

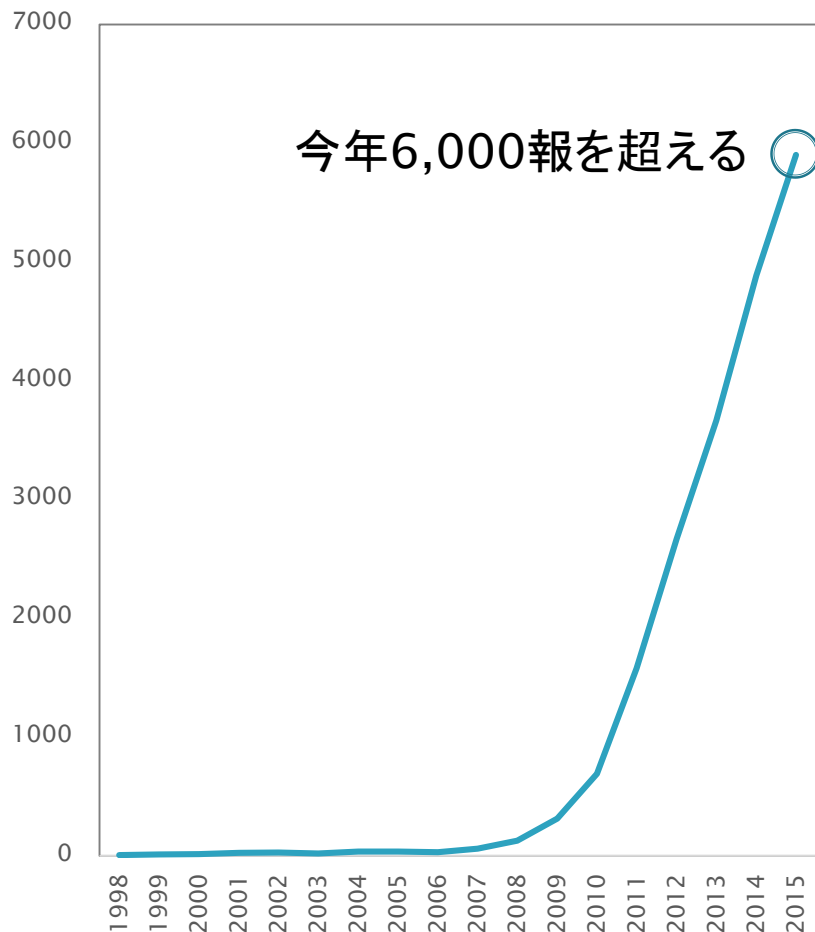
次世代シーケンサー(NGS)解析の 現状と失敗しないための心得

アクシオヘリックス株式会社

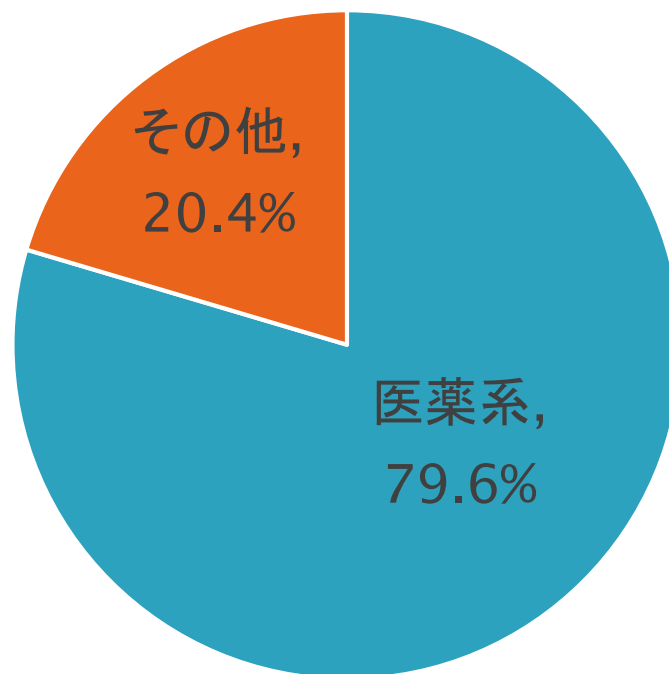
アジェンダ

- ▶ NGSを使った研究状況
- ▶ 失敗事例とその解決法
- ▶ 解析を外注する

次世代シーケンサー(NGS)を使った研究状況



NGS関連論文投稿数の推移



2015年 NGS関連 科研費分野別内訳

NGSデータ解析

★変異解析

シーケンシングした配列と既知の配列情報を比較し、Insertion, Deletion, SNPsを検出することで、ゲノム上のどの位置で、どのような変異が起こったか知ることができます。

