ 27 年度)

課題代表者：中嶋信美

執筆者：中嶋信美、大沼 学、遠藤大二\*、村山美穂\* (\*は所外)

編 者：中嶋信美

## 序

21世紀はゲノム科学の時代と言われています。2006年に次世代シーケンサーの販売が開始されて以来、塩基配列を解読するコストは格段に下がりました。また、次世代シーケンサーの汎用化により、これまで個々の遺伝子の働きを調べ、部分的な情報から全体での働きを推測する方法しかとれなかった研究が、全遺伝子や全一塩基多型情報といった全体情報から必要な情報を抽出し、その細部を解明する研究に移行しつつあります。

これまでに多くのモデル生物（ヒト、マウス、イネ、トマトなど）の全ゲノム配列が解読され、医療や作物の品種改良等の応用研究へ波及しています。モデル生物以外の生物種についても全ゲノム解読が進行中であり、60,000種以上のゲノム解読が概ね終了しています。これらの研究の進展により、創薬、農業、環境など多くの分野でゲノム情報を利用した新たな研究展開が進むものと期待されています。これらの中で特に絶滅危惧生物の保護・増殖研究において研究対象生物種の全ゲノム情報があると、当該種の遺伝的な多様性を拡大する方向で保護・増殖計画を立案できるようになることから、公的な機関による全ゲノム情報の取得と公開が期待されていました。

国立環境研究所では、この流れの中で、ゲノム科学研究をより加速させるため、国内で保護事業が進められている絶滅危惧鳥類3種を選び、そのドラフトゲノム情報を解明し、公開することを目標として、所内公募型提案研究「絶滅過程解明のための絶滅危惧種ゲノムデータベース構築」（課題代表者：中嶋信美）を立ち上げました。この報告書は平成25～27年度の3年間の研究成果をとりまとめたものです。研究成果は全体で5Gbp（50億塩基対）以上のテキストファイルであり、ここではその成果の概要について記載しました。多くの方にこのデータが利用され、絶滅危惧種の保護・増殖の効率化に役立つことを願ってやみません。なお、ゲノム解析の方法論は急速に進歩しているため、この研究で公開したデータは今後適宜アップデートして行く予定です。

平成29年10月

国立研究開発法人 国立環境研究所

理事長 渡 辺 知 保



## 目 次

1	研究の概要	1
1.1	研究全体の目的と目標	1
1.2	研究の概要	1
1.2.1	ヤンバルクイナ	1
1.2.2	コウノトリ	1
1.2.3	タンチョウ	1
2	研究の成果	2
2.1	目的と経緯	2
2.2	材料と方法	2
2.2.1	細胞の培養	2
2.2.2	DNA の抽出	2
2.2.3	RNA の抽出	2
2.2.4	Ion PGM 用のライブラリー作製	2
2.2.5	Miseq 用のライブラリー作製	3
2.2.6	cDNA ライブラリーの作成	3
2.2.7	Ion PGM による塩基配列の取得	3
2.2.8	Miseq による塩基配列の取得	3
2.2.9	Hiseq2000 によるヤンバルクイナ DNA の塩基配列の取得	3
2.2.10	PacBioRSII によるヤンバルクイナ DNA の塩基配列の取得	3
2.2.11	Velvet1.2.1 によるアセンブル	4
2.2.12	SparseAssembler によるアセンブル	4
2.2.13	Genomic Workbench ver 8.5.1 によるアセンブル	4
2.2.14	Trinity による cDNA データのアセンブル	4
2.2.15	Genomic Workbench ver 9.5.1 によるアセンブル	4
2.2.16	染色体数の計測	5
2.3	結果と考察	6
2.3.1	ヤンバルクイナ	6
2.3.2	コウノトリ	6
2.3.3	タンチョウ	6
2.4	まとめ	7
[資 料]		
1	研究の組織と研究課題の構成	11
1.1	研究の組織	11
1.2	研究課題と担当者	11
2	研究成果発表一覧	12
2.1	誌上発表	12
2.2	口頭発表	13





# 1 研究の概要

## 1.1 研究全体の目的と目標

国立環境研究所では平成14年度より、国内外の希少生物の組織を収集し、そこから培養細胞を確立して、恒久的な生物資源を保管する事業（タイムカプセル事業）を行ってきた。平成24年度末の時点で鳥類41種、哺乳類18種、ハ虫類1種について総個体数1,955個体の組織が保管されている。平成24年度10月には、これまで保管した組織や培養細胞を利用して、絶滅危惧種の増殖や保全研究を行うために、野生動物ゲノム連携研究グループが組織された。野生動物ゲノム連携研究グループでは、①遺伝資源保存バンク機能の強化、②保全遺伝学研究の推進、③遺伝資源の有効利用法開発、の3つの研究を進めている。本研究は②を推進するための研究である。タイムカプセル事業で保管している絶滅危惧種のうち、3種（ヤンバルクイナ、タンチョウ、コウノトリ）についてドラフトゲノムの解読を行い、このドラフト配列を家禽類等の全ゲノム塩基配列と比較できるようデータベース化することを目的とした。

