

③メタバーコーディング (MB) 解析の統計解析やビジュアル化

- ・ビッグデータの効果的な統計解析手法の探索
- ・Rなどを用いた統計解析結果のビジュアル化

(2) 定量性の評価

①メタバーコーディング (MB) 解析データの相対的出現割合による定量性の評価

- ・デジタル PCR を用いた 1 個体あたりの標的遺伝子コピー数の絶対定量法の開発
- ・標的遺伝子コピー数の差異の考慮による各出現グループ間の相対的出現割合の補正

②メタバーコーディング (MB) 技術による魚卵の正確な種同定・定量法の開発

- ・DNA 抽出方法の検討
- ・DNA ライブラリー作製工程の簡略化
- ・マサバ・ゴマサバの識別プライマーの開発
- ・黒潮海域魚卵サンプルのメタバーコーディング解析

③環境ゲノムを利用した魚類生息域評価にむけた検討

- ・佐渡周辺海域のトロール調査時に得られる海水を用いた魚類メタバーコーディング

【今年度の成果】

- ・N、S、A、SI、O、CK、IU、紋別ラインにおいて、定期的に生物環境モニタリングを実施した。2018、2019 年度にそれぞれ 681、318 サンプルの DNA ライブラリーを作製 (18S、16S の合計) し、解析を終了した。公的 DB による相同性検索 (Blast) を行い、既知の遺伝子配列対し、99%以上の相同性、かつ 1 種のみがヒットした場合にのみ種同定に成功という基準を設けて解析を行ったところ、816 種を検出・同定することができた。
- ・紋別市の 1 定点から採集した 7 年間の環境 DNA サンプル (n = 394) について、複数の UP を用いて MB 解析を実施した結果、合計 1,938 種の同定に成功した。
- ・動物プランクトンの MB 解析向上のため、約 170 標本の標的遺伝子 (18S および 28S) の配列決定を完了した。本年度分も継続して配列情報を収集し、公的核酸 DB に登録予定である。
- ・渦鞭毛藻、動物プランクトンに特化した各 MB 解析用の 28S UP の開発に成功した。
- ・18S の V7-9 領域の保存性の高い領域にプライマーとプローブを設計することで、デジタル PCR を用いて、1 個体あたりの標的遺伝子 (rRNA 遺伝子) のコピー数を絶対定量できる方法の開発に成功した。これにより、1 個のシステムで多くの種類の分析が可能である。
- ・魚卵から効率的に DNA 抽出する方法を見出した。また、DNA ライブラリーの作製も時間短縮に成功し、96 サンプルの場合、DNA 抽出から DNA ライブラリーの作製にかかる時間としては、10 時間の短縮に成功した。
- ・ミトコンドリア領域にマサバとゴマサバを識別可能なプライマーの開発に成功した。

【実施概要】

1) 環境 DNA 情報の網羅的解析

①環境 DNA 用サンプル採集

対象水域は、北海道紋別市地先、北海道オホーツク海（N, S-ライン）、道東・東北沖（A-ライン）、静岡県御前崎沖（O-ライン）、石垣島周辺海域、東シナ海（CK-ライン）、日本海（SI-ライン）である。紋別市においては、週 1 回を基本として、表層 1 点をバケツ採水し、MB 解析用サンプルを収集した（図 1）。

NS-ラインでは 2019 年のオホーツク海環境調査 2 回（6、8 月）において各調査 7-8 測点でサンプリングを実施した。MB 解析用、植物色素測定用、植物・微小動物プランクトン種査定用には 10m 層を中心に採水を行った。ネット動物プランクトンの MB 解析用には、最深 150m からの NORPAC ネット（目合い 100 μ m）の曳網を行った。また、2020 年 1-3 月にもオホーツク海氷期海洋調査 3 回（計 3 測点）で同様のサンプリングを行う予定である。加えて、ネット動物プランクトンを用いて優占種 30 種を種同定・ソーティングし、個々の遺伝情報分析用のサンプルとした。

A-ラインでは、2019 年度に道東・東北沖の低次生産調査航海、4 回（5 月、7 月、10 月、1 月）において、各 13 測点でサンプリングを実施した。さらに年度内には 3 月にも調査を実施予定である。MB 解析用、植物色素測定用（HPLC）、植物・微小動物プランクトン種査定用には、10m 層を測点として採水を行った。ネット動物プランクトンの MB 解析用には、基本、水深 150-0m の NORPAC 曳網（目合い 100 μ m）を行った。加えて、ネット動物プランクトンを用いて優占種 30 種を種同定・ソーティングし、個々の遺伝情報分析用のサンプルとした。