新規ゲノム配列決定

配列情報が未知の領域や、未知の生物種について、配列を決定するための手法です。

★発現解析

RNAをシーケンシングし発現量を定量する手法で新規RNAの探索にも威力を発揮します。

ChIP-seq解析

免疫沈降法を利用し、タンパク質と結合するDNA配列を解読することで、転写因子が結合する領域を明らかにすることができます。

メチル化解析

遺伝子発現がどう制御されるか調べるためにゲノム配列のメチル化状態を解析する手法です。

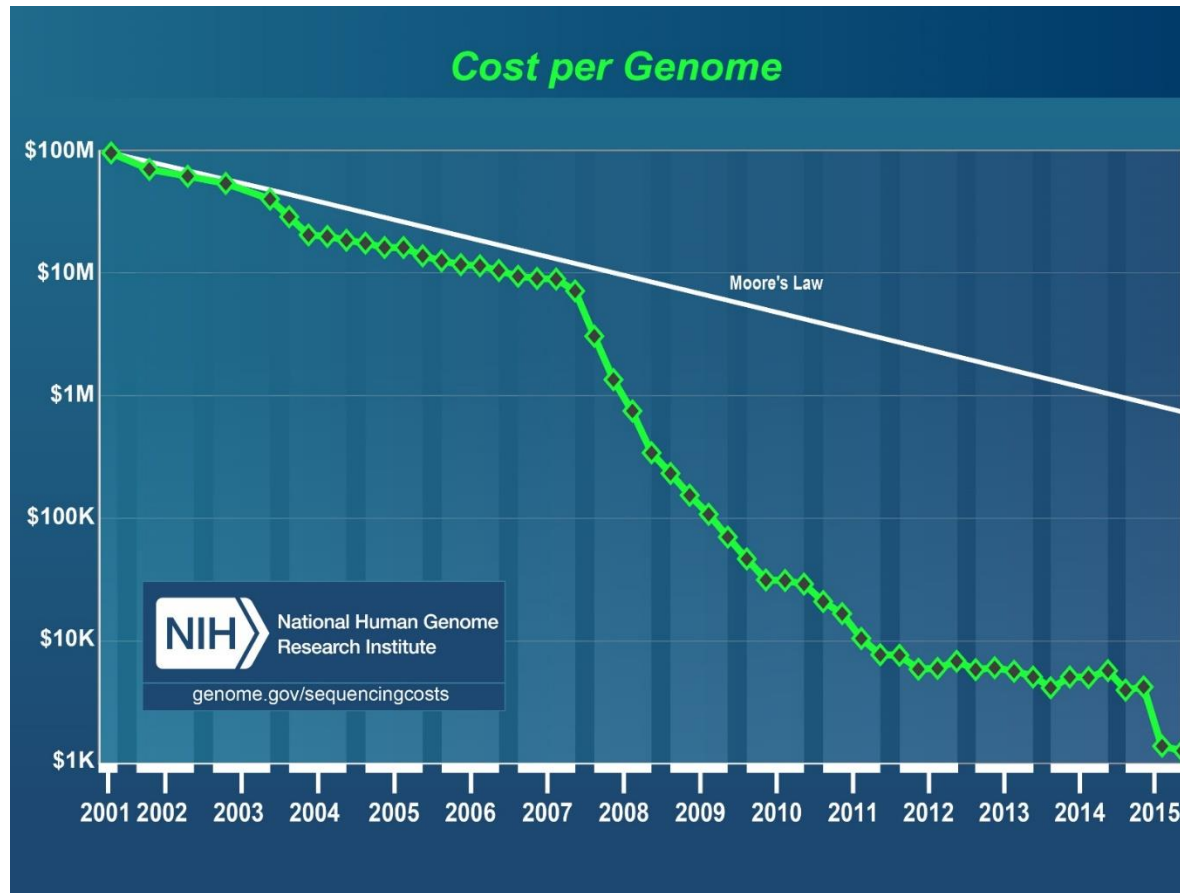
★メタゲノム解析

複数生物種のゲノムをまとめて網羅的にシーケンシングすることで、異なる条件下での生物群の組成を調べたり、物質代謝を推測することが可能になります。

NGS解析を行う上での障害

- ▶ コスト
- ▶ 専門的知識
- ▶ 解析環境の構築

コスト



HiseqX Tenがリリースされたことにより、ヒトゲノム (3.0×10^9 塩基対) が1,000ドルで読めるようになった。

NGSデータ解析作業で必要なこと

専門知識が必要

- ・ 解析ツールを利用するためにバイオ及びITの知識が必要

専用の環境が必要

- ・ シーケンスデータは数千万～数億リード数、容量としては数TB以上のものもある。
- ・ シーケンスデータは大量データの場合もあり、通常のPCでは解析が進められない場合も。
- ・ 専用の計算サーバーが必要となります。

解析に要する時間の確保

- ・ 解析作業に拘束されるため時間の確保ができない。
- ・ 知識を身に付ける時間や人員の確保が難しい。

NGS解析をしても論文が書けない！？

- ▶ 実は今、NGS解析は金食い虫だなんて言われている
→NGS解析をしても論文が書けないから

なぜ？

- ▶ データが出てても解釈ができない
- ▶ 解析する時間がない
 - バイオインフォマティクスの絶対量が足りない
 - 自分で身に付けようとしても時間がない
 - 学生ができるようになるころには卒業してしまう

よくある失敗談

- ▶ シーケンスすればいいと思っていて、解析にお金がかかるとは思っていなかった
- ▶ 見たい結果が得られなかった
 - 研究計画段階でのミス
 - × とにかくやってみればなにか出るだろう
 - サンプルング方法の検討が甘かった
 - × メタゲノム解析で考えていた種が出なかった

失敗しないNGS解析のために

- ▶ NGSから得た大量の配列情報から解析を行い、研究に必要なデータを絞り込むことが重要

