

科学研究費助成事業 研究成果報告書

平成 30 年 6 月 8 日現在

機関番号：10101

研究種目：基盤研究(C) (一般)

研究期間：2014～2017

課題番号：26440193

研究課題名(和文) 多様な無脊椎動物の比較解析基盤としての比較ゲノムデータベース構築

研究課題名(英文) Building Comparative Genome Database for Invertebrate Species

研究代表者

遠藤 俊徳 (Endo, Toshinori)

北海道大学・情報科学研究科・教授

研究者番号：00323692

交付決定額(研究期間全体)：(直接経費) 3,800,000円

研究成果の概要(和文)：10の動物門に含まれる314種の動物ゲノムについて、遺伝子機能のスペクトル(ゲノムスペクトル)を、門、綱、目、科、属の各系統分類において調べ、共通遺伝子を整理することでプロテオームデータと統計量の整備を進めた。全生物系統に共通して存在する遺伝子は1,908で、ほぼ全てが基本生命維持機能のコア遺伝子群であり、系統固有の遺伝子も系統分類ごとに同定した。また、動物系統のみに共通して存在する1,299の遺伝子群を特定した。その多くはミトコンドリアに関係する遺伝子であり動物固有の機構の存在を示唆した。ほかに脊椎動物系統のみ不在の遺伝子など、生物進化過程を解明する重要な手掛かりが得られた。

研究成果の概要(英文)：Gene spectra for 314 animals species across ten phyla was investigated and organized based on the levels of taxonomic classifications; phylum, class, order, family, and genus. The number of genes in common across all investigated phyla was 1,908, most of which were covering fundamental core genes necessary to maintain life activities. A total of 1,299 lineage specific genes were also determined, where the majority were mitochondrial genes, suggesting animal specific mechanism requiring those genes. Interestingly, there were genes absent only in vertebrates. Those findings would provide key information to understand the process of evolution.

研究分野：情報生物学・分子進化

キーワード：比較ゲノム解析 ゲノムスペクトル

1. 研究開始当初の背景

ホヤプロテインデータベース CIPRO は、ホヤの全遺伝子モデルデータを元に、あらゆるバイオインフォマティクス解析・予測を行い、その結果をグラフィカルに統合化したデータベースである。データベース開発は実験系研究者と共同で繰り返しの改善を通して行われ、RNA、タンパク質の発現を含め、ホヤのタンパク質に関する主要な実験情報を網羅し、分子から個体の発生までをカバーする広範な統合データベースとなっている(図は提示データの一部を抜粋)。このデータベース開発においては比較ゲノム解析や分子進化解析も目標項目であったが、近縁生物種のゲノム情報が不完全であったため、極めて典型的なモデル生物との比較ゲノム解析に限られ、分子進化解析も十分に進めることができなかった。ところが、プロジェクトの終わり頃から急速に多種生物ゲノム解読が進み、現在では8種のホヤについて解読が終了しており、また生物多様性の大きな割合を占める他系統の海産無脊椎動物についても、日本の研究者が次々にゲノム解読を進めている(図1)。しかし、プロジェクト規模の問題もあって得られたゲノムデータは一次的解析に留まることが多く、実験研究者にとって有用な下図のような情報を得るには、多くの労力が必要な状況で、特に進化的比較解析情報は不足している。

大量の国産ゲノムデータとデータ置き去りの問題

現在、図1に示す各後生動物門の生物ゲノムが、豊富な海洋資源に恵まれた我が国の生物学研究者により解読中か解読済みである。これらは発生学、系統学、多細胞生物成立過程の理解などに極めて重要な生物群である。しかし3大国際データバンク EBI は脊椎動物以外の非モデル生物については詳細解析を放棄しており、NCBI も大同小異である。我が国の DDBJ は個別生物種の解析を行っていない。その結果、多くの生物種でゲノムデータ産生までは進むものの、宝の山とも言えるゲノム・トランスクリプトームデータは活用が難しい状態のまま、多くの部分が置き去りにされてしまう傾向にある。

2. 研究の目的

原索動物のホヤは脊椎動物進化の鍵となる生物である。我々が過去にバイオインフォマティクス推進事業の支援で開発したホヤプロテインデータベース CIPRO は機能性が高く評価され S 評価を受けた。本申請ではこれをさらに推し進め、系統進化・発生学の比較解析基盤として多種生物統合データベースの開発を行う。このため、本提案では CIPRO DB のバイオインフォマティクス解析処理をパイプライン化すると共に多種間のゲノム比較解析パイプライン確立、分子系