## 1.2 研究の概要

### 1.2.1 ヤンバルクイナ

HiseqとIonPGMで得られた合計104.7 Gbpのデータを高速アセンブラーでアセンブルした結果、N50が6,245 bp、全塩基数が1.1 Gbp、768,680個のコンティグに集約された。この結果をDDBJ（日本DNAデータベース）に登録した。さらにPacBioの平均11kbpで3.8Gbpのロングリードデータを追加し、Genomic Workbench Ver 9.5.1を用いてアセンブルした結果、N50が86,745 bp、43,399個のコンティグに集約された。コンティグの合計長が1.14 Gbpであり、鳥類の平均的なゲノムサイズと同等の大きさであった。mRNAの塩基配列データを取得し、171,023個でN50値が695 bpの転写産物に集約された。染色体数は $n=39$ であることを明らかにした。

### 1.2.2 コウノトリ

MiseqとIonPGMで得られた38.9 Gbpのデータを、ワークステーション2台を用いてアセンブルした結果、N50が13,257 bp、最大198kbpで505,419個のコンティグに集約された。この結果をDDBJに登録した。Genomic Workbench Ver 9.5.1を用いてアセンブルした結果、N50が27,056 bp、94,141個のコンティグに集約された。コンティグの合計長が1.2 Gbpであり、鳥類の平均的なゲノムサイズと同等の大きさであった。mRNAの塩基配列データをアセンブルした結果、190,153個でN50値が477 bpの転写産物に集約された。染色体数が $n=35$ であることを明らかにした。

### 1.2.3 タンチョウ

MiseqとIonPGMで得られた24.1 Gbpのデータを、ワークステーション2台を用いてアセンブルした結果、N50値8,571 bp、最大長107 kbpで434,214個コンティグに集約された。この結果をDDBJに登録した。mRNAの塩基配列データを取得し、アセンブルした結果、102,015個でN50値が446 bpの転写産物に集約された。染色体数が $n=40$ であることを明らかにした。

以上の成果は平成28年6月に国内の公的データベースであるDDBJに3種のドラフトゲノム配列の登録を行い、7月に公開された。

## 2 研究の成果

### 2.1 目的と経緯

国立環境研究所では平成14年度より、国内外の希少生物の組織を収集し、そこから培養細胞を確立して、恒久的な生物資源を保管する事業（タイムカプセル事業）を行ってきた。平成24年度末の時点で鳥類41種、哺乳類18種、ハ虫類1種について総個体数1,955個体の組織が保管されている。平成24年度10月には、これまで保管した組織や培養細胞を利用して、絶滅危惧種の増殖や保全研究を行うために、野生動物ゲノム連携研究グループを組織した。野生動物ゲノム連携研究グループは、①絶滅危惧種組織からの培養細胞の作製と保存法の改良および培養細胞の不老化の研究による遺伝資源保存バンク機能の強化、②DNAマーカー等の開発による保全遺伝学研究の推進、③死亡個体の死亡原因の究明と死亡などの生物資源の有効利用法開発、の3つの研究を進めている。

本研究は②を推進するための研究である。既にヤンバルクイナについては平成23年度よりドラフトゲノム解析に一部着手しており、安価なコンピューターを用いて解析する手法については、ある程度ノウハウの蓄積ができていた。また、平成24年度には次世代シーケンサー IonPGM が導入され、ドラフトゲノム解析が可能となった。本研究ではヤンバルクイナに加え、タンチョウとコウノトリを研究対象とした。これら3種について全ゲノムの塩基配列決定を行い、保護増殖研究等に利用できるように、ドラフトゲノム配列データを公表することを目的とした。また、ドラフトゲノム配列を家禽類等の全ゲノム塩基配列と比較できるようデータベース化する。

### 2.2 材料と方法

#### 2.2.1 細胞の培養

環境試料タイムカプセル棟に凍結保存中の、ヤンバルクイナ (*Gallirallus okinawae*)、タンチョウ (*Grus japonensis*) およびコウノトリ (*Ciconia boyciana*) をゲノム解析の対象とした。ヤンバルクイナについては筋組織（由来は NIES-ID 1991A (20101026)）および培養細胞（由来は 2679A (20120811)）、タンチョウについては培養細胞（由来は 451A (070507)）、コウノトリについては培養細胞（由来は 2881A (K9)）を DNA と RNA の抽出用試料として利用した。

