

論文 / 著書情報
Article / Book Information

題目(和文)	オニヒトデ属(Acanthaster spp.)の比較ゲノム解析
Title(English)	
著者(和文)	湯浅英知
Author(English)	Hideaki Yuasa
出典(和文)	学位:博士(理学), 学位授与機関:東京工業大学, 報告番号:甲第12143号, 授与年月日:2021年12月31日, 学位の種類:課程博士, 審査員:伊藤 武彦,本郷 裕一,立花 和則,二階堂 雅人,山田 拓司
Citation(English)	Degree:Doctor (Science), Conferring organization: Tokyo Institute of Technology, Report number:甲第12143号, Conferred date:2021/12/31, Degree Type:Course doctor, Examiner:,,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

令和3年度 博士論文

オニヒトデ属(*Acanthaster* spp.)の
比較ゲノム解析

指導教員 伊藤 武彦 教授

東京工業大学
生命理工学院生命工学系
生命工学コース

湯浅 英知

目次

第 1 章 序論	1
1.1. 背景.....	1
1.1.1. サング礁の生物多様性	1
1.1.2. オニヒトデについて.....	2
1.2. 目的.....	6
1.3. 本論文の構成	6
第 2 章 サンプル情報	9
2.1. サンプル収集	9
2.2. DNA・RNA 抽出.....	12
2.3. シーケンスデータ	12
第 3 章 ゲノムアセンブル	16
3.1. 方法.....	16
3.1.1. ミトコンドリアゲノム	16
3.1.2. 核ゲノム.....	16
3.2. 結果.....	20
3.2.1. ミトコンドリアゲノム	20
3.2.2. 核ゲノム.....	21
第 4 章 ゲノムアノテーション	24
4.1. 方法.....	24
4.1.1. 構造アノテーション	24
4.1.1.1. ホモロジーベース.....	25
4.1.1.2. RNA-seq ベース	26
4.1.1.3. <i>Ab initio</i> ベース.....	27
4.1.1.4. 予測結果の統合	28
4.1.2. 機能アノテーション	29
4.2. 結果.....	30
4.2.1. 構造アノテーション	30
4.2.2. 機能アノテーション	33

第5章 種間ゲノム比較	34
5.1. 諸言.....	34
5.2. 方法.....	34
5.2.1. ゲノム配列類似度算出.....	34
5.2.1.1. ミトコンドリアゲノム.....	34
5.2.1.2. 核ゲノム.....	35
5.2.2. 系統解析.....	36
5.2.2.1. ミトコンドリアゲノム.....	36
5.2.2.2. 核ゲノム.....	37
5.2.3. 17-mer 頻度分析.....	38
5.3. 結果・考察.....	38
5.3.1. ゲノム配列類似度算出.....	38
5.3.2. 系統解析.....	41
5.3.3. 17-mer 頻度分析.....	44
5.4. まとめ.....	47
第6章 オニヒトデの種分化要因推定	48
6.1 諸言.....	48
6.2. 方法.....	48
6.2.1. 集団史推定.....	48
6.2.2. 遺伝構造解析.....	51
6.2.3. 種間分化領域探索.....	51
6.2.4. 種間遺伝子比較.....	52
6.2.4.1. オーソログ解析.....	52
6.2.4.2. 正の自然選択圧検出.....	52
6.2.4.3. GO/KEGG エンリッチメント解析.....	53
6.3. 結果・考察.....	54
6.3.1. 集団史推定.....	54
6.3.2. 遺伝構造解析.....	58
6.3.3. 種間分化領域探索.....	60
6.3.4. 種間遺伝子比較.....	61
6.4. まとめ.....	74

第7章 結論.....	75
参考文献.....	78
データ登録	89
謝辞	90

第 1 章 序論

1.1. 背景

1.1.1. サンゴ礁の生物多様性

サンゴ礁海域は全海洋表面積の内わずか0.2%程度のごく限られた海域である。しかしながら、サンゴ礁海域には記載されている全海洋生物種の約 3 分の 1 が生息していると見積もられており、サンゴ礁海域の生物多様性は非常に高い[1]。生物多様性の高さからサンゴ礁域は「海の熱帯雨林」と称されるが、サンゴ礁海域の生物多様性の高さは、そこに生息する生物の多くが種分化を経験してきたことを示唆している。サンゴ礁海域の中でもインド洋と太平洋において生物多様性が高く[2]、その要因についてこれまで議論がなされてきた[3]。インドー太平洋に生息するサンゴ礁生物の近縁種間の種分化要因としては、氷河期の海面の低下と海域に干出した大陸棚などによる物理的な生息域の分断および交流の制限[4]や、複雑な生息環境で生じた自然選択などが提唱されている[5]。しかしながら、これらは遺伝子の部分領域など限られた DNA マーカーの結果に基づいた仮説であり、具体的な仮説の検証は行われていないため種分化要因は明らかではない。昨今ではハイスループット DNA シーケンサーによってゲノムスケールでの解析のハードルは低くなりつつあり、ゲノム情報を用いることで、例えば過去の有効集団サイズの変遷（以降集団史と記述）を推定する Pairwise Sequentially Markovian Coalescent（以降 PSMC と記述）モデル[6]のような解析が行えるようになり（詳しくは第 6 章の集団史推定を参照）、現在ではより踏み込んだ仮説の検証が可能である。しかしながら、ゲノム情報を用いてインドー太平洋のサンゴ礁生物近縁種間の種分化要因を検証した研究は未だにない。ゲノムを用いた仮説検証が進まない原因としては、インドー太平洋スケールでのサンプリングの困難さや、サンゴ礁生物の多くが非モデル生物であるためゲノム配列など解析基盤となる情報が整っていないなどの問題が考えられる。

1.1.2. オニヒトデについて

サンゴ礁生物の1つであるオニヒトデ (*Acanthaster spp.*) はインドー太平洋のサンゴ礁海域に生息するヒトデ類である。成熟個体の大きさは直径約25~40cm とヒトデ類の中でも比較的大型で、一般的なアオヒトデなどの様な5本腕を持つ星型のヒトデとは異なり、大体8~21本の腕を持っていることや、体表面に溶血性タンパク質毒を有した棘が多数存在するといった形態的にユニークな特徴を持つ。それ以外に、オニヒトデにおいて最も注目されている特徴は彼らがサンゴ礁形成に寄与する造礁サンゴを好んで捕食する点である(図1-1A)。近年、サンゴ捕食者であるオニヒトデが慢性的に大量発生を繰り返すことで急速かつ広範囲にサンゴ礁の荒廃が起こるため、サンゴ礁保全の観点から問題視されている(図1-1B)。このようにオニヒトデはインド洋と太平洋の両方のサンゴ礁生態系において深い関わりのある生物種であり、サンゴ礁生物の中で最も研究されている生物種と言われるほどオニヒトデ研究はこれまで盛んに行われてきたが、オニヒトデに関する知見は未だ不足していないため大量発生の効果的な管理には至っていないことが指摘されている[7]。

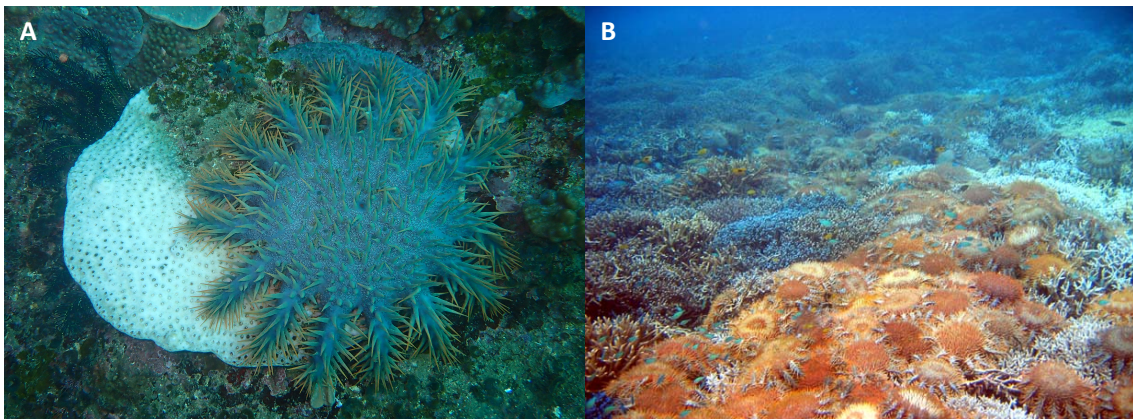


図 1-1：サンゴ食害生物であるオニヒトデ

A サンゴを捕食するオニヒトデ (宮崎県の大島にて美ら海水族館の奥祐太郎博士が撮影)、B 大量発生するオニヒトデの様子 (沖縄県の宮古島にて梶原健次氏が撮影)

オニヒトデ属 (*Acanthaster*) の系統関係は図 1-2 に示した通りである。オニヒトデ属の下には大きく分けてオニヒトデ (学名については後述) とアカオニヒトデ (*Acanthaster brevispinus*) の 2 つのグループが存在する。前者はこれまで説明してきた造礁サンゴ (ハードコーラルとも言う) を好んで捕食し大量発生するサンゴ食害生物であるが、後者はソフトコーラルを好んで捕食し大量発生の報告はない非サンゴ食害生物であるという違いが存在する [8]。本稿ではオニヒトデ属の中でもオニヒトデのグループに注目して研究を行った。オニヒトデのグループは、過去に棘が鋭利ではないアメリカ大陸のカリフォルニア周辺の集団が *Acanthaster ellisii* (Gray, 1840) として別種とみなされていたが [9]、アロザイム解析により太平洋域の全てが 1 種であることが示されたため [10]、インド-太平洋全域にわたってオニヒトデ *Acanthaster planci* (Linnaeus, 1758) 1 種が生息すると認識されてきた。しかしながら、2008 年に Vogler らが行なったミトコンドリアの COI 遺伝子の部分配列を用いた集団遺伝解析から、太平洋で 1 つのクレードとインド洋で 3 つのクレードの計 4 つのクレードに明確に分かれることが明らかとなり (図 1-2)、インド-太平洋のオニヒトデはそれぞれ太平洋種 (図 1-3A)、北インド洋種 (図 1-3B)、南インド洋種、紅海種 (図 1-3C) の 4 種が存在することが示唆された [11]。紅海種のオニヒトデに関してはこれまでに名付けられた学名が存在しなかったものの、各地域のオニヒトデが過去に名付けられた学名から北インド洋種が *Acanthaster planci* (Linnaeus, 1758) であり、その他、太平洋種が *Acanthaster solaris* (Schreber, 1793)、南インド洋種が *Acanthaster mauritiensis* (Loriol, 1885) に該当すると考えられた [9]。これらはインド洋と太平洋のオニヒトデは 1 種という考えのもと進められてきたこれまでのオニヒトデ研究の問題点を示唆する発見であったものの、上述の分類はミトコンドリアの 1 遺伝子の部分配列のみでしか確認されていないという問題も存在する。なぜなら、ミトコンドリア DNA と核 DNA の違いは必ずしも一致する関係ではないため、いくつかの先行研究ではミトコンドリア DNA だけで分類を行うことの危険性が示されている (例: [12][13])。そのため、現在のオニヒトデの分類が複数の核遺伝子やゲノムのような多くの遺伝子情報を用いた

解析からも支持されるかを確認する必要がある。しかしながら、オニヒトデのミトコンドリアと核のゲノムはこれまでに太平洋種のみでしか報告されておらず [14][15]、2020年12月までに他3種のゲノムは発表されていない。

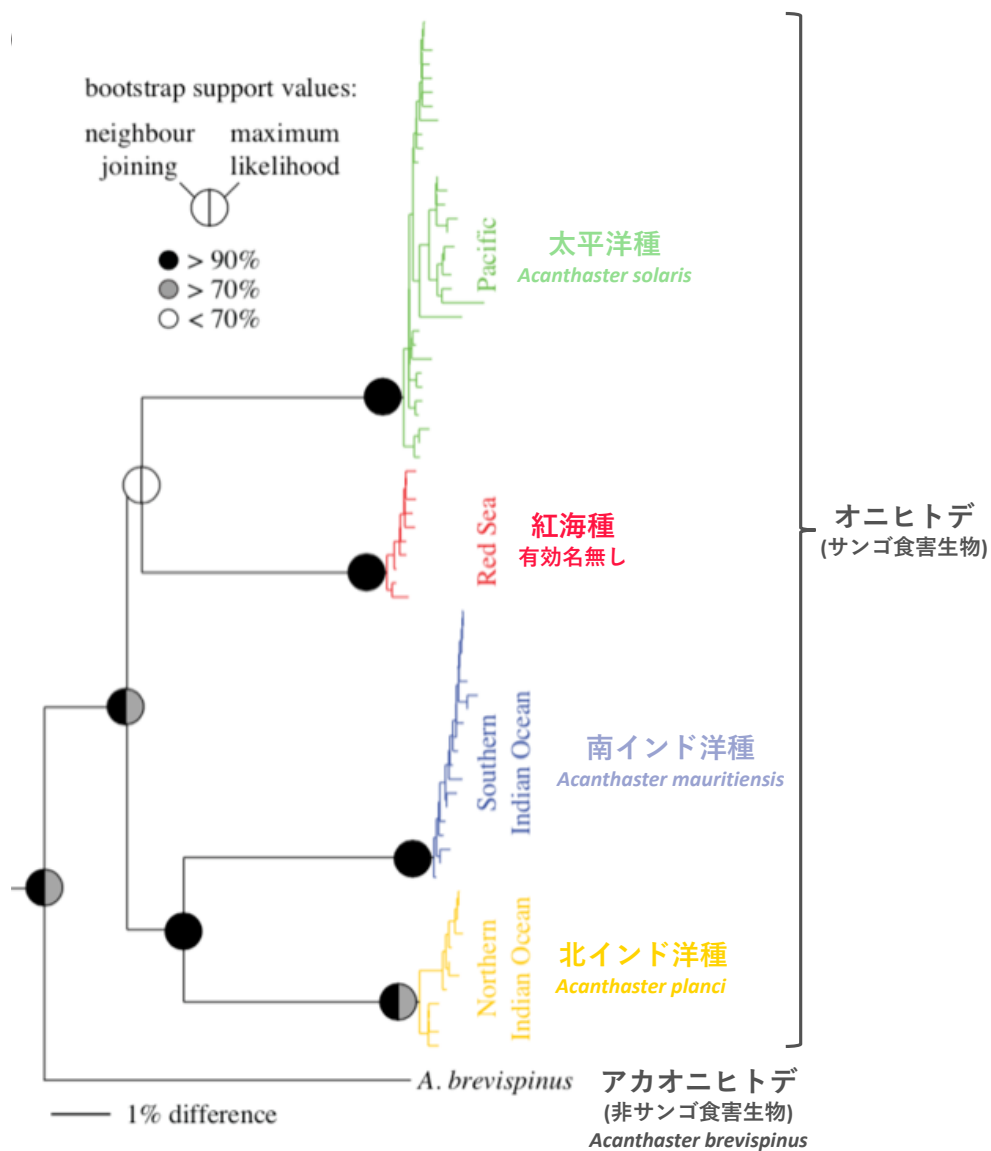


図 1-2：オニヒトデ属 (*Acanthaster*) の系統関係

Vogler et al. 2008[11]の図 1a の一部に著者が手を加えて作成した。学名は Haszprunar & Spies 2014[9]に従った。

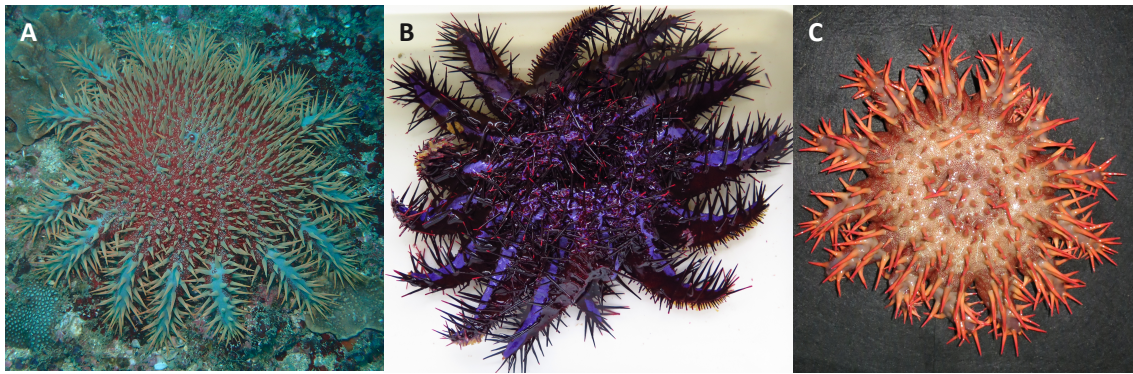


図 1-3：本研究で扱う 3 種のオニヒトデ

A 太平洋種（宮崎県の大島にて宮崎大学の志村晶史氏が撮影）、B 北インド洋種（タイのプーケットにて Phuket Marine Biological Center の Naline Thongtham 博士が撮影）、C 紅海種（イスラエルのエイラトにて University of Queensland の Gal Eyal 博士が撮影）

オニヒトデ種間の違いに関しては、北インド洋種は他種とは体色が異なることに加えて反口側の腕末梢部の棘が欠如していることが報告されており外見から他種との識別が容易であるが[7]、4 種全てを識別できる明確な形態の違いは未だ見つかっていない。形態以外では大量発生パターン、毒強度、行動において種間で違いが存在する可能性が指摘されているものの[11][16][17]、オニヒトデ種間の比較を行なった研究が少なく、3 種以上で比較を行った研究はこれまでに無いため知見は不足している。

種間の遺伝的な違いを提起した Vogler らの研究ではオニヒトデが多様化した時期を配列の相違度から 195 万年から 365 万年前と推定しており、この時期の氷河期による海水面の低下が集団の地理的な隔離を引き起こしたことで種間の異所的種分化を引き起こしたという仮説を提唱している。しかしながら、この仮説は先述のようにミトコンドリアの COI 遺伝子の部分配列から推定された分岐年代のみに基づいており、分岐年代以外に異所的種分化を示すデータなどは存在せず具体的な仮説検証はなされていないため、オニヒトデの種分化要因は未だ明らかにされていない。

1.2. 目的

本研究ではオニヒトデ属の中でもサンゴ食害生物であるオニヒトデを研究対象としている。オニヒトデはミトコンドリア遺伝子の部分配列を用いた系統解析から 4 種存在することが示されているが、核遺伝子やゲノムからは明らかにされておらず不確かさが残る。オニヒトデの分類を明確にすることは各種および種間でのオニヒトデ研究の活性化につながるため、サンゴ食害生物であるオニヒトデのさらなる理解と管理に役立つと考えられる。また、オニヒトデ種がどのように形成されたのか種分化要因は明らかにされていない。生物多様性の根源となる種分化が生じる過程や要因を解明することは生態系を長期にわたり保全・管理する上で重要であるため、オニヒトデの種分化要因の解明は生物多様性が高いことで知られているインドー太平洋のサンゴ礁生態系保全において重要な知見となる。

オニヒトデの種間にどれほどの差異があり、またどのように種分化したか明らかにするにはゲノム情報が最適であると考えられる。ゲノムを用いることで個体や集団間の差異の定量化が容易であることや、遺伝子情報を網羅的に捉えることができるため Incomplete lineage sorting など解釈を歪めるような遺伝子の影響を抑えた評価を行うことができるからである。また、系統推定や PSMC モデルを用いた集団史推定など、集団遺伝学の解析から過去の進化的な背景を捉えることができる。上記を踏まえて、本研究ではハイスループット DNA シーケンサーを用いたオニヒトデ種間の比較ゲノム解析により、現存するオニヒトデ種間のゲノムレベルでの違いとその種分化がどのようにして起きたのか解明することを目的とした。

1.3. 本論文の構成

本論文では、まず第 2 章にて使用サンプルとデータについて説明した後、以下に示す 4 章において本論文の主要部分を述べる。ここではその概要を示す。

第3章では、太平洋種、北インド洋種、紅海種のオニヒトデ3種のゲノム配列のアセンブルについて説明する。3種23個体のミトコンドリアの完全長ゲノム配列と3種3個体の核のドラフトゲノム配列の構築に成功した。北インド洋種と紅海種のミトコンドリアおよび核のゲノム配列が構築されたのは本論文が初めてである。また、太平洋種の核ゲノム配列に関しても、先行研究で報告されたものよりも高品質のドラフトゲノム配列を構築することに成功した。

第4章では、第3章で構築した3種3個体の核ゲノム配列の遺伝子構造予測および遺伝子機能の推定について説明する。本論文では高い精度の遺伝子情報を解析に用いるために、遺伝子構造予測には所属研究室において共同で新規開発した自動遺伝子構造予測パイプラインである LEMON を使用した。その結果、3種3個体の核ゲノム上に先行研究で報告されたものよりも高品質な遺伝子構造を予測することに成功した。

第5章では、ミトコンドリアの限られた情報のみに基づいて分類されていたオニヒトデ種間の違いをゲノムから確認を行った結果を示す。まず、オニヒトデ3種間でゲノム配列の類似度を算出したところ核とミトコンドリアでそれぞれ約92%と約98%であり、既知の別種間であるヒトーチンパンジー間に相当する類似度であることが示された。また、系統解析結果から、ミトコンドリアのみならず核のゲノムにおいてもオニヒトデ3種間に明確な違いが存在することが確認された。ただし、太平洋種内のハワイ個体はミトコンドリアから想定されるよりも大きな分化が核ゲノムに見られることがわかった。最後に、3種23個体のゲノムシーケンスリードを用いた17-mer 頻度分析の結果から、太平洋種とインド洋側2種（北インド洋種、紅海種）の間で重複配列領域に起因すると思われるゲノムサイズ差が存在することが示された。

第6章では、第5章で示された種間の違いがどのように生み出されたのか、その種分化要因をゲノムから検証した結果を示した。集団史推定、遺伝構造解析、種間分化領域探索の結果からオニヒトデ3種は1つの共通祖先集団から分化し、その分化は劇的な海水面の変化による地理的な隔離によって異所的に起こった可能性が示された。種間で分化したゲノム領域を探索したところゲノム全体で種間の分化が進んでいることが確認されたが、極端に分化した領域を見つけることができなかった。そこで、ゲノムスケールから遺伝子スケールに切り替えて種間比較を行ったところ、各種のゲノムから地域適応の痕跡や遺伝子

領域上のゲノム構造変異が見つかり、オニヒトデ種間の違いに繋がる可能性のある特徴を抽出することに成功した。

最後に第7章において本論文の総括として結論を述べた。

第2章 サンプル情報

2.1. サンプル収集

本研究では図 2-1 に示した 7 海域から太平洋種、北インド洋種、紅海種の 3 種のオニヒトデを合計 21 個体収集した。収集した 21 個体の情報は表 2-1 に記載した通りである。DNA 抽出に用いるサンプルはオニヒトデから管足を取り、イシサンゴ類の DNA 抽出に用いられる CHAOS 溶液 (4 M guanidine thiocyanate, 0.1% N-lauroyl sarcosine sodium, 10 mM Tris pH8, 0.1 M 2-mercaptoethanol) [18]に入れて冷蔵保存を行った (家庭用の冷蔵庫を使用)。宮崎の 1 個体 (Miyazaki01) は RNA 抽出のために、生殖腺、肝盲嚢、棘、胃袋、管足の組織を取り分け (図 2-2)、RNA Save 安定化保存液 (Biological Industries 社) に入れて-70 度で冷凍保存した。サンプル収集には以下の方々にご協力を頂いた。

宮崎・石垣：安田仁奈准教授 (宮崎大学)

ハワイ：Zac Forsman 博士 (University of Hawai'i at Mānoa)

タヒチ：Coralie Taquet 博士 (Collège Zéna M'Déré)

プーケット：Nalinee Thongtham 博士 (Phuket Marine Biological Center)

モルディブ：Davide Seveso 博士 (University of Milano-Bicocca)

Enrico Montalbetti 博士 (University of Milano-Bicocca)

イスラエル：Omri Bronstein 博士 (Tel Aviv University)

Gal Eyal 博士 (University of Queensland)

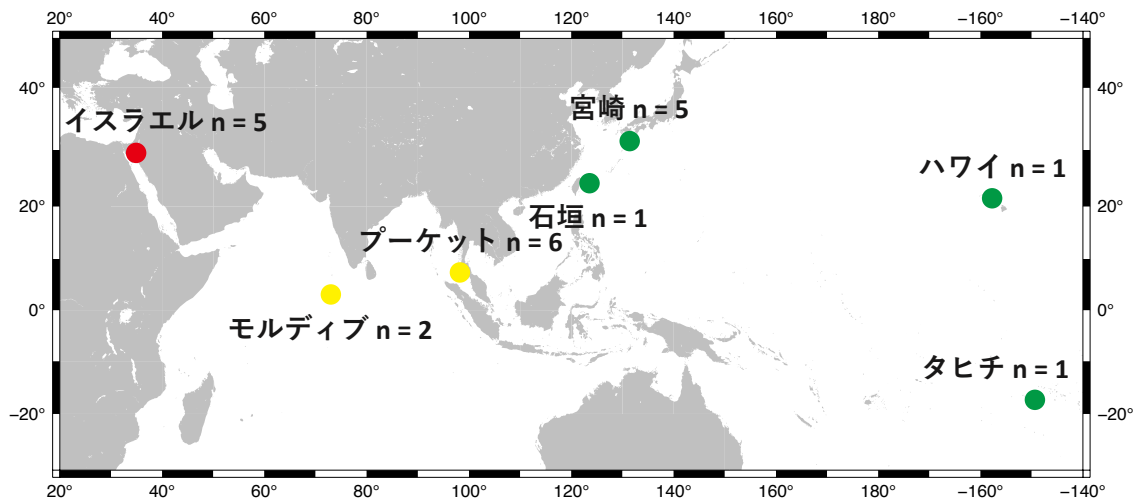


図 2-1：本研究においてサンプルを収集した地点

丸は収集地点を示す。丸の色は先行研究[9][11]で分類された種を示しており、緑は太平洋種、黄は北インド洋種、赤は紅海種を示す。また、n の値は収集したサンプリングの数を示す。地図は Generic Mapping Tools v.6.0.0[19]を用いて作成したものを使用した。

