

厚生労働科学研究費補助金（食品の安全確保推進研究事業）
「食品由来細菌の薬剤耐性サーベイランスの強化と国際対応に関する研究」
平成 26 年度分担研究報告書

分担課題名：薬剤耐性化食中毒菌のヒトおよび家畜由来株のゲノム比較解析

分担研究者	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	第三室
研究協力者	関塚剛史	国立感染症研究所	病原体ゲノム解析研究センター	第三室
研究協力者	竹内史比古	国立感染症研究所	病原体ゲノム解析研究センター	第三室
研究協力者	山下明史	国立感染症研究所	病原体ゲノム解析研究センター	第三室

研究要旨

薬剤耐性食中毒菌の多くが多様なプラスミド伝達により薬剤耐性を獲得していることが明らかとなっている。食品と患者分離株との因果関係について、分離株のゲノム情報を活用した研究が行われるようになってきたが、耐性伝播の根本はプラスミド伝達が主体であり、異なる宿主菌でも同様の薬剤耐性プラスミドを有していることが報告されている。つまり、受け渡しを行う宿主菌のゲノム情報よりも、プラスミド単位での分子疫学のほうが頻繁な耐性授受の過程を追跡するのに好都合と考えられる。本分担研究では、家畜・食肉・ヒト臨床から分離された薬剤耐性菌プラスミドの包括的なデータベースを構築し、菌種間を伝播する薬剤耐性因子の追跡を可能にする解析パイプラインの構築を目的とする。本年度は、昨年度に開発したプラスミド総合解析ツール(GPAT)を改良した。具体的には、quality trimming や配列アセンブリでより多くの解析アルゴリズムを使えるようにし、更に、使い勝手の向上にも取り組んだ。これらの改良により GPAT による解析件数が飛躍的に増加したため、過去の解析結果に容易にアクセスできるデータ管理機能を追加する事が必要となった。データ管理機能はある程度までは実装できたもののさらなる改良が必要である。現在 GPAT は、計算能力の都合上、感染症研究所イントラネット内でのみ公開している。今後は増え続ける解析データをより容易に管理できる仕組みを実装し、Plasmid 配列全体の遺伝子水平伝達を過去・現在・地域に渡り俯瞰的に眺めることができるシステムへと発展させる必要がある。

A．研究目的

ヒト、家畜、食品から分離される薬剤耐性食中毒菌のサーベイランス結果を基盤にし、ゲノム配列レベルおよびプラスミド・レベルでより具体的な耐性化機序と株伝播プロセスを解析する。この解析の基盤を作るために、現在登録されているプラスミド配列のうち、薬剤耐性因子の有無に関係なく全てのプラスミドを抽出し、各配列の特徴（Inc タイプ、薬剤耐性因子、Insertion sequence、Transposon 等）をリスト化する。プラスミド保有菌種の情報（菌種、分離年・国・地域・宿主、各種タイピング結果等）を網羅しデータベース化することで、より具体的な伝播過程が見えてくるものと期待している。

B．研究方法

菌体からプラスミドに相当する DNA を抽出することが先決である。そのために、Pulsed-Field Gel Electrophoresis (PFGE) 用の菌体プラグからプラスミド DNA のみ泳動分離して回収・精製した。PFGE プラグの調整法は菌種に沿ったプロトコルを利用する。プラスミド・サイズに従った泳動距離を得るために、環状 DNA であるプラスミドを S1-nuclease でニックをいれ線状化の前処理を行った。S1-nuclease 処理菌体プラグによる PFGE を S1-PFGE と呼ぶ（図 1）。プラスミドに相当するアガロースバンドを切り出し、DNA を精製した。

精製 DNA を用いて Illumina NEXTERA XT kit にてライブラリー作成した。MiSeq シークエンサーにて解読後、プラスミド解析にかかる一連の情報解析を本年度開発を行った GPAT, iPAT を用いて行った（図 1）。詳細は C. 結果の項目に記述した。

C．結果

プラスミド配列の解読手法の構築

図 1 にプラスミド DNA の回収から情報解析までの一連の流れを示した。MiSeq シークエンサーにて解読後、新規に開発した GPAT を用いて解読リードの quality trimming, (genome subtract), de novo assembly, ORF 推定、推定した ORF の機能推定、および Inc タイプ推定を行う（図 2）。Quality trimming

