大規模ゲノム解析に向けた並列計算環境の比較

安藤瞳 （指導教員：瀬々潤）

1 はじめに

現在，DNA 配列を読むシークエンサは，従来より高速に大量の配列を読むことができるようになった．これらを用いた個人ゲノム時代は到来しており，一例として 2010 年 10 月，理化学研究所のゲノム医科学研究センターが中心となって日本人男性 1 人の全ゲノムシークエンスの解析を発表した [1]．

日本人ゲノムの解析はすでに国際コンソーシアムで読まれたヒトゲノム配列を活用し，次世代シークエンサで日本人のゲノムを読んだ上で，既知のヒトゲノム約 3G 塩基対と対応づけた．シークエンサで読んだ配列はリード配列，既知のゲノムは参照配列と呼び，配列の対応づけはアラインメントと呼ばれる．図 1 に模式図を示す．次世代シークエンサは読み間違いが頻発するため，何重にも読んで精度の高いアラインメントを行う必要がある．

このように，シークエンサ性能の向上によって高速に大量のデータが得られるようにり，より高速な解析が求められている．本研究では，ゲノム解析で抽出するアラインメントに着目し，近年大規模計算センターで用いられる並列分散ファイルシステム Lustre と並列分散環境 Hadoop の分散性能を計測し，安価な複数台の計算機による分散実行によって高速なゲノムデータの解析を目指す．これにより，ゲノム解析に適切な並列処理を提案する．

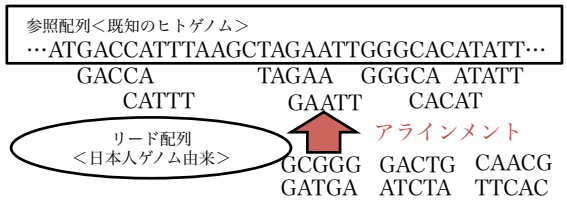

図 1: アラインメントの例

2 実験環境

本研究では，並列分散ファイルシステム Lustre と並列分散環境 Hadoop の実験環境を構築し，bowtie[2] を用いて日本人ゲノムから取得したリード配列のアラインメントを行い，比較を行った．

今回の実験で用いたデータとその容量は以下のとおりである．

＜参照配列＞ Ensembl から既知のヒト 1〜10 番染色体のゲノム（ver.59）を取得し，bowtie で高速にマッピングするためのインデックスをつけたもの．全部で約 1.9GB.

＜リード配列＞ DDBJ の SRA（No.DRA000222）から日本人ゲノム由来のリード配列の一部を取得した約 13.5GB，65,955,311 本.

実験環境は計算機 5 台を用いて構築した．以降，これら 5 台の計算機を F1，F2，F3，M1，C1 として区別する．

表 1: 計算機の仕様

CPU	Intel(R) Atom(TM) D510 @1.66GHz
メモリサイズ	2GB
OS	Linux
カーネル	`2.6.18-164.11.1.el5_lustre.1.8.3`
ネットワーク	Gigabit Ethernet
Lustre ver.	`lustre-1.8.3`
Hadoop ver.	`hadoop-0.20.2`

表 2: 実験一覧

実行環境 / 計算に用いた台数	Lustre	Hadoop	ローカルマシン
1台	実験Lustre1台		実験ローカル1台
3台	実験Lustre3台	実験Hadoop	実験ローカル3台

用いた計算機 5 台は同一の仕様であり，それを表 1 に示す．各環境で実行する計算機の数を変えて表 2 に示すように網羅的に実験を行い，実行中の計算機のディスク，I/O，メモリ，CPU を計測した．また，入力ファイルの準備から出力ファイルをひとつにまとめるまでに要した時間を計測した．

2.1 Lustre 環境の構築

実験 Lustre1 台，Lustre3 台は並列分散ファイルシステム Lustre の環境下で行った．Lustre とはオープンソースソフトウェアの分散ファイルシステムであり，テラバイトクラスのファイルに対応し，実際に大規模計算センターに使われてもいる．Lustre には 2 種類のノードがあり，データの所在を管理する Meta Data Server (MDS) と，実際にデータを格納する Object Strage Server (OSS) である．本研究では，MDS1 台，OSS3 台で Lustre を構築した．ファイルシステム全体のディスクは 590GB になる．図 2 に実験 Lustre1 台，実験 Lustre3 台の実験環境を示す．

・実験 Lustre1 台

実験 Lustre1 台では，1 台のクライアントマシン（C1）で Lustre をマウントし，C1 からアラインメントを行った．

・実験 Lustre3 台

実験 Lustre3 台では，3 台のマシン（C1，F1，F2）で Lustre をマウントし，3 つのクライアントマシンから同時にアラインメントを行った．F1 と F2 は OSS 兼クライアントとして用いた．

2.2 Hadoop 環境の構築

実験 Hadoop は並列分散環境 Hadoop の環境下で行った．Hadoop とは膨大な量のデータを複数のマシンに分散して処理できるオープンソースのプラットフォームである．Hadoop は MapReduce，HDFS という 2 つの仕組みから成り立っている．MapReduce とは並列分散処理をするためのプログラミングモデルであり，