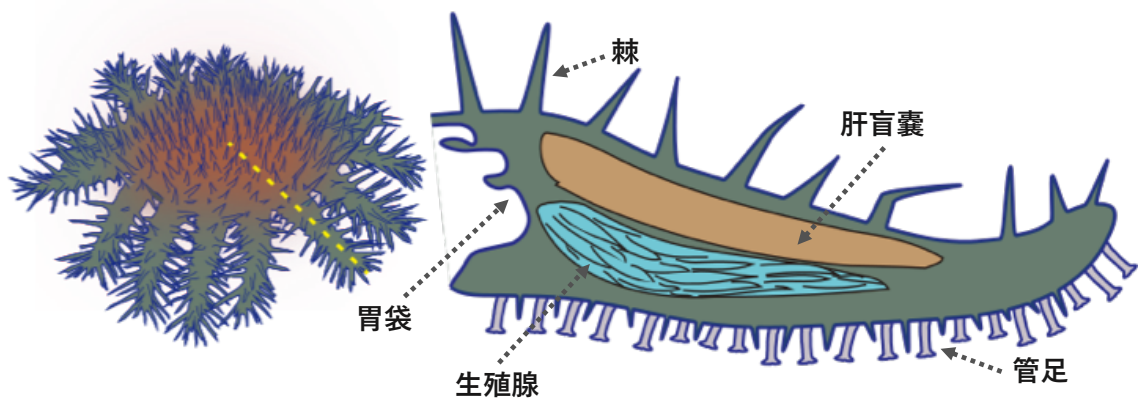


図 2-2：RNA-seq のために取り分けたオニヒトデ組織の模式図

イラストの素材は台湾中央研究院の和田直久博士から提供していただいた。

表 2-1：本研究で用いた 3 種 21+2 個体のオニヒトデのサンプルとゲノムシーケンスリード情報

PE は Paired-end、MP は Mate-pair、PacBio は Pacific Biosciences 社のロングリードを示す。

サンプルID	種名	サンプリング地点	収集年	Biosampleアクセッション番号	ライブラリー	合計シーケンス量	出典
Miyazaki01	太平洋種	宮崎	2014	SAMD00056692	PE, MP, PacBio	533×	本研究
Miyazaki02	太平洋種	宮崎	2018	SAMD00229327	PE	93×	本研究
Miyazaki03	太平洋種	宮崎	2018	SAMD00229328	PE	93×	本研究
Miyazaki04	太平洋種	宮崎	2018	SAMD00229329	PE	91×	本研究
Miyazaki05	太平洋種	宮崎	2018	SAMD00229330	PE	88×	本研究
Ishigaki	太平洋種	石垣	2016	SAMD00229331	PE	195×	本研究
Tahiti	太平洋種	タヒチ	2016	SAMD00229333	PE	124×	本研究
Hawaii	太平洋種	ハワイ	2014	SAMD00229332	PE	196×	本研究
Phuket01	北インド洋種	プーケット	2014	SAMD00229334	PE, MP	558×	本研究
Phuket02	北インド洋種	プーケット	2017	SAMD00229335	PE	41×	本研究
Phuket03	北インド洋種	プーケット	2017	SAMD00229336	PE	42×	本研究
Phuket04	北インド洋種	プーケット	2017	SAMD00229337	PE	41×	本研究
Phuket05	北インド洋種	プーケット	2017	SAMD00229338	PE	41×	本研究
Phuket06	北インド洋種	プーケット	2017	SAMD00229339	PE	41×	本研究
Maldives01	北インド洋種	モルディブ	2018	SAMD00229340	PE	194×	本研究
Maldives02	北インド洋種	モルディブ	2018	SAMD00229341	PE	34×	本研究
Israel01	紅海種	イスラエル	2017	SAMD00229342	PE	43×	本研究
Israel02	紅海種	イスラエル	2017	SAMD00229343	PE	42×	本研究
Israel03	紅海種	イスラエル	2017	SAMD00229344	PE	43×	本研究
Israel04	紅海種	イスラエル	2017	SAMD00229345	PE, MP	182×	本研究
Israel05	紅海種	イスラエル	2017	SAMD00229346	PE	44×	本研究
OKI	太平洋種	沖縄本島	2013	SAMD00054104	PE, MP	139×	Hall et al. 2017
GBR	太平洋種	オーストラリア	2013	SAMD00020546	PE, MP	152×	Hall et al. 2017

2.2. DNA・RNA 抽出

ゲノム DNA の抽出はサンプルを CHAOS 溶液で 1 週間以上冷蔵保存し、組織片が溶液中に充分溶けていることを確認してから行った。組織が溶けた CHAOS 溶液からフェノール・クロロホルム抽出法によってゲノム DNA を取り出し、Tris-EDTA (以降 TE と記述) に溶解させた。ゲノム DNA を溶解させた TE から不純物を取り除くために Blood & Cell Culture DNA Midi Kit (QIAGEN 社) を用いて精製を行った。ゲノム DNA の抽出は宮崎大学の安田仁奈准教授のご協力を頂いた。

続いて、転写産物の配列を取得して遺伝子構造の予測に用いるため、RNA の抽出を行った。最初に RNA Save 安定化保存液中で冷凍させた組織サンプルを取り出し、ホモジナイザーを使用して組織を破碎した。その後、PureLink RNA Mini Kit (Thermo Fisher Scientific 社) を使用して RNA を抽出した。抽出した RNA を安定的に保存し、またシーケンスを行うために SuperScript III First-Strand Synthesis System for RT-PCR (Thermo Fisher Scientific 社) を使用して RNA を逆転写して cDNA を作成した。RNA の抽出は宮崎大学の安田仁奈准教授に行っていた。

2.3. シーケンスデータ

本研究で収集した全 21 個体のゲノム DNA から Paired-end(以降 PE と記述) ライブラリーを TruSeq PCR-free DNA sample Prep kit (Illumina 社) を使用して作成した。PE ライブラリーの DNA 断片長 (インサートサイズ) は 300–650 bp、リード長は 150–250 とし、シーケンスは Illumina 社のシーケンサーを用いて行った。PE リードシーケンスに関するライブラリーとシーケンサーの詳細な情報は表 2-2 に示した。また、PE に加えて、太平洋種、北インド洋種、紅海種のゲノム配列を構築するために各種 1 個体 (Miyazaki01、Phuket01、Israel04) のゲノム DNA から Mate-pair(以降 MP と記述) ライブラリーを作成した。MP のライブラリー作成には Nextera Mate Pair Sample Prep kit (Illumina 社) を使

用し、アガロースゲルから 3 kbp–15 kbp までの DNA 断片を切り出して各サンプルのライブラリーを用意した。シーケンスは Illumina 社の HiSeq 2500 と HiSeq X Ten を用いて行った。MP リードシーケンスに関するライブラリーとシーケンサーの詳細な情報は表 2-2 に示した。さらに、太平洋種については Pacific Biosciences 社（以降 PacBio と記述）の Sequel を用いてロングリードのシーケンスも行った。各 21 個体のゲノムシーケンス情報は先行研究でシーケンスされた太平洋種 2 個体の情報と合わせて表 2-1 に記載している。HiSeq 2500 と Sequel のライブラリー調整とシーケンスは国立遺伝学研究所の豊田敦特任教授に行っていた。Israel04 で HiSeq X Ten でシーケンスを行ったライブラリーと Miyazaki02–05 のライブラリー調整は東京工業大学の流水利恵研究員に行って頂き、シーケンスは（株）マクロジェン・ジャパンに行って頂いた。Ishigaki、Tahiti、Maldives02 のサンプル調整とシーケンスは（株）マクロジェン・ジャパンに行って頂いた。Maldives01 のシーケンスは（株）マクロジェン・ジャパンに行って頂いた。ゲノムシーケンスに加えて、太平洋種 1 個体の 5 組織から RNA-seq を行った。cDNA から Paired-end (PE) ライブラリーを作成し、Illumina 社の HiSeq 2500 を用いてショートリードのシーケンスを行った（表 2-3）。RNA-seq は東京大学の鈴木穰博士に行って頂いた。最後に、Platanus_trim v.1.0.7 (http://platanus.bio.titech.ac.jp/pltanus_trim) を用いてシーケンスリードからアダプター配列と低クオリティ配列を取り除き、以降の解析で用いた。

表 2-2 : ゲノムシーケンスのライブラリーとシーケンサーの情報

PacBio は Pacific Biosciences 社のロングリードを示す。PacBio のシーケンス長はライブラリー全体の平均リード長を示す。

サンプルID	種	ライブラリー	シーケンサー	インサートサイズ (bp)	シーケンス長 (bp)	シーケンス量
Miyazaki01	太平洋種	Paired-end	Illumina HiSeq 2500	300	150	80×
Miyazaki01	太平洋種	Paired-end	Illumina HiSeq 2500	500	150	82×
Miyazaki01	太平洋種	Mate-pair	Illumina HiSeq 2500	3000	150	51×
Miyazaki01	太平洋種	Mate-pair	Illumina HiSeq 2500	5000	150	53×
Miyazaki01	太平洋種	Mate-pair	Illumina HiSeq 2500	8000	150	54×
Miyazaki01	太平洋種	Mate-pair	Illumina HiSeq 2500	10000	150	53×
Miyazaki01	太平洋種	Mate-pair	Illumina HiSeq 2500	12000	150	53×
Miyazaki01	太平洋種	Mate-pair	Illumina HiSeq 2500	15000	150	58×
Miyazaki01	太平洋種	PacBio	Sequel	---	14769	18×
Miyazaki01	太平洋種	PacBio	Sequel	---	14557	17×
Miyazaki01	太平洋種	PacBio	Sequel	---	14324	13×
Miyazaki02	太平洋種	Paired-end	HiSeq X Ten	350	151	91×
Miyazaki03	太平洋種	Paired-end	HiSeq X Ten	350	151	88×
Miyazaki04	太平洋種	Paired-end	HiSeq X Ten	350	151	93×
Miyazaki05	太平洋種	Paired-end	HiSeq X Ten	350	151	93×
Ishigaki	太平洋種	Paired-end	HiSeq X Ten	400	151	195×
Tahiti	太平洋種	Paired-end	HiSeq X Ten	400	151	124×
Hawaii	太平洋種	Paired-end	Illumina HiSeq 2500	400	150	196×
Phuket01	北インド洋種	Paired-end	Illumina HiSeq 2500	400	150	221×
Phuket01	北インド洋種	Mate-pair	Illumina HiSeq 2500	3000	150	56×
Phuket01	北インド洋種	Mate-pair	Illumina HiSeq 2500	5000	150	57×
Phuket01	北インド洋種	Mate-pair	Illumina HiSeq 2500	8000	150	55×
Phuket01	北インド洋種	Mate-pair	Illumina HiSeq 2500	10000	150	56×
Phuket01	北インド洋種	Mate-pair	Illumina HiSeq 2500	12000	150	57×
Phuket01	北インド洋種	Mate-pair	Illumina HiSeq 2500	15000	150	56×
Phuket02	北インド洋種	Paired-end	Illumina HiSeq 2500	600	250	41×
Phuket03	北インド洋種	Paired-end	Illumina HiSeq 2500	600	250	42×
Phuket04	北インド洋種	Paired-end	Illumina HiSeq 2500	650	250	41×
Phuket05	北インド洋種	Paired-end	Illumina HiSeq 2500	650	250	41×
Phuket06	北インド洋種	Paired-end	Illumina HiSeq 2500	600	250	41×
Maldives01	北インド洋種	Paired-end	HiSeq X Ten	400	151	194×
Maldives02	北インド洋種	Paired-end	Illumina NovaSeq 6000	400	151	34×
Israel01	紅海種	Paired-end	Illumina HiSeq 2500	600	250	43×
Israel02	紅海種	Paired-end	Illumina HiSeq 2500	600	250	42×
Israel03	紅海種	Paired-end	Illumina HiSeq 2500	600	250	43×
Israel04	紅海種	Paired-end	HiSeq X Ten	200	151	60×
Israel04	紅海種	Paired-end	Illumina HiSeq 2500	600	250	44×
Israel04	紅海種	Mate-pair	HiSeq X Ten	3000	151	19×
Israel04	紅海種	Mate-pair	HiSeq X Ten	6000	151	18×
Israel04	紅海種	Mate-pair	HiSeq X Ten	10000	151	23×
Israel04	紅海種	Mate-pair	HiSeq X Ten	15000	151	21×
Israel05	紅海種	Paired-end	Illumina HiSeq 2500	600	250	42×

表 2-3 : RNA-seq サンプル情報

サンプルID	組織名	ライブラリー	シーケンサー	シーケンス長 (bp)	シーケンス量	Biosample アクセッション番号
Miyazaki01	生殖腺	Paired-end	HiSeq 2500	100	8 ×	SAMD00231968
Miyazaki01	肝盲囊	Paired-end	HiSeq 2500	100	83 ×	SAMD00231971
Miyazaki01	棘	Paired-end	HiSeq 2500	100	5 ×	SAMD00231972
Miyazaki01	胃袋	Paired-end	HiSeq 2500	100	31 ×	SAMD00231973
Miyazaki01	管足	Paired-end	HiSeq 2500	100	13 ×	SAMD00231974

第3章 ゲノムアセンブル

3.1. 方法

3.1.1. ミトコンドリアゲノム

表 2-1 に示した 3 種 23 個体の PE リードを用いて、各サンプルのミトコンドリアゲノムのアセンブルを実施した。アセンブルには NovoPlasty v.3.7[20]を使用した。NovoPlasty は同種または同属レベルの近縁種の配列を参照してオルガネラのゲノム配列を構築するアセンブラーであるため、リファレンス配列として先行研究で報告された太平洋種のミトコンドリアゲノム配列（GenBank アクセッション番号：NC_007788.1）[14]を入力としてアセンブルを実施した。

3.1.2. 核ゲノム

太平洋種、北インド洋種、紅海種の核ゲノム配列のアセンブルは PE だけでなく MP のライブラリーも調整した Miyazaki01、Phuket01、Israel04 のそれぞれのサンプルから行われた。

まず、核ゲノム配列のアセンブルを行うサンプルの推定ゲノムサイズ、重複領域サイズ、ヘテロ接合度といった基礎的な情報を把握するために k -mer 頻度分析を実施した。最初に Jellyfish v.2.2.10[21]を用いて 17-mer の頻度分布を作成した後、GenomeScope v.2.0[22]を用いて推定ゲノムサイズ、重複領域サイズ、ヘテロ接合度の推定を行った。

続いて、核ゲノム配列のアセンブルを行った。方法の概要は図 3-1 に示した通りである。まず、太平洋種、北インド洋種、紅海種共通で PE リードと MP リードから Platanus v.1.2.5[23]を用いてアセンブルを実施し、scaffold 配列の構築を行った。scaffold 配列から 1 kbp 未満の配列は構築不十分として以降の解析からは除外した。今回の目的は核ゲノム配列を構築することにあるために、コンタミネーション由来と疑われる配列を取り除く必要がある。そのため、コンタミネーション由来の配列を除く作業を実施した。まず、実験段階などでのコンタミネーションが疑われるヒト、バクテリア、ウイルスのゲノム配列を NCBI データベースより取得した（2019 年 8 月にダウンロード）[24]。

次に、オニヒトデの体表面に共在することが報告されているスピロヘータ門のバクテリア[25]も混在している可能性が考えられるため、このバクテリアの染色体配列を NCBI データベースから取得した (Assembly アクセション番号: GCA_013426755.1)。最後に、核ゲノム以外にもミトコンドリア由来の配列が混ざっている可能性が考えられるため、3.1.1.で同じサンプルから構築したミトコンドリアゲノム配列を用意した。核ゲノム以外でコンタミネーションが疑われるこれらの配列をデータベースとして BLASTN[26]で相同性検索を行い、identity>90%かつクエリーカバー率>50%の配列をコンタミネーションとして Platanus の出力結果から除外した。太平洋種についてはさらに、PE と MP のショートリードのアセンブル結果に存在する配列未決定領域 (ギャップ) を可能な限り減らすために PacBio のロングリードのデータを使用した。まず、Canu v.1.7[27]を用いてロングリードのアセンブルを実施し、contig 配列を構築した。Pbalign v.0.3.1 (<https://www.pacb.com/support/software-downloads/>) を用いてロングリードを contig 配列に対してアライメントを行い、アライメント結果から GenomicConsensus package v.2.2.2 の Arrow (<https://github.com/PacificBiosciences/GenomicConsensus>) を用いて contig 配列のエラー修正であるポリッシュを実施した。ロングリードによるポリッシュ済みの contig 配列に対して Bowtie2 v.2.3.5.1[28]を用いてショートリードのアライメントを行い、アライメント結果から Pilon v.1.22[29]を用いてショートリードによる contig 配列のポリッシュを実施した。ショートリードによるポリッシュは contig 配列で修正されるサイト数が減少を続けるまで行い、合計 18 回行った。修正されるサイト数の推移を図 3-2 で示した。エラー修正済みの contig 配列を用いて、ショートリードで構築した scaffold 配列に存在する配列未決定領域の配列を可能な限り決定した。この作業は本研究室の梶谷嶺助教が開発したツールを用いて実施し、その方法の詳細は以下の通りである。また、方法の概要は図 3-3 で示した。

1. 配列未決定領域両端の周辺配列 L bp を切り出した。 L には複数の値が適用される (後述のステップ 5)。
2. Minimap2 v.2.17 [30] (オプション `-c -k 19`) を用いて、切り出した周辺配列をエラー修正済みの Canu の contig 配列にアラインした。
3. アライメント結果から identity<90%、クエリーカバー率<25%、ベストスコアのアライメントが複数ある結果を除外した。

4. 以下 4 条件を満たす周辺配列ペアのアライメント結果を基に、Canu の contig 配列（アライメントペア間の領域）で scaffold 中の配列未決定領域を置換した。
 - ①周辺配列のペアが同じ contig 配列にアライメントされている。
 - ②ペア間の距離が 50 kbp 未満である。
 - ③ペアが同じ方向でアライメントされている。
 - ④対応する Canu の contig の領域中に配列未決定領域が存在しない。
5. 手順 1~4 のステップを合計 8 回繰り返し実施した。この時、ステップ 1 における周辺配列長 L には次の複数の値を順に適用した: 500、1k、5k、10k、20k、40k、80k、160k。配列未決定領域周辺の状況（リピート配列やアセンブリエラーの有無など）によって適切な L の値は異なるが、複数の値を適用することで多くのケースを解決できると期待される。
6. 手順 1~5 のステップを合計 2 回繰り返し実施した。配列未決定領域が短い間隔で密集するケースなどは、このように繰り返しを行うことで多く解決できることが期待される。

最後にコア遺伝子からゲノムの完成度を評価する BUSCO v.4.0.6[31]を用いて (Metazoa Odb10 データセットを使用)、構築した太平洋種、北インド洋種、紅海種の核ゲノム配列の完成度を評価した。

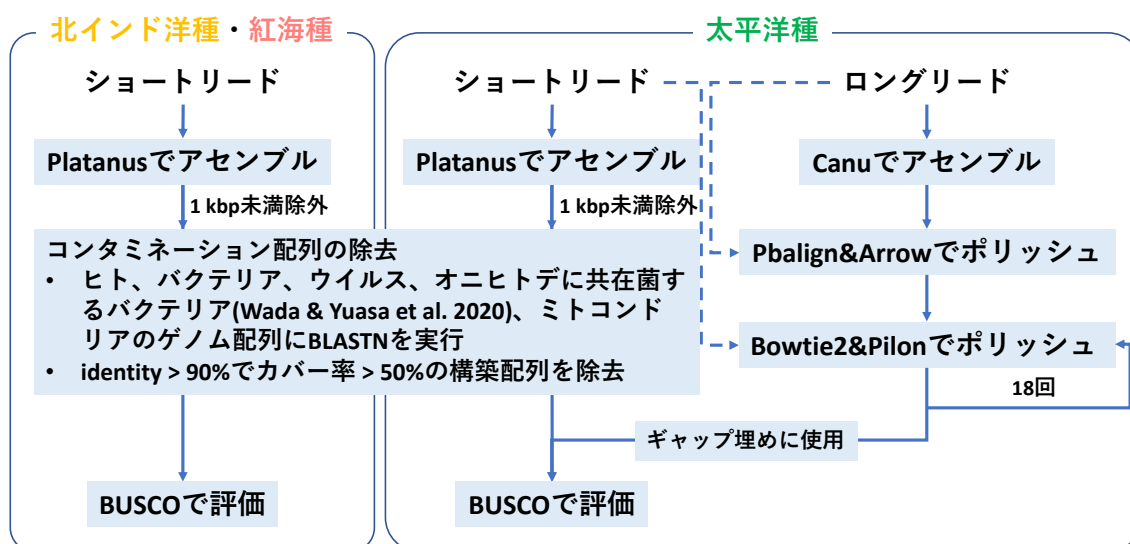


図 3-1：核ゲノムアセンブル方法の概要

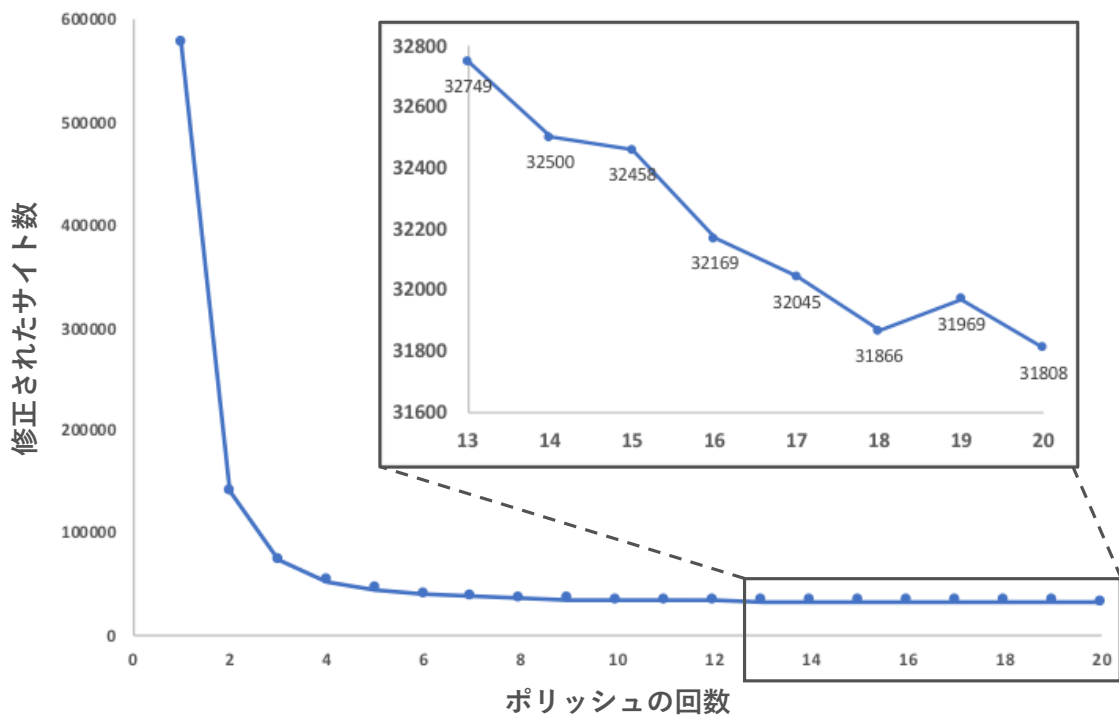


図 3-2：ショートリードによるポリッシュの回数と修正されるサイト数の推移

19回目で修正されるサイト数が増加したことからポリッシュによってエラー修正の限界に達したと考えられ、18回ポリッシュを行った contig 配列を以降の解析で使用した。

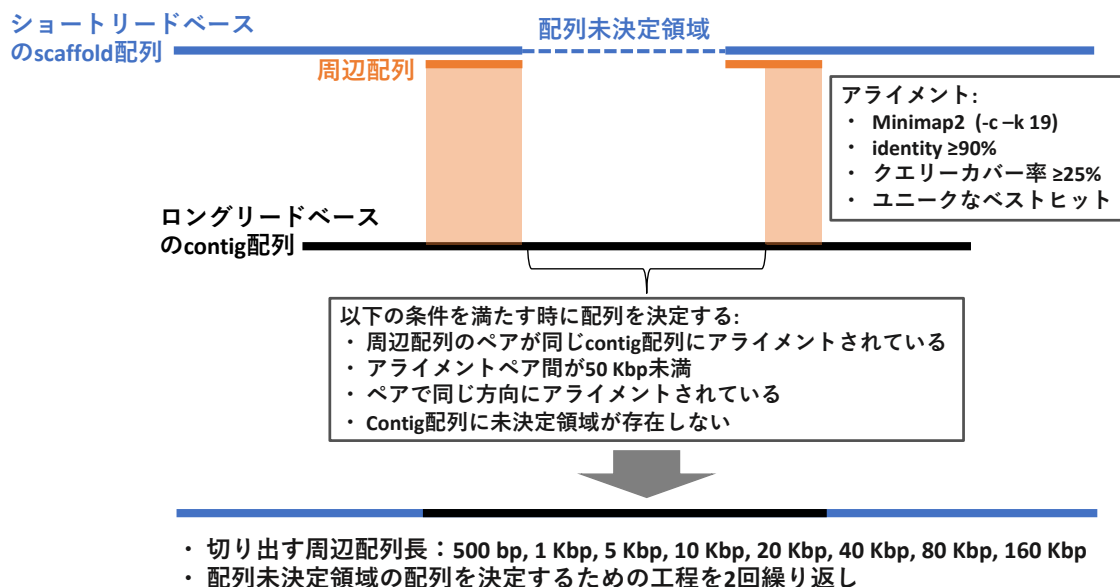


図 3-3：ロングリードでアセンブルした contig 配列を用いてショートリードでアセンブルした scaffold 配列の配列未決定領域の配列決定を行った方法の概要図

3.2. 結果

3.2.1. ミトコンドリアゲノム

本研究で構築した23個体のミトコンドリアゲノム配列とリファレンスに用いたミトコンドリアゲノム配列の情報を表3-1に示した。23個体の全てにおいて完全長のミトコンドリアゲノムを構築することに成功した。リファレンスに用いたミトコンドリアゲノム配列も含め配列長の範囲は16,225–16,239 bpであった。

表3-1：本研究において構築した3種23個体のオニヒトデのミトコンドリアゲノム配列と
先行研究で構築されたオニヒトデのミトコンドリアゲノム配列の情報

サンプルID	種名	配列数 (bp)	配列長 (bp)	Nサイト数 (bp)	アクセッション番号	出典
Miyazaki01	太平洋種	1	16,235	0	LC566225	本研究
Miyazaki02	太平洋種	1	16,239	0	LC566226	本研究
Miyazaki03	太平洋種	1	16,233	0	LC566227	本研究
Miyazaki04	太平洋種	1	16,235	0	LC566228	本研究
Miyazaki05	太平洋種	1	16,234	0	LC566229	本研究
Ishigaki	太平洋種	1	16,237	0	LC566217	本研究
Tahiti	太平洋種	1	16,236	0	LC566236	本研究
Hawaii	太平洋種	1	16,233	0	LC566216	本研究
Phuket01	北インド洋種	1	16,231	0	LC566230	本研究
Phuket02	北インド洋種	1	16,232	0	LC566231	本研究
Phuket03	北インド洋種	1	16,233	0	LC566232	本研究
Phuket04	北インド洋種	1	16,233	0	LC566233	本研究
Phuket05	北インド洋種	1	16,233	0	LC566234	本研究
Phuket06	北インド洋種	1	16,232	0	LC566235	本研究
Maldives01	北インド洋種	1	16,232	0	LC566223	本研究
Maldives02	北インド洋種	1	16,232	0	LC566224	本研究
Israel01	紅海種	1	16,225	0	LC566218	本研究
Israel02	紅海種	1	16,226	0	LC566219	本研究
Israel03	紅海種	1	16,228	0	LC566220	本研究
Israel04	紅海種	1	16,227	0	LC566221	本研究
Israel05	紅海種	1	16,226	0	LC566222	本研究
OKI	太平洋種	1	16,235	0	LC596400	本研究
GBR	太平洋種	1	16,235	0	LC596401	本研究
Okinawa	太平洋種	1	16,234	0	NC007788	Yasuda et al. 2006

3.2.2. 核ゲノム

構築した太平洋種、北インド洋種、紅海種の核ゲノム情報を表 3-2 に示した。本研究で構築した 3 個体に加えて、Hall et al. 2017[15]で報告された太平洋種 2 個体の核ゲノムの情報も比較対象として記載している。まず、本研究で構築した 3 個体と先行研究で構築した 2 個体の結果を比較すると、以下 3 点の理由から本研究で構築した核ゲノム配列の方が先行研究のものよりも高い完成度であることが示唆された。