ではユーザーの選択により skewer または fastq-mcf による処理の後 in house perl script で更に詳細なトリミングを行う。Genome subtract は bwa, samtools, Hydra および in house perl スクリプトを組み合わせた独自パイプラインを用いる。Assemble には A5_miseq, A5 assembler, Spades, の他、p_latanus と p_rice を組み合わせた独自パイプライン (PIPr) を選択することが出来る。ORF 推定には prodigal, gmhmm_heuristic, および gmhmm が選択可能である。ORF の機能推定および Inc タイプ推定には BLAST を用いる。データベースとして、既知の plasmid データベース、COG, ARDB + CARD、および NCBI NR を選択することができる。Inc タイプ推定には PBRT の replicon typing プライマーを用いた *in silico* PCR によって得られた配列データベースを用いる。

結果の概要 (quality trimming 後のリード数、アセンブル後のコンティグ長、コンティグ数、GC 含量、予測される遺伝子数、薬剤耐性遺伝子、Inc type など) は解析トップページの overview 欄で確認することが出来る。

Plasmid データベースは平成 24 年度の plast システム構築時に作成した plasmid データ抽出法を発展させ、NCBI から最新のデータをダウンロードできるシステムを構築した。ただし、解析する日によって結果が変わってしまう事態を防ぐため、GPAT では 2013 年 6 月 10 日時点のデータをもちいている。

GPAT は MiSeq のシークエンス結果だけでなく、fasta format や GenBank format のデータも入力として受け付けることができる。GPAT 内部で使用しているソフトウェア、およびデータベースは全て無償で利用できるものを使用し、利用者および開発者の負担にならないよう配慮した。GPAT はバイオインフォマティクス解析に慣れていない研究者でも簡単に高度な解析ができるよ、操作性に配慮しながら開発を行い、直感的に操作できるユーザーインターフェースを構築した（図 3）。更に、実際に plasmid の解析を行っている研究者に試用してもらい、開発者側と利用者のあいだに認識のズレが生じないようにした。

これらの努力により、GPAT による解析の利便性は劇的に向上し、少数の試験的ユー

ザーのみにより今年度中に通算 1,700 以上の解析が GPAT を使って行われ、一人のユーザーが数百の解析結果を保つケースも現れた。このような多くのデータの扱いを容易にするために、本年度は GPAT に結果リスト表示機能等、データ取り扱い支援機能を追加したものの、利便性向上のためには更なる改良が必要である。

プラスミド間の関係性解析手法の構築

GPAT の開発により、大量の plasmid 配列を容易に解析することが可能になったが、薬剤耐性遺伝子や plasmid そのものの伝播課程を明らかにするためには plasmid 同士の関係性を大規模に解析する必要がある。そのため、plasmid が共通して持つ遺伝子のネットワークを解析するためのソフトウェア iPAT (inter Plasmid Analyzing Tool) を平成 25 年度に開発した。iPAT は GPAT が解析した plasmid の ORF 同士の相同性検索を行い、相同性のあった plasmid 同士をエッジで結合してゆくネットワーク解析を行う。

図には示していないが、臨床で採取され、GPAT で解析した plasmid を iPAT を用いて解析すると、同じ起源を持つと思われる plasmid はネットワーク上でも密接な関係性を示し、起源が違ふと思われる plasmid はネットワーク上も疎な関係性を示した。また、NDM-1 を持つ既知の plasmid のネットワーク解析を行ってみた結果を図 4 に示す。GPAT の持つ GenBank format 読み込み機能を用いて NCBI に登録されている NDM-1 を持つ plasmid 遺伝子を読み込み、iPAT を用いてネットワーク解析を行ったところ、NDM-1 保有 plasmid は 4 つのグループに分類できることが示唆された。iPAT は GPAT が解析したデータだけを扱うことができるため、既知の plasmid との関係性を解析するためには配列を手動で GPAT に入力してゆかなければならない。新たに解読した plasmid と既知の plasmid との関係性の解析を容易にするために、「既知のデータの入力」の過程を自動化できるようにする予定である。この手法は、既知の plasmid 全体の解析にも応用できる可能性がある。今後はこの手法を応用して plasmid による薬剤耐性伝播の網羅的な解析を行う予定である。