1. 研究の計画段階で以下のものを決定しておく。
 - (1) 注目すべきゲノム上領域をできるだけ選出しておく
 - (2) フィルタリングの為に用意するデータの選定
コントロール、比較条件、同一条件 etc...

これらが決定されていないと・・・

- 結果範囲が広すぎて、変異を見つけられない。
- サンプル数が少なく、ターゲットの変異を絞り込めない
- コントロールサンプルが無い場合、目的ピーク以外のものが検出される

2. データの正確性はサンプルの状態に左右される

(1) クオリティの高いデータを取る

比較を意識していないサンプリング、コンタミ etc...は、データが偏る原因となりやすい

(2) 解析の種類、リード長によって、カバレッジを考慮し、必要量のデータを得る

→ミスアセンブルの発生

→サンプル間のデータ量に大きな差があるため、比較解析時に偏った値が出る

→コンタミにより、カバレッジ(データ量)が大幅に減ってしまい、条件が変化してしまう。

NGS解析を失敗しないためには・・・

実験における最大の課題であり、

もっとも重要なのは**サンプリング**

見たい事象のデータを得られるよう、

慎重に**計画立案**をすることが必要

そういったことを相談・解決できるのが、
コンサルティングもついたサービス

NGSデータ解析サービスの種類と特徴

サービス形態	メリット	デメリット
シーケンシング ＋ データ解析	<ul style="list-style-type: none"> シーケンシングから解析までを一括して依頼できる セット価格のため費用も抑えられる。 	<ul style="list-style-type: none"> 納品形式の指定ができない フォローが受けられない。
データ解析	<ul style="list-style-type: none"> シーケンサーを選択できる。 メニュー化された解析なので、コストが抑えられる。 	<ul style="list-style-type: none"> シーケンサーは別途依頼が必要 納品形式を選択することができない。選択できた場合もコストが割高になってしまう。 フォローが受けられない場合が多い
データ解析 ＋ コンサルティング	<ul style="list-style-type: none"> シーケンサーを選択できる。 解析プランについて計画時に相談することができる。 フォローが受けられる。 解析途中で変更依頼ができる。 	<ul style="list-style-type: none"> シーケンサーは別途依頼が必要 コストが割高になってしまう。