- ① scaffold N50 の値が本研究で構築した 3 個体の全てにおいて先行研究の 2 個体よりも大きい。scaffold N50 は scaffold を長いものから並べて全 scaffold 長（構築ゲノムサイズ）の半分に達した時の scaffold 長を示しており、ゲノムアセンブル結果の配列の長さの指標として用いられる。つまり、先行研究よりも本研究で構築した核ゲノム配列の方が、連続性が大きいことを示している。
- ② 配列未決定サイト合計数が本研究で構築した 3 個体の全てにおいて先行研究の 2 個体よりも少なく、配列決定サイト合計数は k -mer による推定ゲノムサイズにより近い値となっている。
- ③ BUSCO による評価値である Complete の割合が先行研究よりも高い。この値はゲノムから完全長で予測できたコア遺伝子の数から算出されており、構築したゲノムの完成度を間接的に示す値である。つまり、先行研究よりも本研究で構築した核ゲノム配列の方が高い完成度であることを示している。

太平洋種に関しては本研究と先行研究で構築した核ゲノム配列が存在するが、本研究の方が先行研究よりも構築ゲノムサイズが推定ゲノムサイズにより近い値を示しており、先行研究では構築できなかった配列を本研究で構築した太平洋種の核ゲノムでは構築できていることを示唆した。そのため、以降の解析において太平洋種の核ゲノム配列は先行研究のものではなく、本研究で新たに構築したものをを用いることとした。

次に本研究で構築された太平洋種、北インド洋種、紅海種を比較すると、構築ゲノムサイズがそれぞれ 420 Mbp、400 Mbp、386 Mbp であり、太平洋種>北インド洋種>紅海種で構築ゲノムサイズに差が確認された。また、17-mer 頻度分析による推定ゲノムサイズにおいても 454 Mbp、419 Mbp、378 Mbp と構築

ゲノムサイズと同様の傾向を示した。推定重複（リピート配列）領域サイズは 183 Mbp、147 Mbp、114 Mbp であり、推定重複領域サイズの差と推定ゲノムサイズの差が近い値を示した。このことからゲノムサイズの差は重複領域に起因していることが示唆された。

表 3-2：本研究の 3 種 3 個体のオニヒトデと先行研究の太平洋種 2 個体のオニヒトデの核ゲノム配列情報

先行研究の構築ゲノム情報とゲノムサイズ推定の各値は元論文[15]に記載されていた値であり、BUSCO による構築ゲノムの完成度評価は本研究で算出した値である。

種		太平洋種	北インド洋種	紅海種	太平洋種	太平洋種
サンプルID		Miyazaki01	Phuket01	Israel04	OKI	GBR
出典		本研究	本研究	本研究	Hall et al. 2017	Hall et al. 2017
構築ゲノム情報 (Scaffold長 \geq 1 Kbp)	ゲノムサイズ (bp)	420,016,408	400,438,766	385,630,944	383,843,944	383,525,304
	Scaffold数	4,526	3,642	3,552	1,765	3,274
	Scaffold N50 (bp)	2,941,429	3,897,460	1,662,177	1,521,119	916,880
	Scaffold最大長 (bp)	10,098,247	11,313,316	8,528,531	11,795,022	3,939,554
	配列未決定サイト合計数 (bp)	4,264,832	7,197,764	7,359,702	9,873,513	10,051,337
	GC含量 (%)	41.3	41.4	41.4	41.3	41.3
ゲノムサイズ推定 (17-mer頻度分析)	ゲノムサイズ (bp)	454,083,373	418,678,427	378,256,433	421,092,086	441,650,244
	重複領域サイズ (bp)	182,858,213	147,396,297	114,220,755	NA	NA
	ヘテロ接合度 (%)	0.97	0.72	1.24	NA	NA
BUSCOによる	Complete (%)	98.1	98.0	97.7	97.2	97.5
構築ゲノムの 完成度評価 (Metazoa 10 odb)	Complete and single-copy (%)	97.6	97.6	97.4	96.9	97.3
	Complete and duplicated (%)	0.5	0.4	0.3	0.3	0.2
	Fragmented (%)	1.4	0.7	0.9	1	0.9
	Missing (%)	0.5	1.3	1.4	1.8	1.6

第4章 ゲノムアノテーション

4.1. 方法

4.1.1. 構造アノテーション

遺伝子スケールで近縁種間の違いを捉えるためには、ゲノム中の遺伝子構造予測の精度は可能な限り高める必要がある。一般的に遺伝子構造アノテーションでは、①近縁種のタンパク質配列を用いたホモロジーベース、②発現遺伝子の情報を用いたトランスクリプトームベース、③学習させた遺伝子情報を元に算出した統計量から予測を行う *Ab initio* ベースといった各手法によって予測された遺伝子構造を統合することで最終的な予測結果を決定している。正確な遺伝子構造予測結果を得るには手作業で統合を行う戦略も存在するが、真核生物のゲノムにはおよそ 2 万もの遺伝子が存在し、手作業による統合は非常に時間がかかる。そのため、可能な限り精度の高い予測結果の統合を自動で行うことにした。各予測結果の統合を自動で行うツールは既に存在し、代表的なツールとして EVIDENCEModeler[32] や MAKER[33] が存在する。オニヒトデ太平洋種のゲノムが報告された論文[15]においても EVIDENCEModeler が採用されている。しかしながら、このようなツールは *Ab initio* ベースの予測結果に重きをおいた統合を行なっているところに問題がある。*Ab initio* ベースは統計量に基づいて遺伝子構造を予測することができるため、ホモロジーベースとトランスクリプトームベースではカバーすることができない遺伝子構造の予測が可能であるという点に強みがあるが、その一方で予測に偽陽性が多いという短所もある。そのため、*Ab initio* を中心とした統合を行う既存ツールでは本研究において重要視している精度の面で問題があると考えられた。そこで、本研究室に在籍していた篠田恭寛氏、高橋和希氏、小林史弥氏、中村優太氏らと共同で開発を行った、*Ab initio* に寄らず複数の予測結果を相補的に用いることで精度の高い遺伝子構造予測を実現した自動構造アノテーションパイプラインである LEMON を本研究の構造アノテーションに採用した。LEMON のアノテーションパイプラインの概要を図 4-1 で示した。

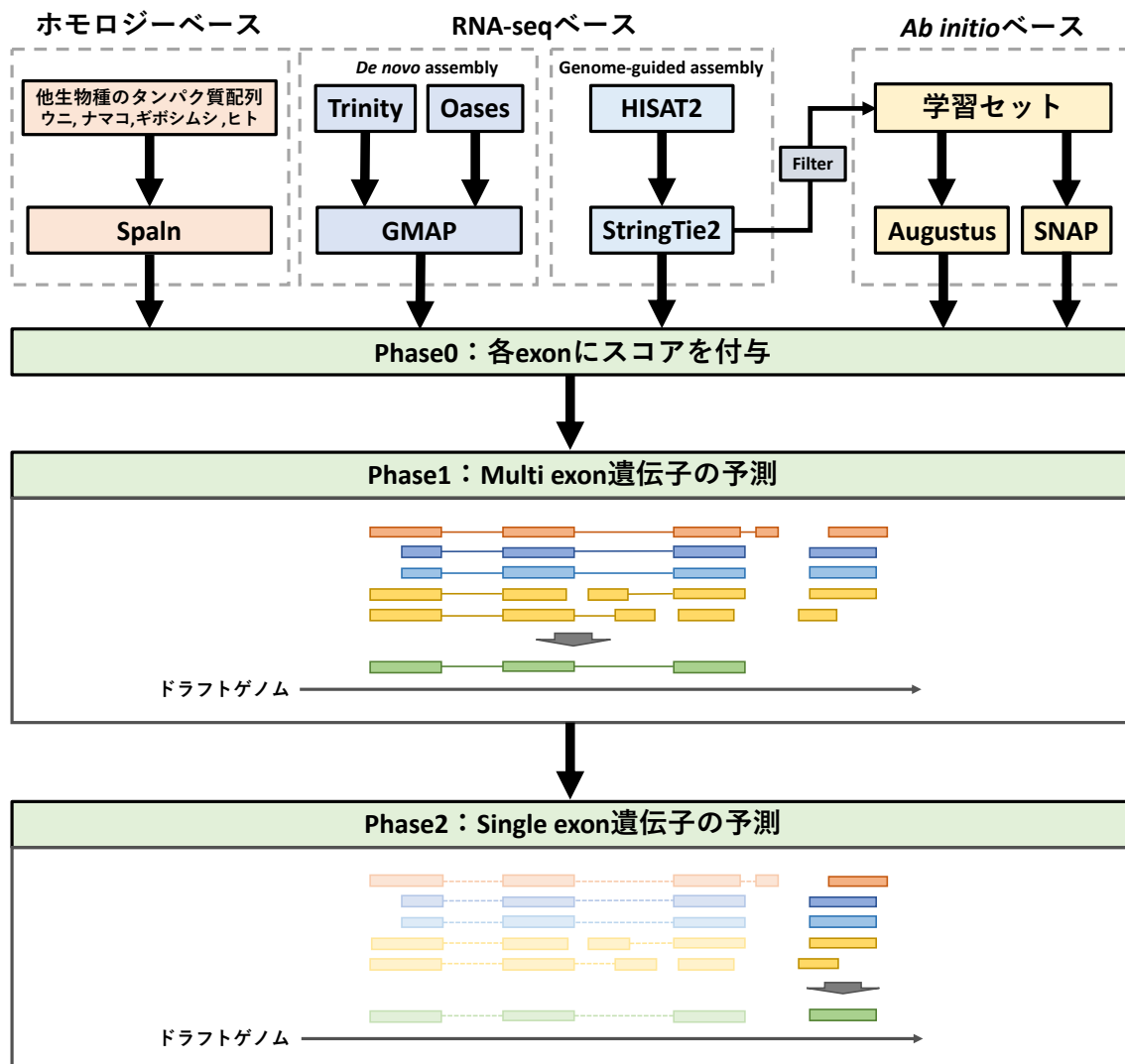


図 4-1 : LEMON のアノテーションパイプラインの概要

所属研究室の元メンバーである高橋和希氏と小林史弥氏の修士論文の図を参考に著者が作成した。

4.1.1.1. ホモロジーベース

ホモロジーベースでは近縁種など他生物種のタンパク質配列を元にゲノムの遺伝子構造の予測を行う。他生物種のタンパク質配列は以下 4 種のものを使用し、タンパク質配列は各データベースより取得した。

1. *Strongylocentrotus purpuratus*: アメリカムラサキウニ
 - オニヒトデと同じ棘皮動物門

- Echinobase より 2018 年 7 月にダウンロード[34]
2. *Apostichopus japonicus* : マナマコ
 - オニヒトデと同じ棘皮動物門
 - v.1.0 を以下 URL のデータベースよりダウンロード[35]
http://www.genedatabase.cn/aja_genome_20161129.html
 3. *Saccoglossus kowalevskii* : ギボシムシ
 - 系統学的に棘皮動物門に近いとされる半索動物門
 - Skow_1.1 を NCBI データベースよりダウンロード[36]
 4. *Homo sapiens* : ヒト
 - アノテーション用のデータが充実しており左右相称動物の中で最も遺伝子構造の正確性が高いと期待される。
 - GRCh38.p12 を NCBI データベースよりダウンロード

Spaln v.2.3.3[37]を用いてゲノム配列に対して上記4種のタンパク質配列のスプライシングを考慮したマッピングを行った。マッピングした結果の中で、フレームシフトなどで配列の途中で終止コドンを含む結果は以降の解析で除外した。

4.1.1.2. RNA-seq ベース

RNA-seq ベースでは転写産物のリードデータを元にゲノムの遺伝子構造の予測を行う。RNA-seq ベースの予測にはリードをゲノム配列にマッピングし、マッピング結果から遺伝子構造を予測する genome-guided assembly と、リードデータをアセンブルすることで得られる遺伝子配列をゲノム配列にマッピングすることで遺伝子構造を予測する *de novo* assembly の2つのアプローチに分けることができる。本パイプラインでは可能な限り高感度の予測を行うために、両方の手法から遺伝子構造の予測を行った。なお、本解析では太平洋種から取得した太平洋種由来の RNA-seq リードを太平洋種、北インド洋種、紅海種の3種に共通して使用した。

■ Genome-guided assembly

1. HISAT2 v.2.1.0[38]を用いて RNA-seq リードをゲノム配列にマッピングした。
2. StringTie2 v.2.0[39]を用いてマッピング結果から遺伝子構造の予測を行った。

3. 予測された遺伝子の向きは各フレームで open reading frame (以降 ORF と記述) を予測し、single exon 遺伝子の場合は最長の ORF の向きを採用し、multi exon gene の場合は StringTie が報告する向きを採用した。
 4. TransDecoder v.5.5.0
(<https://github.com/TransDecoder/TransDecoder>) を用いて ORF 予測を行い、ORF 長 ≥ 90 bp の配列を採用した。
- *De novo* assembly
1. Trinity v.2.8.4[40] と Oases v.0.2.09[41] を用いて RNA-seq リードのアセンブルを行った。
 2. アセンブル結果から配列長 < 300 bp の結果を除外した。
 3. CD-HIT v.4.8.1[42] を用いて identity が 100% である冗長な配列を取り除いた。
 4. GMAP v.2018-07-04[43] を用いてこれまでに得られた配列をゲノム配列にスプライシングを考慮したマッピングを行った。
 5. GMAP の結果より、identity $< 95\%$ の結果を除外した。
 6. TransDecoder を用いて ORF 予測を行い、ORF 長 ≥ 90 bp の配列を採用した。

4.1.1.3. *Ab initio* ベース

Ab initio ベースでは遺伝子情報を学習させて算出した統計量からゲノム配列上の遺伝子構造の予測を行う。*Ab initio* ベースでの予測ではトランスポゾンなどの転写因子由来のリピート配列を遺伝子構造と認識して誤った遺伝子構造を予測する可能性がある。このような誤りを避けるために、まずはゲノム配列上のリピート配列を特定した。次に学習セットを作成し、最後に *Ab initio* による遺伝子構造の予測を行った。

- ゲノム配列上のリピート構造の特定
1. RepeatModeler v.2.0.1[44] 太平洋種、北インド洋種、紅海種の核ゲノム配列から 3 種固有のリピート配列を特定した (オプション: -LTRStruct)。
 2. RepeatMasker v.4.1.0[45] を用いて、手順 1 で作成した 3 種固有リピート配列、RepBase-20181026[46]、Dfam v.3.1[47] の情報を参照し、ゲノム配列上のリピート配列の位置を特定し、N でリピート配列をマスクした。こ

の時、-nolow オプションを用いて転写因子由来ではないと考えられる simple repeat と low complexity repeat のリピート配列はマスク対象から除外した。

■ 学習セットの作成

1. 4.2.1.2 の RNA-seq ベースの Genome-guided assembly の結果から single exon 遺伝子を除外した。学習セットには可能な限り高精度の遺伝子情報を用いる必要があるが、single exon 遺伝子と判定される結果には断片化されたものや偽遺伝子由来のものが含まれる危険性があり、各統計量も特殊である可能性があるためである。
2. RepeatMasker の結果から、exon の一部がリピート領域 (simple repeat と low complexity repeat を除く) と重複する配列は、転写因子由来の誤りの可能性を考慮し、学習セットの候補から除外した。
3. 1,000 配列をランダムに抽出し、学習セットとして採用した。

■ *Ab initio* による遺伝子構造の予測

1. Augustus v.3.3.2[48] と SNAP v.2006-07-28[49]を用いて、それぞれ学習セットから遺伝子構造予測のための統計量を算出した。
2. Augustus と SNAP を用いて、リピート領域 (simple repeat と low complexity repeat を除く) をマスクしたゲノム配列に対して遺伝子構造の予測を行った。

4.1.1.4. 予測結果の統合

本研究で用いた LEMON では、ホモロジーベース、RNA-seq ベース、*Ab initio* ベースでの遺伝子構造予測結果を、*Ab initio* に重点をおいた従来のツールの様な統合を行うのではなく、各手法の結果より exon 単位で精度の高いものを選択し (後述のスコアに基づく)、相補的に使用して統合することで精度を高めている。LEMON における予測結果の統合は phase0、phase1、phase2 の3段階に分かれている。まず、phase0 では各予測結果から算出されたアライメントカバー率、identity、各予測結果が出力する独自指標の情報を基に各 exon にスコアを付与する。次に phase1 では各手法から予測された遺伝子構造のグループ化を行い、phase0 で算出された exon スコアからスコア合計が最大となるような exon-intron のパスを動的計画法により探索して遺伝子構造の予測を行う。そして、予

測された遺伝子のスコア分布から閾値を設定し、その閾値を上回る遺伝子から single exon 遺伝子を取り除いた multi exon 遺伝子が phase1 の最終結果として出力される。最後に、phase2 では single exon 遺伝子の予測を行っている。ゲノム上で遺伝子領域が重なる single exon 遺伝子をグループ化し、その中で最も exon スコアの高い single exon 遺伝子を抽出する。そして抽出された single exon 遺伝子が phase1 で予測された multi exon 遺伝子と遺伝子領域が重複しない場合は phase1 の結果に追加し、最終結果として出力される。

本研究では LEMON v.2.3 の phase0.sh、phase1.sh、phase2.sh を用いて各統合段階を実行し、予測結果の統合を行った。その際、phase1 に対応する phase1.sh を実行する際に閾値の手動設定は行わずにデフォルトで実行した。また、phase2 に対応する phase2.sh では、アメリカムラサキウニの最短 CDS 長が 45 bp であったことを参考にして LEMON が出力する CDS 長の下限を 45 bp に設定して実行した。

最後に BUSCO を用いて Metazoa Odb10 データセットから各種で予測された遺伝子構造アノテーションの完成度を評価した。

4.1.2. 機能アノテーション

構造アノテーションによって遺伝子配列情報を得ることができたが、これらの機能についての情報は無い。以降の解析において生物学的な意味を含めた解析や結果の考察を行うためには、その遺伝子がどのような機能を持っている可能性があるのか情報を付与する必要がある。そこで予測された各遺伝子に機能予測情報を付与する機能アノテーションを実施した。

まず始めに、DIAMOND v.0.9.14.115[50]を用いて予測された遺伝子のアミノ酸配列を UniProt[51]の SwissProt と TrEMBL、NCBI の nr[24]の3つのデータベースに相同性検索を行い、 E 値が 10^{-5} 以下の結果を機能情報として付与した。この時、nr には先行研究で予測された太平洋種の予測結果が含まれており、機能予測の結果が偏ってしまうため、太平洋種にヒットした検索結果は除外した。また、使用したデータベースは SwissProt と TrEMBL は 2020 年 9 月、nr は 2020 年 8 月にダウンロードを行い取得したものである。次に、InterProScan v.5.39-77.0[52]を用いてモチーフ構造の特定と Gene Ontology(以降 GO と記述)情報の付与を行った(オプション:-goterms)。最後に、BlastKOALA v.2.2[53]

と KofamKOALA v.2020-01-06[54]を用いて各遺伝子に Kyoto Encyclopedia of Genes and Genomes (以降 KEGG と記述) データベースの K number の情報を付与した。K number を付ける際は、まず BlastKOALA を実行して KEGG GENES データベースから各遺伝子に K number を振り分けた。BlastKOALA を実行する際は taxonomic group には *Acanthaster* (Taxonomy ID : 133433)、database には genus_eukaryotes を選択した。続いて、BlastKOALA で K number が割り当てられなかった遺伝子を対象に KofamKOALA を用いて Kofam データベースから各遺伝子に K number の振り分けを行った。付与された K number の情報はさらに、KEGG ORTHOLOGY (以降 KO と記述) データベース[55]の情報を元に KEGG PATH と KEGG BRITE の情報を割り当てた。KO データベースの情報は 2020 年 9 月に以下の URL からダウンロードしたものを用いた。

https://www.genome.jp/kegg-bin/get_htext?ko00001

4.2. 結果

4.2.1. 構造アノテーション

構造アノテーションの結果は表 4-1 に示した通りである。比較対象として先行研究で報告されている太平洋種 2 個体の情報も記載した。まず、本研究と先行研究の結果を比較すると、先行研究で予測された遺伝子数は本研究で予測された遺伝子数よりも多かった。先行研究では本研究よりも平均 CDS 長が短い、平均 exon 数が少ない、single exon 遺伝子数が多いといった傾向が確認できた。このことから、本研究よりも先行研究の方が短い遺伝子構造を多数予測していると考えられ、このことが遺伝子数の差を生んでいる可能性がある。次に、BUSCO による予測結果の評価に注目すると先行研究と比べて本研究の予測結果の方が Complete の割合が高く、Missing の割合が低いことが確認された。このことから、本研究 3 種で予測された遺伝子セットは先行研究の遺伝子よりも高い完成度であることが示された。

本研究の 3 種を比較すると、遺伝子数に差が見られた。本研究では太平洋種の RNA-seq しか用意することができず、このことが北インド洋種と紅海種の遺伝子予測に影響を及ぼした可能性が考えられる。しかしながら、BUSCO の評価結果を確認すると Complete の割合がどれも 98%以上であった。このことから、

太平洋種だけでなく北インド洋種と紅海種においても高い完成度で遺伝子セットを構築できたと考えられ、以降の解析でを使用することとした。

表 4-1：本研究の3種3個体のオニヒトデと先行研究で報告された太平洋種2個体の遺伝子構造予測の結果

EVM は EVIDENCEModeler を示す。

種		太平洋種	北インド洋種	紅海種	太平洋種	太平洋種
サンプルID		Miyazaki01	Phuket01	Israel04	OKI	GBR
出典		本研究	本研究	本研究	Hall et al 2017	Hall et al. 2017
アノテーションツール		LEMON	LEMON	LEMON	EVM	EVM
予測遺伝子数		19,397	18,045	18,879	24,323	24,747
予測遺伝子詳細	平均CDS長 (bp)	1,635.2	1,676.8	1,594.6	1,388.1	1,375.3
	全exon数	164,717	158,103	160,000	167,730	167,501
	平均exon数	8.5	8.8	8.5	6.9	6.8
	single exon遺伝子数	1,608	1,395	1,339	5,117	5,264
	合計exon長 (bp)	31,717,709	30,258,364	30,103,960	33,761,920	34,035,711
	平均exon長 (bp)	192.6	191.4	188.1	201.3	203.2
	intron数	145,320	140,058	141,121	143,407	142,754
	合計intron長 (bp)	190,824,129	185,658,358	187,215,836	166,363,337	165,697,219
	平均intron長 (bp)	1,313.1	1,325.6	1,326.6	1,160.1	1,160.7
BUSCOによる評価	Complete (%)	98.1	98.2	98.0	91.4	91.2
	(Metazoa 10 odb) single-copy (%)	97.6	97.8	97.8	90.9	90.8
	duplicated (%)	0.5	0.4	0.2	0.5	0.4
	Fragmented (%)	1.3	1.3	1.3	4.8	5
	Missing (%)	0.6	0.5	0.7	3.8	3.8

4.2.2. 機能アノテーション

機能アノテーションの結果は表 4-2 に示した通りである。どの種においても、少なくとも 51.8%以上の遺伝子に 1 つ以上の機能的な情報を付与することができた。

表 4-2 : 本研究のオニヒトデ 3 種 (3 個体) の核ゲノム上に予測された遺伝子に対する機能アノテーションの結果

GO は Gene Ontology を示す。NCBI-nr の値は太平洋種のヒットを除いた値である。

種	太平洋種	北インド洋種	紅海種
サンプルID	Miyazaki01	Phuket01	Israel04
遺伝子数	19,397	18,045	18,879
機能予測遺伝子数 UniProt-SwissProt	12,621	12,253	12,249
UniProt-TrEMBL	15,672	15,008	15,073
NCBI-nr	17,878	16,949	17,105
InterProScan (モチーフ)	17,956	16,977	17,270
InterProScan (GO)	11,868	11,519	11,500
KEGG (K number)	10,053	9,830	9,799

第5章 種間ゲノム比較

5.1. 諸言

1.1.2.でも述べたとおり、核ゲノムにおけるオニヒトデ種間の違いは未だ明らかにされていない。そこで本研究では、核ゲノムを用いて配列類似度の算出と系統解析を実施し、核ゲノムにおけるオニヒトデ種間の違いの解明を行った。さらに、先行研究[11]がミトコンドリアのCOI遺伝子に特異的な結果でないことも確認するために、ミトコンドリアゲノムにおける種間の違いも合わせて確認した。

ゲノム配列類似度算出と系統解析に加えて、オニヒトデ種間のゲノムサイズに違いがあるのか確認した。3.2.2.の表3-2に示した通り、オニヒトデ3種3個体から構築した核ゲノムのサイズが種間で異なることが確認されたため、オニヒトデ種間の違いの1つにゲノムサイズ差が存在する可能性が考えられた。しかしながら、1種1個体の確認しか行われておらず、これらが個体間の差の範疇である可能性を否定できない。そこで、3種複数個体のゲノムシーケンスリードを用いて17-mer頻度分析を実施することで、ゲノムサイズ差が核ゲノムを構築した個体間の違いではなく種間の差であるのか確かめた。

5.2. 方法

5.2.1. ゲノム配列類似度算出

5.2.1.1. ミトコンドリアゲノム

本研究で構築した3種23個体のミトコンドリアゲノムに加えて、アセンブル時にリファレンス配列として用いた先行研究の太平洋種のミトコンドリアゲノム配列と、オニヒトデ近縁別種のアカオニヒトデのミトコンドリアゲノム配列(アクセッション番号: NC_007789.1)[14]の合計25個体のミトコンドリアゲノム配列を使用して塩基配列の類似度を算出した。

まず、環状のミトコンドリアゲノムの開始位置と終了位置を25個体で統一した。次にMAFFT v.7.3.10[56]を用いてマルチプルアライメントを行い(オプシ

オン:--auto)、ギャップサイトと D ループ領域上のアライメントを取り除いた。本研究では塩基配列の類似度を算出するため、挿入・欠損によって生じたギャップは解析の対象から取り除いた。また、D ループ領域は非遺伝子領域であり、進化速度の早いミトコンドリアゲノムの中でも特に進化速度の早い領域として知られている。このような領域では個体差も大きくアライメントを取るのが難しいため、誤ったアライメントによって塩基配列の類似度に影響することが考えられる。そのため、データベースに登録されていた太平洋種のミトコンドリアゲノム配列(アクセッション番号:NC_007788.1)のアノテーションを元に、アライメント結果から D ループに該当する領域を取り除いた。最後に、ギャップと D ループを取り除いたアライメント結果を基に 2 個体間のミトコンドリアゲノム塩基配列の類似度を全ての組み合わせで算出した。

5.2.1.2. 核ゲノム

核ゲノムにおける塩基配列の類似度を算出するために、2 種間で対応関係が取れる領域の特定を行った。対応関係を取る際にはリピート領域などの冗長な配列によって誤った対応関係が構築される可能性がある。そのような誤りを避けるために 2 種間で双方向に対応関係が取れる領域を本解析の対象領域として、核ゲノムにおける塩基配列の類似度を算出した。2 種間で対応関係が取れる領域の特定には本研究室に所属していた中村優太氏が開発した解析パイプラインを使用した。解析パイプラインは以下の通りである。

1. Minimap2 を用いて、核ゲノムの scaffold 配列を他種の scaffold 配列にアライメントを行った。その際--secondary=no のオプションを使用してトップヒットのみを出力させた。
2. 手順 1 においてアライメントが取れなかった領域を切り出し、再度アライメントを行った。
3. 全ての配列でアライメントが取れなくなるまで手順 2 を繰り返した。
4. 種間双方向のアライメント結果から、双方が共通して支持するアライメントを探索し、その領域を双方で対応関係が取れた領域とした。(図 5-1)

