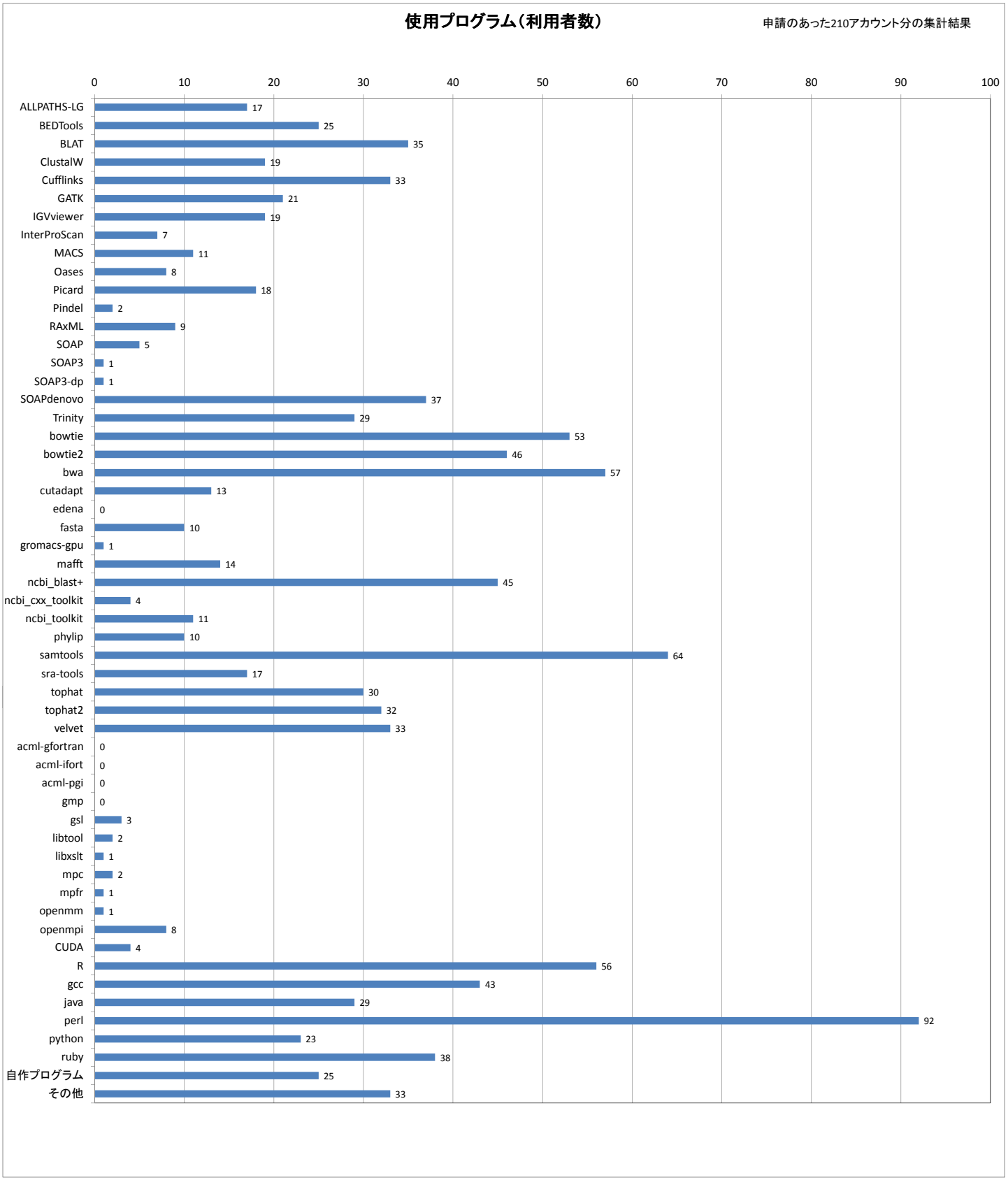
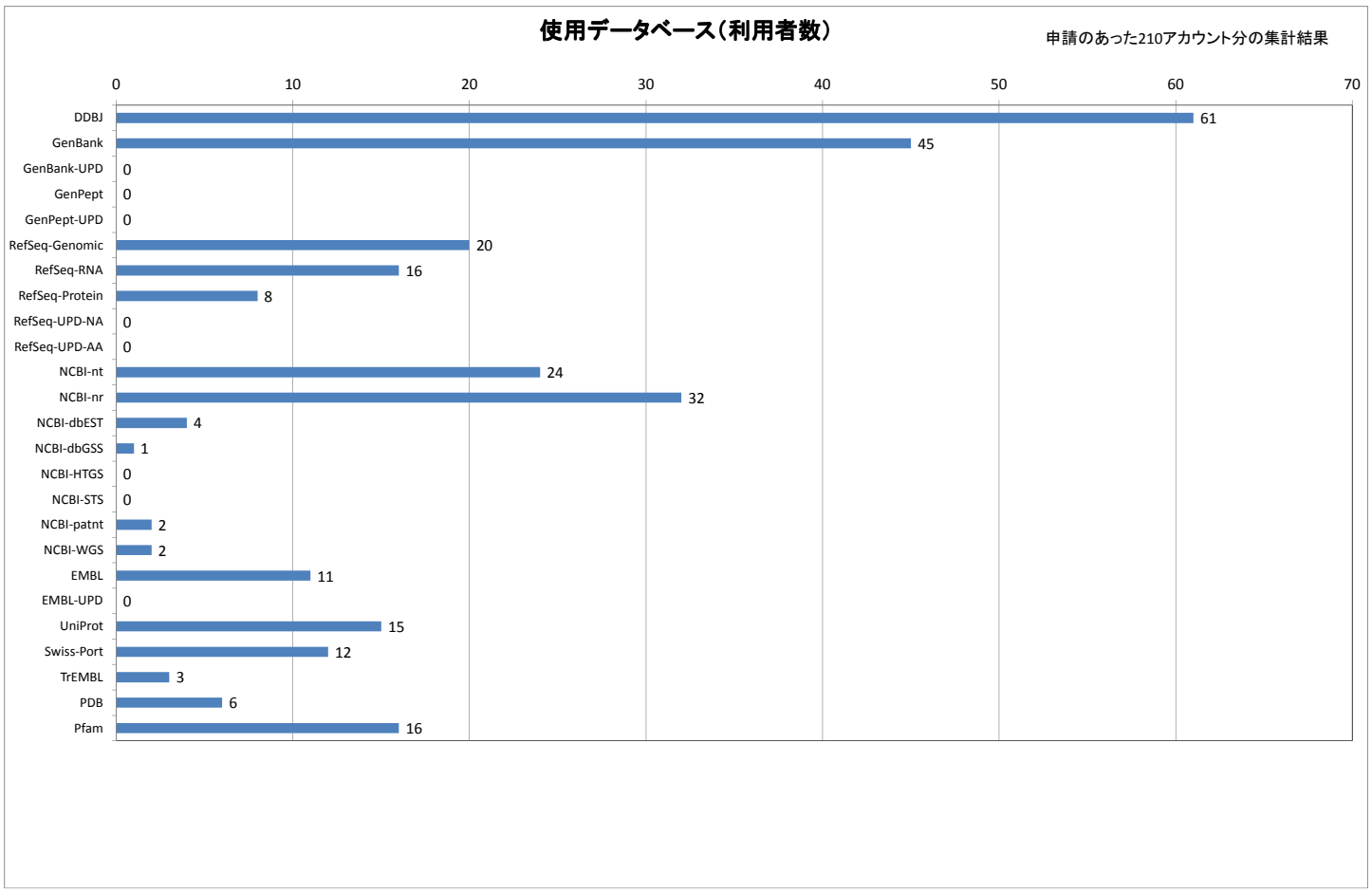
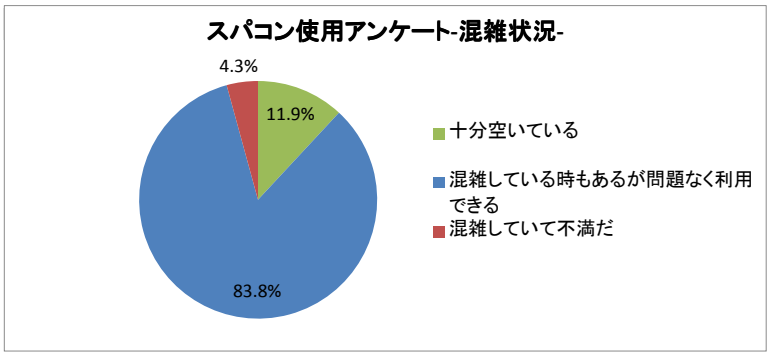
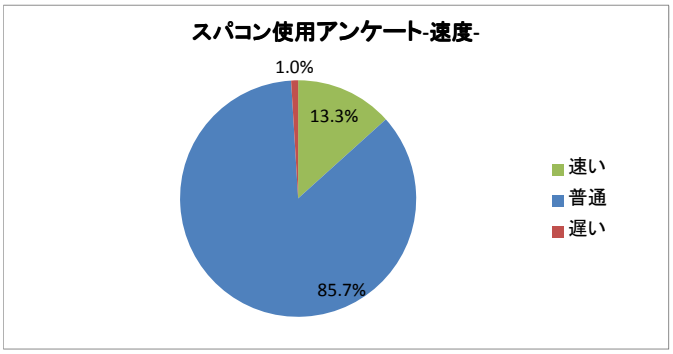
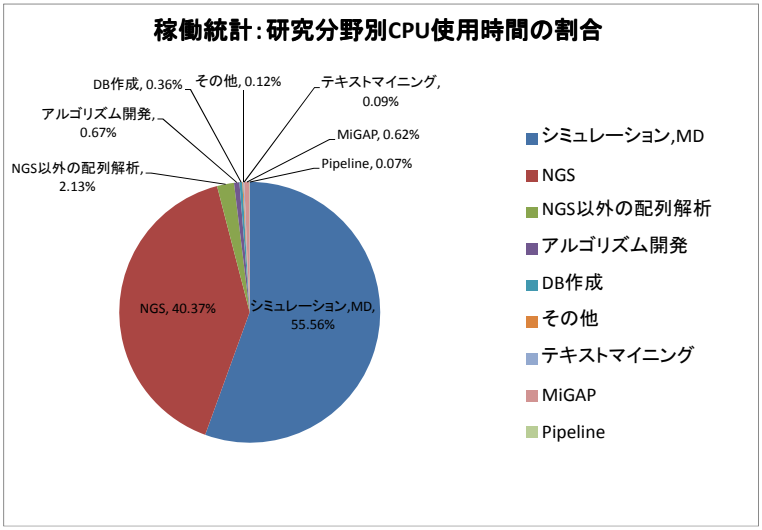
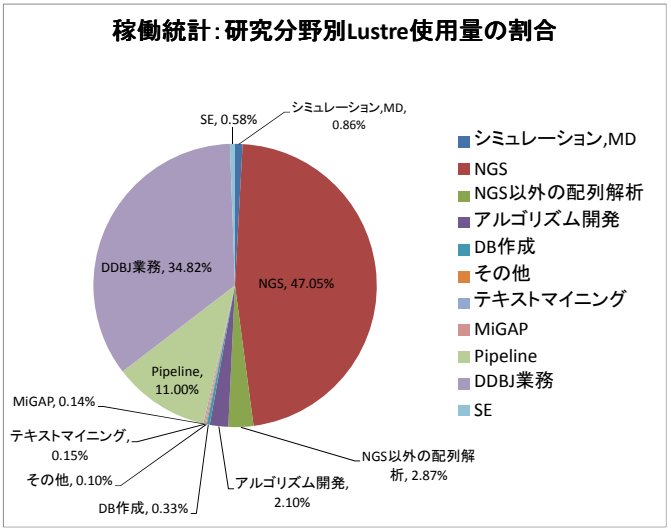

使用プログラム（利用者数）
申請のあった210アカウント分の集計結果
0
10
20
30
40
50
60
70
80
90
100
ALLPATHS-LG 17
BEDTools 25
BLAT 35
ClustalW 19
Cufflinks 33
GATK 21
IGVviewer 19
InterProScan 7
MACS 11
Oases 8
Picard 18
Pindel 2
RAxML 9
SOAP 5
SOAP3 1
SOAP3-dp 1
SOAPdenovo 37
Trinity 29
bowtie 53
bowtie2 46
bwa 57
cutadapt 13
edena 0
fasta 10
gromacs-gpu 1
mafft 14
ncbi_blast+ 45
ncbi_cxx_toolkit 4
ncbi_toolkit 11
phylip 10
samtools 64
sra-tools 17
tophat 30
tophat2 32
velvet 33
acml-gfortran 0
acml-ifort 0
acml-pgi 0
gmp 0
gsl 3
libtool 2
libxslt 1
mpc 2
mpfr 1
openmm 1
openmpi 8
CUDA 4
R 56
gcc 43
java 29
perl 92
python 23
ruby 38
自作プログラム 25
その他 33


使用データベース（利用者数）
申請のあった210アカウント分の集計結果
0
10
20
30
40
50
60
70
DDBJ 61
GenBank 45
GenBank-UPD 0
GenPept 0
GenPept-UPD 0
RefSeq-Genomic 20
RefSeq-RNA 16
RefSeq-Protein 8
RefSeq-UPD-NA 0
RefSeq-UPD-AA 0
NCBI-nt 24
NCBI-nr 32
NCBI-dbEST 4
NCBI-dbGSS 1
NCBI-HTGS 0
NCBI-STS 0
NCBI-patnt 2
NCBI-WGS 2
EMBL 11
EMBL-UPD 0
UniProt 15
Swiss-Port 12
TrEMBL 3
PDB 6
Pfam 16


スパコン使用アンケート-混雑状況-
4.3%
11.9%
83.8%
十分空いている
混雑している時もあるが問題なく利用できる
混雑していて不満だ


スパコン使用アンケート-速度-
1.0%
13.3%
85.7%
速い
普通
遅い


稼働統計：研究分野別CPU使用時間の割合
DB作成, 0.36%
その他, 0.12%
テキストマイニング, 0.09%
アルゴリズム開発, 0.67%
MiGAP, 0.62%
NGS以外の配列解析, 2.13%
Pipeline, 0.07%
NGS, 40.37%
シミュレーション,MD, 55.56%
シミュレーション,MD
NGS
NGS以外の配列解析
アルゴリズム開発
DB作成
その他
テキストマイニング
MiGAP
Pipeline


稼働統計：研究分野別Lustre使用量の割合
SE, 0.58%
シミュレーション,MD, 0.86%
DDBJ業務, 34.82%
NGS, 47.05%
Pipeline, 11.00%
MiGAP, 0.14%
テキストマイニング, 0.15%
その他, 0.10%
DB作成, 0.33%
アルゴリズム開発, 2.10%
NGS以外の配列解析, 2.87%
シミュレーション,MD
NGS
NGS以外の配列解析
アルゴリズム開発
DB作成
その他
テキストマイニング
MiGAP
Pipeline
DDBJ業務
SE

アカウント区分	使用者	所属機関	利用目的	論文名1	論文の状況1	Pubmed ID1	論文名2	論文の状況2	Pubmed ID2	論文名3	論文の状況3	Pubmed ID3	研究成果(分野)	研究成果(分野)(その他)	NGSプロジェクト名	NGSプロジェクト名(その他)	NGSプラットフォーム名	研究成果・経過報告(内容入力)	使用プログラム	使用プログラム(自作プログラム)	使用プログラム(その他)	使用データベース	アンケート-速度-	アンケート-速度-(自由入力欄)	アンケート-混雑状況-	アンケート(自由入力欄)
一般研究用	石川　昌和	国立遺伝学研究所　遺伝情報分析研究室	メタゲノム解析										NGS		RNASeq		Ion Torrent PGM	環境の違いによる、大腸菌のRNA seqをおこなっている。現在、大腸菌のゲノムとトランスクリプトームをcure link genomic kit, RNA kitで抽出し、シーケンスを行う予定である。そこから、ゲノムをassembleして、それを用いて、Transcriptomeをmappingする予定である。	ClustalW fasta			DDBJ	普通		十分空いている	各種のtoolの使い方を統合DBのようにyoutubeにup してくれるとありがたい
一般研究用	馬場崎　勝彦	森林総合研究所	きのこのゲノム解読										NGS		Whole Genome Sequencing		454 GS FLX Titanium Illumina Genome Analyzer Illumina HiSeq 2000	Illumina HiSeq 2000の100b長mate-paired reads 情報量34GbのきのこのゲノムDNAシーケンスを用いて、これまで構築して来ているきのこのcode note genome assemblyを改良し、より詳細なannotationを行うことを計画して進めてきたが、諸般の事情で、計画の達成に至っていない。	velvet perl			DDBJ GenBank	普通		混雑している時もあるが問題なく利用できる	申請書のみで、課金もなく、外部からの利用ができる体制や、スパコンの理解が十分でない利用者にも、比較的容易にアクセス等ができ、大変有り難い存在なので、今後も、システムを余り複雑にしないで、幅広い利用が可能な状態を継続して頂きたい。
一般研究用	森　隆久	国立遺伝学研究所 生命情報研究センター　五條堀研究室	ウイルス配列情報の解析のため										NGS		Whole Genome Sequencing		Ion Torrent PGM	近年、沿岸漁業が営まれる沿岸海域では、赤潮、貧酸、魚病（以下、「赤潮等」という。）など環境由来の漁業被害が発生し、特に八代海において、一昨年と昨年の2年連続で八代海において シャットネラによる大規模な赤潮が発生し、80億円もの漁業被害が生じるなど沿岸漁業経営の安定化、水産物の安定供給の視野に関わる問題となっている。本研究では、海水養殖のデジタルDNAマップあるいは環境評価生物系配列のデジタルDNAチップを設計し、赤潮等の漁業被害の発生リスクを把握可能にする漁場環境評価技術の開発を試みた。	nobi,blast+			GenBank	普通		混雑している時もあるが問題なく利用できる	特になし
一般研究用	鈴木　健介	イルミナ株式会社	次世代シーケンサーデータの解析										NGS		Whole Genome Sequencing		Illumina HiSeq 2000 Illumina MiSeq	新規 de novo アセンブル手法を開発中。ヘテロな状態のゲノムをうまくアセンブルできるアルゴリズムを開発中。ウイルスゲノムなどヘテロな状態のサンプルにおける頑度の低い変異も検出に拾える。大型ゲノムの相同性の高い遺伝子も綺麗に分離できる。SOAPdenovo や velvet との比較を行なう予定。	SOAPdenovo bwa samtools velvet perl			NCBI-nt	速い	データ転送が素晴らしく速い。	混雑していて不満だ	素晴らしく使いやすいサービスで感動しました。個々の研究室がコンピュータを買うのでは無く、NIGのクラウドに資金を集中して増設するのが良いと思いました。
一般研究用	細川　宗孝	京都大学	パラ、ダリア、トウガラシなど各種園芸作物の転写RNA解析。										NGS		RNASeq		Illumina MiSeq	現在次世代シークエンスの解析結果待ちです。解析はNicotiana benthamianaのRNAseq、パラのRNAseq、トウガラシのRNAseqである。解析を外注で行っており、4月の下旬に結果が出るそうです。結果が出次第お世話になります。なお、解析は京都大学の安井助教の助けを借りて行うつもりです。	Trinity			DDBJ NCBI-patnt	普通		混雑している時もあるが問題なく利用できる	まだ使ってません。
一般研究用	黒田　誠	国立感染症研究所・病原体ゲノム解析研究センター	次世代シークエンサーリードの情報解析，blast検索等による病原体検索										NGS		Metagenomics		Illumina Genome Analyzer Illumina MiSeq	11/16からアカウントを取得して利用を開始する予定であったが、ヒト配列を含む解読リードを外部サーバーに転送して良いかどうか思案した結果、連結可能匿名化であっても倫理的に問題が生じる可能性を考え、本スーパーコンピューター一度も利用しなかった。結果的に自前の内部サーバーにて解析することになり、本スーパーコンピューターにおけるパフォーマンスを実感できずに残念である。	nobi,blast+				普通		混雑している時もあるが問題なく利用できる	使用実績がないため、コメントは控えさせて頂きます。申し訳ありません。
一般研究用	岡野　桂樹	秋田県立大学	次世代シークエンサのデータ解析										NGS		Transcriptome Analysis		Illumina HiSeq 1000 Illumina MiSeq	海洋生物フジツボ幼生の各器官に発現する遺伝子のde novo transcriptome解析をめざしている。アカフジツボ幼生を中心におこなっているが、今後共生フジツボ、寄生フジツボのフジツボとホストとの関係にも注目している。現在、HiSEq1000のレーンで データを得たところで、今後Trinityでの解析をめざす。	Trinity			DDBJ	普通		混雑している時もあるが問題なく利用できる	まだ、使用しておりません。データが出始めたので、今後楽しみにしております。
一般研究用	大田　竜也	総合研究大学院大学・先導科学研究科	大容量塩基配列を利用したバイオシステムの進化解析										NGS		RNASeq		Illumina Genome Analyzer II Illumina HiSeq 2000	被子植物の花器官で発現している遺伝子を次世代シークエンサーを用い網羅的に調査した。形態の異なる個体間での異なる発現様式を示す遺伝子を選抜することで、花の形態形成に関係する可能性のある遺伝子のリストを作成している。今後、これらの遺伝子についてDNA配列レベルでの分子進化学的な視点から解析を行い花器官の形態形成で鍵となる遺伝子を探索する予定である。	Oases bowtie samtools velvet gcc				普通		混雑している時もあるが問題なく利用できる	大容量のコンピュータのメモリを使用できる点で重宝している。一方で、時折システムが不定期にダウンする等問題が生じているのは好ましくない。またユーザ間の情報交換の場などがあると問題解決に役立つのではないでしょうか。
一般研究用	清澤　秀孔	高知大学	融合プロジェクトにおける共同研究(城石研と)										NGS		RNASeq		Illumina HiSeq 2000	城石研と共同でMSMマウス129/6のFl1ハイブリッド個体及びハイブリッドES細胞をin vitroで神経細胞へ分化させたサンプルを用いている。またFl1ハイブリッドを用いることにより、母親と父親の染色体を区別したアレル特異的な発現解析を行っている。アレル特異的な発現を見ることによりゲノム刷り込み及びiso の発現制御領域特定を目的としている。	BLAT perl			DDBJ GenBank RefSeq-rna Genomic RefSeq-rRNA NCBI-nt NCBI-nr NCBI-dbEST	速い		十分空いている	データ保存領域の提供も含めてとても感謝しています。
一般研究用	門田　有希	岡山大学大学院　環境生命科学研究科	NGSデータ解析および系統解析	Novel techniques for identifying active retrotransposon family using high-throughput sequencing of PBS sites	In Preparation								NGS以外の配列解析					複数の異作物種において、レトロトランスポゾンの挿入箇所をゲノム網羅的に同定した。この際、品種を区別するためのタグ配列を付加してシーケンス解析をすることにより、一度のシーケンス解析で、100を超える品種について、多数の挿入箇所を決定した。得られた挿入箇所の情報を利用することにより、作物の系統解析および遺伝解析に取り組んでいる。また、品種固有の挿入箇所を利用し、品種判定用のマーカー開発にも取り組んでいる。	BEDTools BLAT ClustalW Cufflinks IGVviewer SOAPdenovo bowtie bowtie2 bwa cutadapt fasta maffft nobi,blast+ ncbi_cxx,toolkit nobi,toolkit samtools velvet mpc R perl			DDBJ NCBI-nt	普通		混雑している時もあるが問題なく利用できる	私の所属機関にはスパコンがなく、遺伝研のスパコンのおかげで、大規模なデータ解析に取り組むことができています。また、登録されたソフトウェアおよびデータベースも充実しており、ほぼ毎日、利用させていただいています。
一般研究用	吉成　晃	北海道大学大学院	シロイヌナズナ変異体のゲノムリシーケンス結果の解析										NGS		Resequencing		Illumina Genome Analyzer IIx	シロイヌナズナの変異体(Col-0背景)と野生型Lerを交配し、F2世代の植物を得た。これらのF2植物のうち変異表現型を示す個体を70~80個体分離し、バルクでDNAを抽出した。このバルクDNAもとにNGS用のライブラリーを調製し、NGSによってゲノムリシーケンスを行った。なお、未だスパコンを利用していないが、今後利用する予定である。	bowtie bowtie2 samtools				普通		混雑している時もあるが問題なく利用できる	未使用
一般研究用	上野　まりこ	京都大学農学研究科	次世代シークエンシングのアウトプット解析										NGS		Whole Genome Sequencing		Illumina HiSeq 2000	予算が不足し、予定していた塩基配列解読を実施できなかったため、解析を行うことができませんでした。データの取得が可能となった時点で再度アカウント申請したいと思います。ご迷惑をおかけして申し訳ございませんでした。	ALLPATHS-LG SOAPdenovo bowtie2 bwa samtools velvet			DDBJ	普通		混雑している時もあるが問題なく利用できる	申し訳ございません、利用できなかったため、アンケートに答えることができません。
一般研究用	葛原　陸	徳島文理大学	公開されているヒトのWholeゲノムシーケンス（1000genomesのシーケンス）から一部配列を抽出し、ヤセンブルを行って、未知有用情報を抽出する。										NGS以外の配列解析					既存の公開されているヒトゲノム配列を基に共通配列・異なる配列などを解析した。グループ間で異なる配列を複数見いだしたのと、共通配列も見いだしている。その中で既存の枠に入らない遺伝子を見いだしつつある。その中には、微生物・ウイルスに近い配列など、興味深い配列が多数存在する。現在、それらをまとめつつある段階である。	BLAT			DDBJ	普通		混雑している時もあるが問題なく利用できる	スパコンを公開して頂いていることに感謝致します。
一般研究用	坪田　博美	広島大学大学院理学研究科附属宮島自然植物実験所	RAxMLなどを用いた分子系統解析	Nots on Aptychella and Gammiella in Japan, a new synonym and a new combination	In Preparation								NGS以外の配列解析					コケ植物の系統・分類学的な研究を行っている。現在、コケ植物蘚類で新属として記載されたYakushimabryumについて、既存の属のシノニムであることを報告すべく論文作成中である。その際、分子系統解析の際、アライメントにmaffft、解析にRAxML等を利用する。また、別の属であるthelieliaについても、その分類学的な所属を明らかにすべく、分子系統解析を行っている。	ClustalW RAxML maffft phylip				普通		混雑している時もあるが問題なく利用できる	震災前は良く利用していましたが、震災後利用回数が減ってしまった。理由としては、以前とシステムが異なること、ローカルのPCの処理速度が遅くなったため、待ち時間や転送時間を考えるとあまり要わらない場合も少なくないことがあげられます。
一般研究用	横井　一仁	個人	ゲノム解析										アルゴリズム開発					Hadoop互換のMapReduceアルゴリズムを用いた並列処理基盤の開発完了した。有効性を検証するため、Wikipediaの全テキストデータに対してワードカウントを行い評価した。評価によりGNU sortとuniqによるワードカウントと比較し、数倍の性能が出せることを確認した。今後はバイオインフォマティクスへの適用を進める。具体的にはスンハリ菌の配列データから、ペプチドマスフィンガープリント法の検索で用いるインデックスを超高速に作成できるかを検証する。	Java perl				普通		混雑している時もあるが問題なく利用できる	快適に使用できています。
一般研究用	宮崎　祐子	岡山大学大学院環境生命科学研究科	NGSデータの解析	Transcriptome analysis of fruiting under an experimental warming in Quercus serrata	In Preparation								NGS		RNASeq		Illumina HiSeq 2000	温度化等の環境変化は、樹木の生理反応を通して個体の維持等に影響を与えると想定される。一方、温暖化に対する樹木の応答反応は様々で、これまでに想定されていない反応も含め、全体像を捉えることは困難である。そこで本研究では、温暖化処理を行った際に、生理反応の初期段階として起こる遺伝子発現の変化を明らかにするため、複数の時期から採取した複数の器官において網羅的遺伝子発現解析を行った。温暖化実験はOTC(Open Top Canopy Chamber: 2m四方の透明アクリル板で樹冠上部の一部を囲う)を設置することで行った。OTCC内部(温暖化処理)と同個体のOTCの外部(コントロール)からそれぞれ葉、芽、幼花、幼花(開花直後のみ)を2012年8月～7月に採取し、遺伝子発現解析試料とした。採取した試料は次世代シーケンサーIllumina HiSeq 2000を用いてde novoトランスクリプトーム解析を行った。マッピングされたリードを用いてDEGseq(Wanget al. 2010)により発現解析を行い、温度化処理と無処理の間で発現に差のある塩基配列を抽出した。さらにBLASTプログラムを用いてデータベース(DDBJ/EMBL/GenBank)上の既知遺伝子との相同性検索を行った。今後、これらの遺伝子の発現挙動を詳細に解析することで、温暖化に対する様々な応答反応の全体像を捉えていく予定である。	nobi,blast+ perl			DDBJ GenBank UniProt	普通		混雑している時もあるが問題なく利用できる	Gene ontology解析が行えるプログラムを含めてくださると嬉しいです。
一般研究用	城石　俊彦	国立遺伝学研究所 系統生物研究センター	新領域融合プロジェクト研究「マウスのRNA-Seq解析と解析手法の検討」を複数の大学および研究所と共同でおこなう。										NGS		RNASeq		Illumina HiSeq 2000	新領域融合プロジェクト研究「マウスのRNA-Seq解析と解析手法の検討」を複数の大学および研究所と共同でおこなっている。本年度から、マウスB6系統とMSM系統のゲノム配列を利用して、B6系統とMSM系統の実験交配系統におけるトランスクリプトーム（RNA-seq）解析を開始した。今後はデータの収集を一層進め、ゲノム機能解析をおこなっていく予定である。	Cufflinks bowtie bwa tophat R				普通		混雑している時もあるが問題なく利用できる	本アカウントを利用することで、情報・システム研究機構　新領域融合プロジェクト研究「マウスのRNA-Seq解析と解析手法の検討」を複数の大学および研究所と共同で円滑におこなうことができている。
一般研究用	城地　保昌	公益財団法人　高輝度光科学センター	X線小角散乱実験と分子シミュレーションを組み合わせてクロマチン構造を解明する。										シミュレーション、MD					分子シミュレーションを援用してクロマチン構造のモデルを構築し、そのモデル構造に基づくX線小角散乱プロファイルを大規模並列計算により解析した。モデルの小角散乱プロファイルを実際のX線散乱実験データの比較することでクロマチン構造に関する考察を行っている。次年度も本解析計算を継続することでクロマチン構造の本質に迫っていきたい。	自作プログラム	mpi			普通		混雑している時もあるが問題なく利用できる	有効に利用させていただいています。ジョブクラスの制限がもう少し緩い方が並列計算を況山使えると思います。
一般研究用	小林　直樹	国立医薬品食品衛生研究所　衛生微生物部	大腸菌を材料に研究をしております。今後、複数検間の全ゲノム比較や次世代シーケンサーデータなどの大規模配列解析をおこないたいと考えておりますので利用を希望します。										NGS以外の配列解析					腸管出血性大腸菌検間の遺伝的多様性と新質的対応を解析するため、新質的異なる検間での配列比較や遺伝子数比較などをおこないたい。予備的な比較として、現在データベースに全ゲノム情報が公開されている菌株について解析を進め、少数の遺伝子について、保有する菌株のデータを含め解析をおこなった。今後全ゲノムレベルでの比較をおこなう予定である。	nobi,blast+ perl			GenBank	普通		十分空いている	今年度は使用頻度が予想より少なく、速度等に問題を感じることはありませんでした。来年度以降、頻度を徐々して使用させていただきたいと思っております。
一般研究用	國枝　武和	東京大学大学院理学系研究科生物科学専攻	大規模遺伝子解析を研究しています。ゲノムデータ・トランスクリプトームデータの解析を行うための利用を希望します。	Genome of anhydrobiotic tardigrade	In Preparation								NGS		Transcriptome Analysis		Illumina Genome Analyzer IIx Illumina HiSeq 2000	トランスクリプトームデータから新規構築した約2万個のヨコヅナアマノムシ遺伝子セットについて、nrデータベースに対して網羅的なBLAST検索を行った。その結果、evalue 1e-5 を閾値とした場合、55%の遺伝子が既知の遺伝子と相同性を示し、注目すべき遺伝子群についてマニュアルでアノテーションの追加・修正を行った。既知遺伝子と相同性が見られなかった残り45%の遺伝子群についてEST データベースと照合した結果、その3分の1が、これらのヨコヅナアマノムシのESTデータと相同性を示すことがわかり、これらはクマムシ類でのみ保存されている遺伝子セットと考えられた。	nobi,blast+			NCBI-nr NCBI-dbEST	普通		十分空いている	大きな演算はあまりおこなっていないせいか、特に混雑は感じませんでした。おかげさまで解析を進めることができました。ありがとうございました。
一般研究用	大川　恭行	九州大学　医学研究院・先端医療医学部門	次世代シークエンサーを用いたエピゲノム解析	A co-localization model of paired ChIP-seq data using a large ENCODE data set enables comparison of multiple samples.	Published	23125362	Direct isolation and RNA-seq reveal environment-dependent properties of engrafted neural stem/progenitor cells.	Published	23072808	Chd2 interacts with H3.3 to determine myogenic cell fate.	Published	22569126	NGS		Epigenetics		Illumina MiSeq	九州大学医学研究院における次世代シークエンサー解析拠点の運用において利用している。九州大学ゲノム解析コンソーシアムとして、学内外のNGS利用者に公開された共同利用施設として活動を行っており、そこで得られたデータの解析をほぼすべてDDBJスパコンシステムを用いている。本年度は特に新たにSNG2解析手法とCODBJスパコンの環境依存的な比較手法を開発し発表を行った。(Maehara et al 2012, N.A.R.)今年度実績で概算してCHIPseq解析350件、RNAseq解析120件、その他エピゲノム解析100年程度の計算処理を行っており、来年度も解析需要の拡大が多く見込まれている。	BEDTools Cufflinks GATK IGVviewer MACS Picard bowtie bowtie2 bwa cutadapt samtools sra-tools tophat tophat2 自作プログラム	mpi, ChIPseq data の網羅的比較解析			遅い		混雑している時もあるが問題なく利用できる	徐々に混雑する頻度が増えているように思います。大変楽得らいい環境ですのでユーザー数は今後も増えると思いますので、もし可能でしたら一部専用キューの発行等も倉めてご検討いただけましたら幸いです。