低

自由度 & コスト

高

データ解析＋コンサルティングサービス 「PictBio」のご紹介

特徴1 研究者様と研究目的に合わせた準備

- 解析方針についての打合せ、ヒアリング
(Skypeなどのテレビ会議を推奨いたします。)
- 解析事例を交えて、解析提案

特徴2 オーダーメイド解析

- 解析アルゴリズムの選択
- 新規解析アルゴリズム、パイプラインの開発

特徴3 作業内容の柔軟性

- 解析途中での結果報告、および作業方針再検討の受付
- 論文などへの掲載実績のない解析への対応

▶ Axiohelix の PictBio とは

◦ 膨大なNGSデータを可視化(Picture)するNGSデータ解析パッケージ

解析作業の流れ(変異解析の場合)

シーケンシングした配列と既知の配列情報を比較し、Insertion, Deletion, SNPsを検出することで、ゲノム上のどの位置で、どのような変異が起こったか知ることができます。

1. リードのクオリティコントロール

クオリティチェックを行い、低クオリティのリードを除去する

2. マッピング作業

シーケンスリードをリファレンスゲノムにマッピングする

3. 重複リード除去

PCR duplicates由来の重複したリードを取り除き、変異検出の際のバイアスを軽減させる

4. マッピング結果の補正

SNVやInDelの正確な位置を求めるために、マッピング結果の補正を行う

5. 変異検出の実施
SNVやInDelを検出する

6. 変異アノテーションの付加

検出された変異に対し、遺伝子のどの領域の変異なのか、もしくは既知の変異であるかの情報を付加する