D . E . 考察・結論

薬剤耐性食中毒菌の多くは薬剤耐性プラスミドによる耐性獲得であり、プラスミド単位で菌種・株間の伝播を追跡できるので、より正確な耐性伝播の様式を明確にできるものと考えている。そのためには、プラスミド配列を利用した詳細な系統分類法の構築が必要である。配列解読から情報解析までの必要な手法がパイプライン化されておらず、配列解読後の解析に時間と負担を要していた。本分担研究において、“解読リードからシームレスにプラスミド解析”が可能な解析パイプラインを構築することを目標にし、誰もが使えて汎用性のある環境整備をはじめた。

本年度は、昨年度構築し、所内イントラネット限定公開した GPAT を改良し、通算 1,700 以上の解析を行った。GPAT は MiSeq による plasmid のシーケンシングからほとんどクリックだけでアセンブリや遺伝子予測・アノテーションまでを通常の web ブラウザのみで行うことを可能にし、1,700 以上の解析実績を積んだ。また、iPAT は GPAT で解析した plasmid 同士の関係性を図示することを可能にした。今後は GPAT/iPAT の利便性を更に充実させるとともに、これらの技術を用いて菌株・株間の plasmid そのものや plasmid 上の遺伝子の水平伝達を俯瞰的に解析できるシステムを構築してゆく予定である。

F . 健康危害情報

なし

G . 研究発表

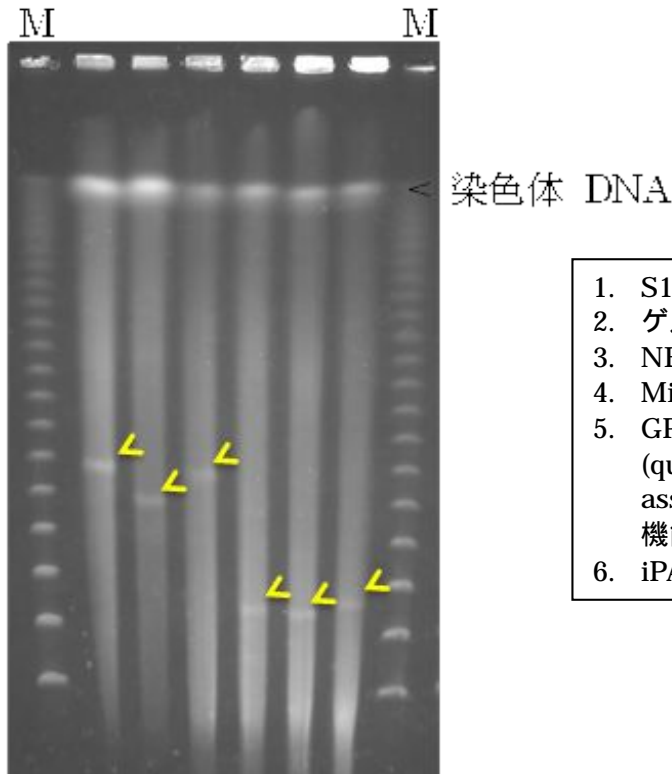
(論文発表)

なし

(学会発表)

第 54 回 Interscience Conference on Antimicrobial Agents and Chemotherapy (ICAAC) (ポスター発表 Akifumi Yamashita, Tsuyoshi Sekizuka, Makoto Kuroda) Plasmidome Community Network Analysis For Antimicrobial Resistance.

2014 年日本細菌学会総会 (ポスター発表 2P-089 山下明史、関塚剛史、黒田誠) Plasmidome network analysis.



1. S1-PFGE
2. ゲル回収
3. NEXTERA XT kit
4. MiSeq 解読
5. GPAT による解析
(quality trimming, de novo assembly, ORF 抽出、遺伝子機能解析など)
6. iPAT による解析