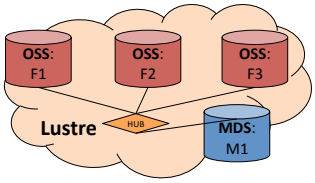

図 2: Lustre 環境

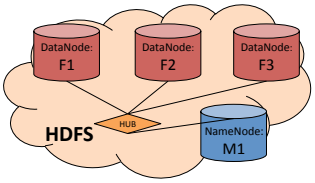

図 3: Hadoop 環境

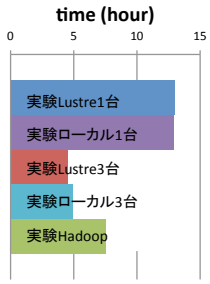

図 4: 実行時間

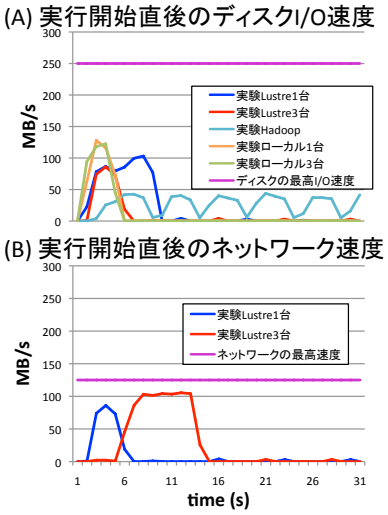

図 5: 各実験の I/O 状況

HDFS とは，MapReduce を行うための Hadoop 独自の分散ファイルシステムである．HDFS は Lustre 同様，データの所在を管理する Name Node と，実際にデータを格納する Data Node からなる．

本研究では，Name Node 1 台，Data Node 3 台で HDFS を構築した．ファイルシステム全体のディスクは 666.9GB になる．図 3 に実験 Hadoop の実験環境を示す．

・実験 Hadoop

実験 Hadoop では，bowtie を Hadoop で並列実行するツール crossbow[3] を用いてアラインメントを行った．

2.3 分散ファイルシステムを利用しない例

・実験ローカル 1 台

比較を行うために，実験ローカル 1 台では 1 台のマシン（C1）のみでアラインメントを行った．

・実験ローカル 3 台

実験ローカル 1 台と同様に比較を行うために，3 台のマシン（C1，F1，F2）を用いて，リード配列を 3 分割して各マシンそれぞれでアラインメントを行った．

3 結果と考察

各実験の実行時間を図 4 に示す．縦軸が各実験，横軸が実行時間を表している．Lustre1 台とローカル 1 台の実行時間はほぼ同じだが，Lustre3 台とローカル 3 台では，実行する計算機の台数は同じでも約 30 分 Lustre3 台のほうが速く実行出来ている．これは，ローカル 3 台で入力ファイルの転送に 30 分以上かかったのに対して Lustre3 台では共有ファイルシステムを利用することによって約 3 分に短縮出来たからであり，bowtie の実行時間自体は 2 つの実験でほぼ同じであった．

また，各実験の bowtie 実行直後のディスク I/O 速度の状況をまとめたものが図 5 の (A) である．縦軸が I/O 速度，横軸が実行からの経過時間を表している．事前に調べた各マシンのディスクの最高 I/O 速度が約 250MB/s であるので，Hadoop 以外で十分な速度が出ており，その速度もすぐに落ちているのが分かる．これは，bowtie は実行直後ゲノムを一度に読み込むので，そこがディスク I/O のボトルネックになったと考えられる．

しかし，Lustre1 台，Lustre3 台では各 OSS が読み込んだリード配列をクライアントとネットワーク間でやりとりして bowtie を実行するので，bowtie 実行直後のネットワークの速度がボトルネックになる可能性がある．Lustre1 台と Lustre3 台の bowtie 開始直後のネットワーク速度の状況が図 5 の (B) である．ネットワーク速度の理論最大値が 125MB/s なので，十分な値が出ているが，速度が落ちるタイミングに注目すると，Lustre1 台はネットワーク速度が落ちてからディスク I/O 速度が落ちているため，ネットワークがボトルネックになっていると考えられる．同じように注目すると，Lustre3 台はディスク I/O 速度が落ちてからネットワーク速度も伴って落ちているため，こちらはやはりディスク I/O がボトルネックになっていると考えられる．

最後に，図 4 が示すように，実行する計算機の台数は同じでも Hadoop の実行時間がローカル 3 台の約 1.5 倍遅かった．図 5 の上段グラフを見ても，ディスク I/O 速度は他の実験に比べても十分に出ていない．そこで Hadoop 実行直後の CPU 使用率を調べると，ディスク I/O の帯域と逆相関の関係にあり，ディスク読み込みが連続的に行われていないことが分かった．

4 まとめと今後の課題

今回の実験では，Lustre を利用し複数台のクライアントでマウントして並列実行することが最も最速な方法であることが分かった．Hadoop を利用した実験では並列分散実行させても速度は速くならなかったが，Hadoop はチューニング次第で改良が可能なので，まだ改善の余地が十分にあると言える．

参考文献

[1] A.Fujimoto *et al.*, Whole-genome sequencing and comprehensive variant analysis of a Japanese individual using massively parallel sequencing, *Nature Genetics, 42*, pp.931—936, 2010.

[2] B.Langmead *et al.*, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome, *Genome Biology, 10(3)*, R25, 2009.

[3] B.Langmead *et al.*, Searching for SNPs with cloud computing, *Genome Biology, 10(11)*, R134, 2009.