

プロジェクト名(タイトル):

アガロースゲル・マイクロカプセルを用いたシングルセルゲノム解析

利用者氏名:

○青木 弘良 (1)

理研における所属研究室名:

(1) 光量子工学研究センター 先端光学素子開発チーム

## 1. 本プロジェクトの研究の背景、目的、関係するプロジェクトとの関係

地球上には、様々な微生物が存在し、我々の環境や生態系を形成している。これらの多くの微生物は培養が難しく、その機能はこれまで未知であった。我々はそれら微生物の機能を、1細胞レベルで解析する技術を確立する技術を確立した。具体的には、それら微生物を微細なゲルのカプセル内に包埋し、機能を司るゲノム DNA を増幅し、DNA シーケンサーで解析する。しかし1つの微生物より数 Gb におよぶ DNA 断片情報が得られ、それを PC 上で再構成するのに(アセンブル)、数時間かかる課題があった。そこでスーパーコンピュータ Hokusai BW によるアセンブルを試みた。

## 2. 具体的な利用内容、計算方法

大腸菌ゲノム DNA データ (3.3 GB)を、Hokusai BW, iMac 2020, および DELL ワークステーション(WS)を用いて、SPAdes アセンブルソフト (Prjibelski A, 2020)による解析時間を比較した (Table 1)。iMac は比較のため、1 および 4 スレッドで計算した。またアセンブル品質として、N50 を求めた。

	iMac		Hokusai	WS
ノード	/		6	/
コア	6	6	240	20
メモリ(GB)	60	60	600	40
スレッド	4	1	480	38
実行時間(min)	81.5	223	80	51
N50 (kb)	2,292	/	2,292	2,345

Table 1 ゲノムデータの解析時間の比較

## 3. 結果

各コンピュータ上でのパラメータと解析結果を Table 1 に示す。解析時間は、WS < iMac(4 スレッド), Hokusai < iMac (1 スレッド)の順に短く、アセンブリ品質である N50 は WS > iMac, Hokusai の順に高かった。

## 4. まとめ

SPAdes は、微生物のゲノム DNA アセンブルに標準的に使用されているソフトウェアで、マルチスレッドによる高速化に対応している。今回 WS の方が Hokusai にくらべ、実行時間が短く、また得られたアセンブル品質も高かった。これは Hokusai での SPAdes パラメータの最適化されていないことや、CPU あたりのメモリ等の影響が考えられる。

## 5. 今後の計画・展望

海外では、SPAdes のようなバイオインフォマティクス用のツールがインストールされているスーパーコンピュータセンターもあり、バイオインフォマティクスに不慣れな生物系の研究者や、十分な解析環境の準備が難しい研究者に、計算環境を提供している。今後 SPAdes のパラメータの最適化により、パフォーマンスの向上と、微生物ゲノム解析の改善が期待される。

## 6. 利用がなかった場合の理由