図1 プラスミド DNA の PFGE 分離から配列解読そして情報解析までの流れ。S1-PFGE 泳動図に泳動分離されたプラスミド断片（黄色矢頭）を示す。

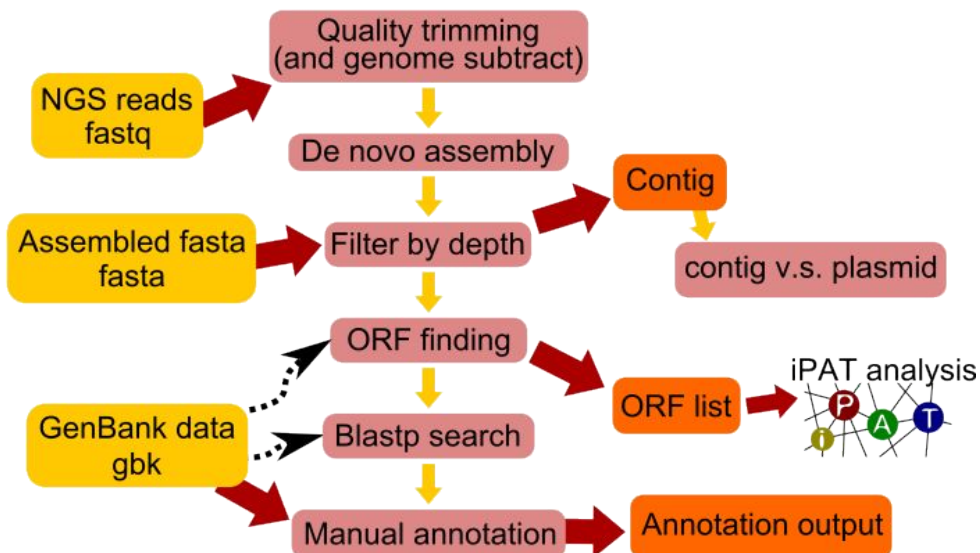


図2 GPAT (Global Plasmidome Analyzing Tool) による plasmid 解析のワークフロー。Quality trimming には skewer または fastq-mcf の後 in house perl スクリプトによる詳細なトリミングを行う。Genome subtract には bwa, samtools, Hydra および in house perl スクリプトを用いる。Assemble には A5_miseq / A5 / spades / PIPr が選択可能である。ORF 推定には prodigal, gmhmm_heuristic および gmhmm が選択可能である。

GPAT: V001-b

http://naruto/cgi-bin/gpat/showStatus.cgi?ID=...

project name: ...

Collection: ...
Organism: Escherichia coli, Host: ...
plasmid size: 110kb, Drug resistance

Current group: ... [Show list](#) Go to another group

ID: ... (2015-01/19-12/23) ...

Overview

Total reads: 117,311
Quality trimming: 91,001 (77.6%)

Assembly: 9 contigs
Length: 1,049-41,350 bp
GC: 50.9 %
Total base: 111,136 bp
N50: 26,645 bp

Number of genes: 137
Detected Inc types: FII, FICF, FIBF

Detected AR genes

id	cov	Detected AR genes	Reference
100	100	AAC(3)-II	41056930
100	100	aminoglycoside 3'-phosphotransferase	17158058
100	100	extended-spectrum beta-lactamase	DQ885477
99	100	tetracycline efflux protein TetA tetracycline resistance protein, class A protein	31795168

Processing Status

Process	Status	Exec
Trimming	Finished	<input type="checkbox"/>
Assemble	Finished	<input type="checkbox"/>
Filter by depth	Finished	<input type="checkbox"/>
Whole Plasmid similarity	Not Analyzed	<input type="checkbox"/>
BLASTN search (inc type)	Finished	<input type="checkbox"/>
ORF finding	Finished	<input type="checkbox"/>
Blastp Search (plasmid)	Not Analyzed	<input type="checkbox"/>
Blastp Search (cog)	Not Analyzed	<input type="checkbox"/>
Blastp Search (ardb)	Finished	<input type="checkbox"/>
Blastp Search (nr)	Not Analyzed	<input type="checkbox"/>

[Download](#)

[Run Process](#)

Prefixed Parameter Settings -- Recommended settings --

--- Advanced settings ---

Trimming params

5' trim length: 5
Trimming method: skewer
Minimum average quality threshold: 15
Trim lower than this q-value: 0 > 15
Minimum remaining sequence length: 80
Maximum length: 0
Remove bacterial genomes: OFF

Assembler settings
Select an assembler program: A5 miseq
Cutoff depth: 0
Cutoff depth=0 means automatic cutoff

ORF finding settings
ORF finding mode: single
Target sequence type: prokidal

BLAST search settings
Gapped search: ON
Filter: OFF
E-value: 1e-10

[go to Annotation Editor](#)
[go to inter Plasmid Analyzing Tool \(IPAT\)](#)
[See your project list](#)

About COG

The cog database was downloaded from NCBI cog database (Mar. 2, 2003 version). We subsequently added class B metallo-beta-lactamase genes obtained from [The Metallo-Beta-Lactamase Engineering Database](#), because the cog does not contain the class B metallo-beta-lactamase genes.

