

課題名 (タイトル) :

ParaHaplo: A program package for haplotype-based whole-genome association study using parallel computing

利用者氏名 : ○三澤 計治*

佐久間 俊広**、滝中 徹**、坂上 和也**、斎藤 実**、浅見 暁**、白戸 幸正**

理研での所属研究室名 :

*社会知創成事業 次世代計算科学研究開発プログラム

次世代生命体統合シミュレーション研究推進グループデータ解析融合研究開発チーム

** (株) NEC 情報システムズ先端技術ソリューション事業部 ハイエンドコンピューティング G

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 DNA 配列決定技術の進歩により、大量のゲノム配列が解析されるようになりました。ゲノム配列は、個人の間には違いがあります。これを利用し、数万人に対して、ゲノム全体に渡り、患者群とコントロール群との遺伝子頻度を統計的に解析することによって、疾患関連遺伝子を発見する手法がゲノムワイド関連解析です。大人数のゲノムデータを解析するためには、高速計算が不可欠です。そこで我々は、京速コンピュータ「京」の上で、ゲノムワイド関連解析のためのプログラム ParaHaplo を開発しています。ParaHaplo は大量のゲノムデータを分割し、ユニット間・コア間のハイブリッド並列により、高速計算を実現しました。この高並列化をテストするために RICC を利用させていただきました。

2. 具体的な利用内容、計算方法

RICC 上でコンパイルし、利用する CPU 数を変えながら速度を測定し、並列度を測定しました。また、ステージングを利用したデータ入出力の並列化も試みました

3. 結果

計算・データ入出力ともに、高並列でも十分な並列性能が出ることを確認されました。

4. まとめ

RICC 上で ParaHaplo が並列コンピューティングの観点から実用に耐えるソフトウェアであることが実証されました。

5. 今後の計画・展望

京速コンピュータ「京」での利用を目指し、今後は RICC 上でさらなるチューニングを行い高速

化を実現する予定です。

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
RICC で ParaHaplo をさらなるチューニングを行いたいと思います。

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

平成 23 年度 RICC 利用研究成果リスト

【論文、学会報告・雑誌などの論文発表】

- Misawa, K. and N. Kamatani. 2011. ParaHaplo 3.0: A program package for imputation and a haplotype-based whole-genome association study using hybrid parallel computing. *Source Code Biol Med* 6: 10.

【国際会議などの予稿集、proceeding】

【国際会議、学会などでの口頭発表】

ParaHaplo 3.0 日本人類遺伝学会第 56 回大会・第 11 回東アジア人類遺伝学会合同大会 千葉県千葉市
2011/11/10-11

ParaHaplo 3.0 A program package for imputation and a haplotype-based whole-genome association study using hybrid
parallel computing K. Misawa, N. Kamatani. International Congress of Human Genetics 2011, at Montreal, Canada
2011/10/10-11

【その他】