一般研究用	岩切　淳一	東京大学　新領域創成科学研究科	RNA－タンパク質複合体の立体構造予測										シミュレーション、MD					RNA－タンパク質複合体予測をRigid body dockingのソフトウェアであるFTdockを用いて行った。FTdockなどの各種dockingソフトウェアはタンパク質－タンパク質の複合体予測向けに作られており、RNA分子を取り扱うことが出来ない。そこで、まず始めにFTdockにてTRNA分子を扱えるように改良を加えてdockingシミュレーションを行った。いくつかのベンチマークデータセットを用いて、性能評価を行ったが、データセットの約半数では、天然構造に近い構造を生成することもできなかった。また、残りの半数についても、生成した10000個規模構造の中に天然構造に近い構造は含まれていたが、これらの構造に対する予測スコアが低いため、候補構造の選択に問題があった。次のステップとして、既知のRNA－タンパク質複合体の相互作用面の情報を基に作成した統計ポテンシャルを用いて、FTdockの生成した候補構造の再スコアリングを行った。その結果、10000個の候補構造の中に天然構造に近い構造が入っている場合は、FTdockにたべて高い確率で、天然構造に近い構造に対して高い予測スコアを与えることに成功した。いくつかの構造については、スコアが最も高い予測構造は天然構造と非常に似ている(i.rMSD 4Å-5 angstrom)ことが確認できた。現在、FTdockと統計ポテンシャルの再スコアリングの組み合わせについて、ベンチマークデータセットを用いて性能評価を行い、この組み合わせの有用性を確認している。	goc perl			PDB	遅い		混雑している時もあるが問題なく利用できる	Thinノードについて、スロット数が十分に空いているにも関わらず、Jobが長い時間実行待ちになることが時々あった。この時は、スロット数ではなく、要求されているメモリの量が既に上限に達していることが多いためだと思われるが、当該時間帯に実際に使用されたメモリ量はさほど大きくないことも多かった(要求量と使用量に大きな差があった)。
一般研究用	山崎　由紀子	国立遺伝学研究所生物遺伝資源センター データベース事業部	バイオリソースに関する解析一般										その他	使用しませんでした。				利用停止申請にも必要でしょうか。私自身は前回申請後一度も利用しませんでしたが、研究室の別のメンバーがBLAST検索に使わせていただき、大変助かりました。	その他		使用しませんでした。		普通		混雑している時もあるが問題なく利用できる	利用しませんでしたので申しあげることはございません。速度や混雑状態もdefaultのままで、私の回答ではございません。
一般研究用	川本　祥子	ライフサイエンス統合データベースセンター	データベース及び解析ワークフローの研究開発										DB作成					所属するライフサイエンス統合データベースセンター(DBCLS)は、JSTの統合データベースプロジェクト基盤技術開発プログラムにおいて、セマンティックウェブ技術を利用したデータベース統合に必要な情報のカタログ化や検索サービスの開発、データベースのRDF化を行っているが、自組織の計算機だけではリソースが不足しているため、DDBJのスーパーコンピュータの利用を申請している。私個人は研究プログラムのとりまとめ役として、研究プログラムの運営管理と支援の目的でアカウントを利用させていただいております。	その他		シェルのみ		普通		混雑している時もあるが問題なく利用できる	ライフサイエンス統合データベースセンターから複数人が利用させていただいておりますが、リクエストに応じていただいたり、情報はこまめにtwitterで配信していただくなど、非常に助かっております。
一般研究用	浦高　啓太	株式会社永和システムマネジメント	RDFによるゲノム情報統合システムの開発（DBCLS統合プロジェクト参加のためのDBJPにあるRDFゲノムデータを参照する必要がある）										DB作成					研究成果・経過報告(内容入力)：文部科学省ライフサイエンスデータベース統合推進事業において、ライフサイエンス統合データベースセンターと連携して、データベース統合化のための基盤技術開発を行うために遺伝研スパコンを利用している。RefSeq、UniProt、ゲノムメタデータなどが格納されたRDFストアから、生物学的な意味単位でデータを取得し、再利用可能なかたちで可視化するためのシステム構築を行なっている。現在、限定利用可能な試作版が稼働している。	ruby			DDBJ GenBank UniProt	普通		混雑している時もあるが問題なく利用できる	root が取れる仮想マシン環境が提供されると嬉しいな、と思いました。
一般研究用	西口　直樹	株式会社 永和システムマネジメント	RDFによるゲノム情報統合システムの開発（DBCLS統合プロジェクト参加のためのDBJPにあるRDFゲノムデータを参照する必要がある）										DB作成					文部科学省ライフサイエンスデータベース統合推進事業において、ライフサイエンス統合データベースセンターと連携して、データベース統合化のための基盤技術開発を行うために遺伝研スパコンを利用している。RefSeq、UniProt、ゲノムメタデータなどが格納されたRDFストアから、生物学的な意味単位でデータを取得し、再利用可能なかたちで可視化するためのシステム構築を行なっている。現在、限定利用可能な試作版が稼働している。	ruby			DDBJ GenBank UniProt	普通		混雑している時もあるが問題なく利用できる	膨大なデータに対し処理ができるため、大変助かります。
一般研究用	岡本　忍	ライフサイエンス統合データベースセンター	ライフサイエンス分野のLinked Datoの作製と加工のため										DB作成					オントロジーの作成、ライフサイエンス分野データの RDF化やデータ加工を行った。また、RefSeq、UniProt、ゲノムメタデータなどが格納されたRDFストアから、生物学的な意味単位でデータを取得し、再利用可能なかたちで可視化するためのシステム構築を行なった。限定利用可能なプロトタイプ版を作成中。	ruby			DDBJ GenBank UniProt Pfam	普通		混雑している時もあるが問題なく利用できる	本年度は、限定的な利用でしたが、便利に使わせていただいています。ありがとうございます。
一般研究用	畠山　貴	株式会社永和システムマネジメント	RDFによるゲノム情報統合システムの開発（DBCLS統合プロジェクト参加のためのDBJPにあるRDFゲノムデータを参照する必要がある）										DB作成					文部科学省ライフサイエンスデータベース統合推進事業において、ライフサイエンス統合データベースセンターと連携して、データベース統合化のための基盤技術開発を行うために遺伝研スパコンを利用している。RefSeq、UniProt、ゲノムメタデータなどが格納されたRDFストアから、生物学的な意味単位でデータを取得し、再利用可能なかたちで可視化するためのシステム構築を行なっている。現在、限定利用可能な試作版が稼働している。	ruby			DDBJ GenBank UniProt	普通		混雑している時もあるが問題なく利用できる	手元に置けないような大きいデータを置けるので助かります。
一般研究用	伊藤　邦彦	株式会社 永和システムマネジメント	RDFによるゲノム情報統合システムの開発（DBCLS統合プロジェクト参加のためのDBJPにあるRDFゲノムデータを参照する必要がある）										DB作成					文部科学省ライフサイエンスデータベース統合推進事業において、ライフサイエンス統合データベースセンターと連携して、データベース統合化のための基盤技術開発を行なうために遺伝研スパコンを利用している。RefSeq、UniProt、ゲノムメタデータなどが格納されたRDFストアから、生物学的な意味単位でデータを取得し、再利用可能なかたちで可視化するためのシステム構築を行なっている。現在、限定利用可能な試作版が稼働している。	ruby			DDBJ GenBank UniProt	普通		混雑している時もあるが問題なく利用できる	SPARQLの実行を快適に行えるため、大変有用に活用させていただいております。

アカウント区分	使用者	所属機関	利用目的	論文名1	論文の状況1	Pubmed ID1	論文名2	論文の状況2	Pubmed ID2	論文名3	論文の状況3	Pubmed ID3	研究成果(分野)	研究成果(分野)(その他)	NGSプロジェクト名	NGSプロジェクト名(その他)	NGSプラットフォーム名	研究成果・経過報告(内容入力)	使用プログラム	使用プログラム(自作プログラム)	使用プログラム(その他)	使用データベース	アンケート-速度-	アンケート-速度-(自由入力欄)	アンケート-混雑状況-	アンケート(自由入力欄)
一般研究用	中村 一成	東京工業大学 生命理工学研究科	比較メタゲノム解析										NGS		Metagenomics		454 GS FLX Titanium Illumina Genome Analyzer IIx Ion Torrent PGM	細菌群集由来の遺伝子配列データをデータベースから取得し、それらの間で配列相同性検索を用いて、比較メタゲノム解析を行なっています。今後は遺伝子機能と系統進化の両方の用いて総合解析を行なう予定です。さらに解析すべき個別ゲノムシーケンスのデータも多数存在するので、それらのデータのアンセンブルや比較ゲノム解析等行なう予定です。	nobi,blast+ R ruby				速い		十分空いている	落雷以外では安定して使えているので良いです。
一般研究用	森 宙史	東京工業大学 大学院生命理工学研究科	ゲノム情報解析を行う上で必要となる種々の並列計算を行うとともに、大容量のメモリを使用したゲノムアセンブリ等の解析を行う。										NGS		Metagenomics		454 GS FLX Titanium 454 GS FLX+ Illumina Genome Analyzer IIx Illumina HiSeq 2000	多数の微生物のゲノム解析やメタゲノム解析のデータを配列相同性検索を用いて比較解析を行い、系統間の遺伝子共有率やメタゲノムサンプル間の遺伝子組成の類似度を計算した。また、Semantic Web技術を用いたDB群築のための、Triple Storeに対してのSPARQLクエリのテスト等を行った。今後もさらに大量なデータの解析に使用したいと考えている。	nobi,blast+ CUDA gcc perl ruby その他		SPARQL	RefSeq-Genomic NCBI-nt NCBI-nr	普通		混雑している時もあるが問題なく利用できる	使用可能なデータベースにSRAを追加してほしい。
一般研究用	河村 秀樹	東京工業大学生命理工学部生命科学科	大規模メタゲノム解析を研究しています。次世代シークエンサーから得られたデータ解析を行うため、利用を希望します。										NGS		Metagenomics		454 GS FLX Titanium Illumina Genome Analyzer IIx	本年度はジェルスクリプトを通じたジョブの実行手順に手間取ったため、シェルスクリプトの扱い方を学ぶことを優先した。今後は、ゲノムの遺伝子配列データや色各遺伝子についての情報などをデータベースから取得し、配列同士の配列相同性検索をはじめとする比較を通じて、遺伝子情報の中の特徴をつかむことに利用しようと考えている。	R gcc				普通		混雑している時もあるが問題なく利用できる	特にありません。
一般研究用	東 光一	東京工業大学生命理工学研究科	微生物比較ゲノムおよびメタゲノム解析のため。										NGS		Metagenomics		454 GS FLX Titanium Illumina Genome Analyzer IIx Illumina HiSeq 2000	シェルスクリプトによるジョブ投入など、スパコンシステムの扱いに習熟し、来年度の本格的な利用に向けての準備を行った。来年度は、現在手もとにある、NGSより得られた大量のメタゲノムデータの様々なデータベースに対しての大規模な配列相同性検索や、複数サンプル間での比較メタゲノム解析などにスパコンの計算能力を利用していく予定である。	nobi,blast+ R ruby				普通		十分空いている	特に意見はありません。スパコン利用に関して不満を感じたことはありません。
一般研究用	富樫 智章	東京工業大学 大学院生命理工学研究科 生命情報専攻	gene annotation(主にMiGaspを使用)、比較ゲノム解析										NGS		Metagenomics		454 GS FLX Titanium	様々なメタゲノムデータをアセンブルして出来たコンティグから遺伝子予測して遺伝子のリストを取得した。それらの遺伝子について様々なデータベースに対して配列相同性検索を行い、その結果から遺伝子アノテーションを行った。遺伝子アノテーションの結果を用いてメタゲノムサンプル間を比較解析し、各サンプルの遺伝子機能の特徴を解析した。	nobi,blast+			NCBI-nr	普通		混雑している時もあるが問題なく利用できる	自分が利用したいときにある程度の数のノードを使用することができ助かりました。各々の機関で実装できるサーバーでは やはり投入できるジョブの数に限りが出てきてしまうので、このような部の公共の計算サーバーがあるのは良いことだと思います。
一般研究用	吉野 弘二	東京工業大学大学院 生命理工学研究科 生命情報専攻 黒川・山田研究室	大規模メタゲノム解析の情報解析										NGS		Metagenomics		454 GS FLX Titanium Illumina Genome Analyzer IIx	公共のデータベースから環境細菌群集に関するメタゲノムデータを大量に取得し、比較解析に利用した。自身でシーケンスしたメタゲノム配列をそれらのデータに対して配列相同性検索を行い、その細菌群集の系統組成と遺伝子機能を推定し、他のサンプルとの違いを統計的に比較解析した。また、メタゲノムデータについてはアセンブルを行い、コンティグを用いた比較解析も行った。	nobi,blast+ velvet perl				普通		混雑している時もあるが問題なく利用できる	job再投入の際、データが上書きされてしまうことに不便さを感じました。
一般研究用	石原 翔真	東京工業大学大学院 情報理工学研究科 計算工学専攻	次世代シーケンサーによって得られたDNA・RNA配列の解析										NGS		Resequencing		Illumina MiSeq	SNPsや構造多型の特定を行い、一個体のゲノムの特定をより高精度に行ちパイプラインを作成しています。複数コアを利用した並列計算により、ショートリードの分類などの処理がより高速に行えるようになっています。解析結果として得られたゲノムは同研究室内で行われているRNASeqの発現量解析で利用されています。	IGVviewer Pindel bowtie2 fasta samtools gcc python ruby				普通		十分空いている	新しく発表された解析ソフトを個人的に導入する際、スパコンの方の設定が特殊で該まる事あります。環境設定の面で注意すべき点についてまとめられたページがあるとてもありがたいです。
一般研究用	Palomares-Rius Juan Emilio	Division of Parasitology, University of Miyazaki	The main use of the supercomputer has been related to bioinformatic work using nematodes.										NGS		RNASeq		Illumina HiSeq 2000	The main use of the supercomputer has been related to bioinformatic work using nematodes. The platform has been used from mapping and assembling Illumina reads in Next-generation sequencing experiments. Two main areas of research have been conducted: i) mapping DNA read to an assembled genome and ii) Differential gene expression in different dauer stages. Both experiments are in data analysis. We will work in similar experiments next year and the use of the supercomputer constitutes an invaluable tool for this research.	GATK Picard bowtie bowtie2 bwa nobi,blast+ samtools sra-tools R			GenBank NCBI-nt	普通		混雑している時もあるが問題なく利用できる	It's fine for me
一般研究用	保田 弘紀	九州大学生体防御医学研究所	毒蛇ハブにおける加速進化機構の解明と創薬シーズ開拓を目指して、ハブのオーミクス解析及び遺伝的多様性の解析を行っている。特に、全ゲノム配列決定のための新規アセンブリに必要な膨大な計算資源を確保するために本システムの利用を申請する。										NGS		Whole Genome Sequencing		454 GS FLX Titanium Illumina Genome Analyzer Illumina MiSeq	日本固有の毒蛇ハブのオミクス研究の一貫として、ハブの全ゲノムシークエンシングを進めている。ショットガンおよびメイトペアのリードを用いた新規アセンブリには膨大な計算資源が必要となるため、NIGのスパコンの登録使用を開始した。しかし2013年2月からの使用開始したばかりであり、報告すべき成果はまだない。	SOAPdenovo bowtie velvet perl				普通		混雑している時もあるが問題なく利用できる	大きなメモリを必要とする計算はそのたびに申請しなくてはならないので簡便化できるとよいと思います。
一般研究用	佐藤 賢二	金沢大学 理工研究域 電子情報学系	ショートリードデータやアミノ酸配列データなどのバイオデータに対する機械学習アルゴリズムの適用可能性について研究を行う										アルゴリズム開発					登録時期が遅かったこともあり、今年度はあまり本格的に利用できなかった。研究室ではPCクラスタを用いてショートリードデータの解析を行っており、今年度はMetavelvetの性能測定や、特殊な gene fusion の発見などについて、予備的な研究を行った。研究成果はまだ論文になっていないが、本学の年末論文や修士論文に実験結果をまとめた。	bowtie bowtie2 bwa ncbi_toolkit samtools velvet			RefSeq-Genomic RefSeq-RNA NCBI-nt NCBI-nr	普通		混雑している時もあるが問題なく利用できる	まだ本格的に使っていないが、使い勝手は良い(東大医科研のスパコンと似ている部分があるからもう慣じるのかもしれない)。
一般研究用	石井 孝司	国立感染症研究所 ウイルス第二部	ウイルス配列情報の解析のため										NGS以外の配列解析					研究の目的とその社会的意義 本研究では、アジアにおけるA型肝炎の流行状況を用いて解析し、これらの国での疫病対策に貢献すること。また、日本へのHAVの流入状況を分子疫学的手法を用いて解析し、日本でのA型肝炎対策に役立てることを目的とした。共同研究先としてスリランカ、ファリピン、韓国を対象国とした。方法スリランカとの共同研究 スリランカでは長年にわたって続いていた内戦が2009年に終結したが、その直後、反政府勢力が去個していたため難民等で多くのC1A型肝炎と見られる急性肝炎の流行が起こった。患者数は1万5千人を超えたが、内戦直後の混乱のため検体の収集も確定診断もできなかった。本研究では、スリランカでのA型肝炎の流行状況の解析や疫病対策への協力、また、比較的抗体保有率の低い地域にウイルスが流入した場合の感染伝播モデルの解析を目的として、スリランカ北部で発生したA型肝炎についてHAV遺伝子の系譜解析を行った。ファリピンの共同研究 2010年に日本で広域流行したhAV株が2007年にファリピンで分離された株と同一クラスターに属していたことから、本流行がファリピンから流入である可能性が示唆された。マニラ市内の河川水を調査、RNA抽出を行い、RT-PCR法でHAVを検出した。HAV株の系譜解析を行った。韓国との共同研究 2006年から韓国では A型肝炎の大きな流行が見られた。その際にgenotype IIIとgenotype IIIA(HGIIIA)にシフトしていることが判明している。日本でも近年は東で流行がみられ、2010年から検出されるようになって来ており、韓国の流行との関連が示唆される。本研究では2009年にソウル市内で得られたHGIIIA株のRT-PCR法による増幅産物を用いて、その大部分を占めるgenotype IIIA全日本でDDBJへ登録された1株と比較し系統解析を行った。成果スリランカとの共同研究 遺伝子型別が可能であったA1株は遺伝子型(HVプロレーブ1-3)、即、IIIAマウプ(ヴァーブ2)にクラスタリングされた。調査の初期は東アジアに発生するIIIAが多く検出されたが、徐々にIA1が優勢に変化した。IA1は調査初期から10日以1件のペースで報告されていたが、30-39日目に急増し、10人の患者のうち、3人がIA1で全ての人は トレーニングセンター勤務中であった。さらにA型肝炎の潜伏期間である約一ヶ月後(30-39日目)に再度患者が増加し、この時の患者の殆どはIIIのエリアに属していたがIIIでは、トレーニングセンターでIA1による A型肝炎が流行し、不顕性感染、もしくは発症者の高さがトレーニングを終えて各地域に配属されたことによって広範囲にIA1が広がったと考えられる。今回の解析は抗体保有率が低い集団において見られる流行拡散パターンを明瞭に示したものと言える。ファリピンとの共同研究 ファリピンのマニラ市内の河川から-からすべてHAV遺伝子が検出され、ウイルスが常在していることが明らかとなった。検出されたHAV株はすべてgenotype IAであり、さらにコトのマプクラスター(1-10)に分類されると考えられる。日本の2010年流行株はIA1に分類され、日本に常在しているIAとは明らかに異なる。本クラスターに属する株はフィリピン河川から1日で検体検体出されるが日本では検出されなくなった。何らかの理由で2010年に日本に流入した本株は、日本で2010年に広域流行を起こした後は定着することなく消滅したと推定される。今後もフィリピンを注目にめぐする A型肝炎流行地からのウイルス流入には注意が必要である。韓国との共同研究 系統解析の結果、韓国のIIIA株と日本のIIIA株とは同一のクラスター上に存在することが証明され、日本のIIIA株は韓国の流行と密接な関係があることを示唆する結果となった。流行地と日本との交流が密接な場合の病原体流入のリスクのモデルとなる結果 が示唆された。今後、詳細な流入経路や流入の時期方法について、分子疫学的手法を用いて解析を行うことを計画している。	ClustalW phylip			DDBJ GenBank	普通		混雑している時もあるが問題なく利用できる	特に問題なく使用できています。
一般研究用	原 雄一郎	産業総合研究所・BIRC	分子進化学および集団遺伝学のための配列解析										その他	その他				本年度はNIG Supercomputerを使用しておりません。並列計算機は研究において必要でしたが、所属機関の並列計算機を用いて研究が進められたため、NIG Supercomputerを使う必要が生じなかったからです。来年度以降に使用していきたいと考えており、アカウントの継続を望みます。よろしくお願いいたします。	perl				普通		混雑している時もあるが問題なく利用できる	来年度以降使用していきたいと考えております。
一般研究用	星山 律哉	国立遺伝学研究所 構造遺伝学研究センター 細胞建築研究室	情報システム研究機構、新領域融合研究センターの研究プロジェクト、データ同化に関する計算を行う。細胞形態の細胞質流動を駆動する力分布を推定する。										シミュレーション、MD					データ同化法を用いて細胞内部既の細胞質流動を引き起こすその分布を推定するプログラムを開発した。このプログラムでは2次元格子法のアルゴリズムを採用することにより、1データあたり数時間で計算を行う事ができるようになった。今後はこの開発済みのプログラムを多くの個体に適用する予定である。これにより細胞内で発生していると仮体ごと個別に推定する事を目指す。	gcc 自作プログラム	mpi			普通		混雑していて不満だ	uのショートカットキーを充実させてほしい。具体的には●uを押してundoを繰り返せるようにしてほしい。●backspaceを受け付けるようにしてほしい。●上下左右のボタンを押した時に上下左右に動くようにしてほしい。
一般研究用	加藤 毅	群馬大学工学部	生命情報分野における大規模データ解析技術の開発										アルゴリズム開発					I have analyzed a data of a huge number of biological sequences which are downloaded from a public database. For the analysis of the huge set of sequences, I used BLAST, to compare these sequences to amino-acid sequences in the NCBI non-redundant protein database. I tried to develop statistical data analysis tools for the results of the BLAST process.	nobi,blast+ gcc java perl python			NCBI-nr	普通		混雑している時もあるが問題なく利用できる	None.