統解析、遺伝子族解析による種間共通基幹遺伝子・遺伝子重複・種固有遺伝子の同定と進化解析を行う。日本で特に盛んな海洋生物の研究により急速にゲノム解読が進む多くの生物種の孤立ゲノム情報を統合化し、同一基盤に載せ進化発生学研究を加速する。

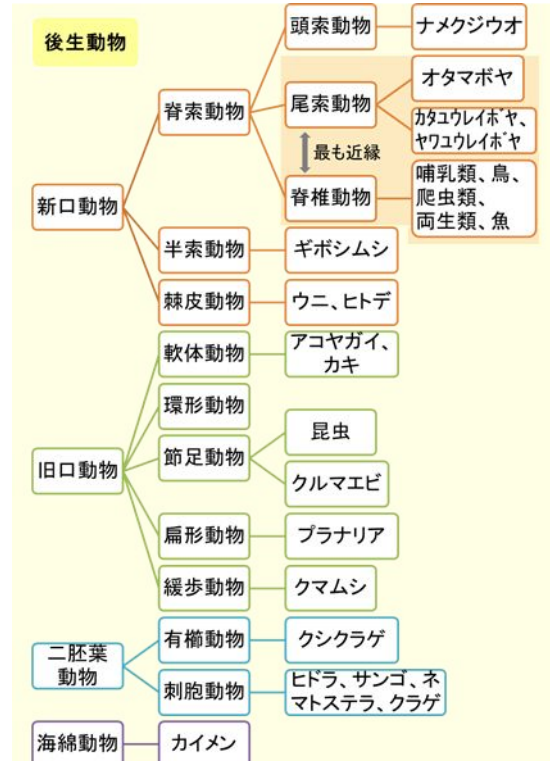


図1 ゲノムが解読されつつある生物と系統

3. 研究の方法

過去に開発したホヤプロテインデータベース CIPRO を基盤とし、この開発スキームを他生物種でのデータベース構築へ適用しやすくするため、関連の固有データベースを含むデータ処理パイプライン統合化を行う。特に遺伝子の命名法と遺伝子オントロジーは種固有のことが多く、特に非主要モデル生物以外の比較解析の妨げとなっているのでその障壁を軽減する。比較解析に不可欠な遺伝子系統関係、すなわち遺伝子ファミリーのオーソロジーとパラロジー、メンバー数あらゆる関係について進化的視点から統一的解析方法の確立を行う。現在、図1の海産生物について急ピッチでゲノム解読が進められているので、その成果を最大活用することを目指す。これらは個々の生物研究にとっても重要で、さらに比較解析によって大きな相乗効果が期待できるため、最終的には汎用的な生物情報解析システムとして全データとシステムツールの公開を目指す。

過去に開発したホヤプロテインデータベース CIPRO を基盤に、共同研究者の開発したゲノム・トランスクリプトーム情報データベース Ghost、発生・解剖学データベース FABA、3次元胚画像データベース 3DPL、化合物応

答データベース ACBD、細胞間相互作用・仮想胚データベース Aniseed のデータを統合化したシステムを開発する。CIPRO 開発での反省点を踏まえ、各種ツールを完全にモジュール化して独立性を高め、開発・維持・機能更新を容易にしつつ、解析ツール間の汎用的入出力データ形式定義を行って、連係化を容易な仕組みとする。統合化基盤として、実験情報と情報解析結果を統合化し、簡潔かつグラフィカルに表現できる CIPRO のエンジンを中心に用いる。実験情報として、ゲノム情報、クローン情報、時間的・空間的遺伝子発現情報(in situ hybridization, RNAseq, EST, Microarray など)、タンパク質発現情報(二次元電気泳動情報、質量分析)、発生学情報、タンパク質修飾情報、文献情報を統合化する。解析情報としては、ゲノム配列情報を基軸とし、機能アノテーション、オントロジー、比較ゲノム情報、グラフィカルな細胞内局在、予測高次構造、膜局在性、修飾予測、同時期・細胞種で発現する遺伝子情報、分子系統樹、ヒト相同遺伝子を介した疾患関連情報を自動生成する仕組みをパイプラインとして整える。これを活用し、ゲノム解読済みのホヤ類 8 系統を近縁種間解析のために用い、ホヤ以外の遠縁系統の各種生物系統と並列的に解析する。各配列にはハッシュタグを付与し、解析効率化を図るとともに、遺伝子の命名混乱軽減を図る。