#### 2.2.2 DNA の抽出

細切した筋組織あるいは回収した培養細胞を Buffer G2 (500ml, GIAGEN)、Proteinase K solution (20 ml) と混合し、56℃で一昼夜インキュベートした。組織あるいは細胞の溶解を確認後、500 μl の PCI(P(フェノール):C(クロロホルム):I (イソアミルアルコール) が 25:24:1 になるように混合) を加え、転倒混和した。11,000 x g、15 分間、4℃で遠心し上層の水層を回収した（この作業を2回行った）。次に水層と等量のクロロホルムを加え転倒混和後、11,000 x g、5 分間、4℃で遠心した。上層の水層回収し、等量のイソプロパノールを加え、DNA を沈殿させた。沈殿した DNA をチップの先で回収し70%エタノール1 ml で2回洗浄した。DNA を室温中で30分乾燥させたあと20 ml の TE に溶解した。

#### 2.2.3 RNA の抽出

ヤンバルクイナ、タンチョウ、コウノトリの培養細胞を回収した。回収した細胞から NucleoTrap<sup>®</sup> mRNA (タカラ) を使用して、polyA RNA の精製を行った。手順についてはメーカーのマニュアル ([http://catalog.takara-bio.co.jp/PDFS/PolyA\\_mRNA\\_j.pdf](http://catalog.takara-bio.co.jp/PDFS/PolyA_mRNA_j.pdf)) に従った。

#### 2.2.4 Ion PGM 用のライブラリー作製

Ion PGM で分析するゲノムライブラリーはキットの説明書に従い以下のように行った。1 μg のゲノム DNA を DNAase で 37℃で8分間切断した。Agencourt<sup>®</sup> AMPure<sup>®</sup> XP Reagent で精製後、切断した DNA の末端を修復後、Agencourt<sup>®</sup> AMPure<sup>®</sup> XP Reagent で DNA を精製した。さらに、ニック修復とアダプターとのライゲーションを行った

後、Agencourt® AMPure® XP Reagent で DNA を精製した。精製した DNA を e-gel で分離し 350-500bp の断片を回収した。Agencourt® AMPure® XP Reagent で DNA を精製後、DNA の濃度と分子量分布をバイオアナライザーで測定した。

#### 2.2.5 Miseq 用のライブラリー作製

2 µg の DNA を 50 µL の水に溶かし、コバリスにより超音波処理を行った。処理条件は Duty factor 20%, Peak Incident Power 50 W, Cycles per burst 200, Duration 30s。この条件で一度処理したあと、容器の水滴を遠心して落としたのち、もう一度同じ条件で処理を行った。切断した DNA を Ampure beads で精製後、TruSeq LT PCR free library preparation kit を用いて末端の修復と 3' 末端の A 付加を行い、アダプターとライゲーションした。Ampure beads で分子量 550 bp 付近の DNA を精製し分子量分布をバイオアナライザーで測定した。さらに、Kapa キットによる Real time PCR でライブラリーの濃度を定量した。

#### 2.2.6 cDNA ライブラリーの作成

エクソンの位置を推定するための情報取得のために、cDNA ライブラリーを作成した。対象種の培養細胞より polyA RNA を oligodT latex を用いて精製した。Ion RNaseq v2 キットを用いて、以下の操作により cDNA ライブラリーを作成した。0.1 µg の polyA RNA を Exonuclease III で 30 °C で 90 秒処理して、RNA を断片化した。アダプターをライゲーションしたのち、逆転写酵素で 1<sup>st</sup> strand を合成した。2<sup>nd</sup> strand を合成した後、PCR によりライブラリーを増幅した。増幅したライブラリーを Ampure beads で精製後、バイオアナライザーで濃度を測定して、cDNA ライブラリーとした。

#### 2.2.7 Ion PGM による塩基配列の取得

Ion PGM による塩基配列の取得は 400 bp キットと 318v2 チップを用いた。2.2.4 で作製したライブラリーを 50 pM に希釈し、25 µL を Ion One touch2 を用いてエマルジョン PCR を行った。増幅された DNA を結合した Beads を精製し、プライマーとアニリング後、DNA polymerase をくわえた。テンプレートに DNA polymerase が結合したものを、318v2 チップに入れた。これを Ion PGM に装着し、850 cycle の条件で塩基配列を解析した。出力結果は fastq 形式に変換後、ハードディスクに保存した。以上の操作を塩基配列にしておよそ 10 Gbp になるまで繰り返した。ヤンバルクイナが 19.5 Gbp、コウノトリが 12.2 Gbp、タンチョウは 13.5 Gbp のデータを得た。また、cDNA の塩基配列も Ion PGM で取得した。

