

気候変動に適応する強靱な新養殖事業

藻類養殖

永田 健・五十嵐洋治¹⁾・柿沼 誠¹⁾

1) 三重大学大学院生物資源学研究科

目的

養殖ヒトエグサの種苗の多くは、現在天然採苗によって確保されている。しかし天然採苗は気象や海況条件に支配されやすい。また種場の遊走子の量を確認する方法も、現在は種網の顕微鏡観察という手法に頼っている。そのため必ずしも安定的に採苗が行えないことから藻体の生産量が安定せず、養殖業者の収入も変動が大きい。したがって、環境水中のヒトエグサ遊走子を DNA レベルで検出することによって安定的に採苗を行う技術の開発が望まれている。そこで本研究では、これまで遺伝学的知見の少ないヒトエグサの全ゲノム DNA 塩基配列を解読し、それら技術開発のための基礎的知見を得ることを目的とする。

方法

1 DNA 抽出と DNA ライブラリーの調製

試料として用いたヒトエグサは、令和 4 年 4 月 19 日に松阪市五主町五主海岸の養殖網から採取した。採集したヒトエグサは、水道水で 10 回程度よく洗った後、滅菌海水中で筆を用いて表面の夾雑物を丁寧に取り除いた。洗浄した試料は、RNAlater (Thermo Fisher Scientific) に浸潤し、4°C で一晩静置後 DNA 抽出まで -20°C で保存した。

試料約 100 mg を液体窒素にて凍結し乳鉢と乳棒を用いて粉末状に破碎し、DNeasy Plant Pro Kit (Qiagen) を用いて DNA を抽出した。操作は付属のマニュアルに従った。得られた DNA は、アガロースゲル電気泳動に供し、DNA の分解度を確認した。また、Qubit 4 Fluorometer (Thermo Fisher Scientific) を用いて濃度を、NanoDrop One (Thermo Fisher Scientific) を用いて純度を確認した。

得られた DNA (500 ng) について、Illumina DNA Prep Kit (Illumina) を用いて DNA ライブラリーの調製を行った。操作は付属のマニュアルに従った。調製した DNA ライブラリーは、Agilent TapeStation 2200

(Agilent) および Qubit 4 Fluorometer でクオリティーチェックを行った。調製した DNA ライブラリーは、マクロジェン・ジャパン株式会社に委託し、HiSeq X シーケンサー (Illumina) にて、次世代シーケンス解析を行っ

た。

2 バイオインフォマティクス解析

Illumina HiSeq X シーケンサーより出力された塩基配列データについて、FastQC v0.11.9 を用いて取得した塩基配列の概要を確認した。続いて、fastp v0.23.1 を用いて、低品質の配列やアダプター配列などを除去するクオリティーフィルタリングを行った。次に、得られた高品質の塩基配列について、CLC Genomic Workbench 8.5 を用いて塩基配列の再構築を行った。再構築された配列は、カバレッジ-コンティグ長グラフ分析および gVolante を用いた保存遺伝子のカバー率の算出により、再構築されたドラフトゲノムの完成度を評価した。

結果及び考察

1 抽出された DNA と DNA ライブラリーの調製

試料より DNeasy Plant Pro Kit を用いて DNA を抽出した結果、濃度 16.5 ng/μL、吸光度比 A260/A280=1.81、A260/230=1.76 と高品質な DNA が得られた。この DNA を用いて、DNA ライブラリーの調製を行ったところ、濃度 4.80 ng/μL、平均断片長 510 bp と受託解析の要件を満たす品質の DNA ライブラリーが調製された。

2 バイオインフォマティクス解析

Illumina HiSeq X シーケンサーより約 509 億塩基対、約 3.4 億リードの塩基配列データを取得した。この塩基配列データについてクオリティーフィルタリングを行ったところ、約 451 億塩基対、約 3.0 億リードの高品質な塩基配列データが得られた (表 1)。

表 1. フィルタリング前後の塩基配列データの概要

	塩基数(bp)	リード数	GC(%)	Q30(%)
生データ	50,889,684,176	337,017,776	50.9	91.6
フィルタリング後	45,169,722,030	303,613,890	50.9	95.4

CLC Genomic Workbench 8.5 を用いて高品質の塩基配列データについて塩基配列の再構築を行ったところ、総塩基数約 1.4 億塩基、スキップフォールド配列数 44,565

本, シーケンスカバレッジ 332.8 fold のドラフトゲノム配列が構築された (表 2)。この内最も配列長が長かったスキファールドは 281,921 塩基であった。一方, 構築されたゲノム配列の完全性の指標となるコンティグの N50 値は 30,303 塩基と短く, 構築されたドラフトゲノムはまだ繋がっていない配列が多く, 短い断片が多く含まれていると考えられた。