PHYLOBASE

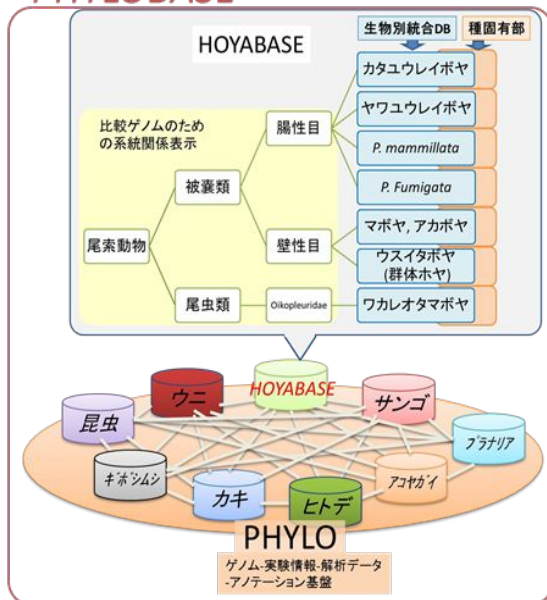


図2 Phylobase 構築の構想

DNA、タンパク質、発現、発生情報等のデータ統合を行い、データ間の多様なアクセス方法を提供・検証と同時に、必要なオントロジーの整備を行う。またこれを軸に、ホヤ類各種と無脊椎動物各種の既存データ統合を行いつつ、進化解析に必要なフレームワークを構築し、受け皿としての仕組みを整備する。これにより新たな知識発見を目指し、生物の進化・発生の共通基盤と多様性の理解を推

進する強力な知識基盤となることを目指す。種による遺伝子名等の違いを加味した命名法整備を行い、また解析基準の整備、必要な解析ツール開発、可視化ツール開発およびそのモジュール化、パイプライン化を行う。これを利用して、進化発生学的な研究を加速させる。

4. 研究成果

多種生物にまたがる比較オミックス統合データベースの開発を目指し、過去にバイオインフォマティクス推進事業の支援を得て構築したホヤ統合プロテインデータベース CIPRO のデータ解析スキームを飛躍的に発展させるため、公共データバンクにて公開されている 314 種の動物ゲノムデータをベースとして系統準備を行い、ゲノム決定された種を含む全動物門を横断する形にプロテオームデータの整備を進めた(表1)。生物ゲノムに含まれる遺伝子セットをゲノムスペクトルとし、真核生物について門、綱、目、科、属の各系統分類において、共通に存在する遺伝子群を整理した(1)。これに先立ち、生物種名から系統・一般名(英語・日本語)・形態を把握できるようにするための辞書を整備、それぞれが持つ遺伝子統計量を一覧化した(2)。また、同一配列を持つ遺伝子群のリストを整備した(3)。全生物系統に共通して存在する遺伝子は 1,908 で、ほぼ全てが基本生命維持機能のコア遺伝子群であった(表1)。系統固有の遺伝子は系統分類群ごとに同定された。また、植物には見られず、動物系統のみに共通して存在する 1,299 の遺伝子群を特定した。多くはミトコンドリアに関係する遺伝子であった。このほか脊椎動物系統のみ不在の遺伝子など、生物進化過程を解明する重要な手がかりが得られたと考えている。これらの結果について論文投稿準備中である。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔学会発表〕(計1件)

遠藤俊徳、小柳量、新里宙也、佐藤矩行「生物種間比較プロテオミクスデータベース構築と遺伝子スペクトル比較解析」日本遺伝学会第 87 回大会(仙台)2015

〔その他〕

ホームページ等

(1) Inter-Phyla Genome Profiler プロテオームスペクトル比較データベース
<http://phylomix.ibio.jp/web/>

(2) Phylomix Genome Master: ゲノム決定済

生物名称・学名・系統・遺伝子統計情報
DB

http://phylomix.ibio.jp/web/genome_aster/tab/2

(3) Phylomix Identical Seqs: 同一アミノ配列を持つタンパク質のデータベース

http://phylomix.ibio.jp/web/spectra/level/identical_seq

6. 研究組織

(1) 研究代表者

遠藤俊徳 (Toshinori Endo)

北海道大学・大学院情報科学研究科・教授

研究者番号: 00323692

表 1 生物の「門」レベルの系統を越えて共通に存在する遺伝子のリスト (抜粋)