#### 2.2.8 Miseq による塩基配列の取得

Miseq による塩基配列の取得は Miseq Reagent v3 600cycle キットを用いた。2.2.5 で作製したライブラリーを 4 nM に希釈し、その 5 µL に 0.2M NaOH を 5 µL 加え混合した。室温に 5 分放置しアルカリ変成を行った。HT1 buffer を 990 µL 加えて希釈し 20 pM にした。同じようにアルカリ変性した ΦX の DNA を 20 pM に希釈して、5 ~ 25% 加えた。HT1 buffer 150 µL と 20 pM のライブラリー 450 µL を混合し、全量を試薬カートリッジに挿入した。Pair-end で 301cycles の条件で塩基配列を解析した。出力結果は fastq 形式に変換後、ハードディスクに保存した。コウノトリについては 2 回、タンチョウは 1 回の解析を行い、それぞれ 26.7 Gbp、10.6 Gbp の pair-end データを得た。

#### 2.2.9 Hiseq2000 によるヤンバルクイナ DNA の塩基配列の取得

ヤンバルクイナの DNA を Hiseq2000 で解析した read data は、環境省沖縄自然保護事務所が実施した保護増殖事業で取得した 85.2 Gbp の pair-end データを譲り受けて使用した。

#### 2.2.10 PacBioRSII によるヤンバルクイナ DNA の塩基配列の取得

ヤンバルクイナについては北海道システムサイエンス株式会社の好意により、PacBioRSII による 3.8 Gbp の Long Read data が無償で提供された。

### 2.2.11 Velvet1.2.1によるアセンブル

コウノトリとタンチョウのリードデータは Velvet 1.2.1 でアセンブルを行った。コウノトリを例として方法を記す。\$ 以下は実施したコマンドの例である。

1. Miseq の pair-end reads の 2 つのファイルを fastq\_quality\_trimmer でクオリティーフィルタリングを実施した<sup>1)</sup>。  
\$ fastq\_quality\_trimmer Q33 -t 20 -l 30 -i 入力ファイル名 -o 出力ファイル名
2. クオリティーフィルタリングした Miseq の pair-end reads の 2 つのファイルをさらに、PRISEQ lite で pair-end フィルタリングを行った<sup>2)</sup>。  
\$ prinseq-lite.pl -min\_len 30 -fastq R1 のリードファイル名 -fastq2 R2 のリードファイル名 -out\_good 出力ファイル名
3. velveth によるハッシュテーブルの作成を行った<sup>3,4)</sup>。Kmer 値 31 ~ 61 の奇数値すべてについて実施した。  
\$ velveth <data dir path> kmer -fastq shortPaired -separate <R1.fastq> <R2.fastq>
4. それぞれの kmer 値のハッシュテーブルについて velvetg を実施してアセンブルを行い、最も N50 値が大きかったアセンブル結果を採用した。  
\$ velvetg <data dir path> -exp\_cov auto -cov\_cutoff auto -unused\_reads yes -scaffolding yes

### 2.2.12 SparseAssemblerによるアセンブル<sup>5)</sup>

ヤンバルクイナの Hiseq2000 出力データの fastq ファイルを R1, R2 ごとに連結して 2 つの fastq ファイルにまとめた。Sparse assembler の Denoiser でクオリティーフィルタリングを実施した。その後、kmer=17 でアセンブルを行った。この作業については独自に ruby によるプログラム assemble3.rb を作成してバッチ処理を行った。

### 2.2.13 Genomic Workbench ver 8.5.1によるアセンブル<sup>6-8)</sup>

コウノトリとタンチョウについては Velvet1.2.1 でアセンブルして作成した contig と Ion PGM のリードデータを De novo assemble で連結した。解析条件は fast mode を用いた。

### 2.2.14 TrinityによるcDNAデータのアセンブル<sup>9)</sup>

cDNA の read data は Trinity によってアセンブルした。

### 2.2.15 Genomic Workbench ver 9.5.1によるアセンブル<sup>6-8,10)</sup>

2016年11月に Genomic Workbench がバージョンアップされ、非常に効率が向上した。以下は本研究期間終了後に行った解析である。ヤンバルクイナのデータを例にして Genomic Workbench 9.5.1 によるアセンブル方法を記した。下記の手順で行った。