O-ラインでは、2019 年度の黒潮低次生産調査航海 4 回（4、8、11、1 月。3 月に 5 回目の調査を予定）において、6-10 測点でサンプリングを実施した。MB 解析用、植物色素測定用、植物・微小動物プランクトン種査定用には、表面および 10m 層を基本とした各測点水深 1000m 以浅の 2-5 層で採水を行った。ネット動物プランクトンの MB 解析用には、水深 150-0m の NORPAC 曳網（目合い 100 μ m）を行った。

SI-ラインでは、2019 年度は日本海低次生産調査航海 4 回（4、6、9、2 月）において、各航海 5 地点でサンプリングを行った。動物プランクトン MB 解析用には NORPAC 曳網（0.1mm、0-150m）を行い、2 層（10m、30m）の採水により MB 解析用、植物色素測定用、植物・微小動物プランクトン種査定用の試料を採取した。加えて、ネット動物プランクトンを用いて優占種 30 種を種同定・ソーティングし、個々の遺伝情報分析用のサンプルとした。予備採集として 11 月に動物プランクトン MB 解析用の NORPAC 曳網（0.1mm、0-150m）を SI-ライン 5 地点で実施した。

CK-ラインおよび隣接する東シナ海域では、2019 年度の陽光丸による調査航海 6 回（4-5、6-7、7、10、2、2-3 月）において、3-5 測点でサンプリングを実施した。MB 解析用の試水は 10m 層およびクロロフィル亜表層極大層（SCM）において採取し、植物色素測定用、植物・微小プラ

ンクトン種査定用の試水は、0, 10, 30, 50, 100 m 層および SCM から採取した。動物プランクトンの MB 解析用試料は目合い 100 μ m の改良型 NORPAC を鉛直曳（150-0m）して採集した。

IU-ライン（石垣島）においては、浦底湾 1 地点で月 1 回、石西礁湖周辺海域 2 地点で年 4 回の観測を行い、表層および中層（浦底湾：10m、石西礁湖：15m）において海水を採集し、MB 解析用、植物色素測定用、植物・微小動物プランクトン種査定用を採水した。ネット動物プランクトンの MB 解析用には、海底直上 1m から表層までの NORPAC ネット（目合い 100 μ m）の曳網を行った。

紋別市では、基本、週 1 回のモニタリングを実施した。MB 解析用、植物・微小動物プランクトン種査定用には表層採水を行い、動物プランクトンの MB 解析用試料は目合い 100 μ m の NORPAC ネットを底層-1m から鉛直曳網して採集した。また、ネット動物プランクトンを用いて優占種 30 種を種同定・ソーティングし、個々の遺伝情報分析用のサンプルとした。

②環境 DNA データの収集

N、S、A、SI、O、CK、IU ラインにて採集した環境 DNA サンプルを用いて、2018、2019 年度にそれぞれ 681、318 サンプルの DNA ライブラリーを作製（18S、16S の合計）し、解析を終了した。公的 DB による相同性検索（Blast）を行い、既知の遺伝子配列対し 99% 以上の相同性を示し、かつ単一種のみがヒットした場合にのみ種同定成功という基準を設けて解析を行ったところ、18S 解析において、816 種を検出・同定することができた。現在、28S を用いた解析も始めており、今後、モニタリングを継続することで、>2,000 種の検出が可能になると予想している。

紋別市におけるモニタリングは、約 8 年前から継続して行っており、7 年間の環境 DNA サンプル（n = 394）を用いて 18S、28S、動物プランクトンを特異的に増幅する 28S（後述）、渦鞭毛藻特異的に増幅する 28S（後述）の 4 個の異なる遺伝子領域を用いて MB 解析を実施した。さらに過去に解析した NORPAC ネットサンプル（2012 年 4 月－2016 年 2 月、n = 164）の 18S および動物プランクトン特異的 28S による MB 解析結果を加えたところ、紋別市の 1 定点において、渦鞭毛藻 196 種、繊毛虫 129 種、珪藻 211 種、菌類 487 種、動物プランクトン 437 種、リザリア 59 種、他の藻類 300 種、合計 1,938 種の同定に成功した（表 1）。ただし、菌類に関しては、ITS 領域を使用するのが一般的であるので、今後、ITS で検出を試み、菌類として何種検出・同定できるか検討を行う予定である。

③動物プランクトン優占種の種名および標的遺伝子配列情報の収集

昨年度より、各モニタリングラインにおいて、NORPAC ネットで採集されたコペポダを中心として動物プランクトン優占種上位 20-30 種について、形態的特徴に基づく種同定を行ったサンプルについて、標的遺伝子の配列収集（18S、28S）を行った。18S については、202 標本中 176 標本、28S については、166 標本中 166 標本の配列収集を完了したので、次年度の