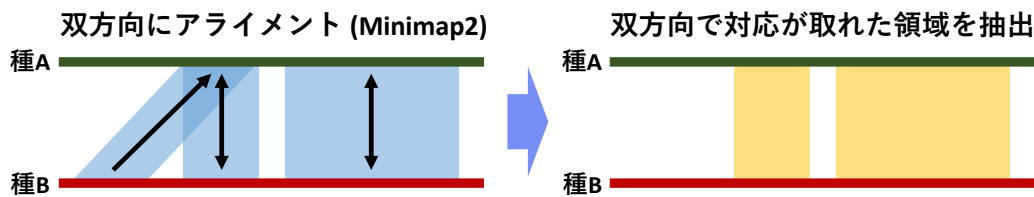


図 5-1：2 種間での対応関係の取り方の模式図

黒い矢印の向きはアライメントの方向を示している。

上記手順 2 においてアライメントが取れなかった領域を切り出して再度アライメントを行う理由としては、本来ならばアライメントが取れる領域が以下 3 つの理由からアライメントが取れなかった領域と判定されてしまうためである。

- ① 対応する配列の断片化によって複数存在する場合、手順 1 のアライメントではトップヒットの配列と対応している領域としかアライメントを取ることができず、その他はアライメントが取れなかった領域と判定されてしまう。
- ② 対応配列中に大きなギャップ領域が存在した場合、Minimap2 ではギャップ領域より下流の配列に本来ならばアライメントが取れる領域が存在しているにもかかわらず、アライメントが取れなかった領域と判定してしまう。
- ③ アセンブルエラーなどによって配列の対応関係がずれてしまった場合、アセンブルエラー箇所からアライメントが取れなかった領域と判定されてしまう。

種間で双方向の対応関係が取れた領域から、2 種間の核ゲノム塩基配列の類似度を全ての組み合わせで算出した。

5.2.2. 系統解析

5.2.2.1. ミトコンドリアゲノム

核ゲノムによる系統推定（手法は次の 5.2.2.2 に記載）では Single Nucleotide Polymorphism（以降 SNP と記述）の情報に基づいて行うため、ミトコンドリアゲノムにおいても核ゲノムと手法を揃えて SNP 情報から系統推定を行った。まず、5.2.1.1 においてミトコンドリアゲノム配列の塩基配列類似度の算出に用い

たアライメントから、自作した Python スクリプトを用いて SNP サイトを抽出した。SNP サイトの塩基配列情報を元に IQ-TREE v.1.6.12[57]を用いて最尤法による系統推定を実施した（オプション：-m MFP -bb 1000）。この時、Bayesian Information Criterion（以降 BIC と記述）から最適なモデルと判断された TIM3+F の置換モデルを使用した。また、最尤法に加えて MEGA X v.10.1.7[58]を用いた近隣接合法による系統推定も実施した。

5.2.2.2. 核ゲノム

核ゲノムの SNP 情報に基づいた系統推定を実施した。始めに、SNP データの作成を以下の手順で行った。

1. 表 2-2 で示した 23 個体のゲノムシーケンス PE リードを、BWA-MEM v.0.7.17-r1188[59]を用いて本研究で構築した太平洋種 (Miyazaki01) の核ゲノム配列にマッピングした。
2. 自作した Python スクリプトを用いてマッピング結果から、 $AS \leq XS$ となるようなリードが複数箇所にマッピングされていて最適なマッピング箇所が識別できない結果は偽陽性 SNP の原因となるため除外した。
3. マッピング結果から挿入・欠失によるアライメントミスを修正する為に、Genome Analysis Toolkit（以降 GATK と記述）v.3.8.1[60]の RealignerTargetCreator と IndelRealigner を用いてローカルリアライメントを実施した。
4. GATK の HaplotypeCaller を用いてマッピング結果から変異情報の抽出を行った。
5. SAMtools v.1.3.1[61]を用いてマッピング結果からリードがマッピングされたサイトを対象に 1 サイトにマッピングされた平均リード数を算出し、平均リード数の 3 倍より多いか平均リード数の 1/3 未満のリード数の変異情報を除外した。
6. GATK の GenotypeGVCFs を用いて 23 個体の各サンプルで別々の変異情報を 1 つに統合した。
7. GATK の SelectVariants から SNP の変異情報のみを抽出した。
8. GATK の VariantFiltration を用いて $QD < 2.0$ 、 $FS > 60.0$ 、 $MQ < 40.0$ 、 $MQRankSum < -12.5$ 、 $ReadPosRankSum < -8.0$ のいずれかの条件に当て

はまる SNP 情報を除外した。

9. フィルターした SNP 情報を vcf2phylip.py v.2.0[62]を用いて各サンプルの塩基配列情報に変換した。この時、SNP をヘテロで持っている場合は International Union of Pure and Applied Chemistry (IUPAC)の定める表記方法に従い、混合塩基として出力した。
10. IQ-TREE を用いて最尤法による系統推定を実施した (オプション: -m MFP -bb 1000)。置換モデルは BIC から最適モデルと判断された TVMe+R2 を使用した。さらに、最尤法に加えて MEGA X を用いた近隣接合法による系統推定も実施した。

5.2.3. 17-mer 頻度分析

23 個体のゲノムシーケンス PE リード (表 2-2) を用いて 17-mer 頻度分析を実施し、推定されるゲノムサイズと重複領域サイズに種間差があるのか確認を行った。23 個体のゲノムシーケンスの PE リードから Jellyfish を用いて 17-mer の頻度分布を作成した。次に GenomeScope を用いてゲノムサイズと重複領域サイズの推定を実施した。この時、北インド洋種の Phuket02 と Maldives02 の 2 つのサンプルはシーケンスリードの量が不足しており、頻度分布から正確な推定を行うことができないと判断されたため本解析からは除外した。

5.3. 結果・考察

5.3.1. ゲノム配列類似度算出

核ゲノムおよびミトコンドリアゲノムにおける個体間の配列類似度の結果をそれぞれ図 5-2、表 5-1 に示した。まず、ミトコンドリアゲノムにおける配列類似度はギャップおよび D ループ領域を除く 15,663 bp から算出されており、オニヒトデ 3 種間での配列類似度は 91.31~92.48%であった。次に、核ゲノムにおける配列類似度は 2 種間で双方向対応が取れた領域 (345,136,804 - 351,141,134 bp、カバー率 87.7-90.4%) から算出されており、オニヒトデ 3 種間での塩基配列の類似度は 97.52~98.11%であった。ヒトとチンパンジーにおけるミトコンドリアゲノムと核ゲノムの塩基配列の類似度はそれぞれ 91.1%と

98.77%であり[63][64]、オニヒトデ 3 種間の配列類似度はヒトーチンパンジー間に匹敵する違いを示していた。このことは、オニヒトデ 3 種間にゲノムレベルで別種に相当する違いがあることを示しており、Vogler らの先行研究[11]で示された結果をゲノムレベルで支持する結果であると言える。ただし、ヒトのヘテロ接合度は 0.1%に満たないのに対して[65]、表 3-2 に示した通りオニヒトデのヘテロ接合度は 0.72~1.24%とヒトよりも高いため、核ゲノムでは種間差だけでなく個体差も大きい可能性があることには留意が必要である。

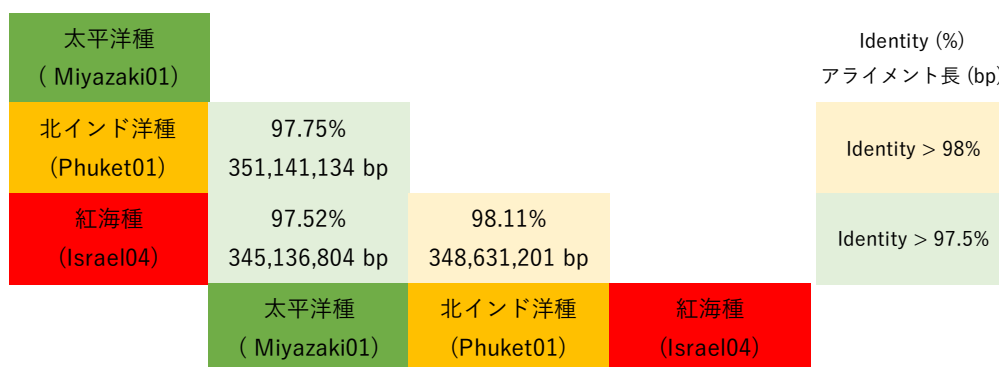


図 5-2：オニヒトデ 3 種 3 個体における 2 個体間の核ゲノム配列の類似度

表 5-1：オニヒトデ 3 種 24 個体とアカオニヒトデ 1 個体における 2 個体間のミトコンドリアゲノム配列の類似度

	太平洋種											北インド洋種						紅海種					アカオニヒトデ NC007789	Identity (%)			
	Miyazaki01	Miyazaki02	Miyazaki03	Miyazaki04	Miyazaki05	Ishigaki	Tahiti	Hawaii	Okinawa NC007788	OKI	GBR	Phuket01	Phuket02	Phuket03	Phuket04	Phuket05	Phuket06	Maldives01	Maldives02	Israel01	Israel02	Israel03	Israel04		Israel05		
太平洋種	Miyazaki01	100.00	99.69	99.96	99.71	99.76	99.61	98.83	99.01	99.38	99.70	99.78	91.66	91.66	91.66	91.64	91.68	91.61	91.63	91.68	91.83	91.82	91.84	91.80	91.86	88.69	100.00
	Miyazaki02	99.69	100.00	99.69	99.74	99.79	99.60	98.84	99.02	99.34	99.74	99.70	91.65	91.66	91.65	91.64	91.67	91.60	91.65	91.67	91.82	91.82	91.83	91.80	91.85	88.65	98.85
	Miyazaki03	99.96	99.69	100.00	99.71	99.74	99.60	98.82	99.00	99.37	99.69	99.77	91.63	91.64	91.63	91.62	91.66	91.59	91.60	91.66	91.82	91.81	91.83	91.79	91.85	88.70	97.70
	Miyazaki04	99.71	99.74	99.71	100.00	99.83	99.62	98.86	99.02	99.41	99.85	99.71	91.65	91.66	91.65	91.64	91.67	91.60	91.64	91.67	91.83	91.82	91.84	91.80	91.86	88.71	96.55
	Miyazaki05	99.76	99.79	99.74	99.83	100.00	99.67	98.93	99.10	99.41	99.82	99.76	91.67	91.67	91.67	91.66	91.69	91.62	91.66	91.69	91.82	91.81	91.83	91.79	91.85	88.71	95.40
	Ishigaki	99.61	99.60	99.60	99.62	99.67	100.00	98.76	98.91	99.25	99.62	99.58	91.58	91.59	91.57	91.57	91.60	91.53	91.57	91.60	91.78	91.78	91.80	91.76	91.82	88.64	94.26
	Tahiti	98.83	98.84	98.82	98.86	98.93	98.76	100.00	99.00	98.46	98.87	98.83	91.52	91.52	91.52	91.50	91.54	91.48	91.51	91.54	91.65	91.67	91.67	91.64	91.71	88.51	93.11
	Hawaii	99.01	99.02	99.00	99.02	99.10	98.91	99.00	100.00	98.63	99.02	99.00	91.69	91.70	91.70	91.68	91.72	91.65	91.65	91.72	91.83	91.83	91.84	91.80	91.87	88.57	91.96
	Okinawa NC007788	99.38	99.34	99.37	99.41	99.41	99.25	98.46	98.63	100.00	99.40	99.41	91.36	91.36	91.36	91.34	91.38	91.31	91.34	91.38	91.54	91.52	91.54	91.50	91.56	88.52	90.81
	OKI	99.70	99.74	99.69	99.85	99.82	99.62	98.87	99.02	99.40	100.00	99.71	91.64	91.64	91.66	91.62	91.66	91.60	91.62	91.66	91.82	91.81	91.83	91.80	91.87	88.70	89.66
GBR	99.78	99.70	99.77	99.71	99.76	99.58	98.83	99.00	99.41	99.71	100.00	91.63	91.64	91.63	91.60	91.66	91.59	91.61	91.66	91.83	91.82	91.84	91.80	91.86	88.67	88.51	
北インド洋種	Phuket01	91.66	91.65	91.63	91.65	91.67	91.58	91.52	91.69	91.36	91.64	91.63	100.00	99.85	99.86	99.83	99.88	99.92	99.41	99.87	92.42	92.44	92.47	92.42	92.46	88.71	
	Phuket02	91.66	91.66	91.64	91.66	91.67	91.59	91.52	91.70	91.36	91.64	91.64	99.85	100.00	99.83	99.82	99.87	99.80	99.39	99.85	92.42	92.44	92.47	92.40	92.43	88.69	
	Phuket03	91.66	91.65	91.63	91.65	91.67	91.57	91.52	91.70	91.36	91.66	91.63	99.86	99.83	100.00	99.81	99.86	99.81	99.40	99.85	92.42	92.44	92.47	92.43	92.48	88.69	
	Phuket04	91.64	91.64	91.62	91.64	91.66	91.57	91.50	91.68	91.34	91.62	91.60	99.83	99.82	99.81	100.00	99.85	99.79	99.39	99.84	92.38	92.41	92.43	92.37	92.41	88.69	
	Phuket05	91.68	91.67	91.66	91.67	91.69	91.60	91.54	91.72	91.38	91.66	91.66	99.88	99.87	99.86	99.85	100.00	99.83	99.44	99.89	92.42	92.45	92.47	92.41	92.43	88.72	
	Phuket06	91.61	91.60	91.59	91.60	91.62	91.53	91.48	91.65	91.31	91.60	91.59	99.92	99.80	99.81	99.79	99.83	100.00	99.37	99.82	92.37	92.40	92.42	92.37	92.42	88.65	
	Maldives01	91.63	91.65	91.60	91.64	91.66	91.57	91.51	91.65	91.34	91.62	91.61	99.41	99.39	99.40	99.39	99.44	99.37	100.00	99.41	92.37	92.41	92.42	92.37	92.42	88.70	
	Maldives02	91.68	91.67	91.66	91.67	91.69	91.60	91.54	91.72	91.38	91.66	91.66	99.87	99.85	99.85	99.84	99.89	99.82	99.41	100.00	92.41	92.43	92.47	92.40	92.43	88.73	
紅海種	Israel01	91.83	91.82	91.82	91.83	91.82	91.78	91.65	91.83	91.54	91.82	91.83	92.42	92.42	92.42	92.38	92.42	92.37	92.37	92.41	100.00	99.73	99.83	99.85	99.76	88.83	
	Israel02	91.82	91.82	91.81	91.82	91.81	91.78	91.67	91.83	91.52	91.81	91.82	92.44	92.44	92.44	92.41	92.45	92.40	92.41	92.43	99.73	100.00	99.73	99.74	99.74	88.85	
	Israel03	91.84	91.83	91.83	91.84	91.83	91.80	91.67	91.84	91.54	91.83	91.84	92.47	92.47	92.47	92.43	92.47	92.42	92.42	92.47	99.83	99.73	100.00	99.86	99.76	88.86	
	Israel04	91.80	91.80	91.79	91.80	91.79	91.76	91.64	91.80	91.50	91.80	91.80	92.42	92.40	92.43	92.37	92.41	92.37	92.37	92.40	99.85	99.74	99.86	100.00	99.80	88.78	
	Israel05	91.86	91.85	91.85	91.86	91.85	91.82	91.71	91.87	91.56	91.87	91.86	92.46	92.43	92.48	92.41	92.43	92.42	92.42	92.43	99.76	99.74	99.76	99.80	100.00	88.84	
アカオニヒトデ NC007789	88.69	88.65	88.70	88.71	88.71	88.64	88.51	88.57	88.52	88.70	88.67	88.71	88.69	88.69	88.69	88.72	88.65	88.70	88.73	88.83	88.85	88.86	88.78	88.84	100.00		

5.3.2. 系統解析

ミトコンドリアゲノムと核ゲノムの SNP サイトを用いて最尤法ならびに近隣接合法で系統解析を行った結果をそれぞれ図 5-3 と図 5-4 に示した。まず、ミトコンドリアゲノム上の 2,996 SNPs に基づいて最尤法と近隣接合法のそれぞれから推定された系統樹では (図 5-3A と図 5-4A)、太平洋種、北インド洋種、紅海種が異なるクレードとして明確に分けられた。このことから、ミトコンドリアゲノムにおいても COI 遺伝子に基づいた先行研究[11]の結果と同様に、オニヒトデ種間の違いが示された。次に、核ゲノム上の 18,535,994 SNPs に基づいて最尤法と近隣接合法のそれぞれから推定された系統樹では(図 5-3B と図 5-4B)、ミトコンドリアゲノムでの結果と同様にオニヒトデ 3 種が異なるクレードに明確に分かれ、核ゲノムにおいても先行研究[11]のオニヒトデ種間の違いを支持する結果が示された。しかしながら、種内においてはミトコンドリアとは異なる 2 つの特徴が核ゲノムの系統樹から確認された。1 つ目は北インド洋種内でモルディブ集団とプーケット集団の違いが確認されたことである。ミトコンドリアに基づく先行研究において、モルディブは北インド洋の東部と西部の遺伝子型特別が付かず、2 つの遺伝子型が混在する海域であったことから、北インド洋の東西の集団から直接または 2 世代を経てモルディブに幼生加入が起きている可能性が示されている[66]。本研究におけるミトコンドリアゲノムの系統樹においてもモルディブ集団の Maldives02 がプーケット集団と区別が付かなかったものの、核ゲノムの系統樹ではモルディブ集団とプーケット集団の間に違いが確認された。このことから、モルディブ集団はミトコンドリアから想定されるほど他集団からの強い幼生加入は受けていない可能性が考えられた。2 つ目は太平洋種内のハワイがミトコンドリアゲノムから想定されるよりも核ゲノムにおいて他の太平洋集団から分化していたことである。アロザイムを用いた研究においても太平洋集団の中でハワイ集団の分化が確認されており[10]、Vogler らの先行研究[11]で用いられたミトコンドリアのデータでは不明瞭であったハワイの遺伝的な分化を核ゲノム配列解析から発見したことには価値があると考えられる。ハワイ集団のオニヒトデが太平洋種の亜種や隠蔽種である可能性が考えられるが、本研究ではハワイは 1 個体しか扱えていないためハワイ集団やその周辺の東太平洋集団のサンプルを増やして確認を行う必要がある。

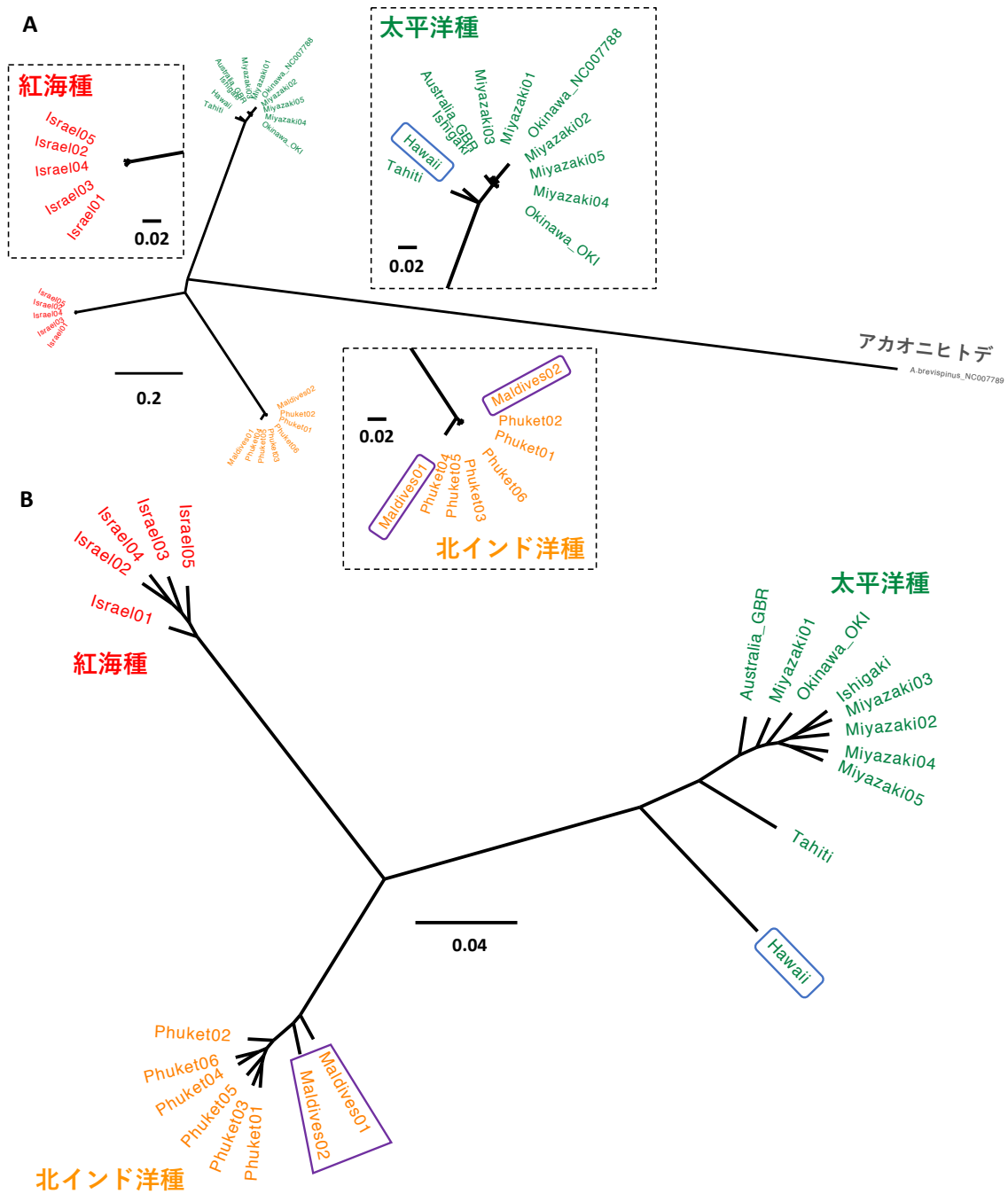


図 5-3：最尤法による系統樹

A はミトコンドリアゲノム、B は核ゲノムの系統樹である。太平洋種クレード内の青枠はハワイのサンプルの位置を示し、北インド洋種クレード内の紫枠はモルディブのサンプルの位置を示している。スケールバーの値は 1 bp あたりの置換率を示している。

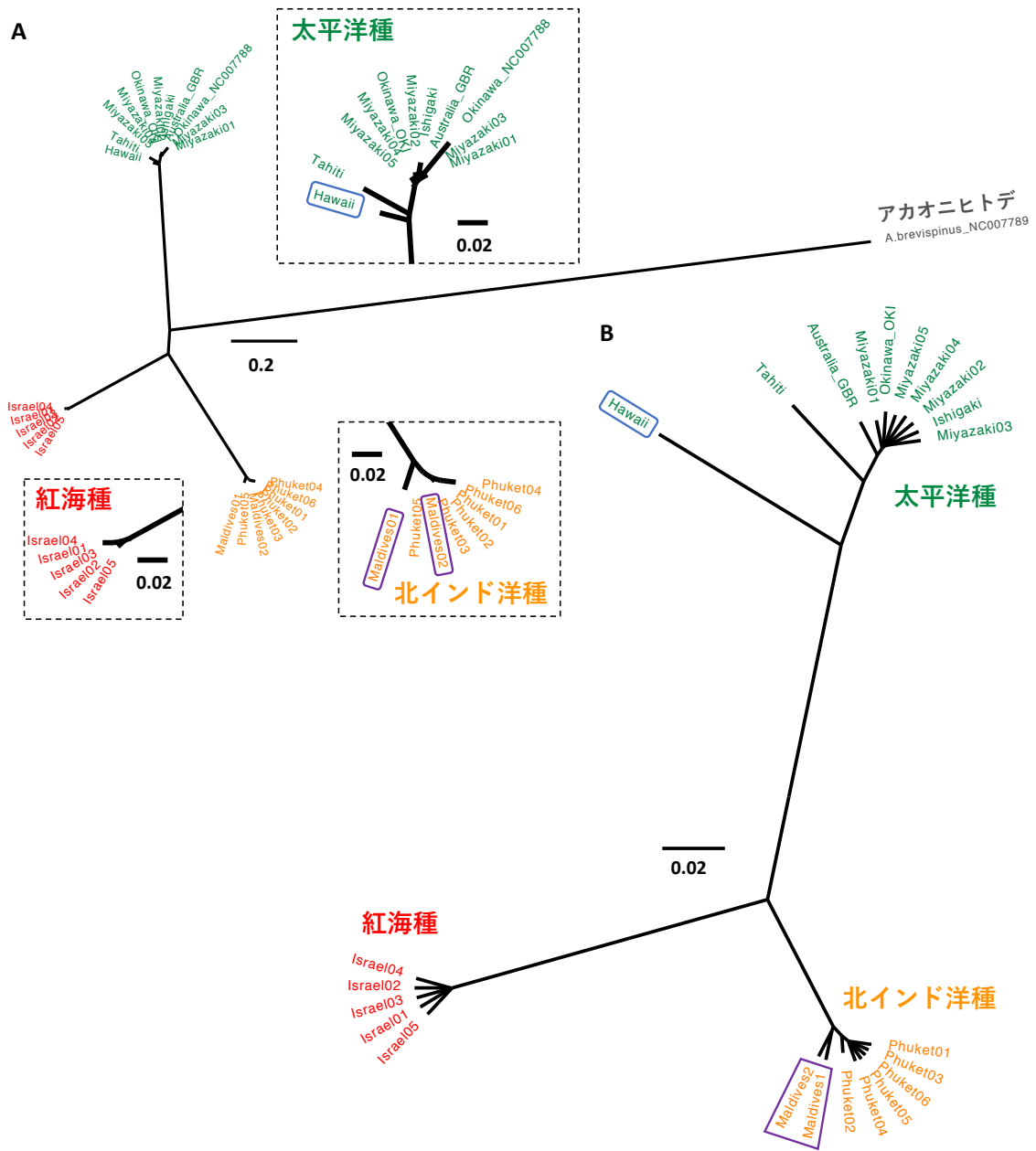


図 5-4：近隣接合法による系統樹

A はミトコンドリアゲノム、B は核ゲノムの系統樹である。太平洋種クレード内の青枠はハワイのサンプルの位置を示し、北インド洋種クレード内の紫枠はモルディブのサンプルの位置を示している。スケールバーの値は 1 bp あたりの置換率を示している。

5.3.3. 17-mer 頻度分析

17-mer 頻度分析から推定される太平洋種、北インド洋種、紅海種のゲノムサイズおよび重複領域サイズの推定値を表 5-2 に示しており、その推定値を図にしたものが図 5-5 である。図 5-5 に示した通り 3 種複数個体で推定ゲノムサイズを確認した場合、太平洋種、北インド洋種、紅海種の順に大きい傾向にあり、太平洋種とインド洋側 2 種（北インド洋種、紅海種）の間で有意差が確認された。このことから、3.2.2. で確認されたゲノムサイズ差は、太平洋種とインド洋側 2 種の間差であると考えられた。また、推定重複領域サイズは推定ゲノムサイズと同様に太平洋種、北インド洋種、紅海種の順に大きい傾向にあり、太平洋種とインド洋側 2 種の間で有意差が確認された。このことから、太平洋種とインド洋側 2 種の間に見られた有意なゲノムサイズ差は重複領域に起因していることが示唆された。