一般研究用	湯野川 春信	株式会社メイズ	徳島文理大学との共同研究、公開ヒトWholeゲノムシーケンス配列から、一部配列を抽出し、アセンブルを行って、未知有用情報を抽出するために利用する。										その他	公開ヒトWholeゲノムシーケンス配列の一部配列から新有用情報抽出				公開されているヒトゲノム配列から一部の配列を抽出し、それをアセンブルするために、配列量とパラメータの調整を行っている。利用しているアセンブラは、インストール済みのVELVETである。その結果、配列量や様々のパラメータの値を変えて、以前行ったプレの解析のおよ倍程度の配列量を、予定された時間内に終了するための条件をみいだした。今後は、このパラメータで、1000genomeプロジェクトのヒトDNA配列の一部を利用して、アセンブルを40セット行う予定であり、さらに、その後に、得られたContig配列群にアノテーションを付与し、データベースに投入して候補の絞込を行う予定である。	velvet				普通		混雑していて不満だ	スパコンが利用できることによって、本研究が可能になりました。たいへん感謝しております。ただ、想定していたよりも、リソースのことを良く検討しなければならなかったことについては、こちらで反省しなければならないことが と思っております。今後も、スパコンが同じような条件で利用できるよう お願いしたいと思います。
一般研究用	竹内 史比古	国立感染症研究所	病原体の次世代シーケンス解析、比較ゲノムおよび de novo アセンブル。										NGS		Whole Genome Sequencing		Illumina Genome Analyzer II Illumina MiSeq	2013年2月28日の時点では、まだ配列解読中であり、スパコンは利用していない。配列解読を終えた後に、アセンブルに取り掛かる予定である。配列解読は、イルミナのGAIIxとMiSeqで行っている。DNA の de novo アセンブルに加え、RNA-seq の de novo アセンブルも行う予定である。	その他		ABySS		普通		混雑している時もあるが問題なく利用できる	今後使わせて頂く予定です。利用の機会を与えて頂き有り難うございます。
一般研究用	奥田 喜弘	国立遺伝学研究所 DDBJセンター	DDBJ業務での大規模データ処理										DB作成					旧スパコン上に置かれていた、特許庁からDDBJへ送付されたJPO形式の塩基配列・アミノ酸配列のファイルをすべて集め、アクセッション番号ごとにエントリーを対応させたデータセットを作成した。その際、DDBJ側でEF作成時に行った修正(アクセッション番号、公報番号/配列番号の変更、その他)をデータに適用した。	perl				遅い		十分空いている	ディスクアクセスが遅いのがありがたいです。
一般研究用	鈴木 郁美	国立遺伝学研究所 生命情報研究センター 発現解析研究室	DNAデータベースのよりよい検索方法の研究を行うため										シミュレーション、MD					DDBJ事業の一環として、DNAデータベースのよりよい検索方法を探索する調査 研究を行った。DNAデータベースは登録件数が大規模なデータであるため、まずデータ構造を俯瞰する調査、研究を行った。各Flatfile featureの有無によりテーブル構造を設えた特徴ベクトルを作成し、意図的な分類基準ではなく、データ構造から客観的に見える基準の探索を行った。	R ruby			DDBJ	普通		混雑している時もあるが問題なく利用できる	特になし
一般研究用	遠野 雅徳	独立行政法人農業・食品産業技術総合研究機構畜産草地研究所	国立遺伝学研究所大量遺伝情報研究室との共同研究推進と、スーパーコンピューターシステムのツールを利用したいため										NGS		Whole Genome Sequencing		454 GS FLX Titanium Illumina Genome Analyzer IIx Illumina HiSeq 2000	有益微生物である乳酸菌の新種のゲノムシーケンス及びリシーケンスを行うことにより、これらの乳酸菌の機能のゲノム情報を抽間させ、有益微生物の発展的利用に繋げることを目的とした。これまでに、複数の手法によりアセンブルやマッピングを実施し最適な条件を十分に検討した後に、アノテーション作業を行っている。これまでの所、研究計画通りに順調に研究が推進されている。	ClustalW Cufflinks bowtie			DDBJ GenBank RefSeq-Genomic	普通		混雑している時もあるが問題なく利用できる	とても快適に利用できており、引き続きサービスの継続を希望したい。
一般研究用	清水 顕史	滋賀県立大学環境科学部生物資源管理学科	作物由来の次世代シーケンスデータの解析	Discovery of high-confidence SNPs from large-scale de novo analysis of leaf transcripts of Aegilops tauschii, a wild wheat progenitor	Published	23125207							NGS		RNASeq		454 GS FLX Titanium	2倍体および4倍体のムギRNA-seqを、それぞれ2系統ずつ用い、de novoアセンブル(newbler)を使用による参照遺伝子ライブラリーOn silicoの構築を行っている。得られた参照遺伝子ライブラリーをreferenceにしたリードマッピングにより、系統間のSNPやInDel変異が検出できる。多型が得られた参照遺伝子配列は、オオムギゲノムにアライメントして仮想地図の位置情報を入手でき る。一部の多型情報に基づいて作成したマーカーは、ムギ類遺伝地図の該当する領域に座乗することも確かめられた。	その他		newbler		普通		混雑している時もあるが問題なく利用できる	newblerのバージョンがよがってから、計算時間が長くなった。スパコンの使用は、計算を投げっぱなしになっているため、こちらの操作にも改善の余地があるかもしれないが快適ができていない。今後は、TrinityやOasisなどを利用してみるかもしれない。
一般研究用	張 昊	国立遺伝学研究所 生命情報研究センター セルイノベーションプロジェクト	セルイノベーション研究										アルゴリズム開発					full analysis of MeDIP-seq data, including sequence alignment, quality control (QC), and determination and annotation of DMRs.I am trying to utilizes many applications in order to perform this task. From alignment (BWA) and subsequent filtering (SAMtools), through generation of numerous quality control metrics (FastQC and MEDIPS), DMR identification (using USeq and in turn DESeq) and finally preliminary annotation of the DMRs (utilizing the capabilities of BEDTools).	ALLPATHS-LG BEDTools bwa samtools openmpi R java 自作プログラム	mpi			普通		混雑している時もあるが問題なく利用できる	とてもいいものです。Thank You Verymuch
一般研究用	香見 忠則	鳥取大学農学部	MiGapの利用	Stimulation of glucoamylase gene expression with starch in Pholiota microspora.	Published								その他	DNA塩基配列の解析				Pholiota microsporaのグルコアミラーゼ遺伝子(PnGlu1) を同定した。アミノ酸をコードする領域は、1743 bpであり、581-amino酸をコードしていた。その内、N末端の最初の17アミノ酸は、シグナルペプチドと考えられた。本タンパク質は、糖質加水分解酵素ファミリー15所有の配列及びC-末端に炭水化物結合モジュール（CBM20）を 見出すことができた。PnGlu1の遺伝子のダイナリオンの菌糸体中の発現量を核々なデンプン（可溶性、トウモロコシ、サツマイモ、ジャガイモ）の入った異少培地で検証し、逆転写定量PCRで調べた。その結果、PnGlu1の転写は、可溶性デンプン、トウモロコシデンプンで顕著であった。	phylip			DDBJ	普通		混雑している時もあるが問題なく利用できる	ほとんど利用していないのでわかりません。
一般研究用	秋山-小田 康子	JT生命誌研究館	次世代シーケンサーによって得られたデータの解析を行うため利用を希望します。ゲモを中心とした脊椎動物のゲノムDe novoアセンブリと発現解析を行う予定です。										NGS		Whole Genome Sequencing		Illumina MiSeq	1月末に登録しましたが、まだ利用できていません。オオヒゲダモ(ゲノムサイズ1.37G)のゲノムシーケンスを行いIllumina MiSeqで4ラン、C400ほどのデータを得ることができました。今後インサートサイズの異なるライブラリーをさらに作製し、同程度の量のデータを取得したのちに、まずアセンブリに利用する計画です。	SOAPdenovo velvet				普通		混雑している時もあるが問題なく利用できる	SOAPdenovoはバージョンアップされているようですが、更新される予定でしょうか？
一般研究用	西村 理	独立行政法人理化学研究所 発生・再生科学総合研究センター ゲノム資源解析ユニット	次世代シーケンエンサーのデータを元にした、バイオインフォマティクス解析ならびに新規ソフトウエアの解析										NGS		Epigenetics		Illumina HiSeq 1000	NGSから産出される大量データに基づき、定量的なChIP-seq解析を可能にするための各種アルゴリズム、データマイニングの条件を検討。非モデル生物におけるリファレンスなしでの定量的なRNA-seqを実現するために、完全de novo アセンブリもリファレンスとする場合とトランスクリプトームデータを加味した場合の比較検討。	BEDTools BLAT Cufflinks GATK IGVviewer MACS Picard SOAPdenovo Trinity bowtie bowtie2 bwa cutadapt maffft nobi,blast+ ncbi_cxx_toolkit nobi,toolkit phylip samtools sra-tools tophat tophat2 libtool R gcc java perl			GenBank RefSeq-RNA RefSeq-Protein NCBI-nt NCBI-nr NCBI-nr dbEST UniProt Swiss-Port TrEMBL Pfam	普通		混雑している時もあるが問題なく利用できる	データ量やユーザー数増加に伴ったディスク容量、利用可能メモリの継続的な増強を期待します。またメンテナンス・障害によるダウンタイムが多い気がします。
一般研究用	友近 裕一	個人	次世代シーケンサーデータ解析										その他	公開DB,オープンアクセス雑誌をもとにしたNGS解析、ゲノム解析				バイオインフォマティクス学習の一環としてNGS解析、ゲノム解析を予定していたが、自宅のLinuxサーバーで計算力が不十分なためDDBJを利用申請した。しかしながら、他の本業や学習に時間をとられ、申請後は数回ログインして、最低限のコマンドを叩いただけでほとんど使っていない。来年度以降にだけ、公開されているNGSデータやオープンアクセス論文を中心に、実験的な一連の解析計画をたてて学習に用いる。	bowtie bwa R perl ruby				普通		十分空いている	まだほとんど利用していませんが、来年度以降、ジョブを投入して実際的な解析作業をやっていく予定です。
一般研究用	太田 博樹	北里大学医学部	NGS解析のため										NGS		Population Genomics		Illumina Genome Analyzer II	昨年10月に申請したばかりで、まだ研究室の体制を整えているところです。ラボではチームで行っており、今ボ内の端末の設置からNGSのためのライブラリー作りなども含めて、全体の体制が70-80%くらい整い、データ上がって来ている状態です。データとはIllumina GA II でフローセル3枚と流したreadのことです。具体的な成果はありません。	BLAT ClustalW bwa fasta			DDBJ GenBank EMBL PDB	普通		混雑している時もあるが問題なく利用できる	お世話になります。
一般研究用	矢野 雅大	東京工業大学大学院 生命理工学研究科 生命情報専攻 黒川山田研究室	大規模塩基配列解析(メタゲノム)GPU使用予定										NGS以外の配列解析					腸内細菌群集より次世代シーケンサを用いて得られた配列サンプル同士の類似度を、DDBJスーパーコンピューターでBLASTを可能な限り並列して走らせることによって、算出した。この算出された類似度を参考にしながら、多数のサンプル間で比較メタゲノム解析を行った。現在、比較解析結果を基に、新たなメタゲノムデータ比較手法を開発中である。	その他		blast		普通		混雑している時もあるが問題なく利用できる	おおむね快適に使用できた。CPUの速度には不満がない。しかし、GPUを用いた計算をしようとした場合に、--monthオプションをつけなければならないなど、手間があった。また、qloginをしない状態でのジョブ投入は、分かっていてもうっかりやってしまうことがあるので、システム側で不可能にしてもらいたい。
一般研究用	丸山 徹	早稲田大学 先進理工学部 生命医科学科	メタゲノムデータの解析に用いる										NGS		Metagenomics		454 GS	海洋サンプルや土壌サンプルに由来するメタゲノムと、その他シーケンスデータの解析を行っている。メタゲノムデータをデータベースに対して相同性検索することで、微生物群が保有する機能遺伝子や、微生物群を構成する種について推定した。また、現在は、ターゲットである微生物群が保有する特徴を抽出するために他のメタゲノムデータとの比較解析を行っている。	BLAT nobi,blast+ python				普通		十分空いている	混雑する事も無く非常に快適に使用する事ができている。マニュアルやホームページも分かりやすく、安心して使用する事ができている。

アカウント区分	使用者	所属機関	利用目的	論文名1	論文の状況1	Pubmed ID1	論文名2	論文の状況2	Pubmed ID2	論文名3	論文の状況3	Pubmed ID3	研究成果(分野)	研究成果(分野)(その他)	NGSプロジェクト名	NGSプロジェクト名(その他)	NGSプラットフォーム名	研究成果・経過報告(内容入力)	使用プログラム(自作プログラム)	使用プログラム(その他)	使用データベース	アンケート-速度-	アンケート-速度-(自由入力欄)	アンケート-混雑状況-	アンケート(自由入力欄)
一般研究用	山崎 千里	産業技術総合研究所 バイオメディシナル情報研究センター	ヒトmRNA配列解析、DDBJへのデータ転送										DB作成					H-InvDB(http://www.h-invitational.jp/)はヒトの遺伝子と転写産物を対象とした統合データベースです。ヒトのすべての転写産物の配列をあらゆる手法で解析することにより、ヒト遺伝子の構造、選択的スプライシング/バリアント、機能性RNA、タンパク質としての機能、機能ドメイン、細胞内局在、代謝経路、立体構造、疾病との関連、遺伝子多型(SNP、マイクロサテライト等)、遺伝子発現プロファイル、分子進化学的特徴、タンパク質間相互作用(PPI)、遺伝子ファミリー などの精査されたアノテーション(注釈付け)情報を提供しています。このH-InvDBのデータを無償でユーザに公開するFTPサイト構築を行いました。	BLAT ClustalW InterProScan fasta perl 自作プログラム	データベース構築	DDBJ	普通		混雑している時もあるが問題なく利用できる	良質なリソースを使用させて頂きありがとうございます。
一般研究用	竹原 潤一	東京工業大学大学院 生命理工学研究科 生命情報専攻	微生物統合データベース「MicrobeDB.jp」に関する研究開発のため										DB作成					微生物統合データベースの構築のために、セマンティックWEB技術を用いて様々な微生物に関するデータを統合し、DDBJスパコン内のSPARQL EndpointにRDF及びオントロジーデータをimportした。それらのデータ及びendpointを用いて、現在セマンティックWEB技術を用いたDB構築のためのデータ構造の最適化を行なっている。	R perl ruby 自作プログラム	SPARQL: RDFデータの検索		普通		混雑している時もあるが問題なく利用できる	SRAのデータを使用可能なデータベースとして提供して欲しい
一般研究用	竹本 龍也	大阪大学・大学院生命機能研究科	ChIP-seq等で得られたデータの解析										DB作成					転写因子による遺伝子発現制御に関する研究は、次世代シーケンサーの活用により、飛躍的に進んだ。しかしながら、その情報量が多いため、その情報を処理するのは困難である。私は、神経板で発現するSox2遺伝子がどういった転写因子によって制御され、また、SOX2が転写因子としてどういった遺伝子を制御しているのかを解析している。これまでに報告されている、ES細胞における遺伝子制御ネットワークのデータを解析したところ、神経板でのSox2遺伝子発現を担うエンハンサーが、ES細胞で複数の転写因子セットによって標識されていることが分かった。このエンハンサーはES細胞ではエンハンサー活性を示さないことから、この転写因子セットによる結合は興味深い。現在、この結合がどういった意味を持つのか解析を行っている。	BLAT			普通		混雑している時もあるが問題なく利用できる	とても満足しています。
一般研究用	佐々木 直文	国立遺伝学研究所 生命情報研究センター 五條研研究室	1. 次世代シークエンサーデータの解析(アセンブリ、マッピング、アノテーション)2. HPCIアーキテクチャー用の配列解析用並列プログラムの開発	The First Symbiont-Free Genome Sequence of Marine Red Alga, Susabi-nori (Pyropia yezoensis)	In Press								NGS		Whole Genome Sequencing		454 GS FLX+ Illumina Genome Analyzer IIx	乾燥ノリ(浅草海苔)の原料の一つである紅藻スサビノリ(<i>Pyropia yezoensis</i>)は、日本における重要な食業生物であり、養殖における品質管理や大独種との品種識別の上でゲノム情報は基礎的に非常に重要な情報である。本研究では、実験的に単離されたsix>P yezoensis</>の培養株を使用し、Roche 454およびIllumina GA 両から得られたリードを用い、Mate-Pairが広い条件でのAb<> novo</> WGS assembly を行った。その結果、約45Mのゲノムのドラフトアセンブリが得られ、これに対するアノテーションの結果から本種における遺伝子セットを決定することに成功した(<>PLoS ONE</> in press)。また、得られた遺伝子の系統樹解析から、葉緑進化に関する知見を得ることができた。	ALLPATHS-LG BLAT Cufflinks GATK Oases Picard RAxML SOAP3 SOAPdenovo Trinity bowtie bowtie2 bwa mafft ncbi.blast+ ncbi.toolkit phylip samtools tophat tophat2 libtool libxslt mpc mpfr openmpi CUDA R gcc perl python 自作プログラム	mpi, p-thread, (gpu)	DDBJ GenBank RefSeq-Genomic RefSeq-RNA RefSeq-Protein NCBI-nt EMBL UniProt Swiss-Prot T-EMBL PDB Pfam	普通	1. 京都大学化学研究所スーパーコンピューターシステム(2001-2004)比較的遅かった。スケーリング性に乏しい。2. 東京大学付属医科学研究所ヒトゲノム解析センター スーパーコンピューター(2006-2010)比較的遅かった。スケーリング性は普通。3. 理研計算科学研究機構 京スーパーコンピューター(2011-)アーキテクチャが独自であるため比較が困難だが、高並列度ならばきわめて速い。	混雑している時もあるが問題なく利用できる	HPCI利用推進の観点から、入り口の制限された、ハイ・パフォーマンスコンピューティングコンソーシアムを設立した計算機資源利用と比較し、広く研究者一般に公開されているのはスーパーコンピュータ校は共用計算機資源の提供は貴重であり、今後とも維持管理、そして充実に期待する。
一般研究用	富所 布紗乃	東京大学・株式会社ダイナコム	大規模遺伝子解析の研究室に所属しています。次世代シーケンサーによって得られたデータの解析を行うため利用を希望します。										NGS		Transcriptome Analysis		Illumina Genome Analyzer IIx Illumina HiSeq 2000	次世代シーケンサーによって得られたデータの大規模遺伝子解析として、RNA−seqやChIP−seq、small−RNA−seq、4Cアッセイの集計などを行いました。また、エキソームシークエンス解析のSNP/INDELコールについて、パイプラインの見直しやパラメーターチューニングをテストデータを用いて行っています。	BEDTools GATK IGVviewer MACS Picard bwa samtools R perl		DDBJ RefSeq-RNA EMBL	普通		混雑している時もあるが問題なく利用できる	東大医科研のスパコンと機成等が似ていて使用感がほぼ同じで使いやすかったです。
一般研究用	坂本 盛宇	国立遺伝学研究所 生命情報研究センター 遺伝情報分析研究室	GTOPデータベース更新作業及び、GTOP解析の依頼に対応するための計算作業										DB作成					GTOPデータベースの更新作業として、主にホモロジー検索を実行しました。今回はファージ、古細菌、原核生物、真核生物合わせて約2000種、シーケンス数約1000万件について解析を実行しています。スパコンとしての計算作業は体は安定しており、今年度、もしくは来年度頃には計算結果を元にデータベースを改定の予定です。まだ未定ですが、来年度も次更新に向けての作業が発生します。	ncbi.toolkit gcc perl 自作プログラム	アミノ酸配列解析(繰り返し予測、2次構造予測など)が目的。シングルで実行され、言語はC。	DDBJ GenBank RefSeq-Protein NCBI-nr UniProt Swiss-Prot PDB Pfam	普通		十分空いている	CPUの実行スピードに問題は無いと思います。arrayを用いたJOB投入といった手法で作業が簡素化されたと思えますが、さて、作業内容として大量のファイルの生成するの要があります。ほとんどのJOBは、なるべく出力ファイルが少なくなるよう対策しましたが、どうしても対策不可能なJOBもあります。もしスパコンで大量ファイルに対応していただけると大変助かります。
一般研究用	鐵田 葉子	総合研究大学院大学・先導科学研究科	シミュレーション実験を行うため										シミュレーション、MD					昨年12月に登録した、これまで、DESKTOPで走らせていた、プログラムは1つのパラメーターの組み合せで、20replication 行うのにおおむね1ヶ月かかった。そこで、このプログラムの実行時間の短縮間化を目指している。現在は、短時間化に必要なプログラム技術の習得と、プログラムの改良を行っている。	自作プログラム	multithreadを使用する予定。		普通		混雑している時もあるが問題なく利用できる	まだ、本格的に使用していないので、特筆することはないです。
一般研究用	重元 康昌	富士通株式会社 TCソリューション事業本部	微生物ゲノム解析(GTPSデータベースの更新)										NGS以外の配列解析					来年度に微生物ゲノムデータベースGTPSの更新に向けて大規模計算をスパコンで実施させていただく予定です。GTPSでは、微生物ゲノム配列に対して再アノテーションを行うためDBLAST やFisherProScanの計算を実施します。今年度はジョブの投入方法や小規模データを用いた解析方法の調査を実施させていただいています。	nobi.toolkit その他	interproscan	DDBJ	普通		十分空いている	とくにありません。
一般研究用	嶋田 透	東京大学 大学院農学生命科学研究科 昆虫遺伝研究室	カイコのBrdおよびsoi変異体は、幼虫の食性が異常となる。変異体の感覚器官のトランスクリプトームをRNA-seqで解析し、両遺伝子の機能発現の機構を解明する。また、寄主植物の異なるカイコ近縁種における感覚器官や消化管のトランスクリプトームをRNA-seqで比較解析する。										NGS		RNASeq		Illumina Genome Analyzer IIx Illumina HiSeq 1000	カイコ食性変異体の成虫触角、幼虫触覚、小腐等のトランスクリプトーム をIlluminaのRNA-seqで解析し、正常種との遺伝子発現の差異を調査した。また、イチジクカマンなど寄主植物の異なるカイコ近縁種における感覚器官 や消化管のトランスクリプトームをRNA-seqで比較解析し、差異のある遺伝子を発見した。	bowtie ncbi.blast+ perl			普通		十分空いている	計算速度、データ許容量とも満足している。
一般研究用	田中 千尋	京都大学農学研究科	植物共生真菌類のゲノムアッセンブリと解析。現在、外生菌根類ホンシメジならびに近縁の非菌根性菌ハタケシメジのゲノムシーケンスを行っております。これらのゲノムデータのアッセンブルを行い、相互に比較するために利用を申請いたします。										その他	ゲノムシーケンスのアッセンブル作業に至る前段階				外生菌根菌ホンシメジLyophyllum shimeji(ならびに近縁の非菌根性菌ハタケシメジLyophyllum decastes)の子実体から単胞子分離株を作出した。それぞれの種において同時培養を行い交配型を調べるとともにコンパチブルな系統を選抜した。それぞれの種において生育の旺盛な1系統ずつからゲノムDNAを抽出した。現在、HiSeq2000によるゲノムシーケンスを行っている最中である。	SOAPdenovo			普通		混雑している時もあるが問題なく利用できる	まだ、十分活用できていないので、フィードバックできません。どうかお許しください。
一般研究用	藤村 啓史	北里大学医学部	NGSを用いたゲノム比較解析のため										NGS		Resequencing		Illumina HiSeq 2000	現在、次世代シーケンサーを用いた全ゲノム比較解析を行うため、リシークエンスデータの取得を待っている段階である。今後、研究対象とする小型魚類の4系統に関して、Illumina HiSeq 2000によって取得されたリシークエンスデータをスパコンへデータ転送する。そして、比較ゲノム解析を行い、対象とする小型魚類の多型情報を得る。	perl			普通		混雑している時もあるが問題なく利用できる	まだ、次世代シーケンサーのリシークエンスデータ取得を待っている段階の為、特に要望等はまだない。
一般研究用	松井 徹	琉球大学熱帯生物圏研究センター	新規分離微生物ゲノムのアノテーション等										NGS		Whole Genome Sequencing		454 GS Junior	環境汚染物質の分解を目的として土壌より分離した放線菌、細菌等のde novo 解析データのアノテーションを行い、ポリケタイド合成酵素、非リボソームペプチド合成酵素、アルカンオキシゲナーゼ、芳香族オキシゲナーゼ等の遺伝子データを抽出した。また、DNA情報に関する遺伝子Pseudo データ抽出を行い、ベクター構築のための基本データとした。	ClustalW		DDBJ	普通		十分空いている	特になし
一般研究用	二階堂 愛	理化学研究所 発生・再生科学総合研究センター	大量DNAシーケンスデータの解析										NGS		RNASeq		Illumina HiSeq 1000 Illumina HiSeq 2000	ChIP-Seqのデータ解析パイプラインを構築し、実際に運用した。特にpeak caller である GPS の性能テストを行った。このパイプラインを利用したChIP-seq の論文は、現在、投稿中でリバイズ中である。また、1細胞RNA-Seqの実験技術の開発が終了し、そのデータ解析パイプラインを構築した。こちらも論文は投稿済みで、現在マイナーリバイズ中である。	BEDTools Cufflinks MACS bowtie2 samtools tophat2 R gcc java perl python ruby		RefSeq-RNA	普通		混雑している時もあるが問題なく利用できる	バイオインフォマティクス、特に配列解析を想定した作りのスパコンで公開されている他になく、このようなユニークなサービスを今後も提供を続けて欲しい。
一般研究用	大島 一正	京都府立大学大学院 生命理境科学研究科	ゲノムデータ (de novo genome assembly, de novo RNA-seq 等) の解析										NGS		RNASeq		Illumina HiSeq 2000	穂食性昆虫における寄主転換の遺伝基盤を解明するため、利用する寄主植物に関して種内多型が見られる昆虫種を利用し、幼虫における発現解析を行った。研究材料には、クルミ科植物を利用するレーンとツツ科を利用するレーンが棲息に見られる蛾昆目ホソガ科のクルミホソガAcrocercops transecta を用いた。あらかじめ、NGSによりドラフトゲノムを構築し、AFLPを遺伝マーカーとした QTL 解析の結果と照らし合わせることで、目的遺伝子近傍のゲノム配列を取得した。RNA-seqのデータも合わせて解析したところ、目的遺伝子近傍のゲノム領域に、2つのレース間で明瞭に発現様式が異なる領域が見られた。	Cufflinks Trinity tophat2		GenBank	普通		混雑している時もあるが問題なく利用できる	大変助かっております。来年度以降、解析量が増えますので、ホームの容量を上げさせていただこうかと思っております.