早期に公的核酸データベースへ登録する予定である。本年度、採集したサンプルについては、現在、DNA 抽出中である。

④瀬戸内海における環境 DNA から魚類 DNA の検出

本課題は、瀬戸内海区水産研究所との共同研究として本年度から始まったもので、魚類の MB 解析による魚種出現情報の網羅的収集と将来の資源量推定に資する情報収集を目的とする。本年度、瀬戸内海水域から採集された環境 DNA 326 個について、現在、魚類検出用プライマーを用いて DNA ライブラリーを作製中であり、次年度の計画検討会で結果を報告予定である。

2) 環境ゲノム解析—定量化に向けた検討

(1) メタバーコーディング解析の高度化

①新規ユニバーサルプライマー (UP) の開発と種同定精度の検討・評価

これまで真核生物（動植物プランクトン、原生動物、菌類）全体を標的とする UP (18S、28S) を用いた MB 解析を行ってきたが、生物群によっては種同定のための解像度が十分ではなく、属レベルまでしか同定できない。更なる高度化を図るために、本年度は、渦鞭毛藻を標的とする 28S 遺伝子解析用 UP および動物プランクトンを標的とする 28S 遺伝子解析用 UP の開発を試みた。手法の概略としては、①NCBI 等の公的核酸 DB から LSU、28S、rRNA 等のキーワードで標的遺伝子の登録配列をダウンロード（約 90 万配列）、②新たにプライマーを設計したい 28S を含む配列をそれより外側に設計された UP を用いて In Silico PCR により取得、③得られた配列をマルチプルアライメント、④保存性の高い領域にプライマーを設計、⑤プライマーの PCR 増幅効率やユニバーサリティーを確認、となる。その結果、渦鞭毛藻を標的とする 28S 遺伝子解析用 UP および動物プランクトンを標的とする 28S 遺伝子解析用 UP の開発に成功した（結果は小課題 1-②で述べた）。

②環境 DNA 解析プラットフォームの改良

現在、使用している MB 解析プラットフォームの改良を実施中である。理由は、解析のステップの一部に解析サーバのメモリ消費が大きくなる工程があり、時折、サーバが停止し、再起動を余儀なくされる事態が度々発生した。このため、メモリを使用しないプログラムに変更中である。加えて、細菌 16S の解析においては、検出された細菌 OTUs をより確実に分類できるよう、QIIME2 で使用されている classifier というプログラムを本プラットフォームに搭載する予定である。これらプラットフォームの修正は、今年度中に完了したい。

③メタバーコーディング解析の統計解析やビジュアル化

・ビッグデータの効果的な統計解析手法の探索

これまで、紋別市のメタゲノム解析や環境データを用いて、主成分分析、クラスター解析、コレスポンデンス解析、MDS 解析などを行ってきた。OTU 間あるいは各 OTU と環境要因

の相関関係を解析する方法としては、eLSA 解析、時系列組み込みネットワーク解析などが知られている。本年度は、ピアソンの相関分析（+Benjamini-Hochberg の補正）に注目して、紋別の MB データを用いて解析した。18S では 1,052 種が同定され、そのうち 18% の種で水温と正か負の相関関係が得られた。硝酸態窒素や無機リン酸とはそれぞれ 14% で相関がみられた（図 2）。有害赤潮種や貝毒原因種の時系列プロットを見ると、明瞭な季節性を示す種、散発的な出現特性を示す種、周年出現する種に区分できた（図 3）。

- ・ R などを用いた統計解析結果のビジュアル化

紋別の 18S による MB データのうち、1,052 種のデータを用いてネットワーク解析を行った（図 4）。種間の関係において、上半分は負の相関を示すペア OTUs（渦鞭毛藻、繊毛虫、動物プランクトン）が多く見られた。種－水温間の関係を見ると、上半分は正の相関を示し、これらが高水温時に出現するプランクトン種であること、また、主－栄養塩（硝酸態窒素や無機リン酸）の関係を見ると、上半分は負の相関を示し、これらは夏場の栄養塩が低く、枯渇環境にあることを示すものである。つまり、ネットワーク解析の結果は、夏季の紋別市地先の海においては、水温が高く栄養塩が低濃度であり、その様な環境の中を好んで出現するプランクトン種が存在し、種間ではニッチの奪い合いの競争関係が激しく、互いに負の相関関係が見られる。一方、冬季は、超低水温、高栄養塩環境にあり、その様な環境に出現するプランクトン種が存在するが、多くの種が競って出現するわけではなく、種間に競争関係はほとんどないことを示す。