About ARDB (ARDB + CARD)

The ARDB used in this service is made of [CARD database](#) plus [ARDB database](#). The original ARDB data is a list of accession number of antibiotic resistance genes, and the sequences were extracted from NCBI NR database by us on Jun. 4th, 2013. Note that we do not hold some of the genes in the original ARDB because of the update of NCBI database that make some accession numbers being expired. This database was updated on Sun Jan 18 23:55:10 JST 2015 according to the update of CARD database.

Collection: Year: Month: Date: Organism: Escherichia coli Host: ...

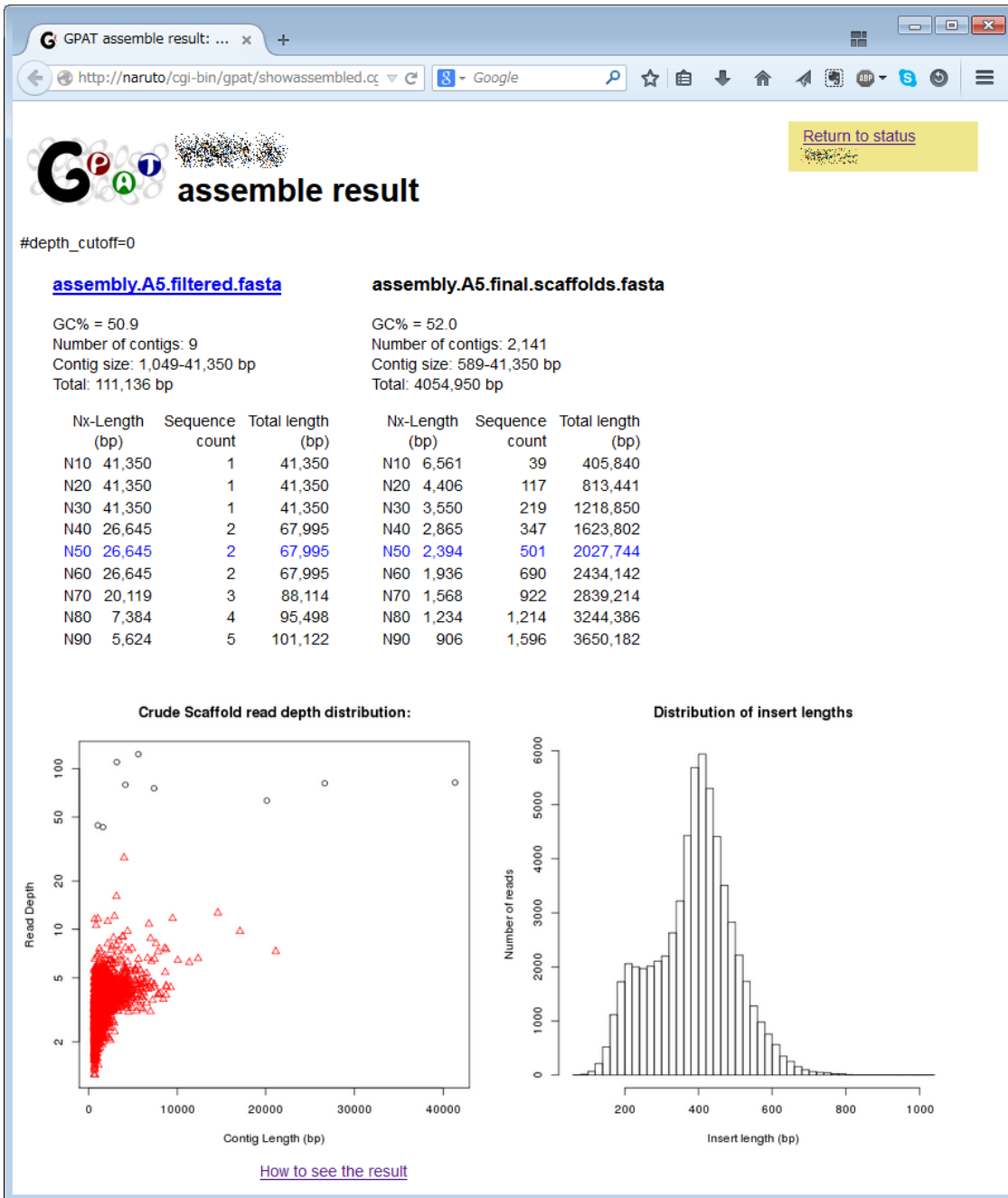
Place: Project Name: ...