一般研究用	川口 翔太郎	金沢大学大学院自然科学研究科電子情報科学專攻	形態素解析を用いたタンパク質モチーフの発見										NGS以外の配列解析					初めに、参考研究の追試を行うために、教師なし形態素解析器のアルゴリズムの実装と自然言語への適用を行いました。期待通りに動作したため、タンパク質に対しても適用してみたところ、精度は落ちるものの概ね期待通りの結果が得られました。今後は、精度向上のために、タンパク質の特性を考慮したアルゴリズムの改良を行っていきます。	gsl gcc perl python			普通		混雑している時もあるが問題なく利用できる	初期パスワードの末尾が「;」だったのですが、パスワードの一部とは思わず、何度も入力に失敗してしまいました。末尾に記号は紛らわしいのでやめた方が良いと思われます。
一般研究用	山本 希	東京工業大学大学院 生命理工学研究科	大規模メタゲノム解析のため										DB作成					自作のDBを用いて、目的の結果が得られるようなクエリの作成、およびSPARQL検索を行った。また、ほかのDBのエンドポイントを使った検索により、他のDBと組み合わせた検索を行うことが可能となった。今後は想定される検索に対応できるように、DB検索時間の短縮や、DBそのものの充実を測って行く予定である。	perl			普通		十分空いている	高速で使いやすい
一般研究用	阿部 貴志	新潟大学 自然科学系	メタゲノム配列を対象にした生物系統推定とタンパク質機能推定のための連続塩基、ならびに、連続アミノ酸の出現頻度に基づく一括学習型自己組織化マップ解析										NGS以外の配列解析					メタゲノム配列を対象にした生物系統推定に使用するために、現在公開されている全生物ゲノムのゲノム配列データを用いて、BLSOM解析を行った。また、植物由来の2次代謝関連酵素タンパク質の特徴抽出目的に、公開されている全植物と草植生物由来のタンパク質アミノ酸配列を対象に、200アミノ酸ごとに断片化した配列の2連続アミノ酸頻度(400次元)に基づくBLSOM解析を行った。	自作プログラム	使用した技術:MPI目的:BLSOM解析プログラム	DDBJ GenBank UniProt Swiss-Prot	普通		混雑している時もあるが問題なく利用できる	マニュアル類も整備されており、快適に利用することができた。
一般研究用	野原 祥夫	三菱スペース・ソフトウエア株式会社	ゲノム・メタゲノム解析										DB作成					ライフサイエンスデータベース統合推進事業統合化推進プログラムの1つであるゲノム・メタゲノム情報を基盤とした微生物DBの統合」において、開発メンバーである東京工業大学、国立遺伝学研究所、基礎生物学研究所、統合データベースセンターとRDFデータベースファイル、及びSPARQLの共有を行った。また、各機関で作成したRDFデータベースを使用したSPARQLの検証をSPARQLTHONにて実施した。	ruby			普通		混雑している時もあるが問題なく利用できる	サービスの停止を少なくしてほしい
一般研究用	平木 秀明	国立遺伝学研究所 系統生物研究センター 生物遺伝資源情報研究室	細胞膜画像処理などの線虫胚発生の四次元解析、ゲノムショットガンのアセンブルなどの配列情報処理。										NGS		Whole Genome Sequencing		Illumina MiSeq	モデル生物C. elegansの近縁線虫D. coronatusは、C. elegansの初期胚では重要である細胞配置とは異なった細胞配置をとって発生する。また単為生殖により世代交代を行う点でもC. elegansと異なっており、どのような遺伝子をもって生きたのか興味深い。このD. coronatusのゲノム配列について、NGS Supercomputerを利用してアセンブルを行い、論文にまとめることができた。さらに、RNA-seqデータの解析も行えた。現在、MiSeqによる配列データを新たに加えてさらなる改善を試みている。	BEDTools BLAT Cufflinks IGVviewer InterProScan Picard RAxML SOAPdenovo bowtie2 bwa maffft ncbi.toolkit samtools tophat2 R gcc java perl ruby その他		NCBI-nr	普通		混雑している時もあるが問題なく利用できる	*/usr/local/pkgにあるプログラムの更新はされないのでしょうか・?・メモリ制限の指定が難しいです。他のユーザーによる停止と区別がつけづらい。いろいろ試していると結局大きめに設定することになりがちです…
一般研究用	寺上 伸吾	農業・食品産業技術総合研究機構 果樹研究所	次世代シーケンサーのデータ解析										NGS		Resequencing		454 GS FLX Titanium 454 GS FLX+ Illumina HiSeq 2000	果樹でも様々な樹種においてゲノムの解読がなされている。そこで世界的に産業の大きいバラ科果樹をメインに複数品種のリシーケンス を行い、品種間SNPを見出した。その結果からゲノムワイドなSNPを選定し、GoldenGate AssayによりゲノムワイドなSNP遺伝子型データを得るとともにアソシエーション解析から形質との関連を見出した。	BEDTools BLAT ClustalW Cufflinks GATK Picard bowtie bowtie2 bwa cutadapt samtools tophat tophat2 R perl ruby		DDBJ GenBank NCBI-nt NCBI-nr Swiss-Prot	普通		混雑している時もあるが問題なく利用できる	今後ともよろしくお願いいたします。
一般研究用	三科 興平	木原生物学研究所 横浜市立大学	コムギRNA-seqのアセンブリ										NGS		RNASeq		Illumina HiSeq 2000	コムギRNA-seqのデータをTrinityによってアセンブルを行っている。試験的にdebugノードでメモリ要求量をテストした場合、配列数コア数増加に伴ってメモリの確保が課題となることがわかったので、FATノードにHiseq1レーン分のデータを投入したが、2月末までの段階でキュー待ちで完了していない。	Trinity tophat			普通		混雑していて不満だ	FATあるいはMEDIUMノードのキュー待ちが解消されず、命令がなかなか実行されない、来年度の増設に期待したい。
一般研究用	高野 知之	明治大学バイオインフォマティクス研究室	新大なメモリを使用するアセンブルなど明治大学バイオインフォマティクス研究室での研究解析に使用。										NGS		Transcriptome Analysis		454 GS Illumina HiSeq 2000	トランスクリプトーム解析の為、illumina及び454リードを用いてアセンブルを行い、複数のトランスクリプトームコンティグを得た。今後、得られたトランスクリプトームコンティグに対し、既知遺伝子に対し、相同性比較などを行い、トランスクリプトームコンティグの妥当性の検証を行う。トランスクリプトーム解析以外でも、明治大学バイオインフォマティクス研究室で行っている解析を行う。	Oases bowtie2			普通		混雑している時もあるが問題なく利用できる	illumina及び、454リードを用いたアセンブルは、膨大なメモリを消費する。しかし、混雑時はどのサーバも残りのメモリ使用量がアセンブルを行うには少ない。
一般研究用	岩嶋 裕貴	長浜バイオ大学	遺伝学研究所より長浜バイオ大学に移られた、池村淑道教授の指導の下で、大規模遺伝子解析を研究しております。本学の計算クラスターでは、本来の研究目的とするゲノムサイズの解析が困難であるため、高原とも利用させていただければと存じます。	Nobel bioinformatics strategies for prediction of directional sequence changes in influenza virus genomes and for surveillance of potentially hazardous strains	In Preparation								NGS以外の配列解析					1万株を越えるインフルエンザウイルス株のゲノム配列を対象に連続塩基の出現頻度に基づいた一括学習型自己組織化マップ(BLSOM) 解析を行った。結果として、インフルエンザウイルスは流行可能な限定ごとに連続塩基の組成が明確に異なるということがみられた。また、流行初期のヒト由来株に注目して、ヒト以外の宿主で流行している株が、ヒトで流行を開始するために重要である連続塩基の推定を行った。	ncbi.blast+ gcc perl		DDBJ GenBank	普通		混雑していて不満だ	並列計算時に、使用できるノード数が極端に限定されていた。
一般研究用	柏木 明博	個人	バイオインフォマティクス ソフトウェア開発・大規模データ処理・分散並列データ処理・等										アルゴリズム開発					HPCを用いたモチーフ発見問題」における計算速度の大幅な向上PCにおいて、約24時間かかっていた処理「モチーフ発見問題」がHPC環境において、Multi Threads, Multi Processesによって、7分に短縮された。※対象遺伝子Mycoplasma Genitalium今後は、CUDA(GPGPU)を用いた高速化も考えていおります。	自作プログラム	gcc,pthreadによる、Multi threads, Multi process技術	GenBank	遅い	Pentium3 933[MHz] 512[MB] : 1 threads: 約24時間Medium node : 500 threads(10 Process × 50 Threads) : 約7分	混雑していて不満だ	thin nodeの利用については、概ね不満はありませんが、Fat nodeはいつも混雑している感がある。
一般研究用	笠原 浩太	大阪大学 蛋白質研究所	プロテインインフォマティクスに関するデータ解析とシミュレーション。										シミュレーション、MD					転写因子およびDNA複合体のMDシミュレーションにより転写因子のDNA結合制御機構の解明を目的として研究を行っている。MDにおいては研究室で開発中の新規方法論での計算を行う。まずは、DNAなどの基本的な構成要素のみからなる単純な系についてと、方法論の妥当性の検証を行った。これからはより複雑な、実践的な複合体モデルのシミュレーションへ展開していく予定である。	R			普通		十分空いている	今のところ、混雑していると感じたことはありません。(まだ大規模な計算をしていないからかもしれません。最大100clos程度)
一般研究用	飯田 哲史	国立遺伝学研究所	DNAの解析										NGS		その他	Repaired Genome sequeincing	Illumina MiSeq	出芽酵母Saccharomyces cerevisiaeの新生DNA鎖を単離・精製したsampleをIllumina Miseqでsequence後、reference genome sequenceにmappingして解析した結果、genome中に新生DNA鎖が高効率で作られる領域を見出した。これらの結果は、S. cerevisiaeのgenome中に高頻度でDNA修復が起きる領域が存在していることを示唆している。	BEDTools bowtie bowtie2 bwa samtools			普通		混雑している時もあるが問題なく利用できる	個人的には、これから使用頻度を上げていきたいです。
一般研究用	山口 康之	日本SGI株式会社	denovo assembly tooloの動作検証										その他	denovo assembly tooloの動作検証				Fatノードにおけるtrinityのメモリ使用量の調査NUMA のマシンとではJAVA の gc スレッドが大量に実行され、メモリやCPUを大量に消費する傾向があることが判明しました。これはjava の gc スレッドのデフォルト値が物理CPU 数になっていることに起因するものと考えられます。Butterfly の実行時には、スクリプト内で -bfly/GCThreads 1 のオプションを指定することによって、CPUメモリなどのリソースの使用量を調整することができます。	Trinity			普通		混雑している時もあるが問題なく利用できる	特にありません。
一般研究用	村上 勝彦	産業技術総合研究所 バイオメディシナル情報研究センター	ヒトマウスにおいて遺伝子発現とシスエレメントとの関連を調べる										NGS以外の配列解析					ヒト、マウスそれぞれで転写開始点を仮決めし、全遺伝子についてプロモーター配列を取り出した。その中のシスエレメントの検索を同定するために網羅的にtoppenの統計情報を出した。これらには既知の転写因子結合部位との対応付けをJASPARおよび文献データを用いて行った。次のステップとして遺伝子発現情報の対応付けを行っている。	gcc perl			普通		混雑している時もあるが問題なく利用できる	e-login関連にまとまった、資料がそこそこ充実していていい。このレポートを書きかけでも保存、修正できるとうれしい。
一般研究用	木村 暁	国立遺伝学研究所 構造遺伝学研究センター 細胞建築研究室	細胞内3次元力学シミュレーション	線虫細胞質流動を駆動する力のパラメータ推定	In Preparation								シミュレーション、MD					線虫の細胞質流動を駆動する細胞内でのカ分布をデータ同化法を用いて推定している。今年度は開発したプログラムを遺伝研のスーパーコンピュータを使って試した。本研究は、当研究室の扇山律教との共同研究で主に扇山が解析を行っているため、私自身は本年度はスパコンを直接は利用しなかったが、来年度は扇山が他研究所へ移動することもあり私自身が解析を行いたい。	自作プログラム	mpi		普通		混雑している時もあるが問題なく利用できる	ユーザー登録はしていますが、今年度は同じ研究室の扇山に解析を任せていたため、私自身は使用していません。

アカウント区分	使用者	所属機関	利用目的	論文名1	論文の状況1	Pubmed ID1	論文名2	論文の状況2	Pubmed ID2	論文名3	論文の状況3	Pubmed ID3	研究成果(分野)	研究成果(分野)(その他)	NGSプロジェクト名	NGSプロジェクト名(その他)	NGSプラットフォーム名	研究成果・経過報告(内容入力)	使用プログラム	使用プログラム(自作プログラム)	使用プログラム(その他)	使用データベース	アンケート-速度-	アンケート-速度-(自由入力欄)	アンケート-混雑状況-	アンケート(自由入力欄)
一般研究用	二橋 亮	独立行政法人 産業技術総合研究所	次世代シーケンサーの結果解析(de novoアセンブリ)のため										NGS		RNASeq		Illumina HiSeq 2000	ゲノムが未知の複数の菌populationに対して行ったRNAseqのde novoアセンブリ(Trinity)を行うことを当初予定していたが、その後テストケースとして遺伝学研究所のMaserを用いた解析を行うことになり、本システムを使用せずに当初の目的が達成できた。そのため、本システムを継続して使用する必要性がなくなったため、アカウント利用の停止を希望する。	Trinity				普通		混雑している時もあるが問題なく利用できる	RNAseq解析については、先月公開されたMaserを利用することで目的が達成できた。
一般研究用	須田 亜弥子	東北大学大学院生命科学研究科	次世代シーケンサIon Torrentで解読されたデータ解析。特定遺伝子解析。										NGS		Whole Genome Sequencing		Ion Torrent PGM	次世代シーケンサーから得られたデータを全ゲノム解析のために、近縁種の全ゲノムとマリシーケンスを行った。しかし、ごく少量の一致性(おとそ20%)しか得られなかったため、解析ソフトを再度検討して解析し直す必要がある。また、現在はデータの中に短い配列が大量に含まれているため、長い断片を解読したデータが必要だと判断した。今後データ量と長さを伸ばしたとで再度解析を試みる予定。	bowtie samtools				普通		混雑している時もあるが問題なく利用できる	まだ数回しか扱ったことがないので未だに慣れないが、今度もっと利用したい。
一般研究用	奥田 裕樹	東京工業大学 大学院情報理工学研究科	大規模ゲノム情報のアセンブリおよび統計解析										NGS		その他	Genome Assembly	Illumina Genome Analyzer II	非モデル生物を対象にしたゲノムアセンブリを行なっている。論文等の成果には至っていない。現状では、NGSにより得られた配列のリミッング等の処理は終了した。ゲノムアセンブリは現在進中段階であり、アセンブラのパラメータを変更して手動での最適化を行なっており、複数パラメータでのアセンブリの並列実行においてスーパーコンピュータのジョブ投入システムを利用している。	SOAPdenovo				普通		混雑している時もあるが問題なく利用できる	特に不満は無い
一般研究用	米本 悠	東京大学大学院 新領域創成科学研究科 情報生命科学専攻	RNA配列解析のため										NGS以外の配列解析					RNAニ次構造予測における確率モデルのパラメータを半教師付き学習により推定する手法を開発した。この手法は、教師付き学習によってパラメータ推定をする既存手法と同等かそれ以上の予測性能を持つことが示された。今後は、学習に用いるデータセットの数を増やし、追加実験等を行う予定である。また、これらの研究成果は論文誌に投稿予定である。	python				普通		混雑している時もあるが問題なく利用できる	まだ使用していません
一般研究用	木村 敦	北海道大学大学院理学研究院	次世代シーケンサーのデータを解析するため										NGS		Gene Regulation Study		Illumina Genome Analyzer II Illumina HiSeq 2000	マウスの生殖器官における noncoding RNAによる遺伝子の活性化メカニズムを探るために、NCBIデータベースに存在するマウス生殖器官のトランスクリプトームのデータを解析した。その結果、いくつかの生殖器官特異的遺伝子が存在するゲノム領域から、生殖器官特異的に noncoding RNAが発現していることがわかった。	bowtie sra-tools			NCBI-nt	遅い		十分空いている	大変優れたシステムでとても助かりたく使っております。私の研究室で必要な解析をする限りは特に大きな問題を感じたことがありません。ただし、落雷や停電などが通常よりも高頻度に起きている印象があり、その点で少しだけ不安を感じております。
一般研究用	吉田 真明	国立遺伝学研究所	申請者と所属研究室が中心となって頒足類de novo genomeプロジェクトを推進しています。次世代シーケンサによって得られた配列のアセンブリ、アノテーションを中心とした大規模配列解析を行うため、スーパーコンピューティングシステムの利用を希望いたします。										NGS		Whole Genome Sequencing		Illumina HiSeq 2000	申請者が進行する頒足類de novo genomeプロジェクトにおいて、トランスクリプトームのアノテーションを目的にTransProtocorを利用した。現在も継続して計算を行なっている。次年度は引き続きアノテーションを目的とした計算と、トランスクリプトームおよびゲノムassemblyの改善を目的としてAllPaths-LGおよびTrinityを用いた計算を計画している。	InterProScan R perl			RefSeq-Protein UniProt Swiss-Prot Pfam	普通		混雑している時もあるが問題なく利用できる	現在のところ混雑を感じることなく利用することができている。またInterProScanのoptionについてフォームから問い合わせたところ速やかに対応いただき、ユーザーのフォローの点においても非常に信頼できると感じている。
一般研究用	沖 真弥	九州大学大学院	次世代シーケンサのデータを解析するため										NGS		RNASeq		Illumina Genome Analyzer II	マウス胚は受精後4.5日後に着床した後、受精後5.5日までにエピブラストと呼ばれるpluripotentな細胞集団を形成する。このエピブラストは未分化なマーカーであるOct3/4やNanogを発現する一方で、Nodalなどのような将来の中胚葉誘導を担う因子も発現しているため、どのようにして分化／未分化の均衡を維持しているのかはよくわかっていない。そこで申請者は受精後5.5日までのマウス胚よりエピブラストのみを単離し、mRNA-seqを行うことで、どのような遺伝子が発現しているのかを網羅的に調べた。また、エピブラストはヒトES細胞と極めて近い性質を有するため、ENCODEプロジェクトで得られたヒトES細胞の発現プロファイルやChip-seqのデータとの比較を行った。	BEDTools IGVviewer MACS bowtie samtools sra-tools tophat R				普通		混雑している時もあるが問題なく利用できる	ゲノム関係のライブラリファイルを共通利用できるようにしていただくと、容量の削減につながると思います。例えば、UCSC の genome.fa や LiftOver の .chain、bowtieのインデックスファイルなどは多くの方が利用されるのではないかと思います。
一般研究用	平尾 章	筑波大学 菅平高原実験センター	倍数体植物を対象としたNGSデータのマッピング										NGS		Population Genomics		454 GS Junior	異質倍数体のシロイヌナズナ近縁野生植物を対象とした塩基多型スクリーニングのため、Pool-Seq法によって得られたNGSデータのマッピングを試みた。また集団遺伝情報を推定するため、DAr法によってゲノム内の分1倍遺伝子座がジェノタイピングされたサンプルを用いて、STRUCTURE解析（ベイジアンクラスタリング法の一種）を行った。集団構造モデルにはadmixture modelを採用し、集団のクラスタリングと同時にゲノム組成の混合構造を推定した。	bwa その他		STRUCTURE		普通		混雑している時もあるが問題なく利用できる	スパコンの使用に慣れておらず、ログインノードにログインした後、インタラクティブノードにログインしないまま、計算機を使用してしまいました。ご迷惑をおかけして申し訳ありません。
一般研究用	森 遼太	東京大学	ゲノム解析										アルゴリズム開発					今年度は利用いたしませんでしたが、高次元経数空間上の確率分布をある特徴量に基いて1次元空間へ射影した分布の計算手法を開発し、さらにそれをRNAニ次構造の分布様態記述へと応用しました。これによって元来RNAのニ次構造予測手法が不可避に抱えていた本質的問題を解決する糸口を見出すことができました。来年度は、本アルゴリズムを実装のエスパコンとで実験し、論文を執筆予定です。	その他		利用せず		普通		混雑している時もあるが問題なく利用できる	高性能なスパコンがこういった形で広く利用可能となっているということは素晴らしいことだと思います。運用・管理は大変な労力であるはずですが、これからもこのような環境が維持されていく事が日本発の生命科学発展の原動力になるものと考えています。
一般研究用	佐藤 健吾	慶應義塾大学理工学部	機能性RNAおよびゲノム解析アルゴリズムの研究開発のため	Bioinformatics (2012) 28(24):3218-3224	Published	23060618	Nucleic Acids Res. (2012) 40(W29-W34	Published	22600734				NGS以外の配列解析					高速かつ高精度なRNA構造アライメントアルゴリズムDAFSの開発を行った。また、RNAニ次構造予測アルゴリズムAnchorFold中のウェブサーバ公開へ向けて開発を行った。今後、RNA構造予測アルゴリズムの開発を引き続き行うとともに、大型ゲノムのためのアセンブリ用ソフトウェア開発を行う予定である。	R gcc ruby 自作プログラム	アルゴリズム開発のため			普通		混雑している時もあるが問題なく利用できる	素晴らしい環境を提供していただきありがとうございます。
一般研究用	沼田 興治	理化学研究所バイオリソースセンター動物実験施設解析技術開発チーム	マウス系統MSM/Msゲノム配列を用いた遺伝子発現解析										NGS		RNASeq		Illumina HiSeq 2000	マウスMM系統のゲノム配列を利用して、B6系統とMSM系統の行雌種マウスにおけるトランスクリプトーム（RNA-seq）解析を試みているところ。精巣におけるデータを解析したところ、他系統を利用した過去の報告と同様に系統間の発現バイアスが観察された。今後、他の組織のデータを解析し、系統間の比較や遺伝子機能推定、ゲノムとの特徴解析などをおこなっていく予定である。	BEDTools bwa samtools perl				普通		混雑している時もあるが問題なく利用できる	大きな処理をおこなう際には大変助かっております．これまでと同様に今後も無償で利用できることを願っております．
一般研究用	中川 草	国立遺伝学研究所 生命情報・DDBJ研究センター	蛋白質翻訳開始機構のゲノム進化解析、そして内在性ウイルスの探索のため	Dynamic Evolution of Endogenous Retrovirus-Derived Genes Expressed in Bovine Conceptuses during the Period of Placentation.	Published	23335512	Identification and functional analysis of three isoforms of bovine BST-2.	Published	22911799	Infectious endogenous retroviruses in cats and emergence of recombinant viruses.	Published	22674983	NGS以外の配列解析					ウシやネコなどの注民種のゲノム中に散在する内在性ウイルスのenvelope遺伝子の相同性探索や系統解析などを行った。RAxML、protrev(JavaのプログラムJ)、MAFFTなどのソフトウエアを使用した。こちらの研究に関して今年度ご報告した。また、ウシのBST-2遺伝子の分子進化解析も行った。また、現在ではマウスのRNA-seq解析、蛋白質翻訳開始機構のためのハクリナリアゲノムの比較進化解析なども行っている。	BLAT RAxML maft ncbi_blast+ samtools R java perl python ruby 自作プログラム	mpi		DDBJ GenBank RefSeq-Genomic RefSeq-Protein NCBI-nt NCBI-nr Pfam	遅い		混雑している時もあるが問題なく利用できる	非常に便利に使っています、特にHDD容量の制限が大きい事はNGSデータを扱うときに重宝します。
一般研究用	谷 重善	浜松医科大学医学部医学科 情報医学	生命情報に関する高速コンピュータライブラリの作成										シミュレーション、MD					生命科学（医学）領域において、従来型のコンピュータで作成したライブラリ類をスーパーコンピュータ向けに改良している途中であり、現時点では、プリコンパイルの試験中で、順次ライブラリ類を充実させていく予定です。また、パラレル化ができるライブラリについては分散処理の評価も行う予定です。スカラーとベクトルが混在しているために、プログラムを管理中です。	gcc java perl ruby				普通	現在、まだ、性能を評価できる利用状況ではありません。	混雑している時もあるが問題なく利用できる	作成しているライブラリ類が、動作する段階で、詳しく報告させていただこうと思います。
一般研究用	永野 幸生	佐賀大学 総合分析実験センター	Whole genome (re)sequencingやRAD-seqを使って、高等植物などの変異を追いかける研究もします。また、RNA-seqを使って、高等植物などの遺伝子発現を調べます。										NGS		Whole Genome Sequencing		Illumina HiSeq 2000	ある樹木のゲノムを解読して、過去100年以上の間におこった体細胞変異を追跡している途中です。自前のサーバーとNIG SuperComputerを併用して行っています。まだ解析を始めたところで粛り、十分な成果は出ていません。しかし、いくつか重要そうなSNVやInDelを発見しており、今後、これらの結果をまとめて、論文発表を予定です。	BEDTools Cufflinks GATK Picard Trinity bowtie2 bwa samtools tophat2				普通		混雑している時もあるが問題なく利用できる	自前のマシンで処理しきれないときに、併用しています。満足しています。
一般研究用	岩崎 渉	東京大学大気海洋研究所	海洋ゲノムに関する幅広い研究を行う										NGS		Metagenomics		454 GS Illumina HiSeq 2000	さまざまな環境に生息する微生物より得られたゲノムデータ、また微生物群集より得られたメタゲノムデータの解析を行っている。特に、様々な環境の変動に対する微生物の応答についてゲノムレベルの情報に基づいた分析を行っている。論文発表に至る研究成果はまだ得られていないが、生態系にとって重要な微生物の働きについて新たな知見が得られつつあり、次年度には成果報告を行いたい。	ncbi_blast+			DDBJ	普通		混雑している時もあるが問題なく利用できる	バイオインフォマティクス研究の推進に大変役だっており、心から感謝申上げます。今後ともどうぞよろしくお願いいたします。