7. 変異情報の加工
要件に合わせてデータを整形する

8. 納品物作成作業（データ、報告書、グラフデータ等）
ご依頼に応じて納品物として解析結果から納品物を作成する

9. 納品とご報告
解析結果を納品物を元にご説明させていただき、次の解析の計画を検討します。

その他の解析支援

アノテーション 情報の付加

- 解析結果に対し公共データベースからのアノテーション付加や統計解析等を行ない、生物学的に意味のあるデータのみを抽出します。

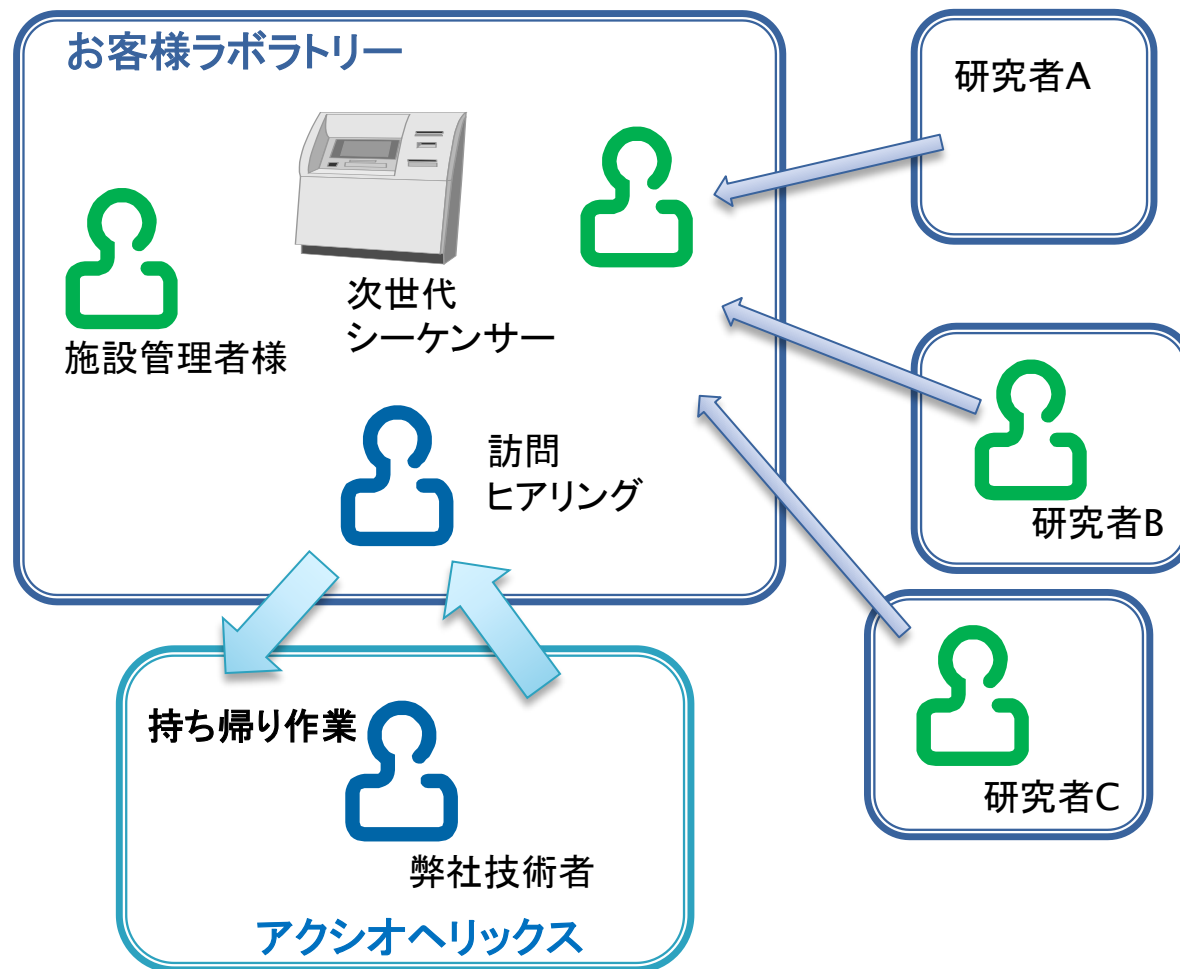
システム構築

- 大量データを管理するためのデータベースや、ラボ内のデータの統合アプリケーション等、ご要望に沿ったシステムを構築いたします。

人材教育

- バイオインフォマティクスの知識・技術を習得したい研究者や学生の方へ向け、プログラミングや解析に関するセミナーを行います。

一括複数件でのご依頼の作業



複数案件に対して効率的なお打ち合わせを実施することで、コストを抑えることができます。

納品物について

- ▶ **ご要望、ご予算に応じた納品物を作成いたします。**
例えば、Excel形式でソートを掛けた状態でのデータ作成や、指定の形式のグラフを使用した比較資料など論文で利用できる形式で作成いたします。
- ▶ **修正要望に対応いたします。**
事前に納品物の内容をご確認いただき、修正などのご要望があれば、それを反映した最終納品物を作成いたします。
- ▶ **納品完了後のフォローアップに対応いたします。**
納品完了後であっても、納品物についての説明などフォローアップを行っております。