Note: ... plasmid size: 110kb, Drug resistance [Change Project Name](#)

Your files: ..._L001_R1_001.fastq.gz
..._L001_R2_001.fastq.gz [Download](#)

[Show/Hide analysis log](#)

[Reset this analysis](#) [Dispose this data](#)



GPAT ORF finding: Sato... x

http://naruto/cgi-bin/gpat/showgeneFinding.cgi?ID=...

ORF finding result

Download predicted genes in [Nucleotide](#) / [Amino acid](#) sequence. Or as a [gene list](#).

Picture Width: 1200 px. Bases for one line: 20000 bp [change](#)

scaffold1.1 length=37262 read depth: 74.07 [to whole plasmid similarity](#)

scaffold2.1 length=9124 read depth: 50.88 [to whole plasmid similarity](#)

scaffold3.1 length=8492 read depth: 79.50 **gene=68, 2269-3048 length=780bp, aminoglycoside adenyltransferase**

http://naruto/cgi-bin/gpat/showblasted.cgi?ID=20141020191240190&gene=68

GPAT result: ... x

naruto/cgi-bin/u/min/showblasted.cgi?ID=...

BLAST result assembly v.s. ardb

Download BLAST result in [raw text](#) / [m8 table](#) / [modified m8 table](#)

gene 70|prodigal|516 aa|+|2952|4499 length= 515 aa

CARD|putative recombinase. Encoded by gene sul1.p01. Orf513. ARO:1000001 process or component of antibiotic ...
59% AF174129.3.gene8.p01 *Escherichia coli*

gene 71|prodigal|274 aa|+|4593|5414 length= 273 aa

CARD|sul1** protein. Encoded by gene sul1**. ARO:1000001 process or component of antibiotic biology or ...
89% AY458224.gene.p01 *Salmonella enterica subsp. enterica serovar Newport*

ARDB|dihydropteroate synthase >g|18426916|gb|AAK02048.2|AF261825_17 sul1 delta fusion protein fusion ...
89% g|487652939|ref|WP_001747810.1| *Proteus mirabilis*

ARDB|dihydropteroate synthase >g|18426916|gb|AAK02048.2|AF261825_17 sul1 delta fusion protein fusion ...
89% g|487652939|ref|WP_001747810.1| *Proteus mirabilis*

ARDB|dihydropteroate synthase >g|18426916|gb|AAK02048.2|AF261825_17 sul1 delta fusion protein

Return to status
Change database:
-- database --
-- database --
plasmid
nr
cog
ardb

GPAT result: [...](#)

naruto/cgi-bin/uhmin/showwholeblasted.cgi?ID=...

BLAST result: assembly v.s. plasmid

Download [raw blast result](#) or [tabular format](#)
[go to plasmid view](#)

[Return to status](#)
[Inverse mode](#)
[Show All](#) [Hide All](#)

scaffold1.1|size48513 length= 48513 bp. Average depth: 367.44 [to ORF](#) [Show this](#) [Hide this](#)

96.0% * [JN883043](#)|pSH111_166|165791|circular| *Salmonella enterica* subsp. e
 92.4% * [AB277724](#)|pP91278|131520|circular| *Photobacterium damselae* subsp.
 92.4% * [AB277723](#)|pP99-018|150157|circular| *Photobacterium damselae* subsp.
 96.0% * [FJ621586](#)|peH4H|148105|circular| *Escherichia coli* strain H4H plasmid
 96.0% * [FJ621587](#)|pAMD4528|158213|circular| *Salmonella enterica* strain AM
 96.0% * [CP000604](#)|pSN254|176473|circular| *Salmonella enterica* subsp. enter
 82.3% * [JN687470](#)|pMR0211|178277|circular| *Providencia stuartii* plasmid pMV
 96.0% * [HQ023862](#)|pUMNK88|160573|circular| *Escherichia coli* UMNK88 plasm
 92.7% * [CP000609](#)|pIP1202|182913|circular| *Yersinia pestis* biovar Orientalis
 92.6% * [JQ824049](#)|pTC2|180184|circular| *Providencia stuartii* plasmid pTC2,
 92.7% * [CP008225](#)|pKPH53|105874|circular| *Klebsiella pneumoniae* subsp. pn
 96.0% * [HQ023863](#)|pAPEC1990_61|161081|circular| *Escherichia coli* strain AF
 93.4% * [JQ010884](#)|pR55|170810|circular| *Klebsiella pneumoniae* plasmid pR55
 88.3% * [JN157804](#)|pNDM-KN|162746|circular| *Klebsiella pneumoniae* strain K
 95.5% * [JX141473](#)|pR148|165906|circular| *Aeromonas hydrophila* plasmid pR1

GPAT result tabular for [...](#)

naruto/cgi-bin/uhmin/showwholeblasted.cgi?ID=...