また、17-mer 頻度分析から推定される重複領域とは異なるが、本論文で決定した 3 種 3 個体のドラフトゲノム配列から検出されたリピート配列の解析結果についても表 5-3 に示す（リピート配列の解析方法は 4.1.1.3. *Ab initio* ベースに記載）。ゲノム配列中のリピート領域の割合は 17-mer 頻度分析から推定された重複配列と同様に太平洋種、北インド洋種、紅海種の順に大きい傾向にあったが、どの種においても検出されたリピート領域の中で分類不明（Unclassified）なりリピート領域が 8 割を占めていた。

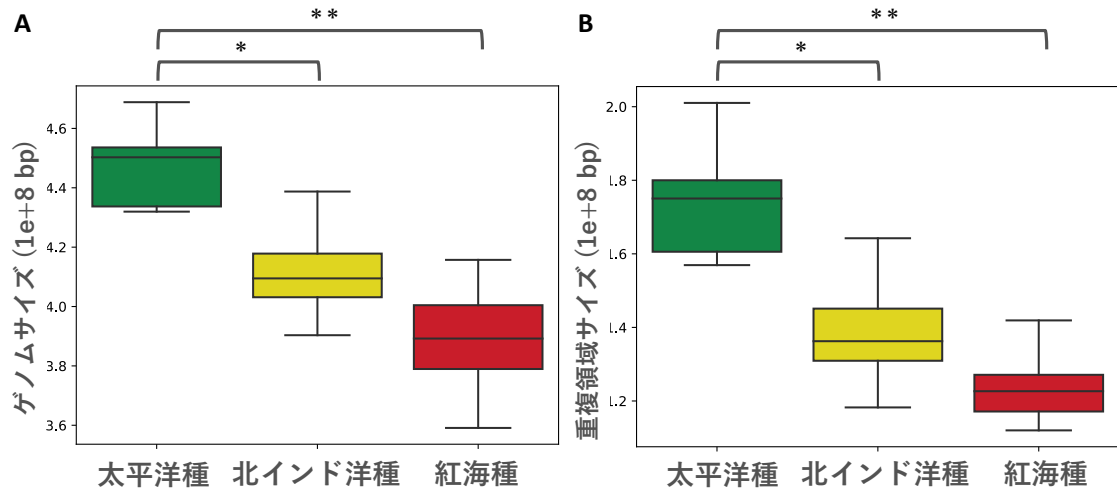


図 5-5 : GenomeScope を用いて算出した各種各個体の推定サイズの箱ひげ図

A 推定ゲノムサイズ、B 推定重複領域サイズ。*と**は t 検定で 2 種間に有意な差が確認されたことを表しており、*は P -value <0.001 、**は P -value <0.0001 を示している。

表 5-2：17-mer 頻度分析から算出された各サンプルの推定ゲノムサイズおよび推定重複領域サイズの値

Phuket02 と Maldives02 はリード量が不十分であったため、本解析から除外している。

サンプルID	種名	推定ゲノムサイズ (bp)	推定重複領域サイズ (bp)
Miyazaki01	太平洋種	454,453,223	181,066,440
Miyazaki02	太平洋種	458,895,635	187,439,114
Miyazaki03	太平洋種	451,033,216	176,806,191
Miyazaki04	太平洋種	450,560,916	175,859,992
Miyazaki05	太平洋種	468,838,717	201,027,300
Ishigaki	太平洋種	433,426,245	159,917,057
Tahiti	太平洋種	431,967,378	156,934,204
Hawaii	太平洋種	449,979,311	174,222,535
Phuket01	北インド洋種	419,206,207	147,221,506
Phuket02	北インド洋種	---	---
Phuket03	北インド洋種	402,420,868	133,734,055
Phuket04	北インド洋種	405,249,853	130,001,657
Phuket05	北インド洋種	390,337,188	118,248,028
Phuket06	北インド洋種	413,701,672	138,770,963
Maldives01	北インド洋種	438,735,683	164,239,162
Maldives02	北インド洋種	---	---
Israel01	紅海種	389,214,239	122,657,975
Israel02	紅海種	415,717,469	141,903,543
Israel03	紅海種	400,448,689	127,114,001
Israel04	紅海種	378,956,169	117,138,513
Israel05	紅海種	359,101,243	112,007,830
OKI	太平洋種	432,350,485	159,556,596
GBR	太平洋種	434,570,287	162,531,620

表 5-3：3 種 3 個体のドラフトゲノム配列から検出されたリピート配列

	大太平洋種 (Miyazaki01)		北インド洋種 (Phuket01)		紅海種 (Israel04)	
	合計長(bp)	割合(%)	合計長(bp)	割合(%)	合計長(bp)	割合(%)
Retroelements	12,866,218	3.06%	11,517,269	2.88%	9,954,971	2.58%
DNA transposons	8,065,824	1.92%	7,406,343	1.85%	6,709,066	1.74%
Unclassified	110,904,477	26.40%	97,039,334	24.23%	88,864,864	23.04%
Small RNA	1,376,356	0.33%	1,278,343	0.32%	1,099,964	0.29%
Satellites	41,783	0.01%	32,466	0.01%	37,641	0.01%
Simple repeats	2,894,100	0.69%	2,350,037	0.59%	2,093,602	0.54%
Low complexity	409,070	0.10%	328,485	0.08%	313,510	0.08%
Total	136,557,828	32.51%	119,952,277	29.96%	109,073,618	28.28%

5.4. まとめ

ゲノムデータに基づく配列類似度算出と系統解析の結果は太平洋種、北インド洋種、紅海種のオニヒトデ 3 種間の違いを示した。ゲノムの塩基配列の類似度はミトコンドリアゲノムと核ゲノムでそれぞれ約 92%と約 98%であり、これは別種の関係にあるヒトーチンパンジー間 (91.1%と 98.77%) に匹敵する値であった。系統解析では核ゲノムを用いてオニヒトデ種間の違いを解析し、オニヒトデ 3 種間に明確な違いが存在することが示された。ただし、種内においてはミトコンドリアゲノムと核ゲノムの系統樹で違いが見られた。北インド洋種のモルディブはミトコンドリアゲノムの系統樹ではプーケットと明確な違いは示されなかったものの、核ゲノムの系統樹では 2 集団間の違いが確認された。このことから、モルディブはミトコンドリアに基づく先行研究[66]から想定されるほど北インド洋種その他海域との強い結びつきはない可能性が考えられた。また、太平洋種のハワイはミトコンドリアゲノムの系統樹から想定されるよりも核ゲノムの系統樹において他の太平洋集団から分化していることが確認された。アロザイムを用いた研究においても太平洋集団の中でハワイ集団の分化が確認されており、ハワイやその周辺を含めた東太平洋集団のサンプルを増やして太平洋種内の亜種や隠蔽種の可能性を検討する必要がある。17-mer 頻度分析からは太平洋種とインド洋側 2 種 (北インド洋種、紅海種) の間で重複領域に起因している可能性のあるゲノムサイズの差が存在することが明らかとなった。

第 6 章 オニヒトデの種分化要因推定

6.1 諸言

第 5 章に示した通り、オニヒトデ 3 種間の違いはミトコンドリアだけでなく核においても確認されたことで、オニヒトデ種の分類が明確になった。本章ではそのオニヒトデ 3 種の違いがどのように形成されたのかについて検証した結果を示す。

オニヒトデ種間の違いを初めて報告した Vogler et al. 2008 の論文[11]では、分岐年代推定の結果からオニヒトデの種分化は鮮新世～更新世前期（195 万～365 万年前）の海水面の変化によって集団間の交流が阻害され、異所的に起こったという仮説が示されている。しかしながら、この仮説は分岐年代推定のみに基づいたものであり、オニヒトデの異所的種分化を示すデータは存在しないため、種分化要因は未だ明らかでない。そこで本研究では、太平洋種、北インド洋種、紅海種のオニヒトデ近縁種のゲノム情報を用いて種分化要因の検証を行うこととした。まず、祖先集団の有効集団サイズの変遷（以降集団史と記述）の推定、個体群の遺伝構造解析、集団間の移住を推定する遺伝子流動解析を行い、各種各海域の進化的な背景を調査した。次に、種間分化領域探索と種間遺伝子比較を行うことで、ゲノムに残る種分化の痕跡を調査した。

6.2. 方法

6.2.1. 集団史推定

PSMC'を用いて 1 個体のゲノムから集団史を推定した。PSMC'は PSMC モデル[6]を改良したもので、より正確に組み替え率などを反映して集団史の推定を行うことができる。集団史の推定は PSMC だけでなく、複数個体のゲノム情報を用いることで PSMC よりも正確に集団史推定を行うことができる Multiple Sequentially Markovian Coalescent（以降 MSMC と記述）モデル[67]が存在する。しかしながら、この MSMC モデルは高精度のハプロタイプのフェージング情報が必要であり[68]、オニヒトデを含めた非モデル生物で高精度のフェージング情報を作成することは難しい。このフェージングの問題を乗り越えて 2 個

体以上から集団史推定を行う SMC++[69]と呼ばれるツールも登場しているが、本研究では各地点で1~6個体とサンプル数にばらつきがあるため、サンプル数による結果の偏りが懸念される。そこで本研究では個体ごとに PSMC'を行うことでフェージングやサンプル数の違いによって生じる影響を回避した。PSMC'の実行は以下の手順で行った。

1. BWA-MEM を用いてゲノムシーケンスを行った 3 種 21 個体の PE リード (表 2-1) を同種のリファレンス配列にマッピングを行った。リファレンス配列には太平洋種は Miyazaki01、北インド洋種は Phuket01、紅海種は Israel04 から構築した核ゲノム配列を用いた。
2. 自作した Python スクリプトを用いてマッピング結果から、 $AS \leq XS$ となるようなリードが複数箇所にもマッピングされていて最適なマッピング箇所が識別できない結果は偽陽性 SNP の原因となるため除外した。
3. マッピング結果から挿入・欠失によるアライメントミスを修正する為に GATK の RealignerTargetCreator と IndelRealigner を用いてローカルリアライメントを実施した。
4. 本解析では連続した配列が求められるため、自作した Python スクリプトを用いてマッピング結果から scaffold 長 < 100 kbp の短い scaffold 配列にマッピングされた結果を取り除いた。マッピング後に短い scaffold 配列を除いたのは、短い scaffold 配列に由来するリードのミスアライメントを避けるためである。
5. MSMC-Tools のマニュアル (<https://github.com/stschiff/msmc-tools>) に従い SAMtools、Bcftools v.1.9 (<https://github.com/samtools/bcftools>)、bamCaller.py を用いてマッピング結果から vcf ファイル (SNP 情報) と mask ファイル (十分にマッピングされた領域の情報) の 2 ファイルを scaffold 配列ごとに生成した。
6. SNPable パイプライン (<http://lh3lh3.users.sourceforge.net/snpable.shtml>) に従い、ショートリードを一意にマッピングするために必要な mappability mask ファイルを scaffold 配列ごとに作成した。
7. MSMC-Tools の generate_multihetsep.py を用いて、これまでに作成した 3 つのファイル (vcf ファイル、mask ファイル、mappability mask ファイル) を PSMC'の実行に必要な入力ファイルに変換した。
8. MSMC2 v.2.1.1[70]を用いて PSMC'を実行した。

手順1において、各種で共通のリファレンスではなく種ごとのリファレンスを用いた理由は、種特異的領域を本解析に含めるためである。PSMCでは非組換え領域のヘテロ接合度から集団史推定を行っており[6]、より正確な推定を行うには種特異的領域上にあるSNPも可能な限り解析対象に含めることが重要であると考えられる。表6-1は種ごとのゲノム配列をリファレンスにした時と太平洋種のゲノム配列をリファレンスにした時に北インド洋種と紅海種の各サンプルから検出されるSNP数を示している。太平洋種をリファレンスにした時に抽出されるSNP数は種ごとのリファレンスを用いた時に比べておおよそ半減していることが確認され、種特異領域上のSNPが対象外になってしまう可能性が示唆された。また、リファレンス配列由来のバイアスが祖先推定を行う集団ゲノム解析に影響を及ぼすという報告[71]も考慮し最終的な手順を決定した。

表 6-1：種ごとまたは大太平洋種のゲノム配列をリファレンスにした時に北インド洋種と紅海種から検出される SNP 数

サンプルID	SNPs (種ごとのリファレンス)	SNPs (大太平洋種のリファレンス)
Phuket01	1,350,317	785,303
Phuket02	654,447	305,457
Phuket03	1,085,290	463,055
Phuket04	1,136,990	480,761
Phuket05	1,121,784	469,202
Phuket06	1,113,323	469,026
Maldives01	1,685,214	1,052,100
Maldives02	1,498,060	879,987
Israel01	1,093,999	510,479
Israel02	1,114,908	505,691
Israel03	1,018,443	438,306
Israel04	1,165,131	487,202
Israel05	1,047,086	487,558

最後に、各個体から推定された集団史の描画は自作した R スクリプトを用いて行った。この時、世代間時間は 3 年、世代間変異率は太平洋種で 9.9×10^{-9} bp/世代、北インド洋種で 9.4×10^{-9} bp/世代、紅海種で 8.8×10^{-9} bp/世代とした。餌環境や生息地域の環境にも依存するが、オニヒトデは発生して約 2 年後から性的成熟が始まり [72]、小さい個体でも 2 年目の冬以降には生殖腺の発達が確認されている [73]。そのため、3 年目の夏の生殖シーズンには大抵の個体が生殖行動に参加すると考えられるため、世代間時間を 3 年に設定した。オニヒトデの世代間変異率に関しては情報が存在しなかったため、本解析でリファレンス配列として用いた各種個体から推定されたゲノムサイズをもとに算出した（太平洋種：454 Mbp、北インド洋種：419 Mbp、紅海種：378 Mbp） [74]。世代間変異率の算出に用いた式は以下の通りである。

$$\text{世代間変異率 (bp/世代)} = 10^{(-0.81+0.61\log_{10}G)} / 10^9 \quad G \text{ はゲノムサイズ (Mbp)}$$

6.2.2. 遺伝構造解析

Bcftools を用いて核ゲノムの系統解析で用いた SNP データセットから二対立 SNP (biallelic SNP) のみを抽出した。連鎖不平衡の影響を軽減する為に、PLINK2 v.2.00a2.3LM [75] を用いて相関係数の 2 乗 (r^2) が 0.5 以上の 200 kbp 以内に存在する SNP ペアを除去した (オプション: --indep-pairwise 200kb 1 0.5)。最後に ADMIXTURE v.1.3.0 [76] を用いて各個体の遺伝構造の推定を行った。

6.2.3. 種間分化領域探索

種間分化領域を行う際に種内多型によって種間の分化が見えづらくなることを防ぐために、各種 1 地点のサンプルを本解析に使用した。太平洋種、北インド洋種、紅海種の中で宮崎、プーケット、イスラエルは 5 個体以上のサンプル数が存在する為、これら 3 地点の 3 種 16 個体のサンプルを本解析に用いる事とし、以下の手順で種間の差異の指標 F_{ST} を算出した。

1. BWA-MEM を用いて 16 サンプルの PE リードを太平洋種 (Miyazaki01) の核ゲノム配列にマッピングした。
2. 自作した Python スクリプトを用いてマッピング結果から、 $AS \leq XS$ とな

るようなリードが複数箇所にマッピングされていて最適なマッピング箇所が識別できない結果は偽陽性 SNP の原因となるため除外した。

3. マッピング結果から挿入・欠失によるアライメントミスを修正する為に GATK の `RealignerTargetCreator` と `IndelRealigner` を用いてローカルリアライメントを実施した。
4. SAMtools を用いてマッピング結果からリードがマッピングされたサイトを対象に 1 サイトにマッピングされた平均リード数を算出し、平均リード数の 3 倍より多いか平均リード数の 1/3 未満のリード数の変異情報を除外した。
5. GATK の `GenotypeGVCFs` を用いて各 16 サンプルで別々の変異情報を 1 つに統合した。GATK の `SelectVariants` から SNP の変異情報のみを抽出した。
6. GATK の `VariantFiltration` を用いて $QD < 2.0$ 、 $FS > 60.0$ 、 $MQ < 40.0$ 、 $MQRankSum < -12.5$ 、 $ReadPosRankSum < -8.0$ のいずれかの条件に当てはまる SNP 情報を除外した。
7. Bcftools を用いて SNP 情報から二対立 SNP の情報のみを抽出した。
8. VCFtools v.0.1.13 を用いて 250 kbp 以上の scaffold 配列を 50 kbp の区画に区切り、区画ごとの平均 F_{ST} を算出した。この時、SNP の数が 10 未満または配列長が 50 kbp 未満の区画は解析から除外した。

6.2.4. 種間遺伝子比較

6.2.4.1. オーソログ解析

SonicParanoid v.1.3.4[77]を用いて、本研究で構築した太平洋種、北インド洋種、紅海種の遺伝子の種間オーソログ関係を推定した。

6.2.4.2. 正の自然選択圧検出

種特異的な遺伝子を検出する為に、3 種間共通シングルコピーオーソログ遺伝子の中で 1 種のみ有意な正の自然選択が検出される遺伝子の検索を行った。まず、MAFFT を用いて 3 種間共通シングルコピーオーソログ遺伝子のアミノ

酸配列をオーソロググループごとにマルチプルアライメントを行った（オプション：--auto）。その後、アライメントしたアミノ酸配列を元に PAL2NAL v.14[78]を用いて塩基配列のコドンアライメントをオーソロググループごとに行った。TrimAl v.1.4.rev22[79]を用いてアライメント結果からミスが疑われる領域を特定し（オプション：-strict, -colnumbering）、自作した Python スクリプトを用いてミスが含まれるコドン領域を除外した。次に、IQ-TREE を用いて 3 種間共通シングルコピーオーソログ遺伝子の各グループの配列から最尤法による系統推定を行った（オプション：-m MFP）。この時、BIC から最適なモデルと判断された GTR+F+I の置換モデルを使用した。最後に、HYPHY v.2.5.1[80]を用いて、3 種間共通シングルコピーオーソログ遺伝子の各グループの配列と系統樹を入力として adaptive Branch-Site Random Effects Likelihood（以降 aBSREL と記述）モデル[81]を実行した。aBSREL モデルは dN/dS に基づいて正の自然選択を検出する branch-site モデル[82]の改良版である。aBSREL モデルで入力する系統樹には 3 種間共通シングルコピーオーソログ遺伝子の各グループの配列を連結させて推定した 1 つの種系統樹を入力に用いる研究例もあるが（*e.g.* [83]）、本研究で扱うオニヒトデ属 3 種は近縁であり各遺伝子に祖先多型（*i.e.* Incomplete lineage sorting）の影響が残っている可能性が考えられるため、樹形が種系統樹と異なることによる誤推定を避ける目的で各遺伝子の系統樹を使用した。また、各オーソロググループにおいて前景情報（正の自然選択を調べる対象の指定）は入力せずに系統樹の全ての枝を対象に有意な正の自然選択を調査した（ホルム・ボンフェローニ法で多重検定の補正、有意水準 5%を採用）。

6.2.4.3. GO/KEGG エンリッチメント解析

他種とオーソログ関係の取れなかった遺伝子グループと種特異的な正の自然選択が検出された遺伝子グループの遺伝子を解析の対象に GO または KEGG（KEGG PATH と KEGG BRITE）の機能ごとに遺伝子をまとめた。その後、フィッシャーの正確確率検定を行い、各グループにおいて有意に濃縮（エンリッチ）されている GO または KEGG の機能を検出した（有意水準 5%）。

6.3. 結果・考察

6.3.1. 集団史推定

PSMC'モデルを用いて集団史推定を行った結果を図 6-1 と図 6-2 に示した。図 6-1 では各海域から代表 1 個体の集団史推定結果を表示しており、図 6-2 では全ての個体の集団史推定結果を海域ごとに表示している。まず、3 種 7 海域の有効集団サイズの増減が 70–125 万年前の Mid-Pleistocene Transition (以降 MPT と記述) 以前まで同調したパターンを示していた。これは各海域の個体が 1 つの共有祖先集団に属していたことを示している。その後、MPT の間に各海域の集団史の傾向に違いが生じ始めており、この時期に各海域で独立した集団が形成され始めたことを示唆していた。ミトコンドリアを用いた先行研究[11]において分岐年代は 195–365 万年前と推定されており、この時期の氷河作用による海水面変動が異所的種分化を引き起こしたのではないかと考えられている。本研究では先行研究よりも 70 万年ほど遅い MPT にオニヒトデ 3 種間が分岐したことを支持したが、この MPT の時期は海水面の変動幅が特に大きく、数百万年ぶりに海水面が 120m も下降した時期であった[84]。数百万年の間で類を見ない程の海水面の低下によって大陸棚が干出し、そのことが各海域間の遺伝子流動を阻害したことで独立した集団の形成に寄与したと考えられる。本研究から MPT の時期の劇的な海水低下で干出した大陸棚による地理的な隔離がオニヒトデの異所的種分化を引き起こしたことを示唆し、先行研究よりも具体的な可能性を示すことができた。また、オニヒトデの有効集団サイズは MPT 頃まで増加して MPT 後に減少する傾向が確認された。この傾向はミドリイシ属 (*Acropora* spp.) のサンゴの集団史推定の結果からも示されている[85][86]。オニヒトデはサンゴの中でもミドリイシ属のサンゴを好んで捕食することが報告されており[87]、餌の増減がオニヒトデの有効集団サイズの増減に影響している可能性が考えられた。さらに、宮崎と石垣の西太平洋では他の 5 海域よりも MPT 後の有効集団サイズの減少が抑えられており、安定した有効集団サイズを示していた。Hall et al. 2017[15]で行われた西太平洋 2 海域 (沖縄本島、オーストラリア) の集団史推定の結果からも同様の傾向が確認されている。西太平洋にはコーラルトライアングルと呼ばれる広さと生物多様性の高さの両方で世界最大のサンゴ礁海域が存在しており[2]、この海域の広大な生息域と餌環境が西太

平洋のオニヒトデのレフュージアとして機能した可能性がある。

最後に世代間変異率を変化させた時の集団史推定結果を図 6-3 に示した。本研究では Lynch 2010[74]を元にゲノムサイズから世代間変異率を算出したが、推定した世代間変異率を 1×10^{-9} bp/世代の範囲で増減させても上記にあげた結果と考察に影響はなかった。また、Hall et al. 2017[15]の先行研究で採用されていた太平洋種の世代間変異率の 1×10^{-8} bp/世代を用いても、太平洋種における上記の結果と考察は変わらないと考えられる。

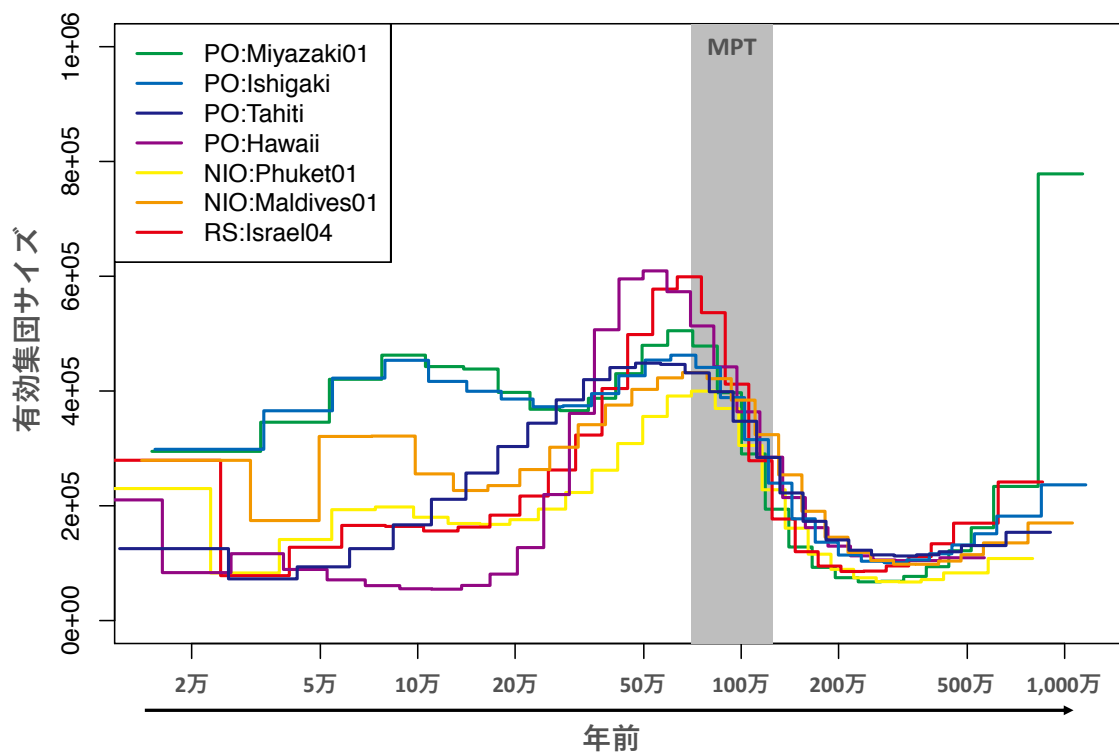


図 6-1 : PSMC'モデルによる 7 海域各 1 個体の集団史推定結果

PO は太平洋種、NIO は北インド洋種、RS は紅海種を示している。灰色塗りの箇所は Mid-Pleistocene Transition (MPT、70–125 万年前) を示している。世代間時間は 3 年、世代間変異率は太平洋種で 9.9×10^{-9} bp/世代、北インド洋種で 9.4×10^{-9} bp/世代、紅海種で 8.8×10^{-9} bp/世代として集団史を推定した。