納品物事例1：発現解析

目的: ある発生段階における遺伝子発現の経時的変化の比較

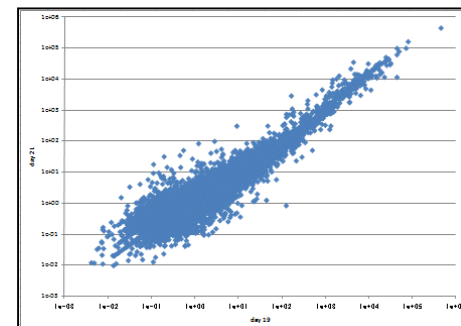
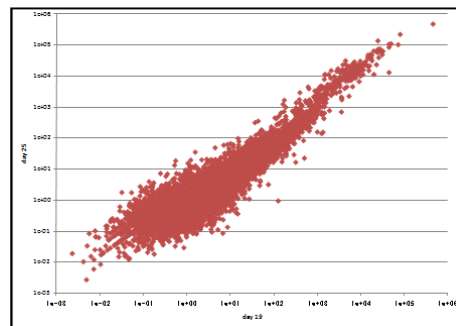
ID	chr	tag d1	tag d2	tag d2!	gene length	norm tag d1	norm tag d2	norm tag d2!	n19 - n2	n19 - n2	n21 - n2	19 / 2	19 / 2!	21 / 2!	21 / 1!	
ENSG000000000001	29	224	144	46	#	1881	12.11250109	11.42530956	4.109532015	0.68719153	8.00296908	7.31577754	1.0601464	2.9474162	2.7801972	0.943265
ENSG000000000002	6	213	138	69	#	864	25.07497261	23.83744056	13.42019049	1.23753205	11.6547821	10.4172501	1.0519155	1.8684513	1.7762371	0.950646
ENSG000000000004	27	131	78	47	#	868	15.35062905	13.41124686	9.099163415	1.93938218	6.25146563	4.31208345	1.1446086	1.6870374	1.4738989	0.873661
ENSG000000000005	9	19	30	4	#	1731	1.116429078	2.586535635	0.388316867	-1.4701066	0.72811221	2.19821877	0.431631	2.8750466	6.6608892	2.316793
ENSG000000000006	29	1	2	0	#	2027	0.050178868	0.147255161	0	-0.0970763	0.05017887	0.14725516	0.3407614	#DIV/0!	#DIV/0!	2.934605
ENSG000000000007	9	132	41	23	#	632	21.24376357	9.681910366	6.115529842	11.5618532	15.1282337	3.56638052	2.1941707	3.4737405	1.5831679	0.455753
ENSG000000000011	27	150	134	142	#	3141	4.857333571	6.366945629	7.597028226	-1.5096121	-2.7396947	-1.2300826	0.7628985	0.6393728	0.8380837	1.310790
ENSG000000000012	X	379	185	119	#	949	40.62071878	29.0937562	21.07191865	11.5269626	19.5488001	8.02183755	1.3962006	1.9277181	1.3806885	0.716229
ENSG000000000015	29	498	375	170	#	17712	2.859804503	3.159787985	1.612889631	-0.2999835	1.24691487	1.54689835	0.9050621	1.7730937	1.9590851	1.104896
ENSG000000000016	1	148	114	46	#	1788	8.419160859	9.515500055	4.323282841	-1.0963392	4.09587802	5.19221721	0.8847839	1.9474	2.2009895	1.130219
ENSG000000000018	6	386	369	88	#	3054	12.85561561	18.03232029	4.842135869	-5.1767047	8.01347974	13.1901844	0.7129208	2.6549473	3.7240426	1.402680
ENSG000000000020	27	60	35	11	#	936	6.520036216	5.580671703	1.97487753	0.93936451	4.54515969	3.60579417	1.1683246	3.3014889	2.8258318	0.855926
ENSG000000000024	27	1	0	0	#	1618	0.062863143	0	0	0.06286314	0.06286314	0	#DIV/0!	#DIV/0!	#DIV/0!	
ENSG000000000029	1	267	164	58	#	981	27.68323634	24.94991784	9.935330492	2.73331849	17.7479058	15.0145874	1.1095522	2.7863428	2.5112318	0.901264
ENSG000000000030	1	84	51	17	#	1044	8.183769595	7.290611506	2.736350684	0.89315809	5.44471891	4.55426082	1.122508	2.9907605	2.6643557	0.890862
ENSG000000000031	1	133	69	30	#	2141	6.318435844	4.809796508	2.354658445	1.50863934	3.9637774	2.45513806	1.3136597	2.6833768	2.0426727	0.761232