（2）定量性の評価

①メタバーコーディング解析データの相対的出現割合による定量性の評価

- ・ デジタル PCR を用いた 1 個体あたりの標的遺伝子コピー数の絶対定量法の開発

18S の V7-9 領域の保存性の高い領域にプライマーとプローブを設計することで、デジタル PCR を用いて、1 個体当たりの標的遺伝子（18S）のコピー数を絶対定量できる方法の開発に成功した。取り急ぎ、16 種の有害赤潮種、有毒渦鞭毛藻の培養株（31 株）を用いてコピー数を算出したところ、50-200 万コピーと種により大きく変動することが判明した（表 2）。18S の配列を公的核酸 DB からダウンロード（約 100 万配列）し、プライマーおよびプローブに設計した領域の保存性を確認したところ、プライマー、プローブともに保存性は十分高く、多くの単細胞性植物プランクトンや動物プランクトン 1 個体の遺伝子コピー数を定量できることが判った（表 3、図 5）。

- ・ 標的遺伝子コピー数の差異の考慮による各出現グループ間の相対的出現割合の補正

MB 解析の長所は、網羅的に出現種を検出することが可能な点であるが（紋別の 7 年間の時系列 MB 解析の結果では約 1,900 種の検出に成功）、欠点は定量性に欠ける点である。MB 解析の各種あるいは OTUs ごとの配列数に影響を及ぼす要因として、①DNA 抽出効率の差異、②PCR 増幅効率の差異、③1 個体あたりの標的遺伝子コピー数の差異などが考えられる。①

や②に関しては、数百種の DNA が同時に混在するため、DNA 抽出効率や PCR 増幅効率を正確に見積もり、補正することはほとんど不可能に近い。しかし、③に関しては、本年度、開発した手法で、各種の遺伝子コピー数を情報を蓄積していけば、ある程度の補正が可能である。紋別市の週 1 回 7 年間の時系列 MB 解析の結果、NGS の相対値において、多くの種で明瞭な出現（ブルーム）のピークが検出され、かつ水温との間に統計学的に有意な正あるいは負の相関関係が認められた（図 2、3）。定量 PCR のような定量性はないが、これらの結果は、時系列 MB 解析の相対値が、それなりに定量性のあることを示すものである。従って、各出現種の 1 細胞/個体当たりの遺伝子コピー数を測定し、データを蓄積することは重要である。

②メタバーコーディング技術による魚卵の正確な種同定・定量法の開発

・ DNA 抽出方法の検討と DNA ライブラリー作製や工程の簡略化

本年度は、DNA 抽出方法の検討と DNA ライブラリー作製の簡略化を試みた。魚卵からの DNA 抽出に適した方法については、DNA 抽出バッファーとして、①水酸化ナトリウム、②Chelex、③水酸化ナトリウムと Proteinase K、④Chelex と Proteinase K の計 4 種類のバッファーを検討した。この時、DNA 抽出効率を高めるため、超音波によって魚卵を破碎した。検討の結果、水酸化ナトリウムに Proteinase K を添加した抽出バッファーが最も DNA 抽出効率が高かった。また、魚卵をピンセットで物理的に破壊することで、超音波破碎の時間を省くことができた。この方法により、作業時間を 2 時間以上短縮することが可能となった。次に、DNA ライブラリー作製工程の簡略化を試みた。DNA ライブラリー作製の工程には、1st PCR→PCR 産物の精製→2nd PCR→PCR 産物の精製の 4 つのステップがある。この工程を簡略化するために、1st PCR 産物を精製する場合としない場合で 2nd PCR を行った。その結果、精製をしない場合でも、問題なく 2nd PCR で増幅させることができた。この工程を短縮するために、PCR 産物の精製を DNA 精製装置で実施した。従来のマグネットプレートを使用した手作業による精製と、KingFisher Flex(Thermo Fisher Scientific)による精製を比較した。PCR 産物の回収率は、手作業では 90%以上 (81~99%)、機械では 80%以上 (76~93%) と、機械でも DNA をほぼ損失することなく精製できることを確認した。作業時間は、96 サンプルを精製する場合、4 時間程度、短縮することができた。以上から、96 個の魚卵から DNA を抽出し、DNA ライブラリーを作製するまでの工程を大幅に短縮することに成功した。

・ マサバ・ゴマサバの識別プライマーの開発

魚類を網羅的に検出・同定する UP が活用されているが、残念ながら、マサバとゴマサバを識別することができない。このため、ミトコンドリア領域に、新に識別用プライマーの開発を行い、識別することに成功した（図 6）。魚卵の MB 解析で識別できない場合でも、分子識別で簡単に同定できるので、今後、有効活用したい。

・ 黒潮海域魚卵サンプルのメタバーコーディング解析

2019年3月に北鳳丸による重要水産資源の産卵生態変動に関する調査が実施され、約50地点で魚卵のサンプリングを行った。残念ながら、今回の調査では魚卵の個数は多くなく、1地点で数個の場合も見られた。1魚卵ずつDNA抽出およびDNAライブラリーを作製し、1地点で>96個の魚卵が採集された3地点については96個の魚卵を用いて、メタバーコーディング解析を行った(合計382個)。解析の結果、23個のOTUsを検出し、18個のOTUs(78%)で種同定に成功した。同定できなかった5個のOTUsのうち、2個のOTUsについては、近縁種の登録があり、属レベルでの同定が可能であった(表4)。以上、魚卵を正確に同定できるメタバーコーディング技術の開発に成功した。今後、本法を用いて各浮遊性魚類の産卵海域、産卵期等の詳細を明らかにできる可能性がある。