1. ヤンバルクイナの Hiseq2000 出力データの fastq ファイルを R1, R2 ごとに連結して 2 つの fastq ファイルにまとめた。ファイル名を yanbaru\_hiseq\_R1.fastq、yanbaru\_hiseq\_R2.fastq とした。  
\$ cat /\*.fastq > ./yanbaru\_hiseq\_R1.fastq
2. 2 つのファイルを fastq\_quality\_trimmer でクオリティーフィルタリングを実施した。  
\$ fastq\_quality\_trimmer Q33 -t 20 -l 30 -i yanbaru\_hiseq\_R1.fastq -o yanbaru\_hiseq\_R1\_good.fastq  
\$ fastq\_quality\_trimmer Q33 -t 20 -l 30 -i yanbaru\_hiseq\_R2.fastq -o yanbaru\_hiseq\_R2\_good.fastq
3. Genomic workbench (GW) De novo assembler でアセンブル。  
GWver 9.5.1 と Fishing module は MacPro (内部メモリー 64G, CPU 12 コア) にインストールした。yanbaru\_hiseq\_R1\_good.fastq と yanbaru\_hiseq\_R2\_good.fastq をペアエンド設定で GW ヘインポートし、word size 30, 40,

50, 60 の 4 つの条件で De novo assemble を実施した。Buble size は auto、mismatch などのコストは 3 , minimum contig length を 500 で実施した。得られたベストのコンティグのファイル名を yanbaru\_hiseq\_word40\_assemble.fa とし、アセンブルに使われなかったリードのファイル名を下記に設定した。

```
yanbaru_hiseq_R1_good (paired) un-mapped reads [yanbaru_hiseq_R1_good] (paired).fastq
```

```
yanbaru_hiseq_R1_good (paired) un-mapped reads [yanbaru_hiseq_R1_good] (single).fastq
```

4. IonPGM によるシーケンス結果 7 回分を一つのフォルダーに入れて cat で連結した。  

```
$ cat /*.fastq > ./yanbaru_PGM.fastq
```
5. yanbaru\_PGM.fastq のクオリティーフィルタリングは fastq\_quality\_trimmer を使って下記のコマンドで行った。  

```
$ fastq_quality_trimmer Q33 -t 10 -l 100 -i ./yanbaru_PGM.fastq -o ./yanbaru_PGM_u100.fastq
```
6. GW Finishing module の Join contig で yanbaru\_PGM\_u100.fastq と yanbaru\_hiseq\_R1\_good\_assembly40.fa を連結した。連結したファイルを yanbaru\_hiseq\_scaffold.fa というファイル名で保存した。
7. GW Finishing module の Join contig で yanbaru\_hiseq\_R1\_good (paired) un-mapped reads [yanbaru\_hiseq\_R1\_good] (paired).fastq と yanbaru\_hiseq\_scaffold.fa を連結した。連結したファイル名を yanbaru\_hiseq\_scaffold2.fa とした。
8. GW Finishing module の Join contig で yanbaru\_hiseq\_R1\_good (paired) un-mapped reads [yanbaru\_hiseq\_R1\_good] (single).fastq と yanbaru\_hiseq\_scaffold2.fa を連結した。連結したファイル名を yanbaru\_hiseq\_scaffold3.fa とした。
9. yanbaru\_PGM\_u100.fastq で yanbaru\_hiseq\_scaffold3.fa の Gap を FGAP で埋めた。Gap を埋めたファイル名を yanbaru\_hiseq\_scaffold4.fa とした。
10. yanbaru\_hiseq\_R1\_good (paired) un-mapped reads [yanbaru\_hiseq\_R1\_good] (paired).fastq で yanbaru\_hiseq\_scaffold4.fa の Gap を FGAP で埋めた。Gap を埋めたファイル名を yanbaru\_hiseq\_scaffold5.fa とした。
11. yanbaru\_hiseq\_R1\_good (paired) un-mapped reads [yanbaru\_hiseq\_R1\_good] (single).fastq で yanbaru\_hiseq\_scaffold4.fa の Gap を FGAP で埋めた。Gap を埋めたファイル名を yanbaru\_scaffold\_final.fasta とし、最終的な scaffold ファイルとした。

## 2.2.16 染色体数の計測

G-staining method (banding) による染色体解析は日本遺伝子研究所 (〒 983-0013 宮城県仙台市宮城野区中野 1-5-28) に依頼した。日本遺伝子研究所における作業手順の概要は以下のとおりである。