カウント

比較

Excelの機能を使用して、選択されたデータに対して比較対象データが自動的にソートされる形式で資料を作成いたしました。

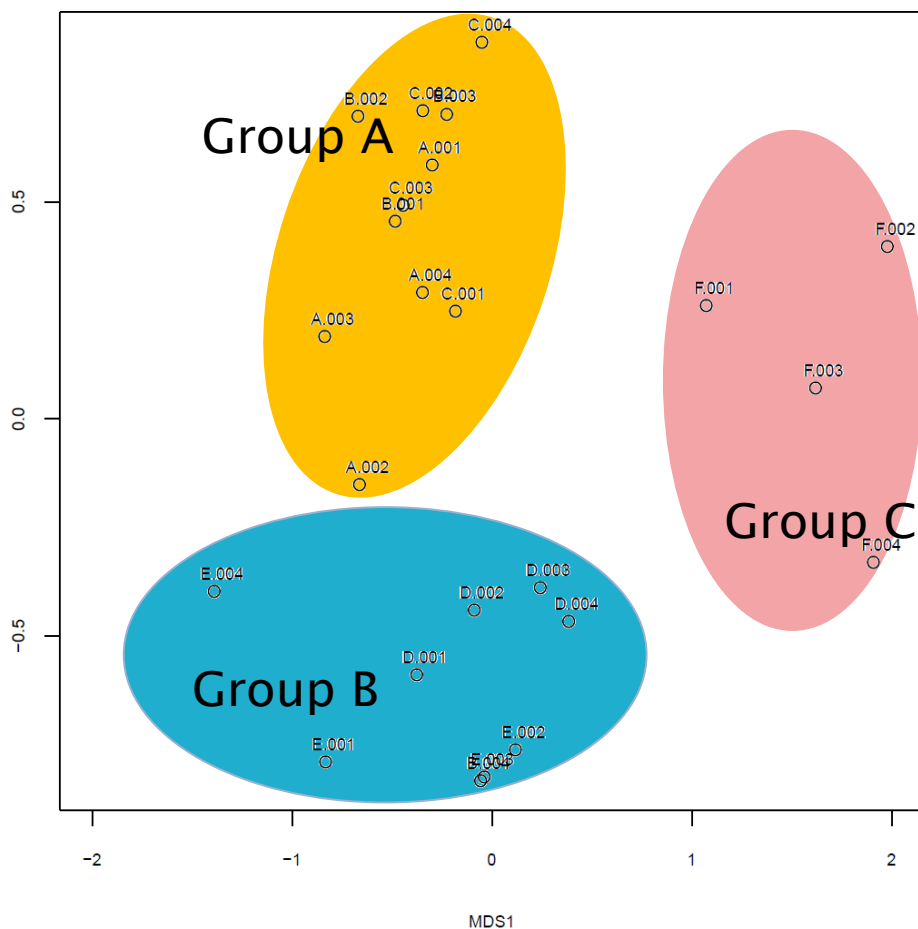
グラフ化



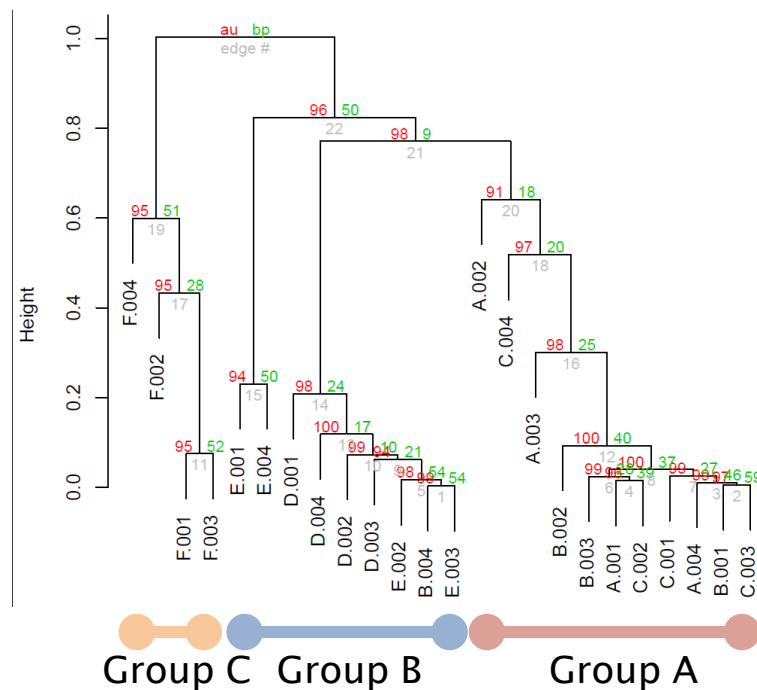
納品物事例2: 集団解析

メタゲノム生物種クラスタリングのグラフ作成例

距離プロット図



クラスタリングツリー



Publications



Contents lists available at [ScienceDirect](#)

Gene

journal homepage: www.elsevier.com/locate/gene

ELSEVIER

Research paper

Influences of diurnal sampling bias on fixed-point monitoring of plankton biodiversity determined using a massively parallel sequencing-based technique

Satoshi Nagai ^{a,*}, Kohsuke Hida ^b, Shingo Urushizaki ^b, Goh Onitsuka ^c, Motoshige Yasuie ^a, Yoji Nakamura ^a, Atushi Fujiwara ^a, Seisuke Tajimi ^d, Katsunori Kimoto ^e, Takanori Kobayashi ^a, Takashi Gojobori ^f, Mitsuru Ototake ^a