③環境ゲノムを利用した魚類生息域評価にむけた検討

本年度、佐渡周辺海域のトロール調査時に得られる海水を採集し、魚類MB解析用にフィルターサンプルを得た。現在、DNA抽出中であり、次年度の計画検討会で成果について報告する。

【図表など】

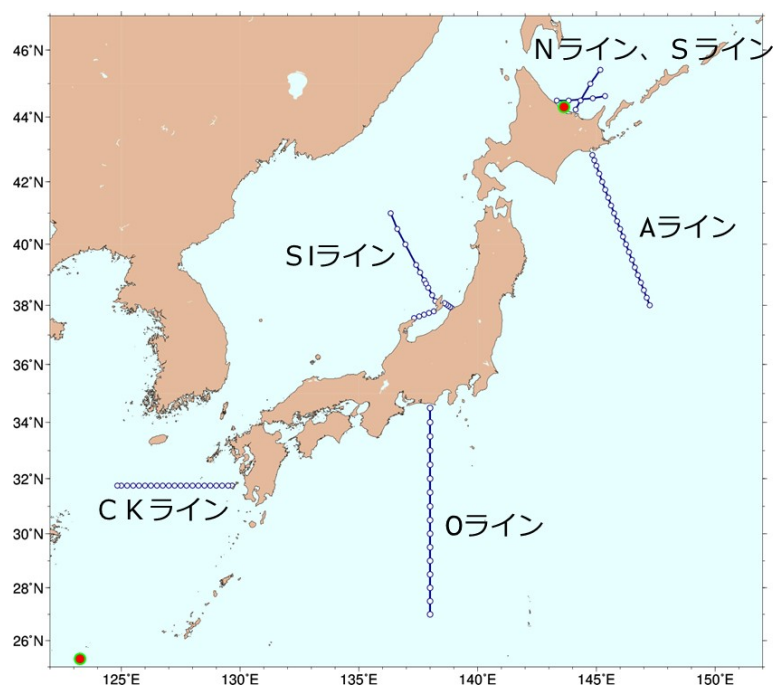


図1 太平洋DB事業におけるモニタリング地点図(紋別市、石垣島を含む)

表 1. 紋別市における MB 解析の結果のうち、種同定に成功した OTU 数を示す。

Supergroup	Phylum	species	Supergroup	Phylum	species
Alveolata	Dinophyceae	196	Viridiplantae	Chlorophyceae	85
	Ciliophora	129		Streptophyceae	57
	Apicomplexa	10			142
	others	2			
		337			
Stramenopiles	Bacillariophyceae	211	Archaeplastida	Florideophyceae	47
	Oomycetes	29		Bangiophyceae	3
	PX clade	33		Compsopogonophyceae	2
	Chrysophyceae	12			52
	Dictyochophyceae	9	Hacrobia	Haptophyceae	27
	Synurophyceae	8		Cryptophyceae	21
	Labyrinthulomycetes	7		Katablepharidaceae	3
	Raphidophyceae	5		Pavlovophyceae	1
	others	17			52
		331			
Opisthokonta	fungi	487	Rhizaria		59
	metazoa	437	Amoezoboa		9
	Choanoflagellida	15	Excavata		5
	others	5	Apusozoa		4
		944	Unclassified		3
in total					1,938