培養細胞の回収を行う前夜 (18 時)、培養液にコルセミドを添加した。T-25 培養ボトルの培養液をコニカルチューブに移し、0.25% トリプシンで 37℃ で 6 分間処理する。細胞をはがし、コニカルチューブの培養液を反応停止として使用した。全量をコニカルチューブに入れ、1,500 rpm、8 分遠心し、上清を除去した。予め 37℃ に保温にした低張液を 5 ml 添加し、駒込ピペットでゆっくりと 37℃ で 20 分間攪拌した。カルノア固定液を 5 滴 (120 ~ 150 µl) 添加、よく攪拌し 10 分静置した。さらにカルノア固定液を 1 ml 添加して攪拌した。1,500 rpm、8 分間遠心し、カルノア固定液を 5 ml 添加して、さらに攪拌した。上清を除去後、「カルノア固定液を 5 ml 添加して、さらに攪拌。上清を除去。」の操作を 5 回繰り返した。最後に上清を除去して残った細胞をサスペンドして、スライドグラスに 1 滴滴下した。高さ 2 センチ程度の高さよりスライドグラス中央に滴下し、乾くまでに 1 分程度で乾燥させ標本とした (室内湿度 55% 前後)。

各標本を5枚作製した。これらの標本を以下の方法でギムザ染色した。ハンクス液、リン酸バッファーを4℃より取り出した。標本を0.2%トリプシン処理(4℃、60秒)後、10%ギムザ液染色を5分間行った。試料を水洗後、乾燥した。

20細胞について、Gバンド分析と染色体構造及び数の構成の分析を行った。30細胞についてMODE分析と染色体数の分析を行った。染色体構成異常があった場合国際標準規約ISCNに基づいた表記を行った。

## 2.3 結果と考察

### 2.3.1 ヤンバルクイナ

Hiseq2000で得られた合計104.7 GbpのデータをSparseAssemblerでアセンブルした結果、N50が6,245 bp、全塩基数が1.1 Gbp、768,680個のコンティグに集約された。この結果をDDBJ(日本DNAデータベース)に登録した。さらにPacBioRSIIの平均11kbpで3.8 GbpのロングリードデータとIonPGMによる19.5 Gbpのデータを追加し、Genomic Workbench ver 9.5.1を用いて再アセンブルし、FGAPでgapを埋めた結果、N50が86,745 bp、43,399個のコンティグに集約された。コンティグの合計長が1.14 Gbpであり、鳥類の平均的なゲノムサイズと同等の大きさであった。mRNAの塩基配列データを取得し、171,023個でN50値が695 bpの転写産物に集約された。染色体数がn=39であることを明らかにした。

### 2.3.2 コウノトリ

Miseqで得られた26.7 GbpのデータをVelvetでアセンブルした結果、kmer=43でもっとも良い結果が得られた。さらに、Velvetで得られたcontigとIonPGMで得られた38.9 Gbpのデータを、Genomic Workbench ver 8.5.1でアセンブルした結果、N50が13,257 bp、最大198 kbpで505,419個のコンティグに集約された。この結果をDDBJに登録した。さらに、Genomic Workbench ver 9.5.1を用いて再アセンブルし、FGAPによってgapを埋めた結果、N50が27,056 bp、94,141個のコンティグに集約された。コンティグの合計長が1.2Gbpであり、鳥類の平均的なゲノムサイズと同等の大きさであった。mRNAの塩基配列データをアセンブルした結果、190,153個でN50値が477 bpの転写産物に集約された。染色体数がn=35であることを明らかにした。

### 2.3.3 タンチョウ

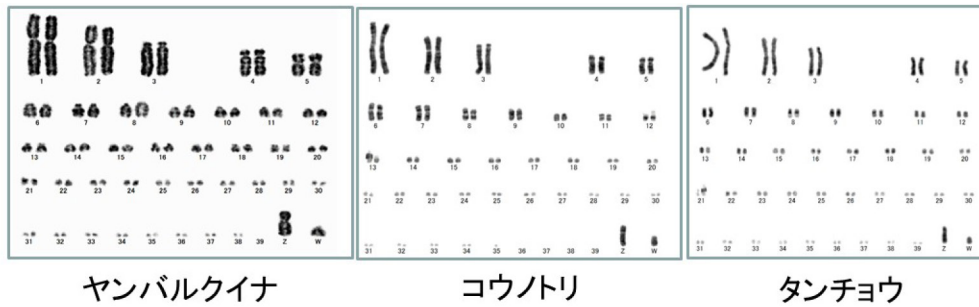
Miseqで得られた10.6 GbpのデータをVelvetでアセンブルした結果、kmer=41でもっとも良い結果が得られた。さらに、Velvetで得られたcontigとIonPGMで得られた24.1 Gbpのデータを、Genomic Workbench ver 8.5.1でアセンブルした結果、N50値8,606 bp、最大長107 kbpで357,545個コンティグに集約された。この結果をDDBJに登録した。さらに、Genomic Workbench ver 9.5.1を用いて再アセンブルし、FGAPによってgapを埋めた結果、N50が8,571 bp、434,214個のコンティグに集約された。mRNAの塩基配列データを取得し、アセンブルした結果、102,015個でN50値が446 bpの転写産物に集約された。染色体数がn=40であることを明らかにした。