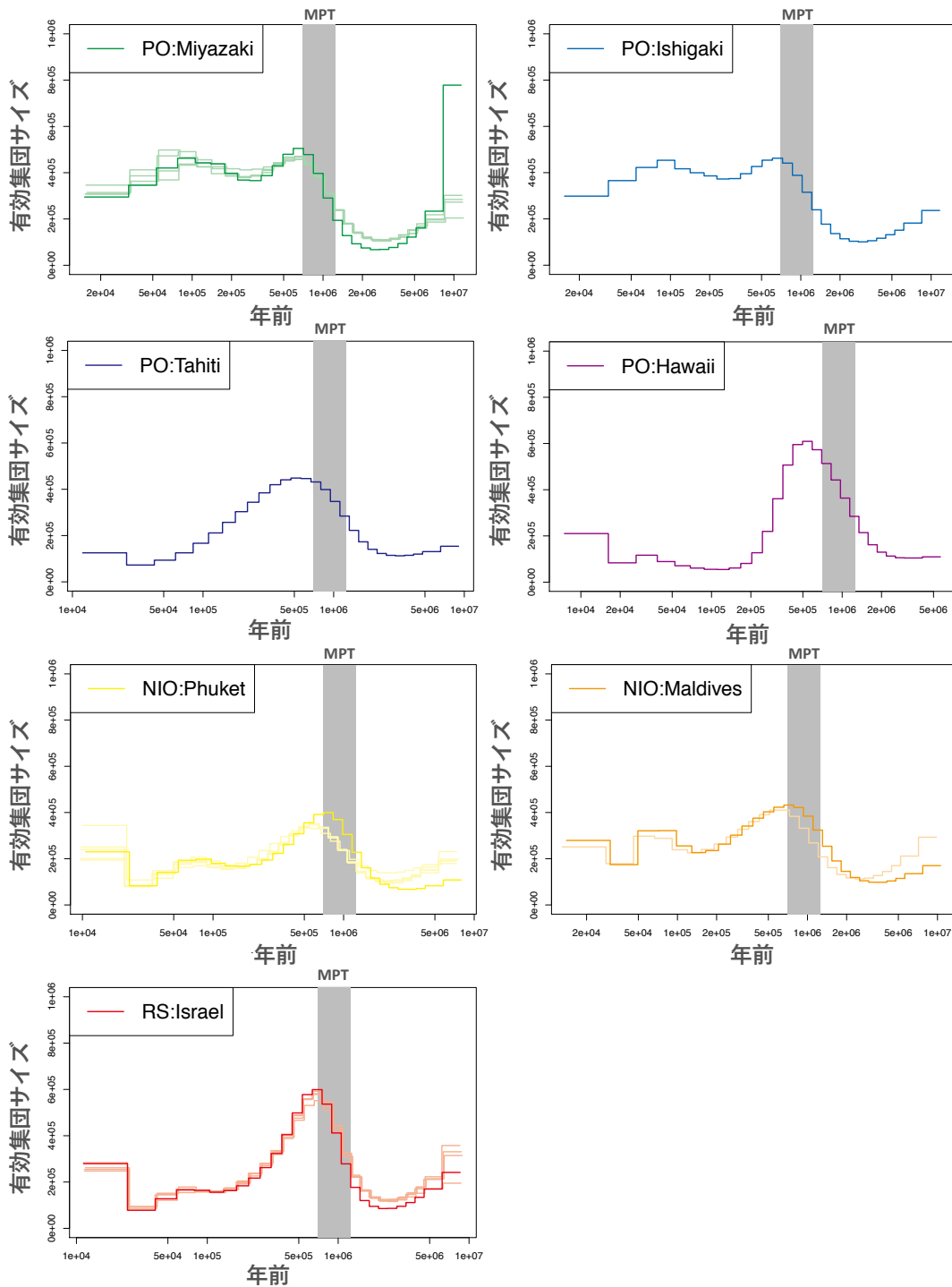


図 6-2：PSMC'モデルによる 7 海域ごとの集団史推定結果

PO は太平洋種、NIO は北インド洋種、RS は紅海種を示している。濃色線は図 6-1 で表示した個体の結果を示し、淡色線はそれ以外の個体の結果を示している。灰色塗りの箇所は Mid-Pleistocene Transition (MPT、70–125 万年前) を示している。世代間時間は 3 年、世代間変異率は太平洋種で 9.9×10^{-9} bp/世代、北インド洋種で 9.4×10^{-9} bp/世代、紅海種で 8.8×10^{-9} bp/世代として集団史を推定した。

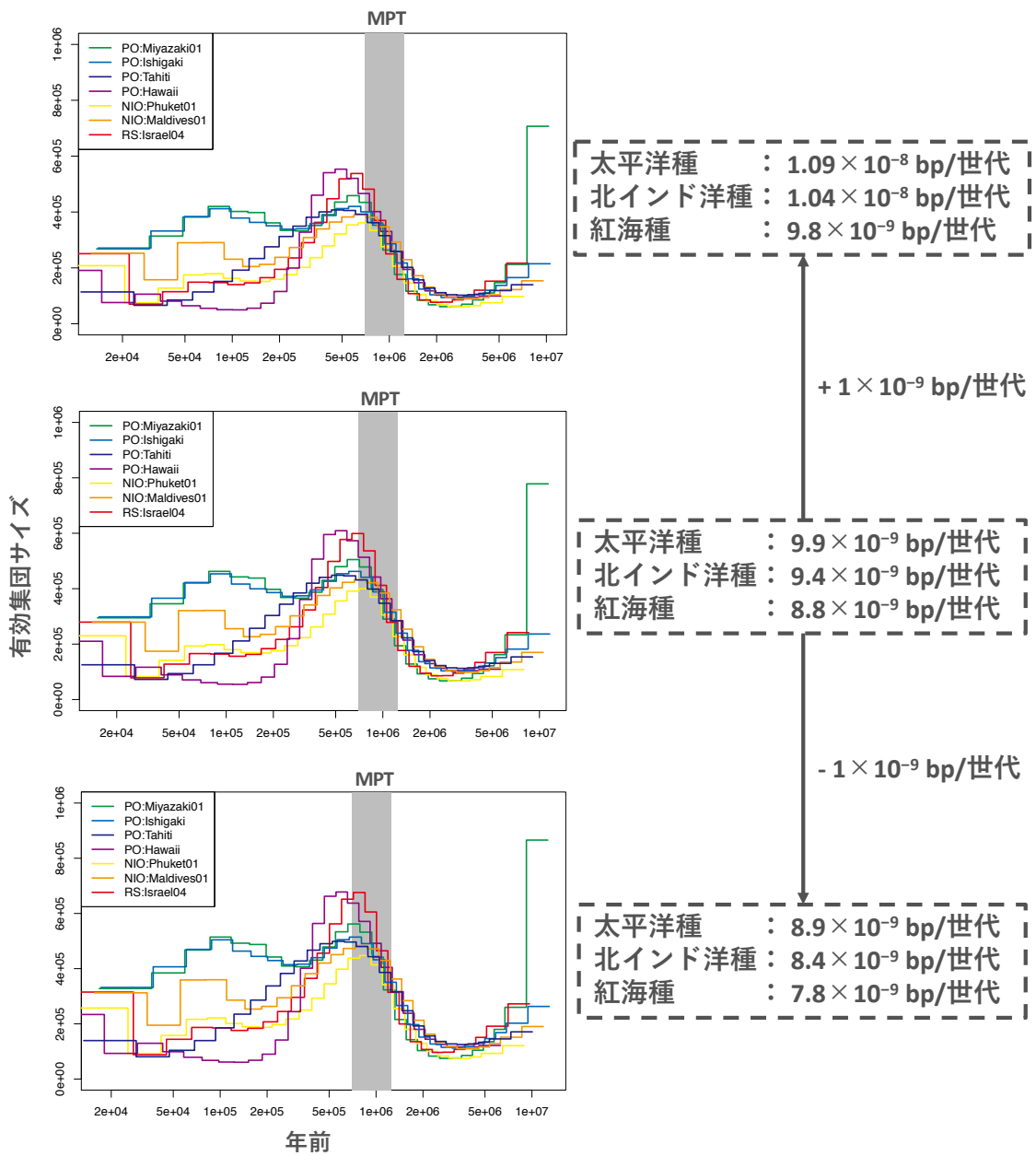


図 6-3 : 世代間変異率を変化させた時の PSMC'モデルによる集団史推定結果

PO は太平洋種、NIO は北インド洋種、RS は紅海種を示している。灰色塗りの箇所は Mid-Pleistocene Transition (MPT、70–125 万年前) を示している。図 6-1 の結果を中段に示しており、上段は世代間変異率を $+1 \times 10^{-9}$ bp/世代、下段は世代間変異率を -1×10^{-9} bp/世代にした時の集団史推定の結果を示している。

6.3.2. 遺伝構造解析

262,994 の中立な二対立 SNP を用いて遺伝構造解析を実施し、各個体群における遺伝構造を解析した。この解析ではまず推定される祖先数 (K) を設定して祖先の組成から遺伝構造の評価を行う。 $K=1-6$ の範囲の交差検証のエラー値を表 6-2 に示した。 $K=2-6$ の範囲で遺伝構造を解析した結果を図 6-4 に示した。表 6-1 の通り、 K が 1 の時に最も低い交差検証のエラー率を示しており、6.3.1 の集団史推定の結果と同様に各個体が 1 つの祖先集団から分化したことが確認できた。また、 K が 3 の時、太平洋種、北インド洋種、紅海種の個体群は明確に異なる遺伝構造を示したことから、オニヒトデ 3 種間の違いが示された。しかしながら、ハワイの遺伝構造は他の太平洋種と比較すると明確ではなく、 K が 4 の時には他の太平洋種から独立した遺伝構造を示した。これはハワイが太平洋種の中でも特異な存在であることを示しており、5.3.2. の系統解析の結果を支持した。また、 K が 5 の時にはモルディブはプーケットから独立した遺伝構造が形成され、5.3.2. の系統解析の結果と同様にミトコンドリアでは見えなかった 2 集団間の違いが核ゲノムにおいて示された。

表 6-2：遺伝構造解析における K の値を変化させた場合の交差検証 (CV) エラー値
CV エラー値が低いものほど適したモデルであることを示す。

K	CV エラー値
1	0.71236
2	0.78549
3	0.88695
4	1.11076
5	1.35808
6	1.82519

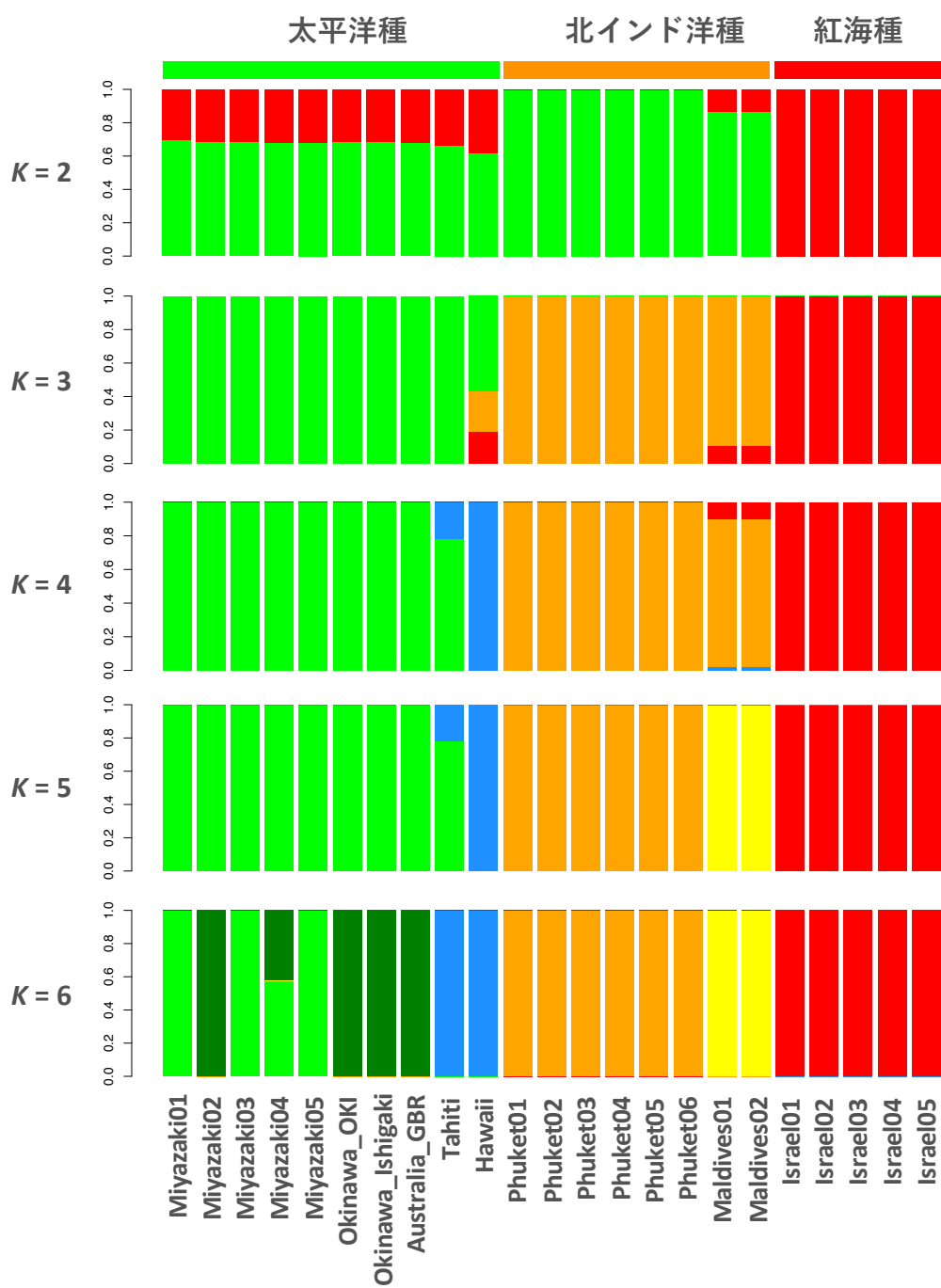


図 6-4 : 遺伝構造解析結果 (K=2-6)。

6.3.3. 種間分化領域探索

オニヒトデは異所的に種分化したという仮説が存在するものの、この仮説を具体的に検証した研究はこれまでに無い。オニヒトデの種分化が従来の仮説通り異所的なものであるのか、それとも生殖隔離によって同所的に起こった可能性も考えられるのか検証するために、核ゲノム配列を一定の長さに区切り、区画内の平均 F_{ST} を算出してその分布を確認した。もしも、オニヒトデが同所的種分化を経験していた場合、生殖隔離に関与する遺伝子が存在するゲノム領域は他のゲノム領域と比べて種間で分化が進み、極めて高い F_{ST} の値を示す強化 (Reinforcement) の痕跡が検出されるはずである (例: [88])。核ゲノムにおける 50 Kbp ごとの平均 F_{ST} の分布を図 6-7 に示した。図 6-5 で示した通り、種間で F_{ST} の値が際立って上昇する領域は存在せず、ゲノム全体で種間の F_{ST} の値が高いことが示された。このことはゲノム全体で種間の分化が進んでいたことを表している。本解析において同所的種分化で予想される強化の痕跡を見つけることができなかったが、オニヒトデは海洋生物として多くの陸上生物より有効集団サイズが大きく、オニヒトデ種間の違いには祖先集団の遺伝的多様性が反映されている (Incomplete lineage sorting) 可能性があり、ゲノム全体の分化が大きくなったことで明確な強化が検出できなかった可能性がある。また、本解析のようにドラフトゲノムで種間分化領域の探索を行なっている研究も存在するが (例: [89][90])、ゲノム配列の断片化で分化領域が分散して強化の痕跡が見えづらくなっている可能性も考えられる。以上の理由から本解析から同所的種分化で見られる強化の痕跡をゲノムから検出することができなかったが、この結果は同所的種分化の可能性を完全に否定しないことは留意すべきである。

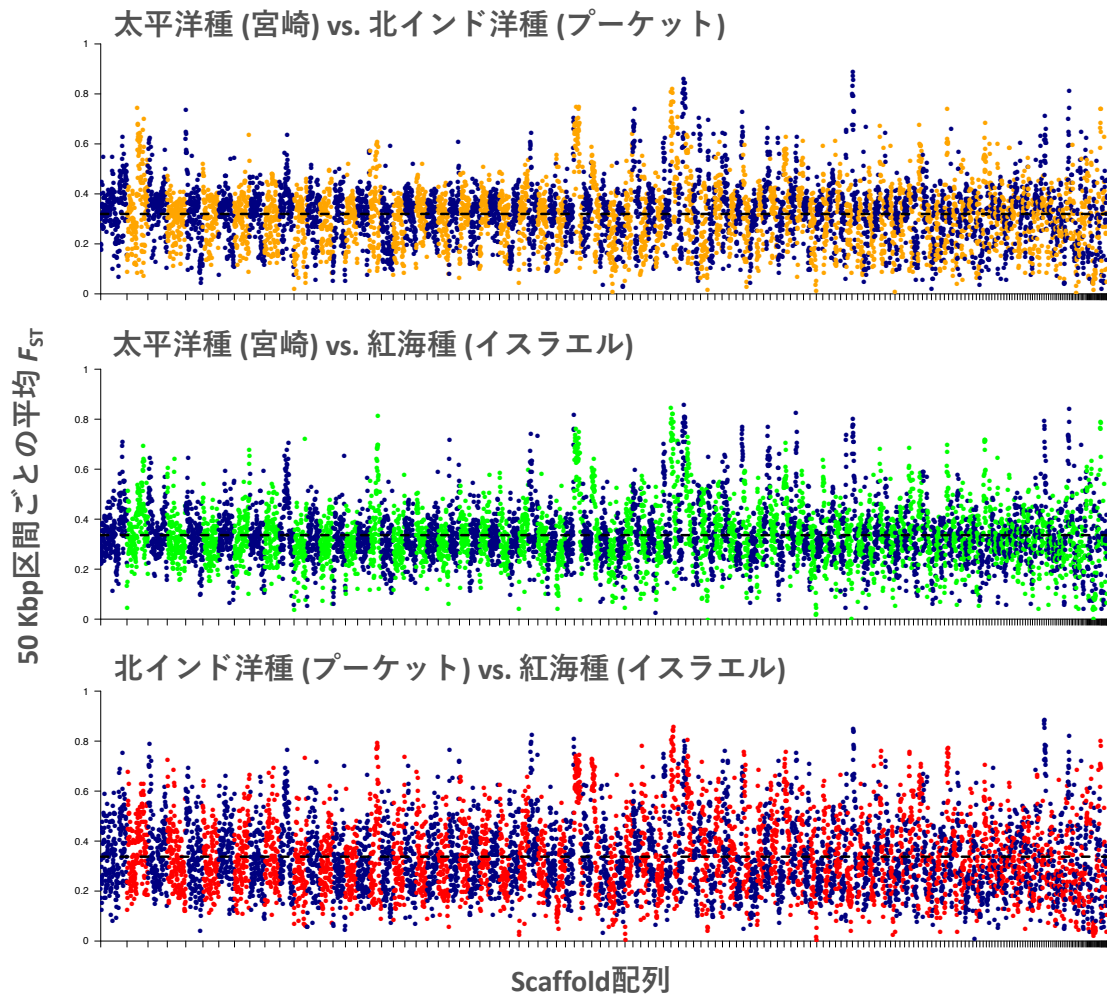


図 6-5 : 核ゲノムにおける 50 kbp ごとの種間平均 F_{ST} の分布

丸は 50 kbp の区画ごとの平均 F_{ST} を示し、破線は全区画の平均 F_{ST} を示している。

6.3.4. 種間遺伝子比較

6.3.4.において生殖隔離のような種間を積極的に分けるような痕跡はゲノムから見つからなかった。しかしながら、3種は異なる海域に生息しており、生息環境ごとに異なる適応戦略を取ってきた可能性が考えられる。種ごとに異なる適応戦略がオニヒトデ種間の違いを生み出しているという考えのもと、ゲノムよりも機能に直接的関わる遺伝子で種間比較を行った。

まず、SonicParanoid を用いて太平洋種、北インド洋種、紅海種の核ゲノムからそれぞれ予測された 19,397、18,045、18,879 の遺伝子の種間オーソログ関係

の推定を行ったところ、合計 16,652 のオーソロググループが構築された。種間オーソログ関係の内訳を表 6-3 に示した。各種 16,054–16,495 の遺伝子が 15,923–16,058 の種間オーソロググループを構築し、1,902–2,902 の遺伝子は種間でオーソログ関係を取らなかった。また、3 種間共通シングルコピーオーソログ遺伝子（3 種で 1:1:1 の対応関係が取れたオーソログ遺伝子のこと）は各種 14,203 であった。さらに、3 種間共通シングルコピーオーソログ遺伝子から種特異的な機能の変化を起こした可能性のある遺伝子を特定するために aBSREL モデルを適用して有意な正の自然選択が存在するかを調べた。3 種間共通シングルコピーオーソログ遺伝子の中で有意な正の自然選択が検出された遺伝子の種間関係を図 6-6 に示した。1 種のみで有意な正の自然選択が検出された遺伝子数は太平洋種、北インド洋種、紅海種においてそれぞれ 174、184、201 であった。

表 6-3 : SonicParanoid による種間オーソログ関係の内訳

種	太平洋種	北インド洋種	紅海種
サンプルID	Miyazaki01	Phuket01	Israel04
全遺伝子数	19,397	18,045	18,879
種間オーソログ遺伝子数	16,495	16,143	16,054
種間オーソロググループ数	16,058	15,928	15,923
3種間共通シングルコピーオーソログ遺伝子	14,203	14,203	14,203
他種とオーソログ関係の取れなかった遺伝子数	2,902	1,902	2,825

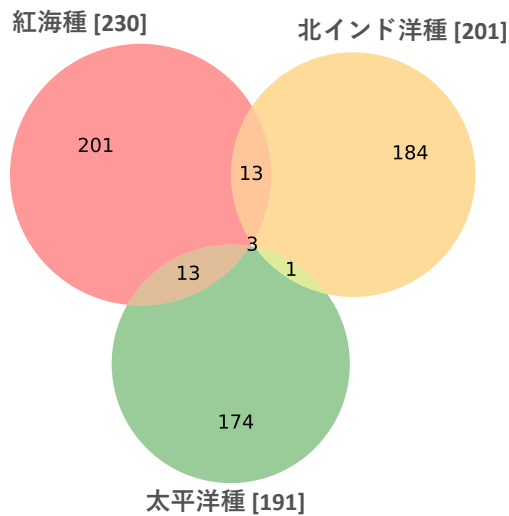


図 6-6：3 種間共通シングルコピーオーソログ遺伝子の中で有意な正の自然選択が検出された遺伝子の種間関係を示したベン図
各数値は遺伝子数を表す。

種特異的な特徴を持つ 2 グループ(種間でオーソログ関係の取れない遺伝子、1 種のみで正の自然選択が検出された遺伝子)のそれぞれを対象に GO と KEGG のエンリッチメント解析を実施し、このグループに属する遺伝子の機能的な偏りから種特異的な特徴の検出を試みた。

まず、他種オーソログ関係の取れない遺伝子グループに対して GO エンリッチメント解析を実施し、各種において有意に濃縮していた GO の機能の一覧を表 6-4 に示した。3 種に共通して GO:0017154 (semaphorin receptor activity) と GO:0071526 (semaphorin-plexin signaling pathway) の 2 つの GO に関わる機能が、他種とオーソロググループが取れない遺伝子において各種で特に濃縮していることが確認された。本研究で用いている遺伝子セットは、精度の高い遺伝子構造予測を行うために開発した自動構造アノテーションパイプラインである LEMON を使用して予測されており、BUSCO による評価も極めて高い値が示されている。しかしながら、自動構造アノテーションでは一定のエラーが含まれることは避けられず、アノテーションエラーにより他種とオーソログ関係が取れないと判定された可能性が十分に考えられるため、この結果を精査することとした。各種において GO:0017154 と GO:0071526 に割り振られたこれらの遺伝子を確認すると、相同性検索においてそのほとんどが plexin ファミリーの遺伝子配列にヒットすることが確認された。そこで、Spaln を用いて先行研究

[91]でキュレーションされていたオニヒトデの plexin ファミリーのタンパク質配列 (Hall et al. 2017[15]で報告された沖縄由来の太平洋種のオニヒトデのゲノムより構築されている) を本研究における 3 種のオニヒトデの核ゲノム配列にアライメントを行い、plexin ファミリーの遺伝子構造予測を行った。その結果、plexin B2-B8 はオニヒトデ 3 種の核ゲノム配列にそれぞれクラスター状に存在しており、先行研究の配列と低い identity を示す plexin が存在することがわかった。Spaln を用いて先行研究の太平洋種の plexin ファミリー遺伝子を各種核ゲノムにアライメントを行った時の identity を表 6-5 に示した (オプション: -Q4)。さらに、MUMmer v.3.1[92]の nucmer を用いて plexin B2-B8 がクラスター状に存在していた領域の種間アライメントを実施し (オプション: -maxmatch -nosimplify)、mummerplot を用いて描画したドットプロット図を図 6-7 に示した。図 6-7 から見て取れるように、種間に対応するゲノム領域に各種で独立した重複や挿入が見られ、この領域のシンテニーの関係が崩れていることがわかった。次に、Spaln を用いたアライメントによってアノテーションが行えた plexin ファミリーの遺伝子の中で、配列の途中で終止コドンが現れない遺伝子を対象に最尤法による系統推定を行い、その結果を図 6-8 に示した。図 6-8 から見て取れるように、図 6-7 でシンテニーの関係が崩れている領域に属する plexin B4-B8 は系統樹においても種間の対応関係が崩れていることがわかった。ゲノムの重複領域に存在する遺伝子構造を正確に予測することは困難であり、系統解析にはさらなる精査が必要であるものの、この領域におけるゲノム構造の変異は種ごとに独立して起きていると考えられ、その領域に plexin ファミリーの遺伝子が確認されたことは興味深い。plexin ファミリーの遺伝子の起源は古く、後生動物の共通祖先よりも以前から獲得していたことがわかっており、発生と生理の様々なプロセスの制御に関与していることがわかっている[91]。このことから、plexin ファミリーが種間の違いを生み出す要因になっている可能性が考えられるが、さらなる解析による検証が必要である。

続いて、1 種のみで正の自然選択が検出された遺伝子に対して KEGG エンリッチメント解析を実施した。各種において有意に濃縮していた KEGG の機能の一覧を表 6-6 に示した。KEGG エンリッチメント解析から紅海種のみで正の自然選択が検出された遺伝子において、ko04927 (cortisol synthesis and secretion) と ko00430 (taurine and hypotaurine metabolism) が濃縮していた。コルチゾール (cortisol) は硬骨魚において高塩分と低塩分の両方の環境における浸透圧調節

に關与していることが報告されており[93]、タウリン (taurine) は細胞において重要な浸透圧調節物質である[94]。紅海は他海域に比べて大きな塩分濃度変化を経験した海域であり[95]、紅海種においてコルチゾールやタウリンに關わる遺伝子群に特徴が見られたことは紅海の特異な塩分環境への適応と關連している可能性がある。また、太平洋種では甲状腺ホルモン (thyroid hormone) に關連した ko4918 (thyroid hormone synthesis) と ko4919 (thyroid hormone signaling pathway) が濃縮していた。甲状腺ホルモンは成長の調節に關わるホルモンである[96]。太平洋におけるオニヒトデの大量発生は他の海域よりも大規模でより頻繁に起こっているという意見があり[11][16]、そのことを踏まえると太平洋種の甲状腺ホルモン關連遺伝子に違いが見られたことは太平洋種の成長率に關わっている可能性がある。遺伝子の機能解析などは今後の課題となるが、本研究から太平洋種と紅海種のゲノムにおいて特徴的な遺伝子を見つけ出したことで、これまで情報が不足しているオニヒトデ種間の差異の解明に繋がる知見を出すことができた。

表 6-4：種間でオーソログ関係の取れない遺伝子グループを対象にした GO エンリッチメント解析を実施した結果 (P -value<0.05)