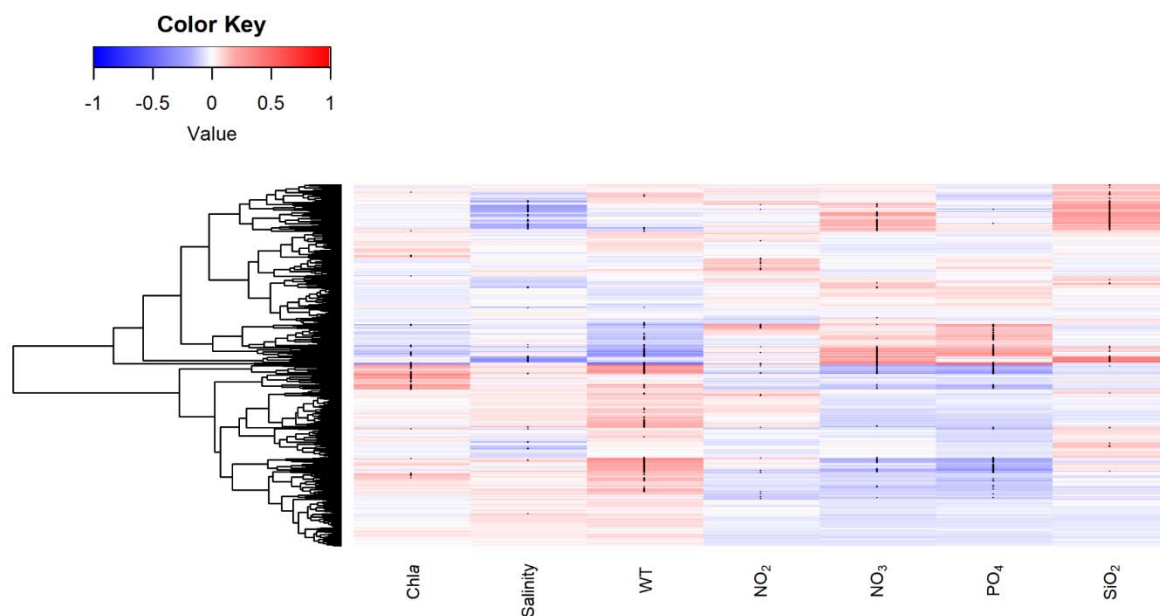


図 2. 紋別市における 18S による MB 解析結果、ピアソンの相関関係と Benjamini-Hochberg の補正（種同定できた 1,052 種を使用、図中の黒丸は有意な相関関係を示す）

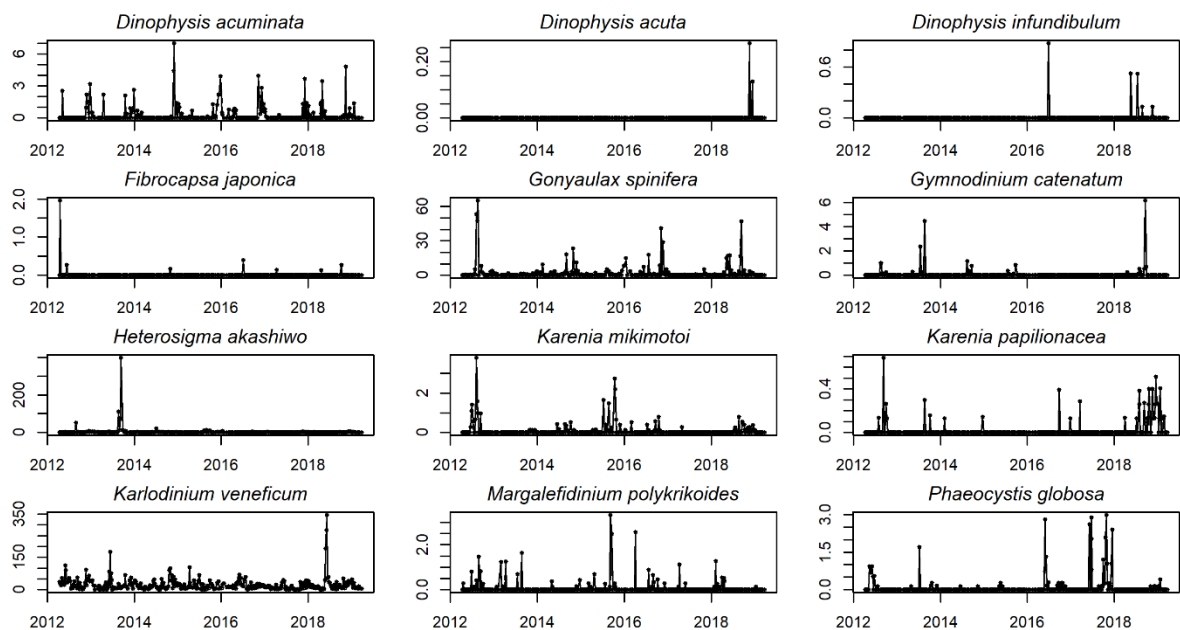


図 3. 紋別市の 18S による MB 解析結果、有害赤潮・貝毒原因藻の時系列データ
(①明瞭な季節性、②散発的出現、③周年出現する種が検出された)