Project: [...](#)
 Query: scaffold1.1|size48513, length = 48513
 Subject: [JN687470](#)|pMR0211|178277|circular, length = 178277
 Annotation: *Providencia stuartii* plasmid pMR0211, complete sequence.

[return to whole plasmid search](#)

Identity	Hit Length	Mismatch	Gap	Query From	To	Subject From	To	E-value	Score
100	88	0	0	1	88	127317	127230	0.0	174.0
96	476	19	0	91	566	108535	108060	0.0	793.0
99	4891	2	0	7219	12109	158119	163009	0.0	9680.0
99	2811	2	0	12104	14914	162984	165794	0.0	5557.0
100	7563	0	0	16869	24431	165790	173352	0.0	14990.0
100	20	0	0	18116	18135	79462	79443	0.0	40.1
99	4931	1	0	24427	29357	173347	178277	0.0	8767.0
100	21	0	0	24843	24863	53920	53900	0.0	42.1
99	19156	2	0	29358	48513		19156	0.0	37960.0
100	19	0	0	35035	35053	36499	36481	0.0	38.2
100	19	0	0	45479	45497	33086	33068	0.0	38.2
100	64	0	0	48450	48513	21155	21082	0.0	127.0

>[JN687470](#)|pMR0211|178277|circular| *Providencia stuartii* plasmid pMR0211, complete sequence.
 Length = 178277

naruto/cgi-bin/uhmin/showwholeblasted.cgi?ID=...

GPAT result

http://naruto/cgi-bin/gpat/showincblast.cgi?ID=...


BLAST result: assembly v.s. incdb

Return to status
Change database: --- database ---

Download BLAST result in [raw text](#) / [m8 table](#) / [modified m8 table](#)

Detected Inc types: **N (scaffold2.1)**

[scaffold2.1 length= 9124 nt](#)



98% N_group_IncN

BLASTN 2.2.18 [Mar-02-2008]

Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Res.* 25:3389-3402.

Query= [scaffold2.1](#) / [go to top](#)
(37,282 letters)

Database: PBRT-inc-seq_ver2.fa
25 sequences; 11,253 total letters

Searching.....done

***** No hits found *****

BLASTN 2.2.18 [Mar-02-2008]

Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Res.* 25:3389-3402.

Query= [scaffold2.1](#) / [go to top](#)
(9124 letters)

Inc type	<i>in silico</i> PCR typing in GPAT	PBRT kit PCR-based replicon typing Carattoli A, Bertini A, Villa L, et al. Identification of plasmids by PCR-based replicon typing. <i>J Microbiol Methods</i> 2005;63:219-28.
A/C	O	O
B/O	O	O
FIA	O	O
FIB	O	O
FIB-M	O	O
FIC	O	O
FII	O	O
FIK	O	O
FIS	O	O
H1	O	O
H2	O	O
HIB-M	O	O
I1	O	O
I2	X	O
K	O	O
L/M	O	O
N	O	O
P	O	O
R	O	O
T	O	O
U	O	O
W	O	O
X1	O	O
X2	O	O
Y	O	O

GPAT

http://naruto/cgi-bin/gpat/showStatus.cgi?group=...

List of: ...

Group	ID	Species Name	Collector name	Organization	Host	Place	Year	Quality	Genome	Number	Contigs	GC%	Contig length	Total size	NOI size	Number	Comment	Inc type	Ref. genes
1	K.pneumoniae	human	N/A	...
2	K.pneumoniae	human	-	...
3	E.coli	human	N	...
4	E.coli	human	PBRT_PAT	...
5	E.coli	human	-	...
6	E.coli	human	-	...
7	E.coli	human	-	...
8	E.coli	human	-	...
9	E.coli	human	-	...
10	E.coli	human	-	...

図3 GPAT 実行結果画面。 パラメータ設定、進捗状況確認など。 de novo assembly 結果。 ORF 検索結果。 ORF の blast 検索結果。 、 既知の plasmid に対する同源性検索結果。 Inc タイプの推定結果。 解析結果のリスト表示。