表1 種ごとのアセンブル結果

生物種	装置	データ量と種類	アセンブラー	N50値(bp)	コンティグ数	全塩基数
ヤンバルクイナ	Hiseq 2000 Ion PGM PacBioRSII	85.2Gbp pair-end 19.5Gbp Single read 3.8Gbp long read	SparseAssembler GWVer 9.5.1+FGAP	6,245	768,680 43,399	1.1Gbp
				86,745		1.1Gbp
コウノトリ	Miseq Ion PGM	26.7Gbp pair-end 12.2Gbp Single read	Velvet+GW ver.8.5.1 GWver 9.5.1+FGAP	13,257	505,419 94,141	1.4Gbp
				27,056		1.2Gbp
タンチョウ	Miseq Ion PGM	10.6Gbp pair-end 13.5Gbp Single read	Velvet+GW ver.8.5.1 GWver 9.5.1+FGAP	8,606	357,545 434,214	1.3Gbp
				8,571		1.2Gbp

GW : Genomic Workbench

N50値 : コンティグを長い方から塩基数を合計し、ちょうど全塩基数の半分を超えたところのコンティグ長を表す値。



	染色体数 n
ヤンバルクイナ	39
コウノトリ	35
タンチョウ	40

図1 ヤンバルクイナ、コウノトリ、タンチョウの染色体写真と染色体数

## 2.4 まとめ

本研究により、絶滅危惧鳥類3種のドラフトゲノム構造を明らかにすることができた。詳しくはプレスリリース (<https://www.nies.go.jp/whatsnew/2016/20160805/20160805.html>) を参照されたい。本研究で公開したデータベースは、個体数減少の原因の推定や病気に対する感受性を推定する研究に利用されていく予定である。なお、De novo assemble のプログラムは発展途上にあり、数ヶ月単位で新たなプログラムが公開されている。ここに示した結果は平成28年12月時点でのものである。将来はアセンブルの方法論の進歩に従って随時更新していく予定である。

## 引用文献

- 1) FASTX-Toolkit : FASTQ/A short-reads pre-processing tools  
[http://hannonlab.cshl.edu/fastx\\_toolkit/index.html](http://hannonlab.cshl.edu/fastx_toolkit/index.html)
- 2) PRINSEQ: Easy and rapid quality control and data preprocessing.  
<http://prinseq.sourceforge.net/index.html>
- 3) Zerbino DR, McEwen GK, Margulies EH, Birney E (2009) Pebble and Rock Band: Heuristic Resolution of Repeats and Scaffolding in the Velvet Short-Read de Novo Assembler. PLoS ONE 4(12): e8407. doi:10.1371/journal.pone.0008407
- 4) Zerbino DR and Birney E (2008) Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Research 18:821-829.
- 5) Ye C, Ma ZS, Cannon CH, Pop M, Yu DY (2012) Exploiting sparseness in de novo genome assembly. The very large memory requirements for the construction of assembly graphs for de novo genome assembly limit current algorithms to super-computing environments. BMC Bioinformatics 13 (Suppl 6):S1
- 6) White paper on de novo assembly in CLC Assembly Cell 4.0 (2012)  
[http://resources.qiagenbioinformatics.com/white-papers/White\\_paper\\_on\\_de\\_novo\\_assembly\\_4.pdf](http://resources.qiagenbioinformatics.com/white-papers/White_paper_on_de_novo_assembly_4.pdf)
- 7) CLC Genomics Workbench  
<http://www.filgen.jp/Product/BioScience21-software/index11-g.htm>
- 8) CLC Genome Finishing Module  
<http://www.filgen.jp/Product/BioScience21-software/index11-Microbial.htm>

- 9) Grabherr, MG et. al (2011) Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat Biotechnol.* 2011 Jul; 29: 644–652.
- 10) Piro VC, Faoro H, Weiss VA, Steffens MBR, Pedrosa FO, Souza EM, Raittz RT (2014) FGAP: an automated gap closing tool. *BMC ResearchNotes*, 7:371. <http://www.biomedcentral.com/1756-0500/7/371>