太平洋種：Miyazaki01			
GO	前景遺伝子数	背景遺伝子数	P -value
GO:0015074 DNA integration	33	89	2.55E-07
GO:0017154 semaphorin receptor activity	8	19	0.004019507
GO:0071526 semaphorin-plexin signaling pathway	8	19	0.004019507
GO:0017040 N-acylsphingosine amidohydrolase activity	4	6	0.005828934
GO:0046514 ceramide catabolic process	4	6	0.005828934
GO:0007224 smoothed signaling pathway	6	15	0.016614653
GO:0006958 complement activation, classical pathway	10	31	0.018693914
GO:0004713 protein tyrosine kinase activity	25	108	0.02150095
GO:0004714 transmembrane receptor protein tyrosine kinase activity	6	16	0.023281285
GO:0007169 transmembrane receptor protein tyrosine kinase signaling pathway	6	16	0.023281285
GO:0019346 transsulfuration	3	5	0.026427384
GO:0016175 superoxide-generating NAD(P)H oxidase activity	5	13	0.033838111
GO:0042554 superoxide anion generation	5	13	0.033838111
GO:0004180 carboxypeptidase activity	6	18	0.041464737
GO:0004715 non-membrane spanning protein tyrosine kinase activity	4	10	0.049574973
北インド洋種：Phuket01			
GO	前景遺伝子数	背景遺伝子数	P -value
GO:0017154 semaphorin receptor activity	15	27	1.08E-08
GO:0071526 semaphorin-plexin signaling pathway	15	27	1.08E-08
GO:0004713 protein tyrosine kinase activity	26	98	7.00E-06
GO:0015074 DNA integration	17	68	0.000551497
GO:0005200 structural constituent of cytoskeleton	9	32	0.004687204
GO:0016758 transferase activity, transferring hexosyl groups	9	35	0.008839203
GO:0005874 microtubule	9	38	0.015309412
GO:0004965 G protein-coupled GABA receptor activity	10	45	0.023545926
GO:0004619 phosphoglycerate mutase activity	2	3	0.030988301
GO:0005078 MAP-kinase scaffold activity	2	3	0.030988301
GO:0006007 glucose catabolic process	2	3	0.030988301
GO:0019432 triglyceride biosynthetic process	2	3	0.030988301
GO:0033227 dsRNA transport	2	3	0.030988301
GO:0051033 RNA transmembrane transporter activity	2	3	0.030988301
GO:1990112 RQC complex	2	3	0.030988301
GO:1990116 ribosome-associated ubiquitin-dependent protein catabolic process	2	3	0.030988301
GO:0007017 microtubule-based process	9	42	0.037811204
紅海種：Israel04			
GO	前景遺伝子数	背景遺伝子数	P -value
GO:0017154 semaphorin receptor activity	17	27	1.74E-08
GO:0071526 semaphorin-plexin signaling pathway	17	27	1.74E-08
GO:0015074 DNA integration	22	53	2.81E-06
GO:0016758 transferase activity, transferring hexosyl groups	14	39	0.001033426
GO:0005200 structural constituent of cytoskeleton	10	26	0.00296427
GO:0035198 miRNA binding	3	3	0.003350736
GO:0006030 chitin metabolic process	5	10	0.009774307
GO:0005176 ErbB-2 class receptor binding	5	11	0.015734655
GO:0005785 signal recognition particle receptor complex	3	4	0.011899642
GO:0010469 regulation of signaling receptor activity	5	11	0.015734655
GO:0030197 extracellular matrix constituent, lubricant activity	5	11	0.015734655
GO:0005874 microtubule	10	33	0.023915385
GO:0005047 signal recognition particle binding	3	5	0.026440347
GO:0008061 chitin binding	5	14	0.046344654

表 6-5 : Spaln を用いて先行研究[91]でキュレートされたオニヒトデ太平洋種の plexin ファミリー遺伝子を各種核ゲノムにアライメントを行った時の identity

太平洋種 : Mlyazaki01						北インド洋種 : Phuket01					
Reference protein ID	Scaffold ID	Protein length (aa)	Identity (%)	Coverage (%)	# exons	Reference protein ID	Scaffold ID	Protein length (aa)	Identity (%)	Coverage (%)	# exons
Aplan_Met-LP1	scaffold_12	1,677	98.3	100.0	27	Aplan_Met-LP1	scaffold_3562	1,677	98.4	100.0	27
Aplan_Met-LP2	scaffold_112	1,410	92.5	100.0	26	Aplan_Met-LP2	scaffold_3465	1,410	82.1	99.4	31
Aplan_Met-LP3	scaffold_647	1,560	88.3	99.9	39	Aplan_Met-LP3	scaffold_20	1,560	86.4	99.6	27
Aplan_Met-LP4	scaffold_647	1,376	89.6	100.0	23	Aplan_Met-LP4	scaffold_1337	1,376	90.6	99.5	23
Aplan_Met	scaffold_128	1,410	99.8	100.0	19	Aplan_Met	scaffold_173	1,410	98.9	100.0	19
Aplan_Plexin-A1	scaffold_453	1,909	99.0	100.0	31	Aplan_Plexin-A1	scaffold_3562	1,909	99.1	100.0	31
Aplan_Plexin-A2	scaffold_112	1,523	96.5	100.0	31	Aplan_Plexin-A2	scaffold_107	1,523	96.3	100.0	30
Aplan_Plexin-A3	scaffold_112	131	99.2	100.0	2	Aplan_Plexin-A3	scaffold_107	131	94.7	100.0	2
Aplan_Plexin-B1	scaffold_12	1,887	99.5	100.0	36	Aplan_Plexin-B1	scaffold_3562	1,887	95.3	100.0	36
Aplan_Plexin-B2	scaffold_647	1,946	98.0	100.0	36	Aplan_Plexin-B2	scaffold_20	1,946	96.1	99.7	36
Aplan_Plexin-B3	scaffold_647	909	88.0	99.8	21	Aplan_Plexin-B3	scaffold_20	909	89.0	99.9	23
Aplan_Plexin-B4	scaffold_647	1,682	89.9	99.8	34	Aplan_Plexin-B4	scaffold_20	1,682	86.4	95.7	32
Aplan_Plexin-B5	scaffold_647	978	87.3	99.9	18	Aplan_Plexin-B5	scaffold_20	978	50.2	59.8	16
Aplan_Plexin-B6	scaffold_647	1,749	91.5	99.7	37	Aplan_Plexin-B6	scaffold_20		91.4	99.8	41
Aplan_Plexin-B7	scaffold_647	1,884	99.3	100.0	35	Aplan_Plexin-B7	scaffold_20	1,884	74.5	98.6	38
Aplan_Plexin-B8	scaffold_647	1,940	89.8	99.4	37	Aplan_Plexin-B8	scaffold_20	1,940	89.0	99.2	37
Aplan_Sema5A	scaffold_238	1,015	99.7	100.0	24	Aplan_Sema5A	scaffold_119	1,015	99.5	100.0	24
Aplan_Sema6A	scaffold_868	872	99.4	100.0	20	Aplan_Sema6A	scaffold_348	872	98.6	100.0	20
Aplan_Sema-S11	scaffold_4477	1,031	99.0	100.0	22	Aplan_Sema-S11	scaffold_348	1,031	97.7	100.0	23
Aplan_Sema-S12	scaffold_4460	711	98.5	100.0	19	Aplan_Sema-S12	scaffold_348	711	94.9	100.0	19
Aplan_Sema-S13	scaffold_4477	710	97.7	100.0	18	Aplan_Sema-S13	scaffold_348	710	95.8	100.0	18
Aplan_Sema-S14	scaffold_4477	714	97.9	100.0	18	Aplan_Sema-S14	scaffold_348	714	97.3	100.0	18
Aplan_Sema-SP1	scaffold_116	720	99.3	100.0	17	Aplan_Sema-SP1	scaffold_193	720	98.2	100.0	17
Aplan_Sema-SP2	scaffold_394	731	99.3	100.0	17	Aplan_Sema-SP2	scaffold_150	731	99.0	100.0	17
Aplan_Sema-SP3	scaffold_62	482	89.2	99.0	15	Aplan_Sema-SP3	scaffold_146	482	90.2	100.0	16
Aplan_Sema-SP4	scaffold_62	712	98.2	100.0	17	Aplan_Sema-SP4	scaffold_146	712	94.7	100.0	17
Aplan_Sema-SP5	scaffold_62	712	98.2	100.0	17	Aplan_Sema-SP5	scaffold_146	712	96.3	100.0	17
Aplan_Sema-SP6	scaffold_62	225	96.4	98.7	5	Aplan_Sema-SP6	scaffold_146	225	95.6	98.7	5
紅海種 : Israel04											
Reference protein ID	Scaffold ID	Protein length (aa)	Identity (%)	Coverage (%)	# exons						
Aplan_Met-LP1	scaffold_247	1,677	85.0	99.9	28						
Aplan_Met-LP2	scaffold_6	1,410	81.5	99.5	31						
Aplan_Met-LP3	scaffold_127	1,560	85.2	99.9	36						
Aplan_Met-LP4	scaffold_127	1,376	89.2	100.0	24						
Aplan_Met	scaffold_154	1,410	98.7	99.8	19						
Aplan_Plexin-A1	scaffold_313	1,909	98.4	100.0	31						
Aplan_Plexin-A2	scaffold_90	1,523	96.5	100.0	30						
Aplan_Plexin-A3	scaffold_90	131	95.4	100.0	2						
Aplan_Plexin-B1	scaffold_247	1,887	98.3	100.0	36						
Aplan_Plexin-B2	scaffold_127	1,946	94.7	100.0	37						
Aplan_Plexin-B3	scaffold_127	909	87.7	99.9	23						
Aplan_Plexin-B4	scaffold_127	1,682	87.3	99.3	37						
Aplan_Plexin-B5	scaffold_127	978	48.9	58.8	14						
Aplan_Plexin-B6	scaffold_127	1,749	94.9	100.0	40						
Aplan_Plexin-B7	scaffold_127	1,884	55.7	98.8	92						
Aplan_Plexin-B8	scaffold_127	1,940	88.0	99.7	37						
Aplan_Sema5A	scaffold_141	1,015	99.2	100.0	24						
Aplan_Sema6A	scaffold_438	872	98.6	100.0	20						
Aplan_Sema-S11	scaffold_408	1,031	95.3	100.0	22						
Aplan_Sema-S12	scaffold_153	711	95.2	99.7	19						
Aplan_Sema-S13	scaffold_112	710	95.5	100.0	18						
Aplan_Sema-S14	scaffold_112	714	96.6	100.0	18						
Aplan_Sema-SP1	scaffold_129	720	97.5	100.0	17						
Aplan_Sema-SP2	scaffold_75	731	96.3	100.0	17						
Aplan_Sema-SP3	scaffold_211	482	90.5	100.0	13						
Aplan_Sema-SP4	scaffold_211	712	95.5	100.0	17						
Aplan_Sema-SP5	scaffold_211	712	95.5	100.0	17						
Aplan_Sema-SP6	scaffold_211	225	97.3	100.0	5						

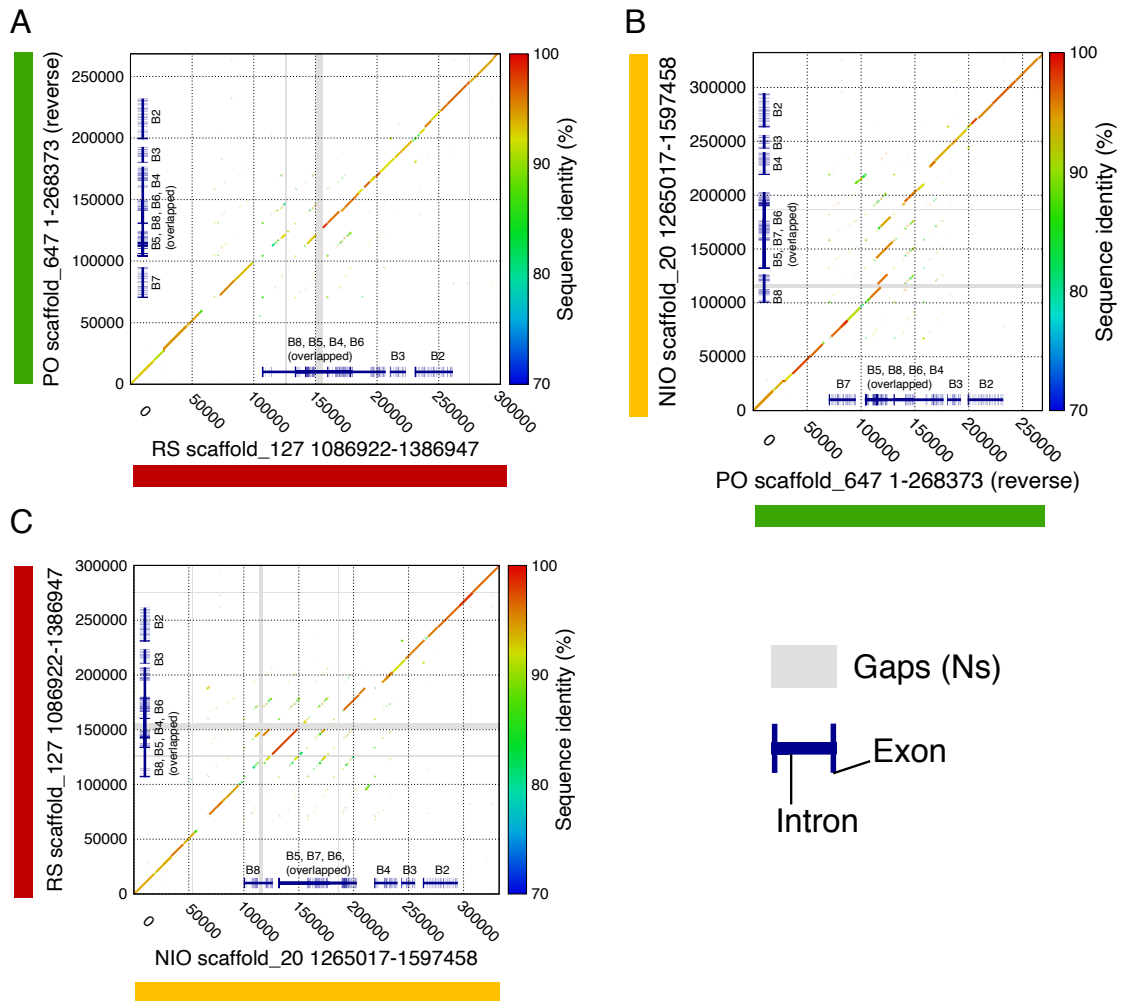


図 6-7 : Plexin B2~B8 がクラスター状に存在していた領域の種間アライメント結果

A. 太平洋種 vs. 紅海種 B. 北インド洋種 vs. 太平洋種 C. 紅海種 vs. 北インド洋種の結果を表示しており、図中の PO は太平洋種、NIO は北インド洋種、RS は紅海種を示す。

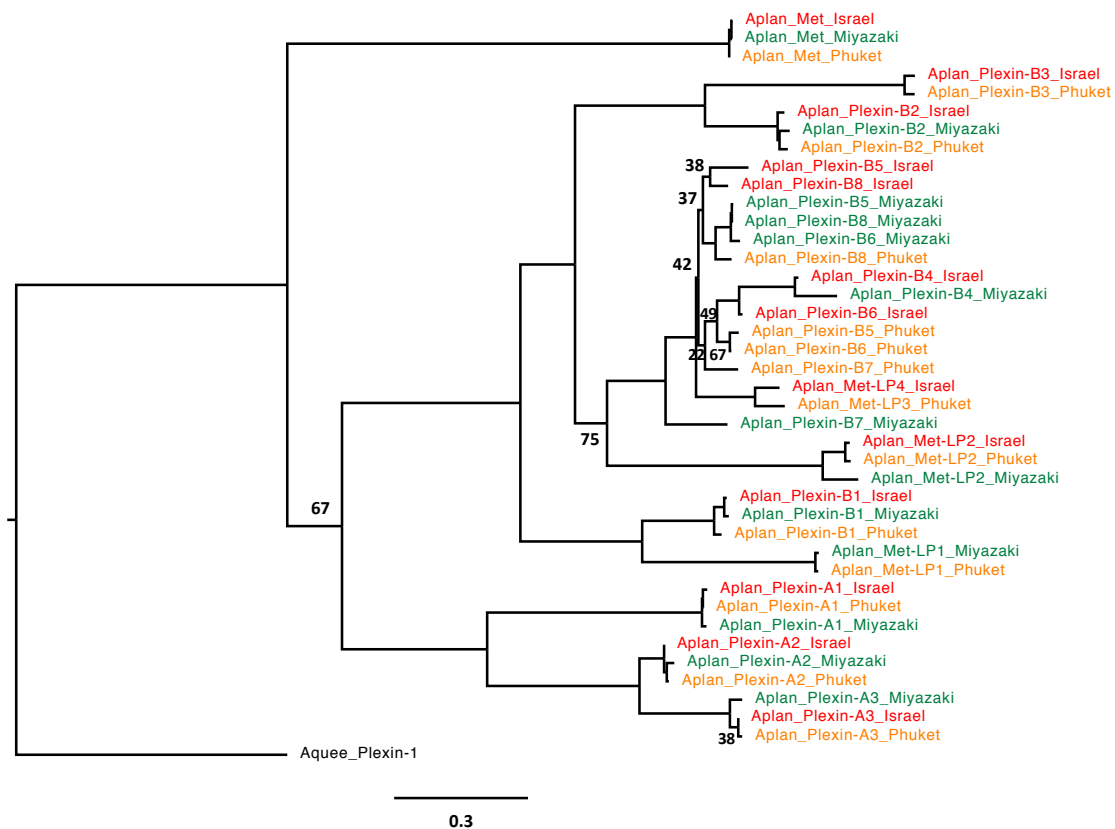


図 6-8 : 3 種オニヒトデの plexin ファミリー遺伝子の系統樹

Spaln を用いたアライメントによって先行研究[76]の plexin ファミリー遺伝子から 3 種オニヒトデの核ゲノムにアノテーションが行えた遺伝子の中で、配列の途中で終止コドンが現れない遺伝子を元に最尤法で推定された系統樹である。アミノ酸配列のアライメント結果から TrimAl を用いてアライメントが貧弱なサイトを取り除き (オプション: -automated1)、IQ-TREE を用いて最尤法で系統推定を実施した (オプション: -m MFP -bb 1000)。最尤法による系統推定の際の置換モデルは BIC から最適モデルと判断された WAG+R3 を使用した。系統樹の緑色は太平洋種 (Miyazaki01)、橙色は北インド洋種 (Phuket01)、赤色は紅海種 (Israel04) の核ゲノムにアノテーションされた plexin ファミリー遺伝子を示している。*Amphimedon queenslandica* (カイメンの一種) の plaxin-1 をアウトグループとして加えた。各節の値は bootstrap 回数が 1,000 回で系統推定を行った時 bootstrap 値であり、80%未満の値のみ表示している。

表 6-6 : 1 種のみで正の自然選択が検出された遺伝子グループを対象にした KEGG エントリー解析を実施した結果

P -value<0.05 の有意に濃縮していた KEGG Ontology の機能のみを表示している。

太平洋種 : Miyazaki01			
KEGG Orthologyの機能	前景遺伝子数	背景遺伝子数	P -value
Thyroid hormone synthesis [PATH:ko04918]	3	42	0.007464286
Biofilm formation - Escherichia coli [PATH:ko02026]	1	1	0.009538541
Wnt signaling pathway [PATH:ko04310]	4	90	0.010845296
Acute myeloid leukemia [PATH:ko05221]	2	20	0.015362301
Insulin resistance [PATH:ko04931]	3	60	0.019646382
Protein export [PATH:ko03060]	2	26	0.025316182
Cell adhesion molecules [PATH:ko04514]	2	28	0.029082236
Thyroid hormone signaling pathway [PATH:ko04919]	3	73	0.032661004
Vitamin B6 metabolism [PATH:ko00750]	1	4	0.037614592
Prostate cancer [PATH:ko05215]	3	79	0.039859872
Transcriptional misregulation in cancer [PATH:ko05202]	3	86	0.049184895
北インド洋種 : Phuket01			
KEGG Orthologyの機能	前景遺伝子数	背景遺伝子数	P -value
Central carbon metabolism in cancer [PATH:ko05230]	4	72	0.007555354
NF-kappa B signaling pathway [PATH:ko04064]	3	39	0.008422085
Phenazine biosynthesis [PATH:ko00405]	1	1	0.010752092
Herpes simplex virus 1 infection [PATH:ko05168]	5	142	0.018774231
Human cytomegalovirus infection [PATH:ko05163]	4	106	0.027586463
Apoptosis - multiple species [PATH:ko04215]	2	25	0.029344688
Human immunodeficiency virus 1 infection [PATH:ko05170]	4	114	0.03471585
Other glycan degradation [PATH:ko00511]	2	28	0.036208969
Viral myocarditis [PATH:ko05416]	2	28	0.036208969
VEGF signaling pathway [PATH:ko04370]	2	28	0.036208969
Prostate cancer [PATH:ko05215]	3	68	0.03686969
Hepatitis C [PATH:ko05160]	3	68	0.03686969
RIG-I-like receptor signaling pathway [PATH:ko04622]	2	31	0.043623268
紅海種 : Israel04			
KEGG Orthologyの機能	前景遺伝子数	背景遺伝子数	P -value
Cortisol synthesis and secretion [PATH:ko04927]	3	47	0.015964219
Protein processing in endoplasmic reticulum [PATH:ko04141]	5	130	0.016071489
Taurine and hypotaurine metabolism [PATH:ko00430]	2	18	0.017219704
Amino acid related enzymes [BR:ko01007]	3	57	0.02655053
Glycosaminoglycan degradation [PATH:ko00531]	2	29	0.042151184
Butanoate metabolism [PATH:ko00650]	2	30	0.044833728

最後に、3種の自然免疫関連遺伝子に焦点を当てて解析を行った。自然免疫系に関わる遺伝子は病原体の感染の影響によって進化速度が早いことが報告されており[97]、各種で生息海域の病原体に対応するために免疫系を独自に進化させた可能性が考えられる。このような免疫関連の適応が起こっているならば、種間でオーソログ関係の取れない遺伝子や、1種のみで正の自然選択が検出された遺伝子のような種特異的な特徴を持つ遺伝子の数に違いとして現れる可能性がある。そのため、これらの遺伝子を対象に自然免疫関連のKEGGのtermについて違いが見られないか確認を行い、結果を表6-7に示した。太平洋種において他の2種よりも自然免疫に関連するko00512 (mucin-type O-glycan biosynthesis)、ko04091 (Lectins)、ko04054 (pattern recognition receptor)のtermが特異的な特徴を持つ遺伝子において有意に多く見られることがわかった(有意水準は5%)。O-グリカン (O-glycan) はムチン (mucin) のコアタンパク質であり、ムチンは微生物からの物理的な感染防御や特定の菌の増殖抑制など免疫の役割を持っている[98]。レクチン (Lectin) は糖鎖結合タンパク質で、グラム陰性菌やグラム陽性菌の細胞壁などの特定の糖鎖構造を認識して結合することから、抗体産生系が欠如している無脊椎動物においてレクチンが感染防御の一役を担っている[99]。また、レクチンやToll-like receptorなどはパターン認識受容体 (pattern recognition receptor) とも呼ばれ、病原体を非自己と検知することで免疫に寄与している[100]。これらに加えて、1種のみで遺伝子数の拡大が見られたオーソロググループに注目して、その遺伝子の機能アノテーションの確認も行った。表6-8に示した通り、有意な差ではないが太平洋種においてTripartite motif-containing protein (TRIM) 遺伝子ファミリーが他2種よりも拡大していることが確認された。このオーソロググループはSwissProtとKEGGでの機能アノテーション結果からTRIM2またはTRIM56に近い機能を有していると考えられるが、それぞれ抗ウイルスと自然免疫制御の役割を果たしていることが報告されており[101][102]、自然免疫に関わる遺伝子であると考えられる。では、なぜ太平洋種において自然免疫関連の遺伝子に特徴が見られたのか？それは、太平洋における多様な病原菌の感染リスクに対応するためであったと考えられる。その理由は、サンゴ礁海域の中でも太平洋は特に生物多様性が高い海域であるからだ[2]。自由生活形の細菌の生物多様性はその環境中に生息する動植物の生物多様性と類似した傾向を示すことが報告されており[103]、生物多様性の高い太平洋のサンゴ礁海域に生息する生物は他

海域よりも多種多様な病原体からの感染リスクに晒されている可能性がある。太平洋での高い感染リスクが太平洋種の自然免疫関連の遺伝子を変化させた可能性が考えられるが、この仮説を立証するにはインド洋と太平洋に横断して生息する他のサンゴ礁生物近縁種間のゲノムでも同様の特徴が見られるのか確認する必要がある。

表 6-7：種間でオーソログ関係の取れない遺伝子と 1 種のみで正の自然選択が検出される遺伝子の中で自然免疫関連の KEGG Orthology に属する各種での遺伝子数
赤字は 1 種で有意に多く見られた KEGG Orthology を示している (P -value<0.05)。

KEGG Orthologyの機能	最多種	太平洋種 Miyazaki01	北インド洋 種 Phuket01	紅海種 Israel04	P -value
Mucin type O-glycan biosynthesis [PATH:ko00512]	太平洋種	5	0	0	0.00598
Pattern recognition receptors [BR:ko04054]	太平洋種	7	2	1	0.03251
Lectins [BR:ko04091]	太平洋種	13	4	4	0.01453
C-type lectin receptor signaling pathway [PATH:ko04625]	紅海種		3	3	6 0.22220

表 6-8 : 1 種で遺伝子数が拡大していたオーソロググループ

最多種	太平洋種 Miyazaki01	北インド洋種 Phuket01	紅海種 Israel04	UniProt-SwissProt	KEGG (K number)	P-value
太平洋種	5	1	0			0.02072259
太平洋種	5	1	0			0.02072259
太平洋種	10	3	4			0.04184668
太平洋種	4	0	1	• Retrovirus-related Pol polyprotein from transposon 17.6		0.05095503
太平洋種	4	1	0			0.05095503
太平洋種	5	1	1	• Tripartite motif- containing protein 2	• K12026 TRIM56; tripartite motif-containing protein 56 [EC:2.3.2.27]	0.05195571
太平洋種	5	1	1	• Probable RNA-directed DNA polymerase from transposon BS • Probable RNA-directed DNA polymerase from transposon X-element	• K00986 ItrA; RNA- directed DNA polymerase [EC:2.7.7.49]	0.05195571
太平洋種	5	1	1			0.05195571
太平洋種	7	1	3	• Rho-related protein racA	• K07868 RHOBTB1_2; Rho-related BTB domain- containing protein 1/2	0.05562321
北インド洋種	1	5	2			0.12080902
太平洋種	5	2	1			0.13385613
太平洋種	4	1	1			0.19081059
太平洋種	4	1	1		• K16817 PLA2G16; HRAS- like suppressor 3 [EC:3.1.1.32 3.1.1.4]	0.19081059
太平洋種	4	1	1			0.19081059
太平洋種i	5	2	2	• Retrovirus-related Pol polyprotein from transposon 17.6 • Transposon Ty3-I Gag- Pol polyprotein	• K07497; putative transposase	0.29052882

6.4. まとめ

集団史推定において MPT 以前の 3 種の集団史のパターンは同調していたこと、遺伝構造解析からインド-太平洋スケールで遺伝構造を評価すると推定祖先数が 1 つの時に最適なモデルになることから、3 種は共通祖先集団から分化したことが確認できた。また、遺伝構造解析による各個体の遺伝構造から 3 種の遺伝学的な違い、太平洋種内におけるハワイの特異さ、北インド洋種内における集団間の違いが示され、系統解析の結果を支持する結果が得られた。集団史推定から MPT の時期に各海域で独立した集団が形成されたことから MPT での劇的な海水面の低下による集団間の物理的な交流阻害がオニヒトデ 3 種の種分化に影響したことが強く示唆された。種間分化領域を探索するとゲノム全体で種間が分化していたことも上記の仮説を支持している。また、ゲノムスケールの比較では極端に種間で分化した領域を特定できなかったが、遺伝子スケールで種間比較を行うと太平洋種と紅海種のゲノムにおいて地域適応の痕跡が見つかった。さらに、3 種で独立して形成されたと思われる構造変異領域が確認され、その領域には *plexin* ファミリーの遺伝子が存在した。これらがオニヒトデ種間の違いに繋がっている可能性が考えられるが、さらなる解析から検証が必要である。

第7章 結論

オニヒトデ種間のゲノムの違いと種分化要因を解明することを目的として、本研究において初めてオニヒトデの北インド洋種と紅海種の核ゲノムのドラフトゲノム配列およびミトコドリアゲノムの完全長配列を構築し、太平洋種、北インド洋種、紅海種のオニヒトデ3種間で比較ゲノム解析を実施した。