一般研究用	山田 拓司	東京工業大学	メタゲノム解析のため										NGS以外の配列解析					土壌や海洋、ヒト腸内細菌叢などから得られた メタゲノムデータに対して、そこに含まれる膨大な遺伝子の機能アノテーションを様々なツールとデータベースを利用して行なっている。ゲノム上の遺伝子の隣接関係を利用することでアノテーション精度の向上を目指す。遺伝子の隣接関係はこれまでは原核細菌の機能アノテーションにのみ利用することができると考えられていたが、近年の報告により真正細菌にも応用できる可能性が示唆された。先行研究において原核細菌のアノテーションパイプラインを確立しているので、今後は真正細菌においても同様のパイプラインを構築する予定である	R ruby			NCBI-nr	普通		混雑している時もあるが問題なく利用できる	複数アカウントが協同で利用できるディスクスペースがほしい。
一般研究用	飯山 和弘	九州大学 大学院農学研究院	微生物ゲノム解析のため。	Draft Genome Sequence of Paenibacillus popilliae ATCC 14706T	In Preparation								NGS		Whole Genome Sequencing		454 GS FLX Titanium Illumina HiSeq 2000	コガネムシ類昆虫に病原性を有する乳化病菌Paenibacillus popilliaeのゲノム解析を行った。hiseq2000および454で解析を行い、現在ギャップクローニングを行っている段階である。ドラフトシーケンスは既にアクセッション番号を取得し、公開している。またドラフトゲノムに関する論文を現在投稿中である。	ClustalW SOAPdenovo bowtie			DDBJ	普通		混雑している時もあるが問題なく利用できる	特に問題はございません。提供されているプログラム等を当方が使いこなせるための基礎的知識が不足していることを実感しています。この点をクリアできれば思っている次第です。自身のスキルを向上させ、今後も使わせていただきたいと思っております。
一般研究用	飯田 直子	国立遺伝学研究所 生命情報研究センター 大量遺伝情報研究室	研究 開発, ゲノムシーケンス解析。										NGS		Whole Genome Sequencing		Illumina MiSeq	スパコン環境を用いて、酵母の変異体の変異点同定をゲノムシーケンス解析により行っている。そのため、bwa、samtoolsを用いている。データの転送、プログラム実行が高速に行うことが出来ている。使用開始から約30個の変異点同定を行った。今後も継続して使用する予定である。解析ツールの向上を目指し、プログラムの実行テストにも利用する。	bwa perl				遅い		十分空いている	データ処理、解析がスムーズに行えている。
一般研究用	高橋 弘喜	千葉大学 真菌医学研究センター	1. 次世代シークエンスデータを対象とした、遺伝子配列解析。2. モンテカルロシミュレーション										NGS		RNASeq		454 GS Junior Illumina HiSeq 2000 Illumina MiSeq	アスペルギルス属真菌のRNA-Seqデータで構築した全contig配列を用いて、相同性検索を行った。その結果、多くの遺伝子に関してアセンブルが問題なく行えていることを確認した。大規模の数理モデルにおいては、微分方程式計算ライブラリーを用いてマルコフ連鎖モンテカルロ法によるシミュレーションを実行し、全パラメーター値の推論に成功した。	nobi_blast+ gsl gcc 自作プログラム	sundialsパッケージによる微分方程式の数値計算を行った。			普通		混雑している時もあるが問題なく利用できる	使い勝手も非常に良く、SEの方の対応も素晴らしかった。計算リソースに恵まれない者にとって、このようなサービスは今後も是非継続していただきたい。
一般研究用	小島 要	東北大学	バイオインフォマティクスにおけるデータ解析手法、アルゴリズム開発										その他	スーパーコンピュータの使用方法の調査				本年度は、昨年12月末からの利用開始であったため、スーパーコンピューターのシステムの使用方法の調査として、ジョブスケジューラーの使用法や、GATKをはじめとする次世代シークエンサーからのデータ解析のためのツールの使用法の確認を行った。来年度より、次世代シークエンサーからのデータを対象とした、解析と解析手法の開発を行う予定。	GATK java				普通		混雑している時もあるが問題なく利用できる	通常のLinuxデスクトップマシン感覚で使用することができるため、非常に使いやすい。また、大規模メモリマシンがあるため、大量のメモリを必要とするデータ解析手法の開発が可能であり、他にはない手法の開発が可能であると感じている。
一般研究用	丸山 史人	東京医科歯科大学・大学院医歯学総合研究科	原核生物のゲノムアセンブル。										NGS		Whole Genome Sequencing		454 GS FLX 454 GS Junior Illumina Genome Analyzer II Illumina MiSeq	4種の異なる細菌種合計、約100株についてIllumina社のGAIIx, MiSeq, Roche 454のGS FLX+, GS Juniorでゲノム配列を取得した。Illuminaについては、ゲノムサイズの50, 454については ゲノムサイズの15程度の量の配列を得た。これらの配列について、velvet, newblerを用いてアセンブルし、metagene, glimmer3でORFを予測し、RNAmmer, tRNAscanの予測データを付加した。nrたいして、Blastをした結果を用いて遺伝子推定を行い、ドラフトゲノム配列を決定した。	bwa mafft phylip samtools velvet R perl			DDBJ GenBank RefSeq-Genomic NCBI-nr	普通		混雑している時もあるが問題なく利用できる	特に問題なし。
一般研究用	菅野 純夫	東京大学新領域創成科学研究科	次世代シークエンサーによるRNAデータ、DNAデータの解析										NGS		RNASeq		Illumina Genome Analyzer II Illumina HiSeq 2000	Maximal-information-coefficient (Reshef, D. N.; Reshef, Y. A.; Finucane, H. K.; Grossman, S. R.; McVean, G.; Turnbaugh, P. J.; Lander, E. S.; Mitzenmacher, M. et al. (2011). Detecting Novel Associations in Large Data Sets". Science 334 (6062): 1518-1524. doi:10.1126/science.1205438)と相互情報量を利用したRNA-seq発現解析を行うための準備としてデータ前処理を行う予定であったが、自作のプログラム作成が間に合わず、今後も見通しが立っていない。	GATK bwa tophat2 自作プログラム	相互情報量（予定）		RefSeq-RNA	普通	比較せず	混雑している時もあるが問題なく利用できる	利用ほとんどせず。
一般研究用	成相 直樹	東北大学	バイオインフォマティクスに関する研究、及び関連する データ解析、アルゴリズム・ソフトウェア開発のために利用										NGS		Transcriptome Analysis		Illumina Genome Analyzer Illumina Genome Analyzer II Illumina Genome Analyzer IIx Illumina HiSeq 2000 Ion Torrent PGM	2012年12月後半からの利用開始であったため、今年度は主にスパコンシステムの使用方法の調査、バイオインフォマティクス関連のソフトウェアの調査・インストール、動作確認等を行った。来年度以降は主にトランスクリプトーム解析、RNA-Seq データ解析に係る新規手法の開発、あるいは既存ツールを使用したデータ解析などの用途に使用していく予定である。	BEDTools BLAT Cufflinks IGVviewer MACS bowtie bowtie2 bwa cutadapt samtools tophat tophat2 R gcc java perl python			GenBank RefSeq-Genomic RefSeq-RNA	普通		混雑している時もあるが問題なく利用できる	汎用的なシステムで使いやすいと思います。
一般研究用	原 一夫	国立遺伝学研究所 生命情報研究センター	利用目的: DDBJ データの全体像を把握しながら検索できるシステムを構築すること. 最終的には、DDBJデータに対するキーワード検索の結果を、検索者の意図に沿って自在にブレーングダウンできるようにしたい。すなわち、検索結果を細分化してグループ表示し、さらに、各々のグループ表示をクリックすると、そのグループをさらに細分化して表示できるようにする。これを「on the fly」で行いたい.										その他	情報学（information retrieval, classification, data mining）				現状と問題点：DDBJ データは、ARSA を用いることにより、パターンマッチおよび「生物群区分」「研究の型別分類」で検索できます。しかし、一件件を超える データ集合からの探し物には、（多くの場合）複数の制約条件の組み合わせが必要ですが、現状ではそれぞれの検索条件の有効性を確認するすべがありません。 経過報告：まず、DDBJ データの明示的な特徴である Feature/Qualifier について、フラットファイル あたりの出現頻度分布を調べました。 http://133.39.115.67/~kazuo/ddbj_parsed/FeatureQualMax/table.html検索の手掲かりとして Feature/Qualifier を用いることで、DDBJ データの全体像をある程度は把握できることが分かりました.	gcc ruby			DDBJ	普通		混雑している時もあるが問題なく利用できる	問題なく使用させていただいております.
一般研究用	杉原 稔	産業技術総合研究所・生命情報工学研究センター	膜タンパク質、RNA 等のMDシミュレーション										シミュレーション、MD					今年度は、結晶構造に基づいてウシopsin とG-タンパク質類互作用のMD計算を実行した。G-タンパク質との相互作用を調べるため、G-タンパク質のミューテーションを行い、その安定性を調べた。現在、結果を発表する論文を準備中である。今後は、関連するシミュレーションを継続する。また、他の GPCRs や RNA のMDシミュレーションを実行する予定である。	その他		namd		普通		混雑している時もあるが問題なく利用できる	何度か質問のメイルをする機会がありましたが、丁寧に解答して頂いて感謝しております。
一般研究用	岡 彩子	新領域融合研究センター	ChIP-seqデータの解析										NGS		Gene Regulation Study		Illumina HiSeq 2000	国立極地研究所との共同研究で、マウス亜種間染色体置換系統（コンゲニック系統）のChIP-seqによるヒストン修飾データ（H3K4me3, H3K4me1, H3K27me3）の解析を行っている。使用しているコンゲニック系統は、標準的実験用系統B6の遺伝的背景に、異なる亜種MSM系統の染色体を導入したもので、現在は、ChIP-seqのreadsをそれぞれの系統のゲノムシークエンスとにマッピングし、終えた段階である。今後は、特にプロモーター領域やエンハンサー領域に存在する亜種間の塩基置換が、近傍の遺伝子の転写に及ぼす影響を調べる予定である。	MACS bowtie bowtie2 tophat				普通		混雑している時もあるが問題なく利用できる	現在、解析は共同研究者にお願いしているので具体的なコメントはできないが、今後はpeakfinderなどを利用したいと思っている。pipelineのような初心者にも使いやすいツールが増えてくれるとさらに利用しやすくなると思う。
一般研究用	山田 明徳	東京工業大学大学院 生命理工学研究科生体システム専攻	シロアリ共生微生物のゲノム解析										NGS		Whole Genome Sequencing		454 GS FLX+ Illumina Genome Analyzer IIx Illumina MiSeq	シロアリ腸内共生原生生物とその原生生物の細胞内共生細菌のシングルセルゲノム解析を中心に行っている。これまでに、細胞内共生細菌の1系統について複数セルから全ゲノム配列を取得することができた。今後、さらに系統を増やして比較ゲノム解析等を通じて、細胞内共生細菌の役割・機能、そしてシロアリ一腸内原生生物一細胞内共生細菌という「マトリョーシカ」型共生系の相互作用や維持機構、進化過程を明らかにしていくつもりである。	BLAT SOAPdenovo Trinity bowtie2 maft ncbi_blast+ samtools velvet java			NCBI-nt NCBI-nr NCBI-WGS	普通		混雑している時もあるが問題なく利用できる	本研究室でも解析サーバーを所有しているが、スパコンと合わせて利用することで効率できこの作業を行うことができている。
一般研究用	荒川 和晴	慶應義塾大学先端生命科学研究所	NGSデータ解析										NGS		Whole Genome Sequencing		Illumina Genome Analyzer IIx Illumina HiSeq 2000	AllPaths-LGやMSR-CA、SOAP denovoなどを利用した薄類新規種ゲノムのde novo assemblyのためにメ大規模メモリが必要だったために申請したが、大容量メモリシステム(RAM 6TB)を新規に研究室で導入したため通伝研のスパコンは事実上ほとんど使っていない。数回テストでログインしたのみである。	ALLPATHS-LG SOAPdenovo Trinity samtools velvet R java perl その他		MSR-CA	GenBank Swiss-Port	普通		混雑している時もあるが問題なく利用できる	上述の通りほとんど使用していない
一般研究用	川股 かおる	東京大学 大学院理学系研究科 生物科学専攻 多様性起源学研究室	バクテリアゲノムを対象に、次世代シーケンサーによって得られたデータのアセンブルと解析を行うために利用を希望します。										NGS		Whole Genome Sequencing		Illumina Genome Analyzer IIx	新規バクテリアのde novo シーケンシングをIllumina GAIIxによって行い（シングルエンドリード・70bp・Too much coverage）、DDBJ Annotation Pipelineを使用して、複数のアセンブルソフトウェアを比較検討し最良コンティグを得ました。結果に対してスパコンアカウントで利用できるMiGaiによりアノテーションを行っている最中です。解析にはGalaxyも利用しています。今後スパコンにはなるアセンブルも行う予定を検討しております。	その他		MiGai	GenBank	遅い		十分空いている	まだスパコンを直接操作した事はないため、使用感はわかりません。スパコンを用いた複数のサービス（DDBJ Annotation PipeLine, Galaxy, MiGai）を用いて解析を行っておりますが、混雑なく、自分の端末ではできない高度な解析を行う事ができるため、大変助かっております。CUIに不慣れなため、現在はWeb画面から扱えるサービスの方をメインに使っておりますが、次年度はスパコンを直接使用する予定です。
一般研究用	堀内 映実	東京大学大学院新領域創成科学研究科	NGSデータ処理の為										NGS		その他	ゲノム支援	Illumina Genome Analyzer IIx Illumina HiSeq 2000	2012年度ゲノム支援にて、菅野・鈴木研でお預かりしている先生方のSequenceQuality確認や、こちらで簡単な1次解析（マッピング、RNAであれば発現量解析やアセンブル、ChIPであればピーク検出、その他視覚化データ作成）を実施する必要がある際に、データ処理をスパコン内で実施させて頂いている。	MACS bwa samtools perl				普通		混雑している時もあるが問題なく利用できる	今のところ不便はない

アカウント区分	使用者	所属機関	利用目的	論文名1	論文の状況1	Pubmed ID1	論文名2	論文の状況2	Pubmed ID2	論文名3	論文の状況3	Pubmed ID3	研究成果(分野) (その他)	研究成果(分野) (その他)	NGSプロジェクト名	NGSプロジェクト名 (その他)	NGSプラットフォーム名	研究成果・経過報告(内容入力)	使用プログラム	使用プログラム (自作プログラム)	使用プログラム (その他)	使用データベース	アンケート-速度-	アンケート-速度-(自由入力欄)	アンケート-混雑状況-	アンケート(自由入力欄)
一般研究用	藤川 高弘	国立極地研究所、新領域融合研究センター	南極氷床アイスコア中の微生物ゲノム解析										NGS		Metagenomics		454 GS FLX Titanium 454 GS FLX+ Illumina HiSeq 2000	南極氷床アイスコア中などにふくまれる微生物からゲノムDNAを抽出し、新型のハイスループットDNAシーケンサーによって得られた大量配列に対して、メタゲノム解析を行った。ゲノム配列のアセンブル、アセンブルされたゲノム配列上のORFの予測を通じて、KEGG、NCBIデーターベースにある広範な微生物ゲノム及び遺伝子情報をおこなっている。	BLAT RAxML ncbi_blast+			普通		混雑していると時もあるが問題なく利用できる	利用できるメモリー容量を増やして欲しい
一般研究用	天野 晃	理化学研究所	DNA塩基配列パターン解析 文献情報解析										アルゴリズム開発					宇宙発表Alan-variance-like plot for DNA sequence data情報メディア宇宙第14回研究会 シグナル解析の手法であるアラン分散プロットをDNAシーケンスエンスデータ解析に適応した。1. 複素数法：A, G, C, Tを1, i, -1, -iに変換し、アラン分散を計算。2. オリゴヌクレオチド法：オリゴヌクレオチド頻度ベクトル間のユークリッド距離を使用してアラン分散を計算。1. において、緩やかな下に凸のプロットが得られた。	BLAT ClustalW GATK bowtie bowtie2 bwa fasta phylip openmpi R gcc java perl python ruby		DDBJ GenBank EMBL UniProt Swiss-Prot PDB Pfam	普通		混雑していると時もあるが問題なく利用できる	MPIのバージョンの違いにより、複数/コンパイルできないバイナリ/コードがあります。複数のバージョンは使えないものでしょうか。
一般研究用	林 俊宏	東京大学大学院医学系研究科	Imaging genetics and biomarker analysis of neurodegenerative diseases.										その他	画像遺伝学(imaging genetics)				神経変性疾患の脳画像統合データベース(米国ADNIプロジェクトなどと提携)のデータを利用し、で、神経画像(MRIなど)とGWAS/バイオマーカーデータの関連解析研究を行っている。2012年11月より使用開始させて頂いており、現在は画像データの前解析(疾患特徴的な情報を抽出するための画像処理と統計解析)を行っている段階である。2013年度半ばをめどに本解析に入る予定である。	その他		FSL	遅い	大規模画像データを並列処理できるので大変助かっております。解析の段階にもよりますが、基本的にはデータ件数(症例数)に比例する計算時間となり、PCでは一件あたりの前処理でも大凡1日必要なので、n>100だと効用は大きいです。	混雑していると時もあるが問題なく利用できる	まだ準備段階なので本格的には使わせて頂いておりませんが、それでも大変助かっております。もともとPCクラスターを作って解析を行っていたので、そのままスケールアップが出来ることろもありがたいです。
一般研究用	徳田 直子	名古屋大学 大学院工学研究科 計算理工学専攻 笹井研究室	遺伝子の発現調節などの染色体が関わる機能と染色体の3次元構造との関連を明らかにすることを目的とする。										シミュレーション、MD					近年Taddeoらにより、出芽酵母のテロメアが核膜に固定されない状況下にある突然変異株では、野生株における遺伝子発現パターンとは異なる遺伝子発現が報告された。しかしながら、テロメアが核膜に固定されない染色体構造のもとでは、なぜ 数々の遺伝子の発現量が野生株において変化するのか詳細なメカニズムは明らかではない。そこで、我々は野生株と突然変異株に対応する核内染色体モデルを構築し、染色体の構造変化・揺らぎという視点からこの問題にアプローチしている。国立遺伝学研究所のスーパーコンピューターを活用、先ほど述べた目的を達成するためのシミュレーションを進行するために用いた、具体的には、テロメアと核膜との相互作用がある条件下とない条件下において、発現量に変化の見られた数々の遺伝子の核内における配置の変化および平均的な配置のまわりの揺らぎを調べた。今後はこれらの遺伝子の核内配置の変化がなぜ発現量の変化に繋がるのか、あるいは、発現量の変化を引き起こす他の原因についても検討する予定である。	goc perl 自作プログラム	ランジェバン動力学シミュレーション		遅い		混雑していると時もあるが問題なく利用できる	メモリー使用率の高いプログラムを実行する際には国立遺伝学研究所のスーパーコンピューターは、搭載されているメモリーが最低でも64GB もあり、とても助かります。ただ、スパコンのシステム稼働状況を見ていると、Thin ノードではコア数は足りているにも関わらず、メモリー使用率が100% 近くに達してしまっているように、さらなる計算が実行できない、という状況がたまに見られます。このような事から、Thin ノードのメモリーをもう少し増加させたほうが、効率よく計算機ノードを活用できるような気がしています。
一般研究用	山川 武廣	国立遺伝学研究所 系統生物研究センター 系統情報研究室	バイオリソースに関する解析一般										DB作成					NIG SuperComputerを使用し、57ライブラリのコムギESTs(679,211seq)についてBLAST解析を行いました。得られた結果は、National BioResource Project(NBRP)のデータベースに登録し、各種関連情報とともにコムギ統合データベース(NBRP KOMUGI)より一般公開しました。	ncbi_blast+ perl		NCBI-nr	普通		混雑していると時もあるが問題なく利用できる	当方の利用できる環境では、非常に時間のかかる解析が短時間で処理できるため、重宝しております。今後も継続して利用させていただきたいです。
一般研究用	相場 厚輝	国立遺伝学研究所 系統生物研究センター 系統情報研究室	バイオリソースに関する解析一般										DB作成					NIG SuperComputerを使用し、57ライブラリのコムギESTs(679,211件)についてBLAST解析を行いました。得られた結果は、National BioResource Project(NBRP)のデータベースに登録し、各種関連情報とともにコムギ統合データベース(NBRP KOMUGI)より一般公開しました。	ncbi_blast+ perl		NCBI-nr	普通		混雑していると時もあるが問題なく利用できる	当方の利用できる環境では、非常に時間のかかる解析が短時間で処理できるため、重宝しております。今後も継続して利用させていただきたいです。
一般研究用	渡辺 拓貴	国立遺伝学研究所 生物遺伝資源センター 系統情報研究室	バイオリソースに関する解析一般										DB作成					NIG SuperComputerを使用し、57ライブラリのコムギESTs(679,211件)についてBLAST解析を行いました。得られた結果は、National BioResource Project(NBRP)のデータベースに登録し、各種関連情報とともにコムギ統合データベース(NBRP KOMUGI)より一般公開しました。	ncbi_blast+ perl		NCBI-nr	普通		混雑していると時もあるが問題なく利用できる	当方の利用できる環境では、非常に時間のかかる解析が短時間で処理できるため、重宝しております。今後も継続して利用させていただきたいです。
一般研究用	渡邊 融	国立遺伝学研究所 系統生物研究センター 系統情報研究室	バイオリソースに関する解析一般										DB作成					NIG SuperComputerを使用し、NCBI-GenBank Release 191.0 のOryza sativa 塩基配列(243,172)についてBLAST解析を行いました。得られた結果は、各種関連情報とともにイネ統合データベース OryzabaseよりNSD (DDBJ/EMBL/GenBank) entry やマップ情報として一般公開しました。	BLAT			普通		混雑していると時もあるが問題なく利用できる	当方の利用できる環境では、非常に時間のかかる解析が短時間で処理できるため、重宝しております。今後も継続して利用させていただきたいです。
一般研究用	木村 学	国立遺伝学研究所 生物遺伝資源センター データベース事業部	バイオリソースに関する解析一般										DB作成					NIG SuperComputerを使用し、NCBI-GenBank Release 191.0 のOryza sativa 塩基配列(243,172)についてBLAST解析を行いました。得られた結果は、各種関連情報とともにイネ統合データベース OryzabaseよりNSD (DDBJ/EMBL/GenBank) entry やマップ情報として一般公開しました。	BLAT			普通		混雑していると時もあるが問題なく利用できる	当方の利用できる環境では、非常に時間のかかる解析が短時間で処理できるため、重宝しております。今後も継続して利用させていただきたいです。
一般研究用	松野 恭兵	国立遺伝学研究所 系統生物研究センター 系統情報研究室	バイオリソースに関する解析一般										DB作成					NIG SuperComputerを使用し、NCBI-GenBank Release 191.0 のOryza sativa 塩基配列(243,172)についてBLAST解析を行いました。得られた結果は、各種関連情報とともにイネ統合データベース OryzabaseよりNSD (DDBJ/EMBL/GenBank) entry やマップ情報として一般公開しました。	BLAT			普通		混雑していると時もあるが問題なく利用できる	当方の利用できる環境では、非常に時間のかかる解析が短時間で処理できるため、重宝しております。今後も継続して利用させていただきたいです。
一般研究用	原田 清令	明治大学 農学部 生命科学科 バイオインフォマティクス研究室	明治大学バイオインフォマティクス研究室での研究解析に使用。NGS解析、アセンブルなど。										NGS		Whole Genome Sequencing		454 GS 454 GS FLX Titanium Illumina HiSeq 1000 Illumina HiSeq 2000	共同研究先、データベースサイトなどから入手したゲノム塩基配列のリードデータのde novo アセンブルなどを行っています。現在進行中の研究内容としては、トマトのMicro-Tomという品種の454ロングリードと、illuminaショートリードを混ぜ合わせ、ゲノムの混合アセンブルなどを行っております。	SOAPdenovo bwa samtools velvet openmpi perl その他		Ray	DDBJ NCBI-nt NCBI-nr UniProt	普通		混雑していると時もあるが問題なく利用できる	今後、ゲノムアセンブルや各種バイオインフォマティクスの解析においては、より大きなデータを扱う機会が増えると思うので、十分な研究を行うには、それに対応して、より多くの計算機リソースが必要になると思われます。
一般研究用	磯口 千洋	個人	NGS解析										NGS		Exome Sequencing		Illumina MiSeq Ion Torrent PGM	ExomeシーケンスデータからのCNV計算や、スプライシングバリアントの検出、変異を考慮したMS同定のためのアミノ酸配列の作成などを行うべく、現在対象データをORAsearchで検索するだけに留まっており、本格的な利用は まだ行えていない。論文で公開されているアルゴリズムやフリーソフトウェア色いろいろと試してみたい。	ncbi_cxx_toolkit openmpi R			EMBL	普通		混雑していると時もあるが問題なく利用できる	これだけのリソースを一般公開していることに感謝します。本格的な利用はまだ行えていないですが、今後目一杯利用させて頂ければと思っています。
