北極海で採集された円石藻 *Emiliania huxleyi* NIES-3366 株のゲノム情報の高精度化

児玉 悠佑 (筑波大学 生物学類)

指導教員：鈴木 石根 (筑波大学 生命環境系)

【背景・目的】

円石藻 *Emiliania huxleyi* は海洋に広く分布し、円石と呼ばれる炭酸カルシウムの結晶を細胞表面に保持する微細藻類である。そのため *E. huxleyi* は有機的と無機的な2つの炭素固定系を通して二酸化炭素の循環に関与し、地球環境に大きな影響を与える。また、*E. huxleyi* を含む一部の藻類はアルケノンという、化石燃料に代わる液体燃料としての利用が期待される長鎖不飽和炭化水素を生産する。このような特徴をもつ *E. huxleyi* の生育挙動の理解や産業利用のためには、ゲノム解読による *E. huxleyi* の遺伝学的な情報の解明が重要である。特に上述のアルケノンの合成機構には未知の部分が多く、アルケノン分子の修飾に関わるデサチュラーゼ遺伝子の同定、その発現メカニズムの解明は、アルケノンの産業利用のポテンシャルを高めることが期待される。

先行研究により北極海で採集された *E. huxleyi* NIES-3366 株は、他の北極海採集株 (NIES-3362, NIES-3365) とともに illumina Miseq のショートリードシーケンスによるゲノム解析が行われている。その結果 NIES-3366 株のゲノムサイズ・予測遺伝子数は、他の北極海採集株と比較して著しく増大していた。先行研究で用いられたショートリードシーケンスは DNA 断片に繰り返し配列が存在すると解析が難しく、正確性に劣る可能性がある。そのため、NIES-3366 株が他の北極海採集株より長いゲノム配列を示す原因を探るには、より精度の高いゲノムデータが望ましい。

そこで本研究では、PacBio Revio/Sequel II によるロングリードシーケンスを行い、NIES-3366 株のゲノム情報の高精度化を目指した。また、予測された遺伝子に illumina のデータで明らかになっていない新規アルケノンデサチュラーゼ遺伝子がないか探索した。

【材料・方法】

1. 微細藻類の培養

E. huxleyi NIES-3366 株は、Marine-art ESM 培地で 18°C、100 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$ の条件で培養し、対数増殖期の細胞を得た。抗生物質 (500 $\mu\text{g mL}^{-1}$ アンピシリン、200 $\mu\text{g mL}^{-1}$ カナマイシン、100 $\mu\text{g mL}^{-1}$ スペクチノマイシン) を培地に加えた。

2. DNA シーケンス

CTAB 法により DNA を抽出し、ProNexBeads によって精製して、シーケンスに必要な品質と量の DNA を得た。シーケンスには HiFi Read を用いた PacBio Revio/Sequel II によるロングリードシーケンスを実施した。シーケンスで得られた raw read からアダプター配列を除去し、1,000 塩基対以下の HiFi Read を削除した後、Hifiasm を用いたアセンブリ、Purge Haplotigs による冗長なハプロティグを除去し、Tiara による真核生物由来 contig を分類した。これらの得られた contig について、RepeatModeler および RepeatMasker を用いて繰り返し配列の分類とマスキングをした。そして BRAKER を用いて遺伝子領域を予測した。予測された遺伝子は Funannotate を用いて、どのような機能をもつ遺伝子か推定する機能アノテーションを行った。ゲノムの品質確認は、配列の連続性を表す N50 値や BUSCO により解析した core gene set の含有率に基づいて評価した。

3. 新規アルケノンデサチュラーゼ遺伝子の探索

今回予測された遺伝子の中から機能ドメイン検索に基づきアルケノンデサチュラーゼ候補遺伝子を抽出し、illumina で予測されたアルケノンデサチュラーゼ遺伝子と照らし合わせた系統的解析を MEGA X を用いて行った。候補遺伝子を絞るため、アルケノンデサチュラーゼに特徴的なヒスチジンボックスの有無、膜貫通領域の予測、立体構造予測、シグナルペプチドを確認した。ヒスチジンボックスの確認ではアミノ酸配列のアライメントを行った。さらに、TMHMM、SOSUI を用いて2つ以上の膜貫通領域をもつ遺伝子を探した。加えて AlphaFold により立体構造モデルを作成した。SOSUSignal と iPSORT Prediction を用いてシグナルペプチドを検索し、アルケノンが存在しないミトコンドリアのシグナルペプチドが見つかった場合は候補から除外して最終的な新規アルケノンデサチュラーゼ候補遺伝子を決定した。

【結果・考察】

今回私が解析した PacBio のロングリードシーケンスと先行研究の illumina ショートリードシーケンスの解析結果の比較を表 1 に示す。N50 や BUSCO の値から、NIES-3366 株のゲノムデータの精度を高めることができた。ゲノムサイズはどちらも概ね同じ値を示した一方で、予測されたタンパク質コード遺伝子の数は illumina に比べて著しく減少した。PacBio のアセンブリに illumina で解析された scaffold および PacBio と illumina で予測された遺伝子配列をマッピングした。その結果、illumina では反復配列で scaffold が分かれ、scaffold を跨ぐ遺伝子が別々ものと予測された箇所があった。このことが PacBio の方が illumina に比べて予測遺伝子が減った原因の1つとして考えられた。

本研究では新たに2つの新規アルケノンデサチュラーゼ候補遺伝子を見出した。1つ目はアライメントによって *E. huxleyi* 北極海採集株では明らかになっていなかった $\Delta 7$ デサチュラーゼ遺伝子 (FUN_000956_T1) である。さらにアルケノンデサチュラーゼの条件を揃えた1つの遺伝子 (FUN_038130_T1) を新規アルケノンデサチュラーゼ候補遺伝子として決定した。

表 1 PacBio と illumina によるゲノムデータ比較

	PacBio	illumina
contig No.	937	229,924
genome size (Mbp)	209	216
protein coding gene No.	40,874	78,388
N50 (bp)	384,086	1,916
BUSCO (%)	66.7	41.2

【今後の展望】

今回ロングリードシーケンスによって明らかになったゲノムサイズの妥当性について、フローサイトメトリーを用いた異なるアプローチによっても検証する予定である。