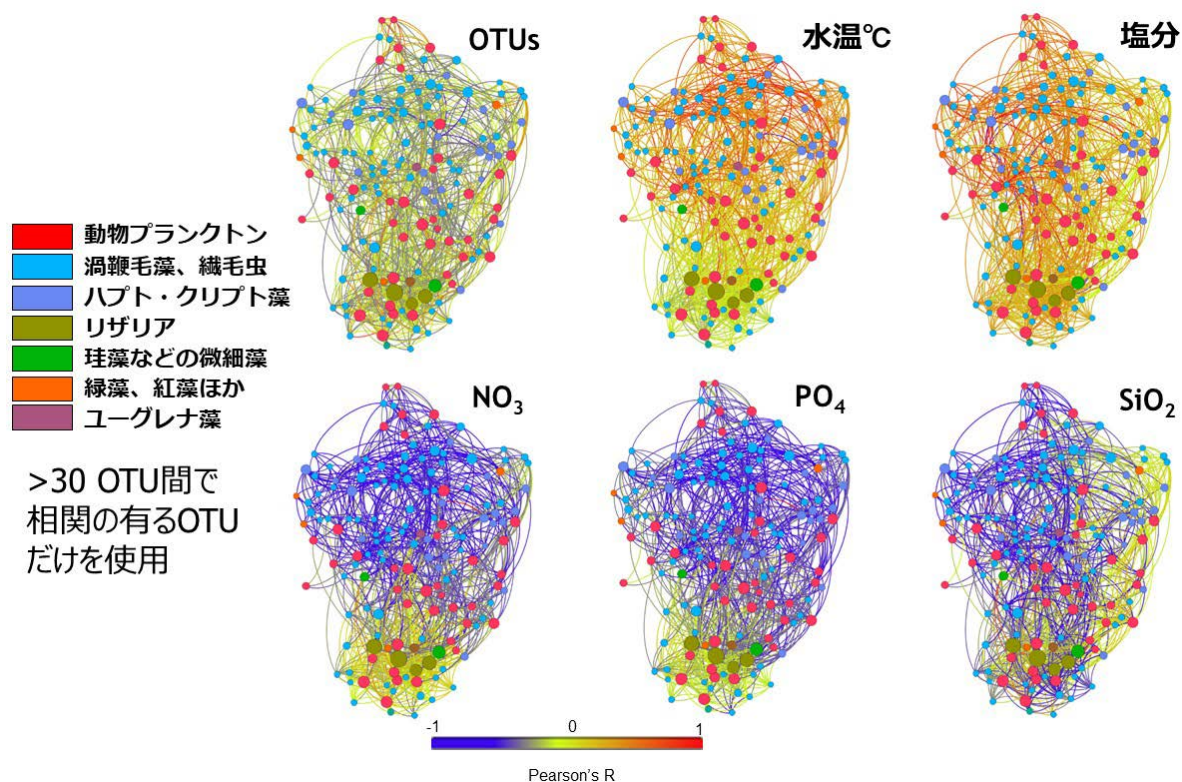


図 4. 紋別市の 18S による MB データを用いたネットワーク解析 (1,052 種間の関係、種間と環境要因の関係を示す)

表 2. デジタル PCR を用いた微細藻類 1 細胞当たりの rRNA 遺伝子コピー数の測定結果

Species	Strain	Average copy nr. cell ⁻¹	SD	MIN	MAX	N
<i>Alexandrium affine</i>	AA2	53,705	530	53,330	54,080	2
<i>Alexandrium catenella</i>	AC_Chile	43,710	721	43,200	44,220	2
	AC02_Japan	69,700	9,617	62,900	76,500	2
	AC10_Japan	42,229	2,818	38,115	44,490	4
<i>Alexandrium pacificum</i>	AP05	254,106	10,189	241,000	263,600	4
	AP20	227,642	60,140	160,175	305,250	6
<i>Alexandrium taylori</i>	AT1	6,390	844	5,550	7,400	4
<i>Chattonella marina</i>	CM1_L	1,227	445	718	1,735	6
	CM1_H	974	273	718	1,213	4
<i>Dinophysis fortii</i>	Miya38	28,925	7,474	23,640	34,210	2
	Mom13	16,140	1,584	15,020	17,260	2
<i>Gambierdiscus toxicus</i>	GT_Se	1,800,100	183,585	1,639,900	2,031,500	4
	GT_La	906,650	107,692	830,500	982,800	2
<i>Gymnodinium catenatum</i>	GC24	2,545	559	2,150	2,940	2
	GC62	2,995	403	2,710	3,280	2
<i>Heterocapsa circularisquama</i>	HC_Le	1,788	11	1,780	1,795	2
	HC_92A	2,573	95	2,505	2,640	2
<i>Karenia mikimotoi</i>	HY	6,853	202	6,710	6,995	2
	NN	7,566	903	6,380	8,420	4
<i>Karenia papilionacea</i>	KP1	6,270	219	6,115	6,425	2
	KP9	5,379	1,414	2,855	7,160	11
<i>Margalefidinium polykrikoides</i>	MP05	3,648	530	3,058	4,168	4
<i>Pseudochattonella verruculosa</i>	PV21_Chile_L	436	62	400	528	4
	PV21_Chile_H	98	15	84	111	4
	PV_Japan	44	2	43	46	2
<i>Pseudo-nitzschia calliantha</i>	PC9	375	71	282	437	4
	PC14	304	47	240	352	4
<i>Pseudo-nitzschia multistriata</i>	PM01	479	59	438	520	2
	PM02	372	14	362	382	2
<i>Pseudo-nitzschia pungens</i>	PP2	1,377	544	634	2,261	17
	PP10	696	32	666	730	4