^a Research Center for Aquatic Genomics, National Research Institute of Fisheries Science, 2-12-4 Fukuura, Kanazawa-ku, Yokohama, Kanagawa 236-8648, Japan

^b AXIOHELIX Co. Ltd., 5-11 Hakozaki, Nihonbashi, Chuouku 103-0015, Tokyo, Japan

^c National Research Institute of Fisheries and Environment of Inland Sea, 2-17-5 Maruishi, Ohno, Hatsukaichi, Hiroshima 739-0452, Japan

^d Kumamoto Prefectural Fisheries Research Center, 2450-2 Naka, Oyamano, Kamiamakusa, Kumamoto 869-3603, Japan

^e Seikai National Fisheries Research Institute, Fisheries Research Agency, 1551-8 Taira-machi, Nagasaki, Nagasaki 851-2213, Japan

^f Center for Information Biology, National Institute of Genetics, 1111 Yata, Mishima, Shizuoka 411-8540, Japan



Contents lists available at [ScienceDirect](#)

Gene

journal homepage: www.elsevier.com/locate/gene

ELSEVIER

MOLECULAR ECOLOGY RESOURCES

Molecular Ecology Resources (2015)

doi: 10.1111/1755-0998.12459

Comparative study of the validity of three regions of the 18S-rRNA gene for massively parallel sequencing-based monitoring of the planktonic eukaryote community

AKIFUMI S. TANABE,^{*} SATOSHI NAGAI,^{*} KOHSUKE HIDA,[†] MOTOSHIGE YASUIKE,^{*} ATUSHI FUJIWARA,^{*} YOJI NAKAMURA,^{*} YOSHIHITO TAKANO^{*} and SEIJI KATAKURA[‡]

^{*}Research Center for Aquatic Genomics, National Research Institute of Fisheries Science, 2-12-4 Fukuura, Kanazawa-ku, Yokohama, Kanagawa 236-8648, Japan, [†]AXIOHELIX Co. Ltd., 5-11 Hakozaki, Nihonbashi, Chuouku, Tokyo 103-0015, Japan,

[‡]City of Mombetsu, Kaiyo-koryukan, Kaiyo-koen, Mombetsu, Hokkaido 094-0031, Japan



Massively parallel sequencing-based survey of eukaryotic community structures in Hiroshima Bay and Ishigaki Island

Satoshi Nagai ^{a,*}, Kohsuke Hida ^b, Shingo Urusizaki ^b, Yoshihito Takano ^a, Yuki Hongo ^a, Takahiko Kameda ^c, Kazuo Abe ^{c,d}

^a Research Center for Aquatic Genomics, National Research Institute of Fisheries Science, 2-12-4 Fukuura, Kanazawa-ku, Yokohama, Kanagawa 236-8648, Japan

^b AXIOHELIX Co. Ltd., 5-11 Hakozaki, Nihonbashi, Chuouku 103-0015, Tokyo, Japan

^c Research Center for Subtropical Fisheries, Seikai National Fisheries Research Institute, 148-446 Fukai-Ohta, Ishigaki, Okinawa 907-0451, Japan

^d National Research Institute of Fisheries and Environmental of Inland Sea, 2-17-5 Maruishi, Ohno, Hatsukaichi, Hiroshima 739-0452, Japan



弊社ホームページの紹介

www.pictbio.com



Twitter account : @PictBio

PictBio

Genome analyses for all researchers

- PictBioについて
- 解析サービス
- パイプライン
- ツール開発
- 実績
- Publications
- 解析のメモ帳**
- お問い合わせ
- 🔍

PictBioとは

バイオインフォマティクスにおけるお悩みを
解決するお手伝いをさせて頂くサービスです

PictBio = Picture + Biology
膨大なバイオデータを可視化いたします

メインラインナップ



次世代シーケンサー関連の解析を中心に
バイオインフォマティクス受託



次世代シーケンサーデータを
自身で検証、解析が可能になる環境を提供



新規のアルゴリズム、ツールの開発や
公開データベースの作成

膨大化したバイオデータにはインフォマティクス技術が必須です。PictBioサービスでは「次世代シーケンサー (NGS) データ解析」を筆頭に、バイオインフォマティクスのノウハウを用いたソリューションを提供しております。
様々なニーズにお応えできるように柔軟性の高いサービスの提供を行っております。

ご清聴いただき有難うございました。