まず、核を含めたゲノムにおけるオニヒトデ3種間の違いは配列類似度算出、系統解析、遺伝構造解析の3つの解析結果から示された。1つ目の配列類似度算出では、オニヒトデ3種間のミトコドリアと核のゲノム配列の類似度はそれぞれ92%と98%であり、これは既知の別種間であるヒト-チンパンジーに相当する類似度であった。2つ目の系統解析ではミトコドリアゲノムと核ゲノムの両方で種毎に明確に異なるクレードを形成していた。さらに、3つ目の遺伝構造解析では現在の分類に基づいて祖先数(K)を3と仮定すると3種間で異なる遺伝構造を示した。これら3つの結果から、ミトコドリアだけでなく核においても太平洋種、北インド洋種、紅海種のオニヒトデ3種間に明確な違いが存在することが示された。しかしながら、種内に注目すると2つの新たな特徴が明らかとなった。1つ目は北インド洋種のモルディブである。ミトコドリアに基づく先行研究[66]と本研究におけるミトコドリアゲノムの系統解析では、モルディブは北インド洋種他集団と遺伝的に区別が付かなかったが、核ゲノムによる系統解析と遺伝構造解析ではモルディブの違いが明確となった。このことから、モルディブ集団は先行研究[66]から想定されているほど他集団と強い結びつきはない可能性が考えられた。2つ目は太平洋種のハワイである。ハワイは核ゲノムでの系統解析においてミトコドリアから想定されるよりも他の太平洋種から分化している傾向が確認された。また、遺伝構造解析では K を3と仮定した時に他の太平洋種に比べて明確な太平洋種の遺伝構造を示さず、 K を4と仮定したときには他の太平洋種から独立した遺伝構造を示した。これらの結果に加えて、アロザイムを用いた先行研究においても太平洋集団の中でハワイ集団が分化していたことが示されており[10]、ハワイ集団は太平洋種の亜種または隠蔽種である可能性が考えられた。しかしながら、本研究ではハワイは1個体しか解析に含めることができず、ハワイやその周辺海域を含めた東太平洋からより多くのサンプルを収集してさらなるゲノム解析を行う必要がある。

次に、オニヒトデ種間の種分化要因の解明を試みた。集団史推定において、各

海域のオニヒトデの有効集団サイズの増減がMPT (70万–125万年前) 以前においては同調したパターンを示しており、各海域の個体が共通した祖先集団に属していたことを示した。このことは、 $K=1$ の時が最適なモデルであることを示した遺伝構造解析結果からも確認することができている。また、MPTの期間から各海域で異なった有効集団サイズの傾向が形成され始めており、この時期に独立した集団が形成され始めたことが推定される。独立した集団が形成された可能性のあるMPTは氷河作用による海水面の変動が激しく、数百万年ぶりに海水面が120mも下降した時期であるため[84]、この時の大陸棚の干出による海域間の分断が各集団の地理的隔離を引き起こした可能性は十分に考えられる。MPTの急激な海水面の低下がオニヒトデ3種の共通祖先集団を物理的に分断したことで異所的種分化に影響した可能性が高いことが示された。

上記の種間分化領域探索において極端に種間で分化したゲノム領域を検出することはできなかったが、知見の不足しているオニヒトデ3種間の違いをより詳細に調べるために、これまでの解析に加えて遺伝子スケールでの比較解析も実施した。その結果、太平洋種と紅海種で地域適応の痕跡と思われる特徴を抽出することに成功し、さらに3種で独自に起こったと考えられる構造変異がplexinファミリーの遺伝子がクラスター状に存在する領域で起こっていることも発見した。これらがオニヒトデ種間の違いを生み出している可能性が考えられるが、より詳細な解析から確認を行う必要がある。

終わりに、本研究の意義について述べる。本研究において北インド洋種と紅海種のゲノム配列を構築したことで、太平洋種以外の2種のオニヒトデ研究に対してゲノム基盤を提供することができた。これまでのオニヒトデ研究のほとんどは太平洋で行われており[11]、太平洋種以外のオニヒトデの知見は不足している。実際に本研究以前にはオニヒトデのゲノムは核とミトコンドリアの両方も太平洋種でしか構築されておらず[14][15]、ゲノムを用いた研究もこれまで太平洋のみで行われている(例:[103][104][105])。本研究によって北インド洋種と紅海種のゲノム配列を構築したことで、太平洋種以外でゲノムを用いたオニヒトデ研究が活性化されることが期待されるため、本研究は太平洋種に偏ったオニヒトデ研究の改善に寄与できると考えられる。本研究によってオニヒトデ3種のゲノム情報が公開されたことになるが、南インド洋種に関してはサンプルを入手することができず、解析に含めることができなかった。オニヒトデ種内と種間の違いをより明確するには南インド洋種を加えてオニヒトデのゲノム

解析を引き続き行う必要がある。また、サンゴ礁生物の遺伝学的な研究の中で、本研究のようにインド洋と太平洋の広域からサンプルを収集し、近縁種間をゲノムから比較した例はこれまでに存在しない。複数サンプルを解析に用いる集団遺伝学的な研究は、全ゲノムシーケンスよりも RADseq[106]や MIG-seq[107]のようなゲノムの部分的な領域を広くシーケンスする比較的安価な手法が用いられることが多い。しかしながら、RADseq や MIG-seq で得られる SNP は数百から数千程度であるのに対して（例：[108][109]）、ゲノム全体を用いた本研究では 262,994–18,535,994 もの SNP を解析に加えることができている。解析対象の生物種のゲノムサイズやヘテロ接合度などによって得られる SNP の数は異なるが、全ゲノムを用いることでより多くのデータが得られることから RADseq や MIG-seq よりも解像度高く解析が行うことができると考えられる。実際にサンゴでは全ゲノムを用いた解析によってこれまでに見えなかった同種集団間の違いを捉えることに成功している研究例もある[4]。また、全ゲノムシーケンスを行うことで、本研究において実施した種間分化領域探索のようなゲノム全体を網羅した解析や PSMC'モデルを用いた集団史推定など解析の幅を広げることができる。さらに、各種近縁種間のゲノムを構築して比較解析を行うことで、種独自の遺伝子の特徴や構造変異なども解析対象に加えることができるようになる。オニヒトデを含めたインドー太平洋のサンゴ礁海域に生息する近縁種間の種分化に関してこれまで限られた遺伝情報に基づいて多くの議論がなされてきたが[4]、具体的な検証まではなされて来なかった。今後はハイスループット DNA シーケンスのコストは下がることでゲノム全体を用いた解析のハードルも低くなると予想され、インドー太平洋のサンゴ礁生物近縁種間の種分化に関する研究においてもゲノムを用いた解析が行われるようになると考えられる。本研究はインドー太平洋のサンゴ礁に生息するオニヒトデ種間の違いと種分化要因をゲノムから解析しており、ゲノムを用いたサンゴ礁生物近縁種間研究の先駆的存在であると言える。

参考文献

1. Reaka-Kudla ML. Known and Unknown Biodiversity, Risk of Extinction and Conservation Strategy in the Sea. In: Bendell-Young L, Gallagher P, editors. *Waters in Peril*. Boston, MA: Springer US; 2001. p. 19–33. doi:10.1007/978-1-4615-1493-0_2.
2. Veron J, Stafford-Smith M, DeVantier L, Turak E. Overview of distribution patterns of zooxanthellate Scleractinia. *Front Mar Sci*. 2015;1. doi:10.3389/fmars.2014.00081.
3. Keyse J, Crandall ED, Toonen RJ, Meyer CP, Trembl EA, Riginos C. The scope of published population genetic data for Indo-Pacific marine fauna and future research opportunities in the region. *Bulletin of Marine Science*. 2014;90:47–78.
4. Carpenter KE, Barber PH, Crandall ED, Ablan-Lagman MaCA, Ambariyanto, Mahardika GN, et al. Comparative Phylogeography of the Coral Triangle and Implications for Marine Management. *Journal of Marine Biology*. 2011;2011:1–14.
5. Price SA, Holzman R, Near TJ, Wainwright PC. Coral reefs promote the evolution of morphological diversity and ecological novelty in labrid fishes. *Ecology Letters*. 2011;14:462–9.
6. Li H, Durbin R. Inference of human population history from individual whole-genome sequences. *Nature*. 2011;475:493–6.
7. Pratchett MS, Caballes CF, Wilmes JC, Matthews S, Mellin C, Sweatman HPA, et al. Thirty Years of Research on Crown-of-Thorns Starfish (1986–2016): Scientific Advances and Emerging Opportunities. *Diversity*. 2017;9:41.
8. Yuasa H, Higashimura Y, Nomura K, Yasuda N. Diet of *Acanthaster brevispinus*, sibling species of the coral-eating crown-of-thorns starfish, *Acanthaster planci* sensu lato. 2017. doi:info:doi/10.5343/bms.2017.1032.

9. Haszprunar G, Spies M. An integrative approach to the taxonomy of the crown-of-thorns starfish species group (Asteroidea: Acanthaster): A review of names and comparison to recent molecular data. *Zootaxa*. 2014;3841:271–84.
10. Nishida M, Lucas JS. Genetic differences between geographic populations of the Crown-of-thorns starfish throughout the Pacific region. *Marine Biology*. 1988;98:359–68.
11. Vogler C, Benzie J, Lessios H, Barber P, Worheide G. A threat to coral reefs multiplied? Four species of crown-of-thorns starfish. *Biology Letters*. 2008;4:696–9.
12. Shaw KL. Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: What mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. *PNAS*. 2002;99:16122–7.
13. Wiens JJ, Kuczynski CA, Stephens PR. Discordant mitochondrial and nuclear gene phylogenies in emydid turtles: implications for speciation and conservation. *Biol J Linn Soc*. 2010;99:445–61.
14. Yasuda N, Hamaguchi M, Sasaki M, Nagai S, Saba M, Nadaoka K. Complete mitochondrial genome sequences for Crown-of-thorns starfish *Acanthaster planci* and *Acanthaster brevispinus*. *BMC Genomics*. 2006;7:17.
15. Hall MR, Kocot KM, Baughman KW, Fernandez-Valverde SL, Gauthier MEA, Hatleberg WL, et al. The crown-of-thorns starfish genome as a guide for biocontrol of this coral reef pest. *Nature*. 2017;544:231–4.
16. Haszprunar G, Vogler C, Wörheide G. Persistent Gaps of Knowledge for Naming and Distinguishing Multiple Species of Crown-of-Thorns-Seastar in the *Acanthaster planci* Species Complex. *Diversity*. 2017;9:22.
17. Burn D, Matthews S, Caballes CF, Chandler JF, Pratchett MS. Biogeographical variation in diurnal behaviour of *Acanthaster planci* versus *Acanthaster cf. solaris*. *PLOS ONE*. 2020;15:e0228796.

18. Fukami H, Budd AF, Levitan DR, Jara J, Kersanach R, Knowlton N. Geographic Differences in Species Boundaries Among Members of the *Montastraea Annularis* Complex Based on Molecular and Morphological Markers. *Evolution*. 2004;58:324–37.
19. Wessel P, Luis JF, Uieda L, Scharroo R, Wobbe F, Smith WHF, et al. The Generic Mapping Tools Version 6. *Geochemistry, Geophysics, Geosystems*. 2019;20:5556–64.
20. Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res*. 2017;45:e18–e18.
21. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*. 2011;27:764–70.
22. Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun*. 2020;11:1–10.
23. Kajitani R, Toshimoto K, Noguchi H, Toyoda A, Ogura Y, Okuno M, et al. Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res*. 2014;24:1384–95.
24. NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*. 2016;44:D7–19.
25. Wada N, Yuasa H, Kajitani R, Gotoh Y, Ogura Y, Yoshimura D, et al. A ubiquitous subcuticular bacterial symbiont of a coral predator, the crown-of-thorns starfish, in the Indo-Pacific. *Microbiome*. 2020;8:123.
26. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009;10:421.
27. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res*. 2017;27:722–36.

28. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature Methods*. 2012;9:357–9.
29. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLOS ONE*. 2014;9:e112963.
30. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;34:3094–100.
31. Seppey M, Manni M, Zdobnov EM. BUSCO: Assessing Genome Assembly and Annotation Completeness. In: Kollmar M, editor. *Gene Prediction: Methods and Protocols*. New York, NY: Springer; 2019. p. 227–45. doi:10.1007/978-1-4939-9173-0_14.
32. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, et al. Automated eukaryotic gene structure annotation using EVIDENCEModeler and the Program to Assemble Spliced Alignments. *Genome Biol*. 2008;9:R7.
33. Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, et al. MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res*. 2008;18:188–96.
34. Kudtarkar P, Cameron RA. Echinobase: an expanding resource for echinoderm genomic information. *Database (Oxford)*. 2017;2017. doi:10.1093/database/bax074.
35. Zhang X, Sun L, Yuan J, Sun Y, Gao Y, Zhang L, et al. The sea cucumber genome provides insights into morphological evolution and visceral regeneration. *PLOS Biology*. 2017;15:e2003790.
36. Simakov O, Kawashima T, Marlétaz F, Jenkins J, Koyanagi R, Mitros T, et al. Hemichordate genomes and deuterostome origins. *Nature*. 2015;527:459–65.
37. Gotoh O. A space-efficient and accurate method for mapping and aligning cDNA sequences onto genomic sequence. *Nucleic Acids Res*.

- 2008;36:2630–8.
38. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*. 2019;37:907–15.
 39. Kovaka S, Zimin AV, Pertea GM, Razaghi R, Salzberg SL, Pertea M. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biology*. 2019;20:278.
 40. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*. 2011;29:644–52.
 41. Schulz MH, Zerbino DR, Vingron M, Birney E. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics*. 2012;28:1086–92.
 42. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*. 2012;28:3150–2.
 43. Wu TD, Watanabe CK. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics*. 2005;21:1859–75.
 44. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. *PNAS*. 2020;117:9451–7.
 45. Smit A, Hubley R, Green P. RepeatMasker Open-4.0. 2013. <http://www.repeatmasker.org>.
 46. Jurka J. Repbase Update: a database and an electronic journal of repetitive elements. *Trends in Genetics*. 2000;16:418–20.
 47. Hubley R, Finn RD, Clements J, Eddy SR, Jones TA, Bao W, et al. The Dfam database of repetitive DNA families. *Nucleic Acids Res*. 2016;44:D81–9.
 48. Stanke M, Waack S. Gene prediction with a hidden Markov model and a new

- intron submodel. *Bioinformatics*. 2003;19 suppl_2:ii215–25.
49. Korf I. Gene finding in novel genomes. *BMC Bioinformatics*. 2004;5:59.
 50. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nature Methods*. 2015;12:59–60.
 51. The UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research*. 2019;47:D506–15.
 52. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics*. 2014;30:1236–40.
 53. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *Journal of Molecular Biology*. 2016;428:726–31.
 54. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, et al. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics*. 2020;36:2251–2.
 55. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res*. 2016;44:D457–62.
 56. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol*. 2013;30:772–80.
 57. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol Biol Evol*. 2015;32:268–74.
 58. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol*. 2018;35:1547–9.

59. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:13033997 [q-bio]. 2013.
<http://arxiv.org/abs/1303.3997>. Accessed 30 Jul 2020.
60. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20:1297–303.
61. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25:2078–9.
62. Edgardo M. Ortiz. vcf2phylip v2.0: convert a VCF matrix into several matrix formats for phylogenetic analysis. Zenodo; 2019.
doi:10.5281/zenodo.2540861.
63. Arnason U, Xu X, Gullberg A. Comparison between the complete mitochondrial DNA sequences of Homo and the common chimpanzee based on nonchimeric sequences. *J Mol Evol.* 1996;42:145–52.
64. Fujiyama A, Watanabe H, Toyoda A, Taylor TD, Itoh T, Tsai S-F, et al. Construction and Analysis of a Human-Chimpanzee Comparative Clone Map. *Science.* 2002;295:131–4.
65. Bryc K, Patterson N, Reich D. A Novel Approach to Estimating Heterozygosity from Low-Coverage Genome Sequence. *Genetics.* 2013;195:553–61.
66. Vogler C, Benzie J, Barber PH, Erdmann MV, Ambariyanto, Sheppard C, et al. Phylogeography of the Crown-of-Thorns Starfish in the Indian Ocean. *PLOS ONE.* 2012;7:e43499.
67. Schiffels S, Durbin R. Inferring human population size and separation history from multiple genome sequences. *Nat Genet.* 2014;46:919–25.
68. Song S, Sliwerska E, Emery S, Kidd JM. Modeling Human Population Separation History Using Physically Phased Genomes. *Genetics.*

- 2017;205:385–95.
69. Terhorst J, Kamm JA, Song YS. Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nature Genetics*. 2017;49:303–9.
70. Schiffels S, Wang K. MSMC and MSMC2: The Multiple Sequentially Markovian Coalescent. In: Dutheil JY, editor. *Statistical Population Genomics*. New York, NY: Springer US; 2020. p. 147–66. doi:10.1007/978-1-0716-0199-0_7.
71. Günther T, Nettelblad C. The presence and impact of reference bias on population genomic studies of prehistoric human populations. *PLOS Genetics*. 2019;15:e1008302.
72. Zann L, Brodie J, Berryman C, Naqasima M. Recruitment, ecology, growth and behavior of juvenile *Acanthaster planci* (L.) (Echinodermata: Asteroidea). *Bulletin of Marine Science*. 1987;41:561–75.
73. Birkeland C, Lucas J. *Acanthaster planci*: major management problem of coral reefs. CRC press; 1990.
74. Lynch M. Evolution of the mutation rate. *Trends in Genetics*. 2010;26:345–52.
75. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4. doi:10.1186/s13742-015-0047-8.
76. Alexander DH, Lange K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics*. 2011;12:246.
77. Cosentino S, Iwasaki W. SonicParanoid: fast, accurate and easy orthology inference. *Bioinformatics*. 2019;35:149–51.
78. Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic*

- Acids Res. 2006;34 suppl_2:W609–12.
79. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009;25:1972–3.
 80. Pond SLK, Frost SDW, Muse SV. HyPhy: hypothesis testing using phylogenies. *Bioinformatics*. 2005;21:676–9.
 81. Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Pond K, et al. Less Is More: An Adaptive Branch-Site Random Effects Model for Efficient Detection of Episodic Diversifying Selection. *Mol Biol Evol*. 2015;32:1342–53.
 82. Yang Z, Nielsen R. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol*. 2002;19:908–17.
 83. Lind AL, Lai YYY, Mostovoy Y, Holloway AK, Iannucci A, Mak ACY, et al. Genome of the Komodo dragon reveals adaptations in the cardiovascular and chemosensory systems of monitor lizards. *Nat Ecol Evol*. 2019;3:1241–52.
 84. Elderfield H, Ferretti P, Greaves M, Crowhurst S, McCave IN, Hodell D, et al. Evolution of Ocean Temperature and Ice Volume Through the Mid-Pleistocene Climate Transition. *Science*. 2012;337:704–9.
 85. Mao Y, Economo EP, Satoh N. The Roles of Introgression and Climate Change in the Rise to Dominance of *Acropora* Corals. *Current Biology*. 2018;28:3373–3382.e5.
 86. Fuller ZL, Mocellin VJL, Morris LA, Cantin N, Shepherd J, Sarre L, et al. Population genetics of the coral *Acropora millepora*: Toward genomic prediction of bleaching. *Science*. 2020;369. doi:10.1126/science.aba4674.
 87. Pratchett MS. Feeding Preferences of *Acanthaster planci* (Echinodermata: Asteroidea) under Controlled Conditions of Food Availability 1. *Pacific*

- Science. 2007;61:113–20.
88. Iijima T, Kajitani R, Komata S, Lin C-P, Sota T, Itoh T, et al. Parallel evolution of Batesian mimicry supergene in two *Papilio* butterflies, *P. polytes* and *P. memnon*. *Science Advances*. 2018;4:eao5416.
 89. Campagna L, Repenning M, Silveira LF, Fontana CS, Tubaro PL, Lovette IJ. Repeated divergent selection on pigmentation genes in a rapid finch radiation. *Science Advances*. 2017;3:e1602404.
 90. Walsh J, Clucas GV, MacManes MD, Thomas WK, Kovach AI. Divergent selection and drift shape the genomes of two avian sister species spanning a saline–freshwater ecotone. *Ecology and Evolution*. 2019;9:13477–94.
 91. Junqueira Alves C, Yotoko K, Zou H, Friedel RH. Origin and evolution of plexins, semaphorins, and Met receptor tyrosine kinases. *Scientific Reports*. 2019;9:1970.
 92. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. *Genome Biology*. 2004;5:R12.
 93. McCormick SD. Endocrine Control of Osmoregulation in Teleost Fish. *Integr Comp Biol*. 2001;41:781–94.
 94. Schaffer S, Takahashi K, Azuma J. Role of osmoregulation in the actions of taurine. *Amino Acids*. 2000;19:527–46.
 95. DiBattista JD, Choat JH, Gaither MR, Hobbs J-PA, Lozano - Cortés DF, Myers RF, et al. On the origin of endemic species in the Red Sea. *Journal of Biogeography*. 2016;43:13–30.
 96. Cabello G, Wrutniak C. Thyroid hormone and growth : relationships with growth hormone effects and regulation. *Reprod Nutr Dev*. 1989;29:387–402.
 97. Adrian J, Bonsignore P, Hammer S, Frickey T, Hauck CR. Adaptation to

- Host-Specific Bacterial Pathogens Drives Rapid Evolution of a Human Innate Immune Receptor. *Current Biology*. 2019;29:616-630.e5.
98. 芦田久. 消化管ムチンを介した微生物と宿主の相互作用. *化学と生物*. 2016;54:901-8.
99. 俊一郎川畑. 自己と非自己を見分ける生体防御レクチン. *化学と生物*. 1999;37:647-52.
100. 米山光俊. 高等脊椎動物における免疫制御. *化学と生物*. 2012;50:243-9.
101. Lokapally A, Neuhaus H, Herfurth J, Hollemann T. Interplay of TRIM2 E3 Ubiquitin Ligase and ALIX/ESCRT Complex: Control of Developmental Plasticity During Early Neurogenesis. *Cells*. 2020;9:1734.
102. Wang J, Liu B, Wang N, Lee Y-M, Liu C, Li K. TRIM56 Is a Virus- and Interferon-Inducible E3 Ubiquitin Ligase That Restricts Pestivirus Infection. *Journal of Virology*. 2011;85:3733-45.
103. Horner-Devine MC, Carney KM, Bohannan BJM. An ecological perspective on bacterial biodiversity. *Proceedings of the Royal Society of London Series B: Biological Sciences*. 2004;271:113-22.
104. Smith MK, Wang T, Suwansa-ard S, Motti CA, Elizur A, Zhao M, et al. The neuropeptidome of the Crown-of-Thorns Starfish, *Acanthaster planci*. *Journal of Proteomics*. 2017;165:61-8.
105. Roberts RE, Motti CA, Baughman KW, Satoh N, Hall MR, Cummins SF. Identification of putative olfactory G-protein coupled receptors in Crown-of-Thorns starfish, *Acanthaster planci*. *BMC Genomics*. 2017;18:400.
106. Lowe EK, Garm AL, Ullrich-Lüter E, Cuomo C, Arnone MI. The crowns have eyes: multiple opsins found in the eyes of the crown-of-thorns starfish *Acanthaster planci*. *BMC Evolutionary Biology*. 2018;18:168.
107. Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and

- Genotyping in Model and Non-Model Species. PLOS ONE. 2012;7:e37135.
108. Suyama Y, Matsuki Y. MIG-seq: an effective PCR-based method for genome-wide single-nucleotide polymorphism genotyping using the next-generation sequencing platform. Scientific Reports. 2015;5:16963.
 109. Takata K, Taninaka H, Nonaka M, Iwase F, Kikuchi T, Suyama Y, et al. Multiplexed ISSR genotyping by sequencing distinguishes two precious coral species (Anthozoa: Octocorallia: Coralliidae) that share a mitochondrial haplotype. PeerJ. 2019;7:e7769.
 110. Combosch DJ, Vollmer SV. Trans-Pacific RAD-Seq population genomics confirms introgressive hybridization in Eastern Pacific Pocillopora corals. Molecular Phylogenetics and Evolution. 2015;88:154–62.
 111. Shinzato C, Mungpakdee S, Arakaki N, Satoh N. Genome-wide SNP analysis explains coral diversity and recovery in the Ryukyu Archipelago. Scientific Reports. 2015;5:18211.

データ登録

ミトコンドリアゲノム配列、核ゲノム配列、サンプル、シーケンスリードに関する情報は DDBJ/NCBI/EBI のデータベースに登録を行った。本研究で構築した 23 サンプルのミトコンドリアゲノム配列のアクセッション番号は LC566216–LC566236、LC596400、LC596401 である。また、Miyazaki01 のサンプルとショートリード (PE と MP) の情報は PRJDB4009、Miyazaki01 のロングリード (PacBio) と核ゲノム配列は PSUB013390、Phuket01 と Israel04 の核ゲノム配列ならびにその他のサンプルとシーケンスリードは PRJDB9937 の BioProject に紐づいて登録されている。

謝辞

本研究は東京工業大学の伊藤武彦教授と梶谷嶺助教の御指導のもとで行われました。心より感謝申し上げます。

宮崎大学海洋分子生態学研究室の安田仁奈准教授にはオニヒトデに関するご助言や多大なサポートを頂きました。心より感謝申し上げます。

本研究に用いた自動遺伝子構造アノテーションツールの LEMON の開発は東京工業大学の伊藤研究室の元メンバーである小林史弥氏と篠田恭寛氏を中心に、中村優太氏、高橋和希氏、奥野未来博士らと共同で行いました。貴重な機会に携わらせて頂き感謝申し上げます。本研究においてゲノムシーケンスを行って頂きました国立遺伝学研究所の豊田敦特任教授と東京工業大学の流水利恵研究員、RNA-seq を行って頂きました東京大学の鈴木穰教授に感謝申し上げます。インドー太平洋の広域サンプリングにご協力いただきました University of Hawai'i at Mānoa の Zac Forsman 博士、Phuket Marine Biological Center の Naline Thongtham 博士、Collège Zéna M'Déré の Coralie Taquet 博士、University of Milano-Bicocca の Davide Seveso 博士と Enrico Montalbetti 博士、Tel Aviv University の Omri Bronstein 博士、University of Queensland の Gal Eyal 博士、石西礁湖のサンゴ礁調査を行われている上野光弘氏、宮崎県グレートダイバーズの福田道善氏、宮崎大学海洋分子生態学研究室の方々に感謝申し上げます。バイオインフォマティクスの手ほどきを頂きました東京工業大学の伊藤研究室の元メンバーである吉村大博士に感謝申し上げます。実験の手ほどきや潜水作業に関する指導を頂きました宮崎大学海洋分子生態学研究室の元メンバーである北野裕子博士に感謝申し上げます。

日本学術振興会特別研究員 DC1(研究課題/領域番号:18J23317)に採用され、経済的支援を受けられたおかげで本研究を遂行することができました。心より感謝申し上げます。

所属研究室である東京工業大学の伊藤研究室のメンバーには日頃より有意義な議論をさせて頂けただけでなく、苦楽を共に乗り越えてきたことで頑張ることができました。心より感謝申し上げます。また、博士課程在学中に温かい応援とサポートをしてくれた家族と山本桂伊氏、宗石瞬氏、谷中絢貴氏ら友人に心より感謝申し上げます。