[資料]



# 1 研究の組織と研究課題の構成

## 1.1 研究の組織

### [A 研究担当者]

生物・生態系環境研究センター	中嶋信美
生物・生態系環境研究センター	大沼 学

(注) 所属・役職は年度終了時点のもの。

### [B 客員研究員]

遠藤大二 (酪農学園大学)	(平成 25 ~ 27 年度)
京都大学 (京都大学)	(平成 25 ~ 27 年度)

## 1.2 研究課題と担当者

### DNA の採取と染色体数計測

生物・生態系環境研究センター	大沼 学
京都大学 野生動物研究センター	村山美穂*

### 塩基配列データの取得とアセンブル

生物・生態系環境研究センター	中嶋信美
酪農学園大学 獣医学部	遠藤大二*

(注) \*客員研究員

## 2 研究成果発表一覧

### 2.1 誌上発表

<雑誌>

---

発表者・(刊年)・題目・掲載誌・巻(号)・頁

---

Yanaka N. Y., Ioki M., Ohkoshi M., Nakajima N., Watanabe M. M. (2014) Herbicide-resistant mutants of *Botryococcus braunii* race B (strain BOT-22). *J. Appl. Phycology*, 26 (1), 25-28.

Yamaguchi H., Suzuki S., Sano T., Tanabe Y., Nakajima N., Kawachi M. (2016) Draft genome sequence of *Microcystis aeruginosa* NIES-98, a non-microcystin-producing cyanobacterium from Lake Kasumigaura, Japan. *Genome Announcements*, 4 (6), e01187-16.

---

## 2.2 口頭発表

発表者・(暦年)・題目・学会等名称・予稿集名・頁

---

Ioki M., Nakajima N., Watanabe M. M. (2014) Fueling the Molecular Biology of *Botryococcus braunii*. Asia Oceania Algae Innovation Summit 2014. Daejeon, Nov-14.

Ioki M., Houmpheng T, Nakajima N., Watanabe M. M. (2014) Ultraviolet-B radiation triggers hydrocarbon biosynthesis in *Botryococcus braunii*. Asia Oceania Algae Innovation Summit 2014. Daejeon, Nov-14.

熊倉圭子, 遠藤大二, 浅川満彦, 牛山喜偉, 長 雄一, 大沼 学, 中嶋信美 (2014) EPIC (Exon-Primed Intron-Crossing) - PCR による北海道の鳥類の種判別方法の開発. 第 20 回日本野生動物医学会つくば大会, 第 20 回日本野生動物医学会大会講演要旨集, 43

諸岡佳恵, 遠藤大二, 大沼 学, 中嶋信美 (2014) ヤンバルクイナ (*Galirallus okinawae*) の全ゲノム解析のためのコンテイング間ギャップ PCR 法の検討. 第 20 回日本野生動物医学会つくば大会, 第 20 回日本野生動物医学会大会講演要旨集, 116

中嶋信美, 大沼 学, 小林麻衣, 遠藤大二 (2016) 国立環境研究所のタイムカプセル事業で保存している絶滅危惧野生鳥類 3 種のドラフトゲノム解説. 第 22 回日本野生動物医学会宮崎大会, 第 22 回日本野生動物医学会宮崎大会講演要旨集, 119

小林麻衣, 羽賀 淳, 中嶋信美, 大沼 学, 遠藤大二 (2016) ヤンバルクイナゲノムデータを用いた飛翔能力関連遺伝子分析. 第 22 回日本野生動物医学会宮崎大会, 第 22 回日本野生動物医学会宮崎大会講演要旨集, 120

---

国立環境研究所研究プロジェクト報告 第 124 号  
NIES Research Project Report, No.124

(SR - 124 - 2017)

絶滅過程解明のための絶滅危惧種ゲノムデータベース構築  
(所内公募型提案研究)  
平成 25 ～ 27 年度  
Establishing genome database to understand extinction process  
FY2013 ～ 2015

---

平成 29 年 10 月 10 日発行

編 集 国立環境研究所 編集分科会

発 行 国立研究開発法人 国立環境研究所

〒305-8506 茨城県つくば市小野川 16 番 2

E-mail: pub@nies.go.jp

Published by the National Institute for Environmental Studies

16-2 Onogawa, Tsukuba, Ibaraki 305-8506 Japan

October 2017

---

組 版 株式会社 イ セ プ

〒305-0005 茨城県つくば市天久保 2 丁目 11-20

---

無断転載を禁じます

国立環境研究所の刊行物は以下の URL からご覧いただけます。  
<http://www.nies.go.jp/kanko/index.html>