一般研究用	峯田 克彦	北海道大学大学院情報科学研究科	NGSデータの解析、および分子進化解析										NGS		RNASeq		Illumina HiSeq 2000	本年度中に、RNA-seqデータを用いた解析、およびその解析結果を利用した比較解析を実施することを想定して、事前にスパコンの利用申請をさせていただいたが、実験サンプルの調整に時間を要したため、当初予定していた状況にならなかった。そのため、今年度は実際にスパコンを利用してのRNA-Seq解析は行っていない。	perl			普通		混雑していると時もあるが問題なく利用できる	上述の理由により、本年度は実際にスパコンを使用しての解析を行っていないため、速度・混雑状況へのアンケートも含めて、回答は控えさせていただきます。
一般研究用	中谷 明弘	新潟大学・研究推進機構	遺伝子発現解析										NGS		RNASeq		Illumina HiSeq 2000	国立遺伝学研究所を中心とした複数機関の間でのモデル動物(マウス)のNGSによるゲノム配列情報に関する共同研究のための情報共有を行うために主に外部記憶装置を介したサイズの大きなデータの共有のために使用しました。東京国立遺伝学研究所の関係の網羅的な計算手法等の解析に関しては引き続き継続して検討していく予定です。	その他		UNIXコマンド群		普通		混雑していると時もあるが問題なく利用できる	「スパコン」ならではの重たい計算には活用できていませんでした(申し訳ありません)。
一般研究用	稲垣 宗一	国立遺伝学研究所 育種遺伝研究部門	次世代シーケンサーデータの解析										NGS		Epigenetics		Illumina HiSeq 2000	公開されているクロイネナズナゲノムバイサルファイトシーケンスによるDNAメチルセデータおよび、自分で行ったタイリングアレイによるクロマチン修飾データを解析し、各遺伝子内におけるパターンの抽出を行った。様々なパラメーターを変化させて、クロマチン修飾パターンを詳細に調べて、DNAメチル化との関連を見出すことが出来た。	bowtie samtools perl			遅い		十分空いている	まだスパコンのパワーを十分に生かせるような仕事をしていないが、速度が早いので条件を色々変えて解析を行うのに有用である。
一般研究用	鈴木 克之	近畿大学工学部生物化学工学科	次世代シーケンサーによって得られた細菌ゲノム塩基配列データの解析を行うため。										NGS		Whole Genome Sequencing		Illumina HiSeq 2000	中等度放射線耐性菌K. rosea NBRC15344株のゲノム配列データをIlumina社の次世代シーケンサーHiSeq2000を用いて解析した結果、総コンティグ数203、総塩基数3,781,154を得た。これらデータに対してアノーテーションを実施するためにMGAPプログラムを適用した結果、3,586のタンパク質コード配列、47のtRNA、2377のリボソーム結合部位(RBS)、およびひとつのrRNAの存在を推定することができた。	その他		MGAP	DDBJ	遅い		混雑していると時もあるが問題なく利用できる	解析速度の速さに驚きました。ただし、全投入DNAデータに対するアノーテーションの結果を一つの表にしようとすると、各項目別にデータを揃えるのが困難でした。
一般研究用	川原 善浩	農業生物資源研究所 農業生物先端ゲノム研究センター	次世代シーケンサーのデータ解析										NGS		RNASeq		Illumina Genome Analyzer II Illumina HiSeq 2000	Trinity、Oases、ABySS、SOAPdenovoなどのアセンブルツールを利用して、de novo transcriptome assembleの評価を行った。自分でツールをインストールするのが困難な場合でも、すでにサーバにインストールされているものを利用して、使い勝手や動作を確かめることが出来て便利であった。	Cufflinks GATK Oases Picard SOAPdenovo Trinity bowtie2 bwa cutadapt samtools tophat2 velvet R perl			遅い		混雑していると時もあるが問題なく利用できる	NGS関連ツールはアップデートが頻繁なのでフォローするのが大変だと思いますが、更新をもう少し早くしてくれると助かります。
一般研究用	高岡 勝吉	大阪大学	マウス初期発生期に必要な遺伝子の探索										NGS以外の配列解析					過去に報告されているいくつかの他研究室のchipseqデータを解析を行った。マウス初期胚を用いたchipseqのデータはまだ少ないが、マウス塞床前後の胚と性質が似ているとされるES細胞や、EPiSTEM細胞のデータは充実しており、それらのデータを用いて解析することで、マウス初期胚発生に必要な遺伝子の候補を絞り込むことができた。	IGVviewer MACS bowtie		DDBJ NCBI-dbGSS	普通		混雑していると時もあるが問題なく利用できる	非常にありがたいです。
一般研究用	岩林 一彦	山梨大学 医学部	オルガネラ遺伝子および長寿遺伝子の遺伝情報を用いた医学的研究										その他	総合・複合・ゲノム科学				日本人の集団が外国に移動して寿命が変わった報告に基づいて、日本人の余命と外国人の余命とのいろいろな比較を発表したが、動物の寿命に関連する報告されている Sirtuin・Klotho・Foxo などの遺伝子のメイン機能をヒトのほかの遺伝子と関連づけて解析したい。 エネルギーの無制限な取り込みが寿命を短くすることは確かなのでエネルギー関連遺伝子の制御を解析したい。	BLAT			DDBJ GenBank	普通		混雑していると時もあるが問題なく利用できる	
一般研究用（大規模）	野口 英樹	国立遺伝学研究所 先端ゲノミクス推進センター	ゲノム配列解析全般										NGS		Whole Genome Sequencing		454 GS FLX+ Illumina HiSeq 2000 Illumina MiSeq PacBio RS	本年度は主に、シーラカンスのゲノム・RNA-seq解析（主として遺伝子予測）に本計算機を利用した。ab initio遺伝子予測のほか、RNA-seqのde novoアセンブルや既知脊椎動物遺伝子との相同性検索を行い、シーラカンスの全遺伝子のアノテーション情報を作成した。また、作成した遺伝子配列をもとに脊椎動物間で比較解析を行った。これらの成果は、現在論文にまとめ投稿中である。	BLAT ClustalW Cufflinks SOAPdenovo Trinity bowtie2 nobi_blast+ tophat2 velvet goc perl 自作プログラム	c++とopenmpを用いて、シーラカンス用の遺伝子予測プログラムおよび、RNA-seq用のアセンブラ（補助的なもの）を開発。		GenBank RefSeq-Genomic RefSeq-RNA RefSeq-Protein NCBI-nr Pfam	普通		混雑していると時もあるが問題なく利用できる	基本的には快適に使えていますが、急ぎの時（論文のリバイスなどで期間が限られている時）にはあてにせずないのが現状です。本当に混んでる場合は仕方ないと思いますが、必要以上のメモリを確保していると（CPUも実メモリも空いてる）がいるとイラッときます。
一般研究用（大規模）	前原 一満	九州大学 医学研究院 エピジェネティクス分野	次世代シークエンサーによるエピゲノム解析	A co-localization model of paired ChIP-seq data using a large ENCODE data set enables comparison of multiple samples.	Published	23125363	Direct isolation and RNA-seq reveal environment-dependent properties of engrafted neural stem/progenitor cells.	Published	23072808	Chd2 interacts with H3.3 to determine myogenic cell fate.	Published	22569126	NGS		Epigenetics		Illumina Genome Analyzer IIx Illumina HiSeq 1000	多能性幹細胞は、再生医療の可能性の幅を飛躍的に広げることに成功した。しかしながら、iPS細胞やES細胞等の実用化のために、特定の組織への分化を決定する調節システムが必要な点は、未だ解決されていない。この分化調達系確立のためには、細胞分化における一群の遺伝子がいかに選択され発現が調達されるのか理解することが不可欠だと考えられる。エピジェネティクスの進展に伴って、個々の遺伝子発現制御解析から、遺伝子発現の"群"としての解析への転換を踏まえて遺伝子群全体を制御する引き金を探索することで幹細胞システムの応用を図る試みを進めている。次世代シークエンサーやスーパーコンピューターとの連携を図りながら、骨格筋分化をモデル系として分化に伴って発現する一群の遺伝子の発現制御メカニズム解明を行った。論文として成果を3編報告した。	BEDTools IGVviewer MACS Picard bowtie bowtie2 cutadapt samtools tophat tophat2 R		DDBJ	遅い		混雑していると時もあるが問題なく利用できる	専用キューの設定があれば素晴らしいですね。
一般研究用（大規模）	松波 雅俊	北海道大学 地球環境科学院	エゾサンショウウオの表現型可塑性を引き起こす遺伝子を同定するためのNGSを用いたRNA-seqによる発現量解析	Vertebrate Paralogous Conserved Noncoding Sequences May Be Related to Gene Expressions in Brain	Published	23267051							NGS		RNASeq		Illumina HiSeq 2000	エゾサンショウウオの表現型可塑性の分子機構を解明するために経時的なRNA-seqをおこなった。得られたreadをTrinityによりアセンブルした後、readをcontigにmappingすることで各遺伝子の発現量を見積もった。これにより、異なる表現型で有意に発現量が異なる遺伝子を同定した。今後は、さらに繰り返し実験を行う予定である。	ClustalW Cufflinks Picard Trinity bowtie bowtie2 mafft nobi_blast+ nobi_cxx_toolkit nobi_toolkit R gcc java perl ruby		RefSeq-Genomic RefSeq-RNA NCBI-nt NCBI-nr	普通		混雑していると時もあるが問題なく利用できる	ソフトウエアアップデートをしてほしい。
一般研究用（大規模）	広瀬 修	金沢大学電子情報学系	次世代シーケンサデータ解析、特に、RNA-seqデータの解析法に関する研究に利用する.										NGS		RNASeq		Illumina Genome Analyzer II	近年RNA-seqと呼ばれる技術によって、細胞内に存在するエクソンの網羅的な観測が可能になった。この技術では、エクソン自体を観測するのではなくエクソンを構成するRNA断片を観測する。したがって、スプライシングバリアントが存在する場合には、そのバリアントの構成比、すなわち、産生されつつあるタンパク質の比率を容易には特定できないという問題がある。観測されるRNA断片は、使用するシーケンサーの種類によって、単にRNA断片を観測する single read と RNA断片の両端から観測する paired-end read に分けられる。このなかでも、paired-end read の場合には、現在のところ、paired-end の特徴を効果的に利用できていない。本年度の研究では、paired-end に固有の生成モデルを構成し、その逆問題を解くことによって、スプライシングバリアントの存在比の推定精度を向上させることを目指した。その結果、精度を顕著に向上させることができ、現在論文準備中です。	BEDTools Cufflinks IGVviewer bowtie bowtie2 bwa samtools srat tools tophat tophat2 velvet gsl R gcc perl ruby				普通		混雑していると時もあるが問題なく利用できる	高速かつ大容量のリソースを無料で利用することができ、大変良い研究環境でとても満足しています。今後ともなにとぞよろしくお願い申しあげます。
一般研究用（大規模）	藤澤 貴智	国立遺伝学研究所 生命情報研究センター 文量遺伝情報研究室	次世代シークエンサ等の配列解析に利用する。										DB作成					文部科学省ライフサイエンスデータベース統合推進事業 統合化推進プログラム「ゲノム・メタゲノム情報を基盤とした 微生物DBの統合」において モデル微生物のゲノムアノテーションリファレンス情報整備を目的として、Blast、InterPro解析データの生成およびデータ形式変換のために遺伝研スパコンを利用している。シアノバクテリア、放線菌データについて、ゲノム情報を中心に各種データの RDFへの変換を実施し、試作版データベースのリソースを整備している。	InterProScan nobi_blast+ perl ruby		DDBJ GenBank RefSeq-Genomic UniProt Swiss-Port	遅い		混雑していると時もあるが問題なく利用できる	大量のデータ解析に有益に利用させていただいてます。
一般研究用（大規模）	土田 博子	明治大学農学部生命科学科	大規模データのアセンブル、マッピング。										NGS		RNASeq		Illumina HiSeq 2000 Illumina MiSeq	高速シーケンサーのデータを用いた高精度なSNP探索の方法を確立を目指して、現在は、高速シーケンサーデータの準備を行っている段階です。開発点では、小規模のテストデータでプログラムの検討やオプションなどの条件検討行い、何通りかの方法を試し、高精度なSNPの探索法を検討しています。大規模解析は、もう少し先になる予定です。	bowtie bowtie2 bwa samtools tophat tophat2				普通		混雑していると時もあるが問題なく利用できる	これから使用させていただきます。
一般研究用（大規模）	夏目 俊	公益財団法人 岩手生物工学研究センター	ゲノムのde novo assembly										NGS		Whole Genome Sequencing		Illumina HiSeq 2000	我々の研究グループでは、いくつかの植物ゲノムのデノボ・アセンブリを行なっている。主にALLPATHS-LGを用いた処理を行なっているが、このプログラムは多大な時間とデータ容量領域が必要であるため、本スーパーコンピュータシステムの利用は当研究に欠かせないリソースとなっている。来年度も本年度同様の作業を継続して行い研究成果を出したい。	ALLPATHS-LG Trinity			DDBJ	普通		混雑していると時もあるが問題なく利用できる	停電や障害なども頻繁に発生するようだが、その都度適切に丁寧な対応をしていただき感謝している。オンラインでの利用のため、サービス提供者と利用者のコミュニケーションが重要と考えるので、今後も活発な情報提供をお願いしたい。

アカウント区分	使用者	所属機関	利用目的	論文名1	論文の状況1	Pubmed ID1	論文名2	論文の状況2	Pubmed ID2	論文名3	論文の状況3	Pubmed ID3	研究成果(分野)	研究成果(分野)(その他)	NGSプロジェクト名	NGSプロジェクト名(その他)	NGSプラットフォーム名	研究成果・経過報告(内容入力)	使用プログラム	使用プログラム(自作プログラム)	使用プログラム(その他)	使用データベース	アンケート-速度-	アンケート-速度-(自由入力欄)	アンケート-混雑状況-	アンケート(自由入力欄)
一般研究用(大規模)	笠原 雅弘	東京大学大学院新領域創成科学研究科	次世代シークエンサーのデータ解析ソフトウェア開発およびそのための並列計算基盤ソフトウェアの解析。	Genome Sequencing and Assembly Using Only Long Reads from Single-Molecule Sequencers with High Error Rate	In Preparation								NGS		Whole Genome Sequencing		Illumina Genome Analyzer Ex Illumina HiSeq 2000 Illumina MiSeq PacBio RS	全ゲノムショットガン法を用いた新規(de novo)ゲノムアセンブリでは、ゲノム中に含まれるリードまより反復配列を原因として結果にミスアセンブリが多く含まれる。あるいは生成コンティグ長が短くなる問題が知られている。また、近年の技術改良が目覚ましい第三世代シーケンスセンサーであるPacBio RS は、広く出回っている商用シークエンサーの中では最も長いリードを出力することができる。そこで、我々は新規ゲノムアセンブリにおいて PacBio RS のロングリード(CLRリード)を用いることで反復配列の影響をより抑え、より高精度なゲノム配列を復元するアルゴリズムの開発を目指した。既存技術として、454 や Illumina などの短(高精度なリードを PacBio RS のロングリードにアラインメントを行いシークエンシングエラーを補正する PacBioToCA や LSC などが知られているが、454 や Illumina 等のシークエンシングプラットホームはそれぞれゲノム中の特定配列コンテクストに応じて系統的なシークエンシングエラーが発生したり、特定の配列が欠如または偏りの固有の問題があり、454 や Illumina などのリードを補正に用いることでそのようなエラーを原理上持たないと思われている PacBio RS ロングリードに問題を持ちふんでいた。この問題に対処するために、我々は PacBio RS のロングリードのみを用いて(PacBio RS のロングリード vs PacBio RS のロングリード)アラインメントを実行し、得られたアラインメントからシークエンシングエラーを打正するアルゴリズム Sprai を開発し、エラー打正後のリードを Celera Assembler (CA) を用いてアセンブリを行った。Sprai + CA の組み合わせを用いて E coli K-12 MG1655, E coli SE11, M. ruber, P. heparinus DSM2366 から得られた PacBio RS のロングリードのみを用いて実際にエラー打正・ゲノムアセンブリを行ったところ、何れのバクテリアに対しても数個のコンティグが得られ、ロングリードのみから染色体レベルの良いコンティグが得られることを示した。また、Sprai 開発中に PacificBioscience 社から発表された類似アルゴリズム HGAP とのお総も行い、ゲノムサイズに対して 100x 以上のリードを入力とした場合に同等の、40x 程度のリードを入力とした場合にはより大きなサイズのコンティグを出力できることを見いだした。論文投稿準備とともに、現在はより大きなゲノム(70Mb～2Gb)に対してテスト・改良を行っている。また、数百Mb以上の大きなゲノムに対しては Illumina リード等とハイブリッドセンブリを行いシークエンシングコストを抑えるための研究開発も継続して行っている。	bwa nobi_blast+ gcc java perl python ruby 自作プログラム	multithreadを主に用いプラインメント・ゲノムアセンブリを行った。			普通		混雑していう時もあるが問題なく利用できる	ディスクのアクセス帯域が非常に広く、また、次世代シークエンサー向けのサードパーティソフトウェアが数多くインストールされており(ゲノム解析をメインに)行うユーザーにとっては非常に使いやすい。スパコンとのディスク容量を節約するために fuse (File System in User Space) などをユーザー権限で使えるようになっていると(ラボのディスクをネットワーク経由で使えるので)より良い。
一般研究用(大規模)	大柳 一	国立遺伝学研究所 系統生物研究センター 植物遺伝研究室	高速シーケンサーを利用した野生イネ多様性に関する研究										NGS		Whole Genome Sequencing		Illumina Genome Analyzer Ex Illumina HiSeq 2000	主にOryza属野生イネのゲノム解読(主にリシーケンシング)に利用させていただいています。数種以上の近縁野生イネ系統について、高精度ゲノム解読済みの栽培イネゲノムにマッピングすることで多型を検出し、育種学的・分子遺伝学的な追加解析や考察をおこなっています。現在は解析はほぼ終了し論文執筆を進めています。	Cufflinks bowtie bowtie2 bwa cutadapt phylip samtools tophat tophat2 R perl				普通		混雑していう時もあるが問題なく利用できる	・時々混雑しているなと感じることがご最近増えました。頻繁ではないです。・ディスクについて、大現精ユーザとして今のところ自由に使わせていただいています。ありがとうございます。スパコンが本格稼働したら、コストダウンのために研究室のファイルサーバ(100TB強)を導入して完全にスパコンディスクに依存する時にしようかと思っていたのですが、最近のディスク高速のメールなど拝見すると、それは難しいのかと思い始めています。
一般研究用(大規模)	亀田 倫史	産業技術総合研究所 生命情報工学研究センター	生体高分子(主にRNA)の立体構造予測。具体的にはMD計算およびdocking計算を行う。										シミュレーション、MD					1)RNA立体構造予測に関して(RNA分子は、通常のMDシミュレーションを行うには大きすぎるので、1塩基をSPD球で表現した粗視化モデルを構築した。MD用計算ソフトCHARMMを改造し、この粗視化モデルを組み込み実行した。さまざまなRNA配列に対して、立体構造予測計算を実行したところ、短め(塩基長50以下)であれば、高確率で立体構造を正しく予測できることがわかった。現在、さらに多数の配列に対して是解構造を予測できるようにモデル改良中である。2)遺伝性アルツハイマー病変症をもたらすAβペプチド変異体の立体構造解析本年度は、難陀賞GM1からなるミセルと、WTのAβペプチドとの相互作用をMDシミュレーションを用いて研究した。まず、GM1を表現する分子モデルは、既存のMDソフトに用意されていないので、量子化学計算ソフトGaussianを用いて、分子モデルを構築した。次に作成した分子モデルを元にミセルを構築し、MDシミュレーションを行ったところ、安定な構造を保つことから、分子モデルは適切であると判断した。現在、GM1からなるミセルと、WTのAβペプチドを共存させた系のMDシミュレーションをすることで、研究を進めている。計算結果と実験データが合致することが確認できた時点で、遺伝性変異が入ったAβペプチドとの共存シミュレーションを行う予定である。	gromacs-gpu openmm openmpi CUDA R gcc java perl python				遅い		混雑していう時もあるが問題なく利用できる	これまで様々なスパコンを利用してきましたが、スペック、利用環境、サポートなど、非常に優れていると思います。運営には、多くの予算と手間が必要だと思いますが、是非維持していただきたいと思います。今後ともよろしくお願いします。
一般研究用(大規模)	菊地 泰生	宮崎大学医学部寄生虫学	寄生虫ゲノムの解析										NGS		Whole Genome Sequencing		Illumina HiSeq 2000	寄生性線虫Bursaphelenchus xylophilusのゲノムをde-novoでアセンブルを行い、リファレンスゲノムを完成させた。遺伝子予測はEST情報とManually curated genesを用いて行った。現在はいくつかのステージ、コンディションにおけるRNAseqを用いて遺伝子予測の向上を手掛けるとともにDifferential Expression解析を行っている。	ALLPATHS-LG BEDTools BLAT GATK IntelProScan Picard RAxML Trinity bowtie2 bwa mafft nobi_blast+ samtools tophat2 velvet R perl python			DDBJ GenBank RefSeq-RNA NCBI-nt NCBI-nr EMBL Swiss-Port Pfam	普通		混雑していう時もあるが問題なく利用できる	今のところ目立った混雑もなく快適である。
一般研究用(大規模)	藤澤 貴智	国立遺伝学研究所 生命情報研究センター 大量遺伝情報研究室	プロジェクト用アカウント										DB作成					文部科学省ライフサイエンスデータベース統合推進事業において、ライフサイエンス統合データベースセンター、東京工業大学、基礎生物学研究所、遺伝研・DDBJと連携して、データベース統合化を目的とした基盤技術開発を行なうために遺伝研スパコンを利用し、ゲノム、環境、生物種情報等についてまとにRDFデータを共有している。	その他		2次グループ作成目的のアカウントのため特に利用なし		普通		混雑していう時もあるが問題なく利用できる	2次グループ作成を目的としたアカウントのため、特にありません。
一般研究用(大規模)	藤澤 貴智	国立遺伝学研究所 生命情報研究センター 大量遺伝情報研究室	開発用SPARQLエンドポイント構築および公開のため										DB作成					文部科学省ライフサイエンスデータベース統合推進事業において、ライフサイエンス統合データベースセンター、東京工業大学、基礎生物学研究所、遺伝研・DDBJと連携して、データベース統合化を目的とした基盤技術開発を行なうために、SPARQLエンドポイントのインストールを行い、ゲノム情報、タンパク質情報、リンク情報のRDFデータを格納を実施した。UniProt等大規模公共データベースのRDFインポートやSPARQLによる問い合わせの実行速度の検証など開発フェーズで試行している。	その他			OWLIM, Virtuoso		遅い		混雑していう時もあるが問題なく利用できる	大変有用です。
一般研究用(大規模)	川島 秀一	情報・システム研究機構 ライフサイエンス統合データベースセンター	ゲノム配列データの解析										NGS		RNASeq		Illumina Genome Analyzer	現在、ネッタイシマカ(Aedes aegypti)のデングウイルス感染実験、および ステフエンサイハマダラカ(Anopheles stephensi) の Plasmodium yoelii 感染実験の RNA-seq 解析を行なっている。ネッタイシマカはドラフトゲノムが公開されていたためショートリードをリファレンス配列へマッピングできる。一方、A. stephensi はゲノム配列が公開されていないため、Trinityアセンブラでde novoアセンブルを行った転写産物配列をリファレンス配列として、マッピングを行うという手順をとっている。	BEDTools ClustalW Cufflinks SOAPdenovo Trinity bowtie bowtie2 bwa fasta mafft nobi_blast+ phylip tophat tophat2 velvet R gcc perl ruby			DDBJ GenBank RefSeq-Genomic NCBI-nt NCBI-nr Swiss-Port TrEMBL Pfam	普通		混雑していう時もあるが問題なく利用できる	貴重なリソースを利用させていただいて感謝しています。今後とも、ソフトウェアやデータベースの更新等、対応していただければと思います。
一般研究用(大規模)	牧野 能士	東北大学大学院生命科学研究科	次世代シークエンサー由来の大量データを解析する(RNAseqやPassemblyなど)。										NGS		RNASeq		Illumina HiSeq 2000	Hiseq2000で読んだアノールトカゲ配列のアセンブリを実施した。また、同じHiseq2000で読んだアノールトカゲのRNAseqの解析も実施した。各遺伝子の個体間の発現量のばらつきが小さく、期待した通りのアセンブリデータが得られた。今後は、他種の配列データを用いたアセンブリを実施していきたい。	BLAT ClustalW bowtie nobi_blast+ samtools				普通		混雑していう時もあるが問題なく利用できる	有用に活用させて頂いている。
