

課題名 (タイトル) :

全ゲノムシーケンスデータ解析パイプライン並列化

利用者氏名 : ○角田 達彦 藤本 明洋  
阿部 哲雄 中村 英二

所属 : 横浜研究所 ゲノム医科学研究センター  
統計解析・技術開発グループ/情報解析研究チーム

1. 本課題の研究の背景、目的、関係するプロジェクトとの関係

近年の著しいシーケンス技術の発展により、個人ゲノムシーケンスが可能となった。超並列シーケンサーの発展は著しく、現在では 55 Gbp (ヒトゲノムの約 17 倍) の塩基配列データが約 10 日で得られている。全ゲノムシーケンスは世界中で活発に行われており、世界的には数千人規模で全ゲノムシーケンスが行われている。また、今後もシーケンサーのデータ産出量は増大していくことは確実であり、全ゲノムシーケンスは次世代の疾患研究において、極めて重要な役割を担っていくと考えられる。しかしながら、超並列シーケンサーはデータ量が多く、正確な解析のためには膨大な計算量を必要とし、次世代計算機「京」での実行が期待される。我々の解析パイプライン (解析ソフトウェア群) の「京」での実行に向けてのチューニングのために RICC を利用した。

2. 具体的な利用内容、計算方法

解析パイプラインの移植を開始した。

3. 結果

解析パイプラインの移植を開始した。マッピングプログラムのコンパイルを行い、正常に実行できることを確認した。

4. まとめ

開始したばかりだが、RICC の能力の高さが実感できた。

5. 今後の計画・展望

解析パイプラインを完全に移植し、並列数を上げて実行する。

6. RICC の継続利用を希望の場合は、これまで利用した状況 (どの程度研究が進んだか、研究におい

てどこまで計算出来て、何が出来ていないか) や、継続して利用する際に行う具体的な内容

解析パイプラインを完全に移植し、並列数を上げて実行し、さらなる速度向上を狙う予定である。

7. 一般利用で演算時間を使い切れなかった理由

8. 利用研究成果が無かった場合の理由

上記の目的に使用したため、論文という形での成果はない。