表 3. デジタル PCR で使用したプライマーとプローブ領域の保存性の確認 (n = 4, 982)

F_primer-seq	T	G	G	A	G	Y	G	A	T	H	T	G	T	C	T	G	G	T	T	D	A	T	T	C	C	G
A	0.1	0.2	1.5	90.1	2.9	0.1	1.1	91.5	0.5	0.1	0.1	0.3	0.0	0.0	7.2	0.4	0.0	0.0	0.0	82.8	99.8	0.0	0.0	0.1	1.1	0.4
C	0.0	0.0	0.0	0.2	0.5	29.1	0.1	6.3	9.0	7.3	0.8	0.1	0.2	98.3	1.4	0.0	9.2	0.6	0.6	0.2	0.0	1.2	0.3	90.0	98.1	0.1
G	0.1	99.6	98.1	7.6	86.8	0.1	97.1	1.7	0.4	0.1	0.0	99.2	0.0	0.1	0.1	98.4	90.6	0.0	0.0	10.9	0.1	1.5	0.0	9.1	0.1	99.1
T	99.7	0.0	0.1	0.6	9.1	69.3	0.1	0.4	89.9	91.1	97.6	0.1	99.3	1.2	89.1	1.1	0.1	99.3	99.3	6.0	0.0	97.2	99.6	0.6	0.7	0.2
others	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	
gap	0.1	0.2	0.3	1.5	0.6	1.5	1.5	0.1	0.2	1.5	1.5	0.3	0.5	0.4	2.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
entropy	0.04	0.05	0.16	0.57	0.73	0.99	0.23	0.51	0.54	0.50	0.20	0.09	0.07	0.15	0.64	0.14	0.47	0.07	0.06	0.85	0.02	0.22	0.04	0.52	0.17	0.09

probe-seq	A	T	A	A	C	A	G	G	T	C	T	G	T	G	A	T	G	C	C	C
A	99.7	0.6	98.2	95.8	0.1	99.8	0.3	1.2	0.0	0.1	2.1	0.3	0.1	0.3	99.7	0.0	0.1	0.0	0.1	0.1
C	0.1	4.2	0.0	0.1	99.1	0.0	0.0	0.1	0.1	99.4	1.4	0.1	0.8	0.1	0.0	0.1	0.0	99.1	99.5	99.5
G	0.0	0.3	1.3	3.6	0.0	0.0	99.4	98.6	0.0	0.0	0.1	99.2	0.1	97.9	0.0	0.0	99.7	0.1	0.0	0.1
T	0.0	94.6	0.2	0.3	0.7	0.0	0.1	0.1	99.7	0.2	96.2	0.0	98.7	0.1	0.1	99.7	0.0	0.7	0.2	0.0
others	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.1
gap	0.2	0.2	0.2	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.3	0.3	1.5	0.2	0.1	0.1	0.1	0.2	0.2
entropy	0.04	0.37	0.15	0.29	0.08	0.03	0.06	0.12	0.04	0.07	0.29	0.08	0.12	0.17	0.03	0.04	0.03	0.09	0.06	0.06

R_primer-seq	T	C	A	C	C	T	A	C	G	G	A	W	A	C	C	T	T	G	T	T	A	C	G
A	0.0	0.1	98.1	0.1	0.1	0.0	100	0.0	7.0	2.9	92.4	87.6	98.8	0.1	0.0	0.0	0.0	0.4	0.0	0.0	99.0	0.0	0.4
C	0.1	99.6	1.6	99.4	99.3	0.0	0.0	100	0.2	0.0	2.5	0.4	0.0	99.8	99.6	0.2	0.0	0.0	0.1	0.0	0.6	98.3	0.0
G	0.0	0.1	0.0	0.2	0.1	0.0	0.0	0.0	91.0	96.9	0.8	2.8	0.9	0.0	0.0	0.7	0.1	99.2	0.2	0.1	0.1	0.0	99.4
T	99.9	0.2	0.3	0.3	0.2	100	0.0	0.0	0.1	0.0	2.7	4.5	0.0	0.0	0.2	98.7	99.7	0.2	99.5	99.7	0.0	1.2	0.0
others	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.1	0.1	0.1
gap	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.3	0.1	1.7	4.7	0.2	0.1	0.1	0.3	0.1	0.1	0.1	0.2	0.2	0.3	0.1
entropy	0.01	0.04	0.15	0.06	0.07	0.00	0.00	0.00	0.54	0.21	0.53	0.76	0.10	0.03	0.04	0.12	0.03	0.08	0.06	0.04	0.10	0.15	0.07