一般研究用(大規模)	長崎 英樹	国立遺伝学研究所 生命情報研究センター 大量遺伝研究室	次世代シークエンサ等の配列解析に利用する。	DDBJ Read Annotation Pipeline: a cloud computing-based pipeline for high-throughput analysis of next-generation sequencing data	In Preparation								NGS		Whole Genome Sequencing		AB SOLiD System Illumina Genome Analyzer Illumina Genome Analyzer II Illumina Genome Analyzer Ex Illumina HiScanSQ Illumina HiSeq 2000	パラゴムゲノム解析ではお世話になりました。http://www.bridgestone.co.jp/corporate/news/2012071002.htmlhttp://car.watch.impress.co.jp/docs/news/20120711_546034.htmやビニコウケゲノム解析でも利用させていただいています。https://pag.confex.com/pag/xxi/webprogram/Paper6485.html	ALLPATHS-LG bwa samtools gcc perl その他		ローカルでplatanusとか	DDBJ GenBank	普通		混雑していう時もあるが問題なく利用できる	大変お世話になっております。今後ともよろしくお願いいたします。
一般研究用(大規模)	谷澤 靖洋	国立遺伝学研究所 生命情報研究センター 大量遺伝情報研究室	高速シーケンサーの配列データ解析										その他	DDBJ Pipelineの開発・運用、共同研究によるNGSデータ解析				DDBJ Pipelineの開発、運用を行いました。http://g.ddbj.nig.ac.jp/pipelineプリザストンとの共同研究によるパラゴムのゲノム配列を行い、下記のプレスリリースを発表しました。http://www.bridgestone.co.jp/corporate/news/2012071002.html その他、遺伝研スパコンを利用した解析結果を、下記のデータベースに公開しました。http://tga.nig.ac.jp/ /http://tga.nig.ac.jp/dnapod/	ALLPATHS-LG BLAT SOAPdenovo bwa samtools perl python その他		Platanus, Blast	DDBJ GenBank	普通		混雑していう時もあるが問題なく利用できる	FATノードMediumノードについては一部のユーザーにとる語点が目立つときがあります。これらノードについてはhuノードとは別の制限も必要ではないでしょうか。また必要以上のメモリを要求しているケースもあるようですので、ユーザーへの啓蒙が必要です。また、無料で解放しているので、強制的な接続をもっと強めてもよいのではないかと思います。(一時期に比べ高点検部が減ってきたように思いますので、すでにやられていて効果が出てきているのではないかとも思います)が
一般研究用(大規模)	水口 洋平	国立遺伝学研究所 生命情報研究センター 比較ゲノム解析研究室	NGSデータ解析										NGS		Whole Genome Sequencing		454 GS FLX+ Illumina HiSeq 2000	ゲノムが既知、および未知の生物種に対して、NGSを利用した解析を行ったが、ゲノムが未知の生物種に対しては、GCバイアスなどでNGSでは配列が得られないものがあるなど、コンタイン サートのペアエンド配列を一定量得られるように、実験で開発するなどの対応が必要だった。また、ヒトゲノムサイズのゲノムに対して、一般的な多型探索をWhole Genomeで行うとメモリ使用量がボトルネックになり、データ分析や配列補正などを複数の処理が必要になる。これらの経験や結果を反映した成果を作成中である。	BLAT Cufflinks GATK Picard SOAPdenovo Trinity bowtie2 bwa phylip samtools tophat2 perl その他		last		普通		混雑していう時もあるが問題なく利用できる	10などの影響により、一つのノードで複数のジョブが投入されるときに、たまに負荷が高すぎて、結果が得られないときがあるので、整理していない何が起こっているのかが不明なことがある。また、JAVAのメモリ不足でジョブが終了したときに、エラーファイルに何も書かれなく、qacctで異常終了だったと推定するのは、手間。
一般研究用(大規模)	小池 麻子	(株)日立製作所 公共システム事業部	JST 統合化推進プロジェクトの一環として、ヒトゲノムの公開データを利用した解析に利用する										NGS		Exome Sequencing		Illumina Genome Analyzer II	ヒトの公開whole genome 及びexomeデータを用い、コンセンサス配列を生成する際の、アセンブル、マッピング、リアライメント、変異抽出などの計算手法依存性の検討を行うとともに、vcfデータを用いた手法の妥当性検討を行った。また、多検性の集団問題異の体系化を行うとともに、疾患関連変異データの集団間差異との関係性の解析を実施した。	Cufflinks GATK Picard bowtie bowtie2 bwa cutadapt samtools tophat openmpi R gcc その他		ABYSSなど多数		普通		混雑していう時もあるが問題なく利用できる	fat nodeは1ヶ月のジョブのみでなく、一週間ジョブがあってもよいように思います。
一般研究用(大規模)	筈木 将司	国立遺伝学研究所 生命情報研究センター 比較ゲノム解析研究室	次世代シーケンサーからのデータ処理全般										NGS		その他	パイプライン構築	454 GS FLX+ Illumina HiSeq 2000 Illumina MiSeq PacBio RS	Illumina社製の次世代シーケンサ(HiSeq2000,HiSeq2500,GA Ex,MiSeq)の解析パイプラインを構築した。本パイプラインはシーケンサから生成された生データのクオリティチェック、リファレンス配列へのマッピング、マッピング結果を基にしたエラーレートの推定情報を求める、レーン単位で生成される生ファイルを分割し、クラスター上で並列処理を行うことにより高速化を実現している。	gcc perl				普通		混雑していう時もあるが問題なく利用できる	大量のデータをコピーしているので、転送速度が気になる。
一般研究用(大規模)	坂本 亜子	国立遺伝学研究所 生命情報研究センター 大量遺伝情報研究室	次世代シーケンサ等の配列解析に利用する。										NGS		Whole Genome Sequencing		unspecified	DDBJ-Pipelineアカウントでの作業が、主となっており、(現在、pipelineサイト内にはTaxonomy Ver. Up等の更新項目の修正作業、pipelineユーザー入力データの確認及び、依頼による変更作業などのため、不定期な作業が発生しております。)今年度は自分アカウントでの作業がほとんどありませんでしたが、来年度以降、作業が発生する場合を見込み継続申請致します。	SOAP SOAPdenovo bowtie bowtie2 bwa samtools tophat velvet				普通		混雑していう時もあるが問題なく利用できる	特に問題ありません。
一般研究用(大規模)	山本 泰智	ライフサイエンス統合データベースセンター	生命科学分野の文献情報処理										DB作成					生命科学分野の文献情報データベース、MEDLINE全体から抽出した略語とその展開形の組をデータベース化したAllie について、RDF化を行うとともに、WikipediaをRDFにしたDBpediaのエントリと展開形をリンクする作業を行っている。これにより、Allieの利用者は効率的に文献中に出現する略語に関するWikipediaにおける記事を取得できるようになる。	gcc java perl				普通		混雑していう時もあるが問題なく利用できる	hadoopも利用できると嬉しい。
一般研究用(大規模)	寺田 愛花	東京工業大学 大学院情報理工学研究科	細胞の状態制御には複数の転写因子の組合せが重要である。申請者は、単一の転写因子ではなく 複数の転写因子が組合わさる事で初めて起きる現象発見のため、貴研究所のスーパーコンピュータシステムを利用する。										アルゴリズム開発					異なる転写因子と作用することで遺伝子の発現を制御する 転写因子の組合せの発見を目指し、新たな多重検定補正手法を開発した。本手法により、既存手法の過剰な多重検定補正と 膨大な実行時間の2つの問題を同時に防ぐことができる。その結果、全遺伝の転写因子の組合せから、遺伝子発現に有意に関わるものを発見することが可能となった。現在は、組合せで働く転写因子の発見を目指し、データベースに蓄積されている 転写因子の結合部位の網羅的な観測データやマイクロアレイなどの遺伝子発現データを、本手法を用いて解析している。	python				普通		十分空いている	スパコンを利用することで、多くのデータを同時に解析できるようになった。その結果、多数なデータに対して、開発した手法を公用することが可能になった。新たな知見の発見に向けた解析だけではなく、本手法の将来・不得手とするデータを知ることでも、スパコンシステムがとても役立っている。
一般研究用(大規模)	北住 愛	University of Maine, Orono	イネのNGS解析										NGS		Transcriptome Analysis		Illumina HiScanSQ	植物遺伝研究室:倉田先生のもと共同研究として、Oryza officinalisの器官別mRNA-seqを解析しています。この後、同じ植物体から得たライブラリーを追加し、それぞれの組織特異的な遺伝子発現のパターンを探ります。同時に、栽培イネであるOryza sativaとの比較も行います(そのためにさらにTBにも容量を引きとげる申請をし、受理いただきました)。現在、Oryza officinalisにおいてのマッピング環境やバラター調節中です。	Cufflinks Trinity bowtie2 bwa cutadapt nobi_blast+ samtools tophat R perl python			DDBJ GenBank Pfam	普通		混雑していう時もあるが問題なく利用できる	いつもお世話になっています。ありがとうございます。
一般研究用(大規模)	赤間 悟	東京工業大学 大学院情報理工学研究科	遺伝子解析を始め、その他バイオインフォマティクス的解析全般										NGS		Transcriptome Analysis		Illumina HiSeq 2000	植物胞質雄性稔株について、環境変化に伴う発現変動の解析を行っている。まず、ショートリードをアセンブルし、遺伝子領域予測並びに遺伝子アノテーションを行うことで、新規のトラフトゲノムの作成を行った。次に、RNA-seqデータから、ゲノムの由来を同定しつつ発現量を解析する手法を構築した。現在、異なる環境下で得たRNA-seqデータについて、発現量解析を進めている。	bowtie2 samtools tophat tophat2 R perl python				遅い		混雑していう時もあるが問題なく利用できる	容量、同時実行ジョブ数ともに非常に多くのリソースを使わせて頂き感謝しております。
一般研究用(大規模)	安井 康夫	京都大学農学研究科	新型シーケンサのアウトプットの解析(ALLPATHS-LG, SOAPdenovo等を用いたゾルゲノムアセンブル、およびSOWTIE, BWA, samtoolsを利用したマッピング(による変異解析)	Survey Sequences of Buckwheat Genome	In Preparation								NGS		Whole Genome Sequencing		Illumina HiSeq 2000	(1)ソバゲノム解析 ソバゲノム(約1.2Gb)の約3倍からなる160-200bpの両端末配列の解読、および3・10kbのジャンピングライブラリーから得た塩基配列をデータとして、SOAPdenovoおよびALLPATHS-LGによるアセンブルを行った。両アセンブル結果を比べたところ、SOAPdenovoでは3Kbp(ジャンピングライブラリーの最小距離)に足りなくがあるようで、多くの時間(30日)と膨大なメモリ(4T)を必要とするが、ALLPATHS-LGのアセンブルがより信頼できるであることがわかった。ただ、個々のアセンブル結果を比べていないので、今後更なる検証が必要と思われる。現在、40kbpロングジャンピングライブラリーの作成と塩基配列解読を予定しており、即ちれば4月中に両解析を行いたいと考えている。さらに、GBSによるマッピングを行い、bin mapとGSSを組み合わせた形で発表したいと考えている。GSSマッピング(2ヶ月)に実施予定であり、今年の夏もしくは秋を目指に発表したいと考えている。この際、遺伝研のプラットフォームを利用させていただければ、非常に嬉しいです。 (2)ゲノムサブトラクション重イオンビーム照射により得られた自家不和合性変異体のゲノム全50倍のカバレッジのHiseq2000を用いて解読し、さらに上記で得られたアセンブル結果をリファレンスとして用いてBWAによるマッピングを行った。現在、PCRによるウェットでの確認作業中であるが、十分にゲノムサブトラクションを実施できる。以上のように本研究ではALLPATH-LGによるゲノムアセンブルやBWAを用いたマッピングを実施しており、遺伝研のスパコン利用なくては成立しえない課題となっている。今後、ますます多くの作物でゲノム解析が実施されると考えられ、遺伝研でのスパコン利用は日本のゲノム育種に大きく寄与すると考えています。	ALLPATHS-LG Gases SOAPdenovo Trinity bowtie bwa cutadapt velvet perl			DDBJ GenBank	普通		混雑していう時もあるが問題なく利用できる	この時代、ウエットだけでは研究が成り立たなくなっております。特に遺伝学分野ではゲノム解析が簡便になってきたため、ドライを使いきれるかどうかが研究の鍵となっていると思います。私はコンピューターはずぶの素人なのですが、スパコンチーム様からも熱切丁寧なご指導を頂き、なんとか研究を進めることができております。(とくにALLPATHS-LGの起動ではお世話になりました。)遺伝研スパコンの大容量のメモリはもちろん大きな魅力ですが、利用者に対するバックアップ体制も非常に大きな魅力です。今後とも何卒よろしくお願いいたします。
一般研究用(大規模)	小笠原 理	国立遺伝学研究所 生命情報研究センター 遺伝子発現解析研究室	GPUを利用した集団遺伝学のシミュレーションおよび 各種データベースの作成										シミュレーション、MD					分子進化の速度はおおよそ一定であるが、形態進化の速度は一定ではないことが知られている。分子進化と形態進化を橋渡しする理論として遺伝子発現調節の進化的変化が重要なのではないかと考えられており、従来のトランスクリプトームの測定技術の発達により、ようやく遺伝子発現の進化的変化の一般的な性質が明らかになってきたところである。申請者はこれまでの測定データの性質を統一的に説明できる理論モデルを提案しているが、これまでは各個体の適応度の計算に時間がかかるため、十分なサイズのシミュレーションができていなかった。特に寄連する遺伝子座の数が十分でなかった。今年度は遺伝研スパコンを利用して遺伝子発現進化に対する自然淘汰の影響の詳細なシミュレーションを行っており、現在結果のデータの解析を行っているところである。この結果はのちに論文として発表する予定である。	CUDA R gcc java perl python 自作プログラム	集団遺伝学的なシミュレーションプログラム(センチュウ:ロンミュレーション)Java multi-threadプログラムやGPUを用いて並列化している。			普通		混雑していう時もあるが問題なく利用できる	ノード間の並列化のために現在はMPIしか使えないが、それだけでやFortranでコードを書かないといけない。Javaやスクリプト言語でもノードを超えた並列計算ができるようにRabbitMQなどのAMQPを提供したらよい思う。
一般研究用(大規模)	望月 孝子	国立遺伝学研究所 生命情報研究センター 大量遺伝情報研究室	次世代シーケンサ等の配列解析に利用する。										DB作成					今年度は、DRAデータベースからNGSのゲノム配列を収集し、均一な精度でDNA多型を検出し、検出したDNA多型に遺伝子構造、同義置換/非同義置換などの注釈情報を付与したDNA Polymorphism Annotation Database (DNApod) の構築 (http://tga.nig.ac.jp/dnapod/) を行った。	BLAT GATK bwa samtools sra-tools perl				普通		混雑していう時もあるが問題なく利用できる	時期的に非常に混雑しているときがあり、解析が進まず苦労したこともありました。が、容量、速度に大変満足し、活用させて頂いております。今後とも、どうぞよろしくお願い致します。

アカウント区分	使用者	所属機関	利用目的	論文名1	論文の状況1	Pubmed ID1	論文名2	論文の状況2	Pubmed ID2	論文名3	論文の状況3	Pubmed ID3	研究成果(分野)(その他)	研究成果(分野)(その他)	NGSプロジェクト名	NGSプロジェクト名(その他)	NGSプラットフォーム名	研究成果・経過報告(内容入力)	使用プログラム	使用プログラム(自作プログラム)	使用プログラム(その他)	使用データベース	アンケート-速度-	アンケート-速度-(自由入力欄)	アンケート-混雑状況-	アンケート(自由入力欄)
一般研究用(大規模)	舘々 潤	東京工業大学 大学院情報理工学研究科	高速シーケンサによるゲノム解析、発現量解析手法の開発、及び、たんぱく質相互作用情報の解析												NGS	Whole Genome Sequencing	Illumina HiSeq 2000	植物のゲノム7種に関して、ゲノムアセンブリ、RNA-seq等を含めた遺伝子構造・発現解析と、遺伝解析を行なっている。一般に植物のゲノムは非常に高次でリピートが多いと言われているが、今回対象としている種は、いずれも200-500Mbp程度、ラボ内で掛けあわせが行われていたり、遺伝的にheterozygosityが低いものであったりと、比較的アセンブルが容易な環境下の個体を選定して行なっているため良好な結果を得ている。	ALLPATHS-LG BLAT Cufflinks SOAPdenovo Trinity bowtie2 bwa samtools tophat tophat2 R gcc java perl python ruby			DDBJ NCBI-dbEST	普通		混雑している時もあるが問題なく利用できる	medium ノードが混んでいることが多いように感じる。thinノードのメモリが足りないことが多い(コアを全部使う前に、メモリの上限が問題になる事が多い)
一般研究用(大規模)	千葉 啓和	基礎生物学研究所 ゲノム情報研究室	大規模比較ゲノム解析												アルゴリズム開発			ドメイン単位のオーソロググループが与えられたとき、ドメインの進界をマルチプルアライメントント上で再評価して修正することのできるシステムを構築している。オーソログクラスター毎に、マルチプルアライメントを作成して、スコア計算と妥当性の評価を繰り返し行った。スコアの最適化によって、修正ができるシステムの開発につなげる。									
一般研究用(大規模)	西山 智明	金沢大学 学際科学実験センター	真核生物のde novo assemblyなど												NGS	Whole Genome Sequencing	Illumina HiSeq 2000 PacBio RS	約100 MBの真核生物ゲノムについてPacBio RSのリードデータとPBJellyをもちいてGap filling を行った。報告のGapを埋めることができたが、QVは必ずしも期待されるほど上がっておらず、先にError Correcsion を行うか、さらにリードをマッピングしてコンセンサスをとるなどの工夫を加えることを今後試みる。ヒメミカヅキモ、クルミホソガ虫についてはIllumina リードデータをゲノムにマッピングする解析を行った。	ALLPATHS-LG GATK SOAPdenovo bowtie bwa その他			PBJelly, RSEM, PacBioToCA	普通		混雑している時もあるが問題なく利用できる	Fat, Mediumノードは混んでいて使いにくい。Thinノードは大抵十分に空いている。
一般研究用(大規模)	三浦 理	早稲田大学大学院 先進理工学研究科	間期染色体三次元構造の構築機構の解析	間期染色体三次元構造の構築機構の解析											シミュレーション、MD			本研究では、間期染色体の三次元構造の構築原理を解明することを目的として、出芽酵母(Saccharomyces cerevisiae)の間期染色体の全長を、スクレオソームの配置が識別できる分解能でモデル化した。具体的には、スクレオソームの配置、リンカーDNAの物理的特性、および細胞核の大きさのパラメータを用いてモデルを構築した。予測された間期染色体構造と核内における実際の構造の比較を行なったところ、予測値と実験値とが非常によい一致を示した。このことから、モデルの構築に用いた3つの因子(スクレオソームの配置、リンカーDNAの物理的特性、細胞核のサイズ)が、間期染色体の三次元構造を規定する本質的な因子であり高可能性が強く示唆された。	R gcc				普通		混雑している時もあるが問題なく利用できる	スパコンを使用させて頂くことで、高効率に、統計的に十分な量の計算を行なうことが可能となりました。特に、DNA物性や細胞核内環境などを変化させた様々なパラメータセットを用いた計算を行なうことが容易になり、研究を進行するにあたり、非常に助かっております。今後とも、どうぞ宜しくお願い致します。
一般研究用(大規模)	大田 達郎	情報・システム研究機構ライフサイエンス統合データベースセンター	SRAにご登録されている全NGSデータのシーケンスクオリティ統計値の計算												DB作成			Sequence Read Archiveに登録された全てのアクセス可能なfastq配列データに対して、最も広く利用されているクオリティ・コントロールを行うソフトウェアの1つであるFastQCを実行し、公共NGSデータのクオリティを集計し、DBの構築を行っている。現在、構築したDBのビューワーをベータ版として公開しており、正式公開のための準備中である(http://g86.dbcls.jp/sra)。新規に追加されたデータに関しても順次計算を続けており、破損しているデータに対する結果のリカバリも行なっている。	sra-tools ruby その他			FastQC	遅い		混雑している時もあるが問題なく利用できる	おおむね満足しています。スパコンの度々運営を行なっているチームの皆さんに感謝したいと思います。まだ時間はかかるとも思いますが、SRAのデータがどのノードからでも見えるようになり、転送のスピードも上がるとより助かります。
一般研究用(大規模)	高田 豊行	国立遺伝学研究所 系統生物研究センター	情報・システム研究機構 新領域融合プロジェクト研究において、マウスのRNA-Seq解析と解析手法の検討を複数の大学および研究所と共同でおこなう。特に本アカウントは、複数のプロジェクトの関係者が解析に必要なデータや結果を共有するために必要であるとともに、申請者が解析手法の検討と比較を行うために重要である。												NGS	RNASeq	Illumina HiSeq 2000	情報・システム研究機構 新領域融合プロジェクト研究で進行している、マウスのRNA-Seq解析と解析手法の検討を複数の大学および研究所と共同でおこなった。これまでに、解析に必要なリファレンス情報やHiSeq2000により得られた複数の組織から得られたサンプルによる遺伝子発現データを格納し、解析を開始することができた。	Cufflinks bowtie tophat R			RefSeq-Genomic RefSeq-RNA	普通		混雑している時もあるが問題なく利用できる	プロジェクトに参加している他のスパコンユーザと、効果的に解析の連携を行うことができ、大変ありがたく利用させていただいております。
一般研究用(大規模)	高田 豊行	国立遺伝学研究所 系統生物研究センター	マウスのゲノム機能解析を行う。具体的には塩基配列のアセンブル、ゲノム多型抽出、遺伝子発現、各種アノテーション、系統分類、統計解析、画像解析等を行う。												NGS	RNASeq	Illumina HiSeq 2000	マウスのゲノム解析として、新領域融合プロジェクト「遺伝機能システム学」(責任プロジェクトディレクター:ロサブテーマ)である。マウス遺伝子発現データの解析を行っている。これまでに、リファレンス配列の整備や、必要な情報の取得を行うとともに、HiSeq2000、2ラン分のデータを取得し遺伝子発現解析のためのデータ整備を進めた。	bowtie bwa tophat R			DDBJ GenBank RefSeq-Genomic	普通		十分空いている	汎用プログラムの導入など、時間とともに使用感が向上しており、大変ありがたく感じております。今後ともよろしくお願いいたします。
一般研究用(大規模)	佐藤 和広	岡山大学資源植物科学研究所	オオムギゲノム配列のアッセンブル												NGS	Whole Genome Sequencing	454 GS FLX Titanium Illumina HiSeq 2000	オオムギ品種「はるなこ茶」の454ゲノムとHiseqでのハイブリッドアッセンブルを試みている。Newbler2.7をインストールし、まず、454単独ならびHiseq単独でのアッセンブルを実施したが、コマンドラインが不適切であったためか、演算に多くの時間を要し、結局満足のいく結果は得られなかった。引き続き、ジョブ投入によってHiseqのリードのアッセンブルを試みたが、現在のところ成功していない。商用ソフトウエアの使用も試みているが、スパコンへのインストールは難しいと感じている。	SOAPdenovo velvet その他			Newbler	普通		混雑している時もあるが問題なく利用できる	引き続き、利用を試みて成果を得たいと思っているのでご支援いただきたい。
一般研究用(大規模)	小柳 亮	沖縄科学技術大学院大学 マリンゲノミックスユニット	次世代シークエンサデータを用いたゲノムのde novoアセンブル	MarinegenomicsDB: An integrated genome viewer for community-based annotation of genomes	In Press										NGS	Whole Genome Sequencing	454 GS FLX Titanium 454 GS FLX+ Illumina Genome Analyzer IIx Ion Torrent PGM	青所スーパーコンピュータを採用して行ったアコヤガイゲノムの解析についてまとめた上記の論文がZoological Science誌に採択されました。私どものユニットにはアセンブルの必要なゲノムシークエンスデータが他にも数種あり、OIST以外資機では処理しきれない状況が今後も起こると予想されますので、引き続き使用許可をいただければ幸いです。	Trinity その他			celera assembler	普通		混雑している時もあるが問題なく利用できる	OISTにはない大容量メモリの計算機が利用できるので、大変感謝しております。
一般研究用(大規模)	藤田 雅文	国立遺伝学研究所 系統生物研究センター 植物遺伝研究室	高速シークエンサーのデータ解析												NGS	Transcriptome Analysis	Illumina Genome Analyzer IIx Illumina HiSeq 2000 その他	2013年1月21より利用を開始したばかりである。NGSのデータファイルは1個50GBを超えるため、初期設定の1TBでは ディスク容量が足りず、2月20付でリソース拡張申請を行った。現在はNGSのデータ解析に用いるcatadapt、bowtie2等の各種ツールの使用方法を指導者の指示の下で学んでいる段階であり、研究成果と呼べるものはまだ無い。	Cufflinks bowtie bowtie2 cutadapt samtools tophat2 perl その他			fastQC	普通		十分空いている	スパコンの「速度」については、他機の使用経験がないため、回答不能。「NGSプロジェクト名」では、Transcriptome Analysisの他にWhole genome sequencingにも関与している。複数選択できるようにしてほしい。
一般研究用(大規模)	五十嵐 香理	明治大学 農学部 生命科学科	大規模データアセンブル、マッピング												NGS	その他	Transcriptome analysis, Whole genome sequencing	共同研究で3つの研究課題の解析のため、ゲノムとトランスクリプトームのde novo assembleを行いました。1つの研究課題はアセンブル結果から機能アノテーションなどを行い、現在論文投稿中です。他、2つの研究課題は一通りのアセンブルは終了しましたが、シーケンスリードの追加と入れ替えがあったため、現在も継続してアセンブルの検討を行っています。	Dases SOAPdenovo Trinity velvet openmpi perl その他			Ray	普通		混雑している時もあるが問題なく利用できる	混雑するようになってから確かに少し使いづらくはなりましたが、使用メモリ量が500G程度の解析でしたらほとど問題なく使用できました。