表 2. 再構築されたドラフトゲノムの概要

総塩基数 (nt)	139,917,335
スキホルド配列数	44,565
最長配列長 (nt)	281,921
最短配列長 (nt)	200
平均配列長 (nt)	3,140
N50 配列長 (nt)	30,303
L50 配列数	1,066
1K nt 以上の配列数	11,497 (25.8%)
10K nt 以上の配列数	2,735 (6.1%)
100K nt 以上の配列数	148 (0.3%)
GC 含量(%)	52.8
シーケンスカバレッジ	×332.8

続いて再構築された塩基配列にシーケンスリードをマッピングして, カバレッジ-コンティグ長グラフ分析を行った (図 1)。グラフの横軸は各コンティグのカバレッジ, グラフの縦軸はコンティグ長を示し, 各プロットは塩基配列が再構築されたそれぞれの長さのコンティグに, 高品質なリードをマッピングしたときのシーケンスカバレッジが示されている。

図 1 で示されるように, カバレッジが 100 fold を超えたあたりに 2 つのピークが確認された。この内, 最もカバレッジの大きいピークがヒトエグサゲノムのコンティグ, カバレッジ 100 fold 程度のコンティグはヒトエグサの葉緑体やミトコンドリア等のオルガネラゲノムであると推察された。一方で, カバレッジが 10 fold 程度のあたりに複数のピークが確認された。これは, 共生細菌やコンタミネーションしたバクテリア由来のゲノム配列が再構築された結果であると推察される。したがって, 今回は塩基配列データのクオリティフィルタリングを行っているものの, シーケンス中からバクテリア由来の配列は除去されておらず, より完成度の高いドラフトゲノム配列構築のためには, バクテリア由来の塩基配列をターゲットにしたさらなるフィルタリングが必要であることが示された。

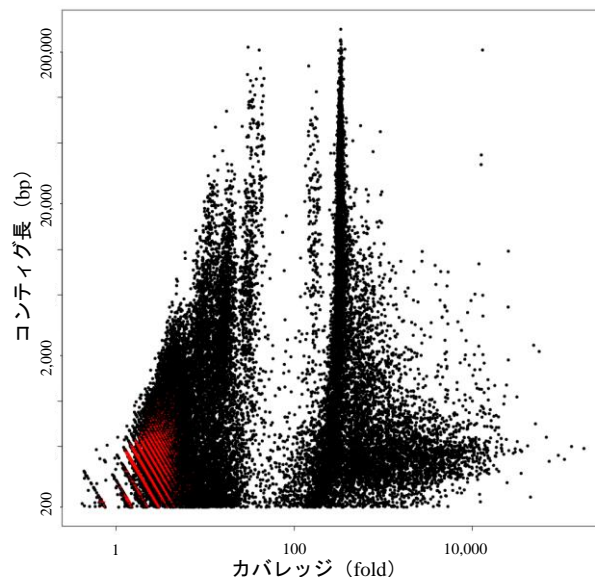


図 1. カバレッジ-コンティグ長グラフ分析の結果

最後に, gVolante にて BUSCO スコアを算出し, ゲノムの完全性を評価した。BUSCO スコアとは, コア遺伝子セット (保存遺伝子セット) が再構築された配列の中にどれだけあるか調べることで, ゲノムシーケンスがどの程度の精度でできているかを評価する指標である。その結果, 表 3 に示すように緑藻類のデータベースに登録されている 1,519 個のコア遺伝子中 1,268 個の完全長の遺伝子が検出され, BUSCO スコアは 83.5% となった。完全性の高いゲノムが再構築されると, BUSCO スコアが 90% を超えてくるが, 現在用いているショートリードシーケンサーのみのデータでは, gVolante を用いた解析においても塩基配列の再構築がやや不完全なものであることが確認された。この問題を解決するためには, 用いるソフトウェアの見直し, また現在解析を行っているロングリードシーケンサーのデータで塩基配列の再構築を補完することで, より完成度の高いヒトエグサのドラフトゲノム配列が構築できるのではないかと考えられた。

表 3. gVolante によるゲノムの完全性の評価

緑藻類データベース中のコア遺伝子の数	1,519
検出されたコア遺伝子の数	
完全長	1,268 (83.5%)
完全長+部分配列	1,320 (86.9%)
検出されなかったコア遺伝子の数	199 (13.1%)
コア遺伝子あたりの平均オルソログ遺伝子数	1.17
1 つ以上のオルソログ遺伝子を持つ割合(%)	5.99
BUSCO スコア	83.5%