Phyla	Gene Function
Ch He Ec Pr	1,2-dihydroxy-3-keto-5-methylthiopentene dioxygenase
-- Ar -- Br	1,4-alpha-glucan-branching enzyme
Mo -- Cn --	1-phosphatidylinositol 3-phosphate 5-kinase
Po -- -- St	10 kda heat shock protein, mitochondrial
	116 kda u5 small nuclear ribonucleoprotein component
	15-hydroxyprostaglandin dehydrogenase [nad(+)]
1908	2-amino-3-ketobutyrate coenzyme a ligase, mitochondrial
	2-aminoethanethiol dioxygenase
	2-methoxy-6-polyprenyl-1,4-benzoquinol methylase, mitochondrial
	2-oxoglutarate dehydrogenase, mitochondrial
	26s protease regulatory subunit 10b
	26s protease regulatory subunit 6b
	26s protease regulatory subunit 7
	26s protease regulatory subunit 8
	26s proteasome non-atpase regulatory subunit 1
	26s proteasome non-atpase regulatory subunit 10
	26s proteasome non-atpase regulatory subunit 12
	26s proteasome non-atpase regulatory subunit 13
	26s proteasome non-atpase regulatory subunit 14
	26s proteasome non-atpase regulatory subunit 2
	26s proteasome non-atpase regulatory subunit 3
	26s proteasome non-atpase regulatory subunit 4
	26s proteasome non-atpase regulatory subunit 5
	26s proteasome non-atpase regulatory subunit 6
	26s proteasome non-atpase regulatory subunit 7
	26s proteasome non-atpase regulatory subunit 8
	26s proteasome non-atpase regulatory subunit 9
	28 kda heat- and acid-stable phosphoprotein
	28s ribosomal protein s16, mitochondrial
	28s ribosomal protein s29, mitochondrial
	28s ribosomal protein s33, mitochondrial
	28s ribosomal protein s5, mitochondrial
	28s ribosomal protein s9, mitochondrial
	3'-5' exoribonuclease 1
	3-hydroxy-3-methylglutaryl-coenzyme a reductase
	3-hydroxybutyrate dehydrogenase type 2
	3-hydroxyisobutyrate dehydrogenase, mitochondrial
	3-hydroxyisobutyryl-coa hydrolase, mitochondrial
	3-ketodihydrospingosine reductase
	3-oxoacyl-[acyl-carrier-protein] synthase, mitochondrial
	3-phosphoinositide-dependent protein kinase 1
	39s ribosomal protein l11, mitochondrial
	39s ribosomal protein l19, mitochondrial
	39s ribosomal protein l22, mitochondrial
	39s ribosomal protein l24, mitochondrial
	39s ribosomal protein l28, mitochondrial
	39s ribosomal protein l45, mitochondrial
	39s ribosomal protein l46, mitochondrial
	39s ribosomal protein l47, mitochondrial
	4-coumarate--coa ligase 1
	4-hydroxybenzoate polyprenyl transferase, mitochondrial
	4-hydroxyphenylpyruvate dioxygenase
	40s ribosomal protein s10
	40s ribosomal protein s11
	40s ribosomal protein s12
	40s ribosomal protein s13
	40s ribosomal protein s15
	40s ribosomal protein s16
	40s ribosomal protein s17
	40s ribosomal protein s18