一般研究用(大規模)	片山 俊明	大学共同利用機関法人 情報・システム研究機構 ライフサイエンス統合データベースセンター	ゲノム解析とゲノムデータベース構築のためのデータ作成												DB作成			RDFによる大規模データベースの統合に必要となる、オントロジーの整備、RDFデータの生成、SPARQL検索の実装や検証などを行った。扱う対象が巨大データであるためSemantic Webに関する難技術の成熟が必要な部分もあるが、将来性は高いと判断し、これらをもとにゲノムアノテーション情報を統合化するプリケーションの開発を継続している。	Cufflinks bowtie2 bwa ncbi.blast+ tophat2 R gcc java ruby			DDBJ GenBank RefSeq-Genomic NCBI-nr UniProt Pfam	遅い		混雑している時もあるが問題なく利用できる	生命情報科学のコミュニティのために自由で使いやすい環境が整えられておりありがたいです。
一般研究用(大規模)	近藤 伸二	国立極地研究所新領域融合研究センター	次世代シークエンサが出すゲノム、RNA及びメタゲノム配列データの処理、解析	Metagenomic analyses of ice core samples collected in south pole	In Preparation										NGS	Transcriptome Analysis	454 GS FLX Titanium Illumina Genome Analyzer Illumina Genome Analyzer IIx Illumina HiScanSQ	454 GS FLX Titanium Illumina Genome Analyzer Illumina Genome Analyzer IIx Illumina HiSeq 2000 Illumina MiSeq	BLAT ClustalW Cufflinks SOAP SOAPdenovo samtools tophat velvet R gcc			DDBJ RefSeq-Genomic RefSeq-RNA RefSeq-Protein NCBI-nr Pfam	普通		混雑している時もあるが問題なく利用できる	fatノードが混雑の為か、利用が困難な時があった。
一般研究用(大規模)	鳥谷 恵子	東京大学大学院・新領域創成科学研究科・メディカルゲノム官野・鈴木研究室	次世代シーケンサーから産出されるヒト熱帯熱マラリア原虫のRNA_seqデータの1次加工(mapping,SNP抽出等)	Interactive Transcriptome Analysis of Malaria Patients and Infecting Plasmodium falciparum	In Preparation										NGS	RNASeq	Illumina Genome Analyzer II	NGSによりマラリア患者116 samples、健常者28 samplesのRNASeq tag(30万個ほど)を得てヒトとマラリアゲノムにそれぞれ(cmap,発現量をヒト、マラリアそれぞれ遺伝子毎に算出した。またGATKにより、SNPを抽出した。上記データより以下の解析を行った。1.重症度(マラリアデクの刺合)、体温、性別、年齢についてサンプルの2群の平均値解析により各カテゴリ間連遺伝子を推定した。さらに各群にenrichされたGO term,PATHWAY (を得た。2.ヒト・マラリア遺伝子の全組み合わせのサンプル間の発現パターンの相関を算出し系統樹作成。正・負相関のクラスターにenrichされたGOを得た。3.SNPについて:遺伝子毎sample間のSNPの状態と発現量から系統樹を作成したが何も注目すべき事柄は無かったので破棄とした。マラリアの薬剤耐性遺伝子のSNPを調べた。4.その他統計値算出以上を論文にまとめ、現在投稿準備中。3月投稿予定。	GATK bowtie bwa samtools tophat R perl				普通		混雑している時もあるが問題なく利用できる	今のところ概ね満足している
一般研究用(大規模)	横山 幸治	明治大学 農学部 生命科学科 バイオインフォマティクス研究室	アセンブル・マッピング等で使用												NGS	Transcriptome Analysis	454 GS FLX Illumina HiSeq 2000	SOAP・SOAPdenovo・Oases・velvetような様々なアセンブルラを使用させていただきています。当研究室では、リード数が多くてメモリ不足となり、実行できないアセンブルが存在します。そういった時には、大量のメモリが存在するこちらのスパコンを使用されていただいています。ですので、引き続き使用させていただければと思います。	Oases SOAP SOAPdenovo velvet				普通		混雑している時もあるが問題なく利用できる	混雑している感がありますが、JOBを投げられれば、問題なく利用できます。できれば、混雑しているのを解消していただければと思います。
一般研究用(大規模)	浜田 達昭	東京大学	自ら開発したルゴリズム・ソフトウェアを利用してがんなどのNGSデータの大規模解析を行う	PBSIM: PacBio reads simulator---toward accurate genome assembly	Published	23129296									NGS	Whole Genome Sequencing	PacBio RS	PacBioの2種類のリード(CLR・CCS)をシミュレートするシミュレーターPBSIMを開発し、PBSIMを用いて、既存のPacBioアセンブリソフトウェアPacBioToCAの評価実験を行った。シミュレーターはhttp://code.google.com/p/pbsim/ より無償でダウンロード可能である。	自作プログラム	mpi			普通		混雑している時もあるが問題なく利用できる	とても快適に利用しています
一般研究用(大規模)	野澤 昌文	国立遺伝学研究所 生命情報研究センター	次世代シーケンサーデータの解析	Reevaluaton of dosage compensation in Drosophila species	In Preparation										NGS	RNASeq	Illumina Genome Analyzer Illumina Genome Analyzer II Illumina HiSeq 2000 Ion Torrent PGM	ショウジョウバエにおいて遺伝子量補償が本当に存在するのか、またその程度を調べるため、さまざまな種のRNA-seqデータを解析し、既知遺伝子の発現量を種間で比較した。また、自ら実験して生成したRNA-seqデータについても解析を行った。その結果、まだ予備的な結果ではあるものの、ショウジョウバエにおいては確かに遺伝子量補償が広く存在していることが明らかとなった。しかしながら、遺伝子量補償を受けていない遺伝子群も確かに存在することが明らかになった。現在、個々の遺伝子の補償の程度を決定する要因を探っている。	bowtie bowtie2 bwa ncbi.blast+ ncbi.toolkit sra-tools tophat2 perl				普通		混雑している時もあるが問題なく利用できる	今のところ、ジョブを投入するたびにいつの場合すぐにジョブが走るので快適に使用させていただいています。今後、特にストレージが増強されるとさらに良いと思います。ご検討よろしくお願いいたします。
一般研究用(大規模)	山口 淳一	東京大学新領域	カイコ・ブナヶ月ちうのゲノム、及び、トランスクリプトーム解析において、次世代シーケンサーから出力される大規模シーケンスデータを解析するための。												NGS	RNASeq	Illumina HiSeq 2000	今週のはじめにアカウントを新送して頂き、比較トランスクリプトーム解析のために使用を開始した。現在、Trinity を使った Hiseq2000、1レーン分の de novo アセンブルのジョブが順調に稼働中である。このアセンブルデータは、ゲノム配列がオープンになっていないアゲハチョウ科のトランスクリプトーム解析には必須であり、大きなコンピューターの稼働が、研究を進めるとその最大の恩恵事項であった。今回の遺伝研のスパコン利用に関するスムーズな承認と丁寧なWEBとの解説のおかげで、研究を滞りなく進められる目処が付いた。今後、さらにRNAseqの実験データを追加し、近縁異種間の比較トランスクリプトーム解析から、脱皮目 昆虫の体表鱗移形成に関わる遺伝子群の進化に関する論文を発表していく予定である	Trinity bowtie ncbi.blast+ samtools java perl			NCBI-nr	普通		混雑している時もあるが問題なく利用できる	大型計算機の利用は経験がなかったため躊躇していましたが、WEB上のスライドのPDFファイルの説明がとてもわかりやすかったため、思い切って申請に踏み切り、順調に使えるようになりました。大変、感謝しています。
一般研究用(大規模)	橋 靜佳	東京大学大気海洋研究所	海洋ゲノムに関する幅広い研究を行う												NGS	Metagenomics	454 GS FLX+	さまざまな環境に生息する微生物より得られたゲノムデータ、また微生物群集より得られたメタゲノムデータの解析を行っている。特に、様々な環境の変動に対する微生物の応答についてゲノムレベルの情報に基づいた分析を行っている。論文発表に至る研究成果はまだ得られていないが、生態系にとって重要な微生物の働きについて新たな知見が得られつつあり、次年度には成果報告を行いたい。	bwa nobi.blast+ samtools perl				普通		混雑している時もあるが問題なく利用できる	バイオインフォマティクス研究の推進に大変役だっており、心から感謝申し上げます。今後ともどうぞよろしくお願いいたします。
一般研究用(大規模)	新美 輝幸	名古屋大学 大学院生命農学研究科	高速シークエンサーにより得られたデータ解析のため。												NGS	RNASeq	Illumina HiSeq 2000	Illumina HiSeq 2000により得られた配列ファイルをサーバにアップロードして、解析用にファイルの統合等を行った。また、RNA-Seq解析に必要なアプリケーションのダウンロードなど、解析環境を整備した。レファレンスゲノムが存在しないため、今後得られた配列のDe novoアセンブリを行う予定である。	ClustalW Trinity bowtie ncbi.blast+ perl				普通		混雑している時もあるが問題なく利用できる	ウェブマニュアルが充実していてとてもよい。解説動画(講習会の様子？)があるとより分かりやすいかも。
一般研究用(大規模)	町山 麻子	東京大学大気海洋研究所	海洋ゲノムに関する幅広い研究を行う												NGS	Metagenomics	454 GS FLX 454 GS Junior	さまざまな環境に生息する微生物より得られたゲノムデータ、また微生物群集より得られたメタゲノムデータの解析を行っている。特に、様々な環境の変動に対する微生物の応答についてゲノムレベルの情報に基づいた分析を行っている。論文発表に至る研究成果はまだ得られていないが、生態系にとって重要な微生物の働きについて新たな知見が得られつつあり、次年度には成果報告を行いたい。	BLAT samtools velvet perl			DDBJ GenBank	普通		混雑している時もあるが問題なく利用できる	問題なく利用できる。
一般研究用(大規模)	櫛 康広	京都大学重長類研究所	次世代シーケンスデータ解析												NGS	RNASeq	Illumina HiSeq 2000 Illumina MiSeq	霊長類のWhole genome shotgun (resequencing), transcriptome (RNAseq), exome (exome-seq)の解析をスパコンとで行なっています。具体的には、チンパンジーのトリ/ゲノムとトランスクリプトーム解析、マカクザルのエキソーム解析、霊長類鱗トランスクリプトームの解析を行なっています。	BEDTools Cufflinks GATK IGVviewer Picard SOAP SOAPdenovo bowtie bowtie2 bwa samtools sra-tools tophat tophat2 R gcc java perl python ruby				普通		混雑している時もあるが問題なく利用できる	手持ちの計算機資源では解析が間にあわなくなりつつあり、DDBJのスパコンを利用させていただいています。混雑している場合もありますが、現状ではストレスなく使えています。今後ともこのようなサービスを引き続き提供いただけますようよろしくお願い致します。大規模利用申請を別途行ないました。そもらが認められればー般研究用のアカウントは利用を停止する予定です。
一般研究用(大規模)	神沼 英里	国立遺伝学研究所 生命情報研究センター 大量遺伝情報研究室	次世代シークエンサ等の配列解析に利用する。	DDBJ Read Annotation Pipeline: a cloud computing-based pipeline for high-throughput analysis of next-generation sequencing data	In Preparation										NGS	Population Genomics	454 GS FLX Titanium AB SOLiD System 3.0 Illumina Genome Analyzer Illumina Genome Analyzer IIx Illumina HiSeq 2000 Illumina MiSeq	今年度はDDBJ Pipelineのテストや、科研費の「次世代シークエンサーカイブ配列を用いた植物系統間多型の解析」に利用しました。具体的にはカンキツ11品種の多型解析を行い、denovoアセンブリや参照配列へのアライメントを行いました。また、DDBJ Pipelineは論文に投稿し、現在査読中です。	BEDTools BLAT IGVviewer SOAPdenovo bowtie bowtie2 bwa nobi.blast+ samtools sra-tools tophat tophat2 velvet gcc java perl ruby			DDBJ GenBank RefSeq-Genomic NCBI-nt NCBI-nr EMBL Pfam	遅い		混雑している時もあるが問題なく利用できる	利用し易いです。外部の方からも、バイオツールが便利だと意見を頂く事がありますよ。
一般研究用(大規模)	柴田 朋子	基礎生物学研究所	ゲノムアセンブリ												NGS	Whole Genome Sequencing	Illumina HiSeq 2000 PacBio RS	ゲノム解析手法の開発を目的とし、Illumina HiSeq及びPacBio RSを用いて新規バクテリアのゲノムシークエンシングを行い、両データを合わせてアセンブリを行った。いくつかのプログラムを用いて結果の比較を行ったほか、データ量の組み合わせを様々に変えることでアセンブリに必要充分なデータ量を探った。	ALLPATHS-LG ruby その他		Celera Assembelr		普通		混雑している時もあるが問題なく利用できる	通常は所属する基生研のコンピューターを利用して解析を行っていますが、必要メモリ量が多いときなどに遺伝研スパコンを利用させていただいています。解析が迅速に進み、大変助かっています。
一般研究用(大規模)	福多 賢太郎	国立遺伝学研究所生命情報研究センター比較ゲノム解析研究室	比較ゲノム解析研究に関するアルゴリズム及びアプリケーションの開発等												NGS	Whole Genome Sequencing	Illumina HiSeq 2000	インド産アカザザルをリファレンス(rheMac2)として、ニホンザル、中国産アカザザル、中国産カニクイザルの全ゲノムを対象とした比較ゲノム解析を行っている。【結通】各サルについてのSNPを割り出した(ニホンザルでは約1000万箇所のSNPを同定)。また、それぞれのサルについての特異的なSNPを見つけた(精度は今後の予定)。以上のSNP情報を元に4種を含めた系統樹の作成を行った。今後はDel関りの処理を固めた後に、ニホンザルに焦点を絞り、アノテーション情報を活用してニホンザルに特異的な生物学的な知見を明らかにしたいと考えている。この段階を一区切りとして論文にまとめる予定である。平行して、ニホンザルの標準ゲノム配列の作成にむけたアセンブルを行う。なお、南極シュードモナスの解析結果は依頼主に提出済み(Xbh SCARにて報告されるとのこと)。	GATK IGVviewer Picard samtools velvet 自作プログラム	C++, mpi: マッピングプログラム。Matlab(及びシェル,awk): パイルアップデータ(BAM)の処理、SNPの抽出、各種統計値の算出。		DDBJ NCBI-WGS EMBL	普通		混雑している時もあるが問題なく利用できる	thinノードの使用については大変快適と感じる。medium以上についてはかなり混雑しているように見える(特にメモリ)。現在のジョブの状況を把握し、自分のジョブがいつ実行されるかといった現状を把握しやすくするために、medium以上については「誰がどのようなリソース多いつまで確保しているのかが分かるような、タイムテーブルのような情報(カレンダーのように閲覧できる情報)」を用意するといったことは考えられないでしょうか。

アカウント区分	使用者	所属機関	利用目的	論文名1	論文の状況1	Pubmed ID1	論文名2	論文の状況2	Pubmed ID2	論文名3	論文の状況3	Pubmed ID3	研究成果(分野)	研究成果(分野)(その他)	NGSプロジェクト名	NGSプロジェクト名(その他)	NGSプラットフォーム名	研究成果・経過報告(内容入力)	使用プログラム	使用プログラム(自作プログラム)	使用プログラム(その他)	使用データベース	アンケート-速度-	アンケート-速度-(自由入力欄)	アンケート-混雑状況-	アンケート(自由入力欄)
一般研究用(大規模)	神沼　英里	国立遺伝学研究所　生命情報研究センター　大量遺伝情報研究室	次世代シークエンサ解析用のWebサービスDDBJ Read Annotation Pipelineを提供する	DDBJ Read Annotation Pipeline: a cloud computing-based pipeline for high-throughput analysis of new-generation sequencing data	In Preparation								NGS		Population Genomics		454 GS FLX Titanium AB SOLiD System 3.0 Illumina Genome Analyzer Ilumina Genome Analyzer Ix Illumina HiSeq 2000 Illumina MiSeq	今年度はDDBJ Pipelineのテストや、科研費の「次世代シークエンサアーカイブ配列を用いた植物系統間多型の解析」に利用しました。具体的にはカンキツ1品種の多型解析を行い、denovoアセンブリや参照配列へのアライメントを行いました。また、DDBJ Pipelineは論文に投稿し、現在査読中です。	BEDTools BLAT IGVviewer SOAPdenovo bowtie bowtie2 bwa nobi.blast+ samtools sra-tools tophat tophat2 velvet gcc java perl ruby			DDBJ GenBank RefSeq-Genomic NCBI-nt NCBI-nr EMBL Pfam	遅い		混雑している時もあるが問題なく利用できる	利用し易いです。外部の方からも、バイオツールが便利だと意見を頂く事があります。
一般研究用(大規模)	神沼　英里	国立遺伝学研究所 生命情報研究センター　大量遺伝情報研究室	NIGスパコン上でDDBJパイプライン高次解析を受け持つGalaxyを実行させる	DDBJ Read Annotation Pipeline: a cloud computing-based pipeline for high-throughput analysis of new-generation sequencing data	In Preparation								NGS		Population Genomics		454 GS FLX Titanium AB SOLiD System 3.0 Illumina Genome Analyzer Ilumina Genome Analyzer Ix Illumina HiSeq 2000 Illumina MiSeq	今年度はDDBJ Pipelineのテストや、科研費の「次世代シークエンサアーカイブ配列を用いた植物系統間多型の解析」に利用しました。具体的にはカンキツ1品種の多型解析を行い、denovoアセンブリや参照配列へのアライメントを行いました。また、DDBJ Pipelineは論文に投稿し、現在査読中です。	BEDTools BLAT IGVviewer SOAPdenovo bowtie bowtie2 bwa nobi.blast+ samtools sra-tools tophat tophat2 velvet gcc java perl ruby			DDBJ GenBank RefSeq-Genomic NCBI-nt NCBI-nr EMBL Pfam	遅い		混雑している時もあるが問題なく利用できる	利用し易いです。外部の方からも、バイオツールが便利だと意見を頂く事があります。
一般研究用(大規模)	保坂　碧	国立遺伝学研究所　総合遺伝研究系　育種遺伝研究部門	次世代シーケンサーを用いた発現解析等										NGS		Epigenetics		Illumina Genome Analyzer Ix Illumina HiSeq 2000	これまでにマッピングツールなどを駆使し、シロイナズナのエピジェネティック制御因子変異体においてゲノム高構成がおきていることを確認した。今後はさらに詳細な解析を行うための解析の条件検討などを行う予定である。また、現在は変異体における発現パターンやエピジェネティック修飾の異常がどのように表現型に影響を及ぼすかの包括的に解析するための手法を構築中である。	Cufflinks Trinity bowtie bowtie2 fasta samtools tophat perl				普通		混雑している時もあるが問題なく利用できる	特に不自由なく使用させていただいております。
一般研究用(大規模)	成田　岳雄	京都大学　医学研究科	RNA-seqデータを用いたtranscriptome解析										NGS		RNASeq		Illumina HiSeq 2000	RNA-seqのdataを用い、ヒト、マウスでの発現解析を行っている。種および組織依存的なisoformを考慮した遺伝子発現量の変化を調べている。今後、この結果をiPSの細胞株に適用し、トリ特異的な遺伝子発現パターンおよび新規遺伝子の同定に繋げ、その細胞株の特徴をヒト細胞株でも再現させたいと考えている。	Cufflinks SOAP3-db SOAPdenovo Trinity			NCBI-nt NCBI-nr	普通		混雑している時もあるが問題なく利用できる	Trinityを最新のバージョンにして欲しいです。
一般研究用(大規模)	森川　洋匡	大阪大学免疫学フロンティア研究センター	免疫細胞、特にT細胞のGenome, Transcriptome や Epigenomeのシーケンスデータを用いた細胞内遺伝子ネットワーク解析。対象はマウスとヒト。										NGS		Transcriptome Analysis		Helicos HeliScope Illumina HiSeq 2000 Ion Torrent PGM	現在、解析のためのサンプルを準備中。必要に応じて関係するpublish済みのSRAのdata、GEOのdatを解析しています。具体的にはRegulatory T-cellとConventional T-cell の比較を(Transcriptome, Epigenomeの両面から比較しています。Illuiffer-seqのデータが準備できればそちらを進める予定です。	BEDTools BLAT Cufflinks MACS SOAP bowtie bowtie2 bwa samtools sra-tools tophat tophat2 R perl			DDBJ RefSeq-RNA EMBL	普通		十分足いている	十分満足しております。環境設定などで困ったときにサポートしていただけるのもありがたいです。
一般研究用(大規模)	吉武　和敏	国立遺伝学研究所　生命情報研究センター　遺伝情報分析研究室	de novoアセンブル。特にゲノムのアセンブルをいくつかのソフトウェアで行いたい。										NGS		Whole Genome Sequencing		Illumina HiSeq 2000	菜種生物/無脊椎動物のゲノムアセンブルを行い、今年度はメイトペアなしのデータしか得られていないため、そこまでのデータでアセンブルし、アノテーションなどの解析を行った。その結果、微生物のコンタミが多かったため、再度シーケンスを行っている。次年度はメイトペアを含めたゲノムアセンブルを行いたいと考えている。	ALLPATHS-LG BEDTools BLAT Cufflinks GATK IGVviewer Oases Picard Pindel SOAPdenovo Trinity bowtie bowtie2 bwa nobi.blast+ samtools sra-tools tophat tophat2 velvet R java perl			NCBI-nt NCBI-nr	遅い		混雑している時もあるが問題なく利用できる	ゲノムアセンブルには500GB程度のメモリーが必要ですが、そのようなサーバを無料で利用できる場所はほかに知らないので、とても感謝いたしております。サーバの負荷状況もリアルタイムで更新されていて、混雑状況が一目でわかりとても使いやすいシステムだと思います。
一般研究用(大規模)	矢野　健太郎	明治大学・農学部	DNA配列解析、発現解析など	The Tomato Genome Consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485: 635-641.	Published								NGS以外の配列解析					各植物のゲノムアノテーション解析や遺伝子発現解析などを行った。ゲノム配列解析では、SNP探索などを行うとともに、ゲノム構造から、同義置換・非同義置換の頻度などを明らかとした。また、遺伝子発現解析では、これまでに開発している遺伝子発現解析手法を応用し、多くの条件下(サンプル、遺伝子型、環境条件など)での発現プロファイルを用い、特異的な発現遺伝子の探索とそれらの機能解析を実施している。	perl			DDBJ GenBank	普通		混雑している時もあるが問題なく利用できる	たいへん助かっております。今後も、このサービスの継続をどうぞよろしくお願いいたします。
一般研究用(大規模)	尾崎　遼	東京大学 大学院新領域創成科学研究科 情報生命科学専攻	多種体のNGSデータの処理および解析。										NGS		Transcriptome Analysis		Illumina HiSeq 1000	魚類A 8種の時系列トランスクリプトーム解析から、特定の条件下で発現の時間変化する遺伝子群を同定する手法を開発した。これを用いて、特定条件への適応に関連していると考えられる複数の未知遺伝子を含む遺伝子群を同定した。また、アノテーションのない種について、トランスクリプトームアセンブリ(de novo)を行うことで、遺伝子アノテーションを作成中である。	BEDTools Cufflinks IGVviewer cutadapt samtools sra-tools tophat2 R,java perl			DDBJ	普通		混雑していて不満だ	スパコン管理チームの方々の素早い対応に感謝しております.
一般研究用(大規模)	内藤　雄樹	ライフサイエンス統合データベースセンター	転写産物を中心とした大規模配列データの検索技術開発のため。										DB作成					DDBJリリースの各種データ情報を含む大規模塩基配列データについて、接尾辞配列(suffix array)によるインデックスを非常に高速に作成することができた。作成したインデックスを用いて、塩基配列を高速に検索することができるサービスを構築中である。構築したサービスはすべてのユーザが自由に利用できるようにする予定である。	自作プログラム	特になし		DDBJ	遅い		混雑している時もあるが問題なく利用できる	ディスクへのアクセスがかなり高速で計算がはかどった。
一般研究用(大規模)	谷口　丈晃	株式会社　三菱総合研究所	次世代シーケンサ等を対象にした大規模データの解析										NGS以外の配列解析					プロモータ配列における物理化学的な統計量や、モチーフ検索結果などを変数とした発現頻度推定モデルの作成・精度評価。特に、ステップワイズ変数選択や正則化項を加えた線型モデルを構築する。テストデータでは、少数の変数で高い精度のモデルを作成することはできているが、実データのオープンテストでは思い通りの結果が得られていない。	R			DDBJ	遅い		混雑している時もあるが問題なく利用できる	マニュアルも充実しており、快適に使わせていただいています。