Phyla	Gene Function
	40s ribosomal protein s19
	40s ribosomal protein s2
	40s ribosomal protein s20
	40s ribosomal protein s21
	40s ribosomal protein s23
	40s ribosomal protein s24
	40s ribosomal protein s25
	40s ribosomal protein s26
	40s ribosomal protein s3a
	40s ribosomal protein s4
	40s ribosomal protein s6
	40s ribosomal protein s7
	40s ribosomal protein s8
	40s ribosomal protein s9
	40s ribosomal protein sa
	5' -3' exoribonuclease 2
	5-oxoprolinase
	52 kda repressor of the inhibitor of the protein kinase
	6-phosphogluconate dehydrogenase, decarboxylating
	60 kda heat shock protein, mitochondrial
	60s acidic ribosomal protein p0
	60s acidic ribosomal protein p1
	60s acidic ribosomal protein p2
	60s ribosomal export protein nmd3
	60s ribosomal protein l10
	60s ribosomal protein l11
	60s ribosomal protein l12
	60s ribosomal protein l13
	60s ribosomal protein l13a
	60s ribosomal protein l14
	60s ribosomal protein l15
	60s ribosomal protein l17
	60s ribosomal protein l18
	60s ribosomal protein l18a
	60s ribosomal protein l19
	60s ribosomal protein l22
	60s ribosomal protein l23a
	60s ribosomal protein l26
	60s ribosomal protein l27
	60s ribosomal protein l27a
	60s ribosomal protein l28
	60s ribosomal protein l3
	60s ribosomal protein l30
	60s ribosomal protein l31
	60s ribosomal protein l32
	60s ribosomal protein l35
	60s ribosomal protein l5
	60s ribosomal protein l6
	60s ribosomal protein l7a
	60s ribosomal protein l8
	60s ribosomal protein l9
	60s ribosome subunit biogenesis protein nip7
	7,8-dihydro-8-oxoguanine triphosphatase
	85/88 kda calcium-independent phospholipase a2
	a-kinase anchor protein 17a
	abc transporter g
	abl interactor 2
	abnormal spindle-like microcephaly-associated protein
	acetyl-coa acetyl transferase, mitochondrial
	acetyl-coenzyme a transporter 1
	actin cytoskeleton-regulatory complex protein pan1
	actin-binding protein ipp
	actin-related protein 2/3 complex subunit 1a
	actin-related protein 2/3 complex subunit 3
	actin-related protein 2/3 complex subunit 4
	actin-related protein 3
	actin-related protein 5
	actin-related protein 6
	activating signal cointegrator 1
	activating signal cointegrator 1 complex subunit 1
	activating signal cointegrator 1 complex subunit 2
	activating signal cointegrator 1 complex subunit 3
	activator of 90 kda heat shock protein atpase
	acyl-coa dehydrogenase

Phyla	Gene Function
	<p>acyl-coa-binding domain-containing protein 6 acyl-coa-binding protein acyl-coenzyme a thioesterase 13 acylamino-acid-releasing enzyme acylpyruvase fahd1, mitochondrial adenylate kinase isoenzyme 6 adenylosuccinate lyase adenylyl transferase and sulfurtransferase mocs3 adipocyte plasma membrane-associated protein adp-ribosylation factor 1 adp-ribosylation factor-like protein 2 adp-ribosylation factor-like protein 5b adp-ribosylation factor-related protein 1 adrenodoxin-like protein, mitochondrial af-9 afg3-like protein 2 alanine--trna ligase, cytoplasmic alcohol dehydrogenase class-3 aldose reductase alkylated dna repair protein alk b alpha-1,3-mannosyl-glycoprotein 2-beta-n-acetylglucosaminyl transferase alpha-1,3/1,6-mannosyl transferase alg2 alpha-ketoglutarate-dependent dioxygenase alk b alpha-methylacyl-coa racemase alpha-soluble nsf attachment protein alpha/beta hydrolase domain-containing protein 13 aminoacyl trna synthase complex-interacting multifunctional protein 1 aminomethyl transferase, mitochondrial amn1 amp deaminase 2 an1-type zinc finger protein 4 anamorsin anaphase-promoting complex subunit 1 anaphase-promoting complex subunit 10 anaphase-promoting complex subunit 2 anaphase-promoting complex subunit 4 anaphase-promoting complex subunit 7 angio-associated migratory cell protein angiogenic factor with g patch and fha domains 1 ankyrin repeat and sam domain-containing protein 3 ankyrin repeat and sam domain-containing protein 6 ankyrin repeat and zinc finger domain-containing protein 1 ankyrin repeat domain-containing protein 17 ankyrin repeat domain-containing protein 29 ankyrin repeat domain-containing protein 39 ankyrin repeat domain-containing protein 50 ankyrin repeat domain-containing protein 54 ankyrin repeat, ph and sec7 domain containing protein secg ankyrin repeat, sam and basic leucine zipper domain-containing protein 1 ankyrin-1 ankyrin-2 ankyrin-3 annexin a7 ap-1 complex subunit gamma-1 ap-1 complex subunit sigma-2 ap-2 complex subunit mu ap-2 complex subunit sigma ap-3 complex subunit delta-1 ap-5 complex subunit beta-1 apobec1 complementation factor apoptosis inhibitor 5 apoptotic chromatin condensation inducer in the nucleus arginine and glutamate-rich protein 1 arginine n-methyl transferase 7 arginine--trna ligase, cytoplasmic arginine-glutamic acid dipeptide repeats protein arginine/serine-rich protein pnisr arginyl-trna--protein transferase 1 argonaute-2 armadillo repeat-containing protein 3 armadillo repeat-containing protein 4 armadillo repeat-containing protein 7 armadillo repeat-containing protein 8 (以下略)</p>