

エキスパートがキュレートしたtRNA遺伝子データベース

阿部貴志¹、池村淑道¹、小原康雄¹、上原啓史¹、中泉友紀¹、平野晋也¹、木ノ内誠²、金谷重彦³、山田優子¹、武藤昱⁴、井口八郎¹

(¹ 長浜バイオ大学、² 山形大学、³ 奈良先端科学技術大学院大学、⁴ 弘前大学) <http://trna.nagahama-i-bio.ac.jp>

要旨 微生物ゲノム解析の進行に伴い、国際塩基配列データベースから1,000種を超えるゲノム配列が公開されている。さらに、多様な環境中で生息する難培養性微生物類を対象としたメタゲノム解析が開発され、これまで未活用であった環境由来ゲノム資源の網羅的な理解が可能となりつつある。

我々は、タンパク質合成で重要な役割を持つtRNA遺伝子に着目し、現在公開されているバクテリアの543の完全長ゲノム、394のドラフトゲノム、ならびに約650万の環境由来DNA断片配列を対象に、tRNA遺伝子の網羅的な探索を行い、データベース化した(Abe et al., Nucl. Acids Res. 2009 Database Issue, in press.)。tRNA遺伝子の検索をより完全なものにするため**3種のtRNA遺伝子領域予測プログラム(tRNA Scan-SE, Aragorn, tRNAfinder)**を併用して探索を行った。各プログラムにより予測領域が一致しない場合が全予測tRNA遺伝子の約5%あったが、これらについては、**tRNA研究に携わってきたシニア世代の専門家がマニュアルにて精査**を行った。このシニア世代研究者の専門知識の活用は、ライフサイエンス分野における統合データベース整備事業の人材養成プログラムの一環として行われたものである。

予測プログラムとマニュアル・チェックによって、完全長ゲノムからは30,544遺伝子、ドラフトゲノムからは20,975遺伝子、環境由来DNA配列からは92,542遺伝子の合計142,061のtRNA遺伝子を探索・格納した。本データベースでは、各tRNA遺伝子の塩基配列やクローバ葉モデル、各Phylum/Species等のコドンとアンチコドン表が閲覧でき、また各tRNA間のホモロジーやtRNA遺伝子のゲノム上での位置や近接するtRNA遺伝子を容易に検索できる。精査された膨大な数のtRNA遺伝子からなるこのデータベースは、tRNAの分子生物学的研究や微生物の進化を論ずる上に極めて有用なものとなるであろう。

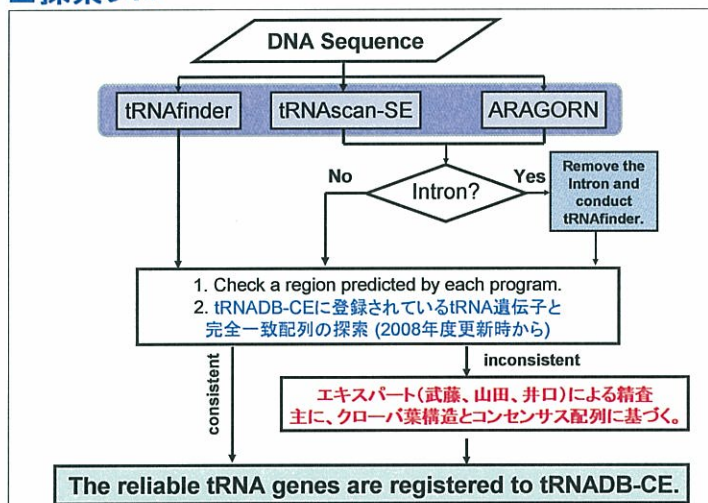
探索方法

■データセット

- 1) **完全長ゲノム (tRNAScan-SEの解析はある)**
2008年度 真正細菌 496種 古細菌 47種
更新中 真正細菌 226種
- 2) **ドラフトゲノム**
2008年度 真正細菌 394種
更新中 真正細菌 223種
- 3) **環境サンプル由来メタゲノム解析の配列群**

650万件のエントリー

■探索フロー

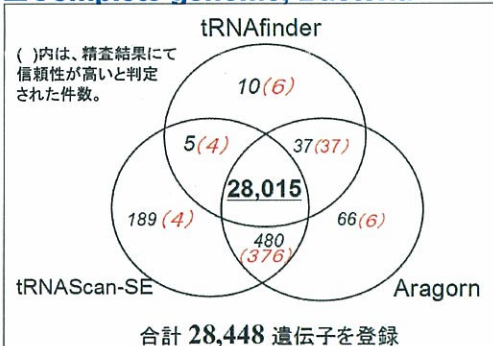


■エキスパートによる精査の流れ



- ◎ tRNA遺伝子の検索
- ◎ マニュアルチェック用データの発信
- チェック済みデータの返信
- ★ 送られてきたデータの集計
- ★ 判定会議(於:長浜バイオ大)
金谷(奈良先端大)、木ノ内(山形大)が加わる
- ★ データベース「IRNADB-CE」に登録

■ Complete genome. Bacteria



■不足必須tRNA遺伝子探索

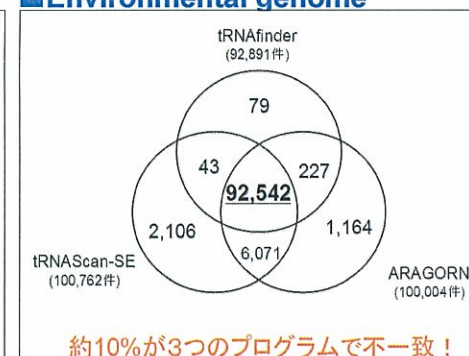
496種類の真正細菌ゲノム中、必須と考えられるtRNA遺伝子が不足しているケース
70種、76遺伝子見出された！

1. イントロンがあつてそのtRNA遺伝子は検出されていなかった。(38件)
 - IRNAScan-SEとAragornで実施。
 - Burkholderia gladii 14株にてtRNA^{Arg}(TAA)はイントロン構造を持つ。
 - *Syntrophocystis* sp. PCC 6803のtRNA^{Met}(CAT)については、FlatFischより取得。
 2. 菌細胞が保有するプラスミドにそのtRNA遺伝子がコードされていた。(9件)
 3. マニュアルチェックで除外したものの中にそのtRNA遺伝子を見出した。(5件)
- 計52件のtRNA遺伝子を追加登録**
- 全く検出することが出来なかったケース
- *Mycoplasma* 8種(tRNA^{Arg}(CCG) or tRNA^{Arg}(TGC))
 - *Haemophilus influenzae* PiliGG (tRNA^{His}(TGT)) は近縁4種では見つかったため、相同性検索を実施したところ、配列ミスの可能性が高い。
 - *Escherichia coli* 536 (TAC) は、他の大腸菌では見つかったが、相同性検索で、*Escherichia coli* 536は、tRNA^{His}は4種は検出された。

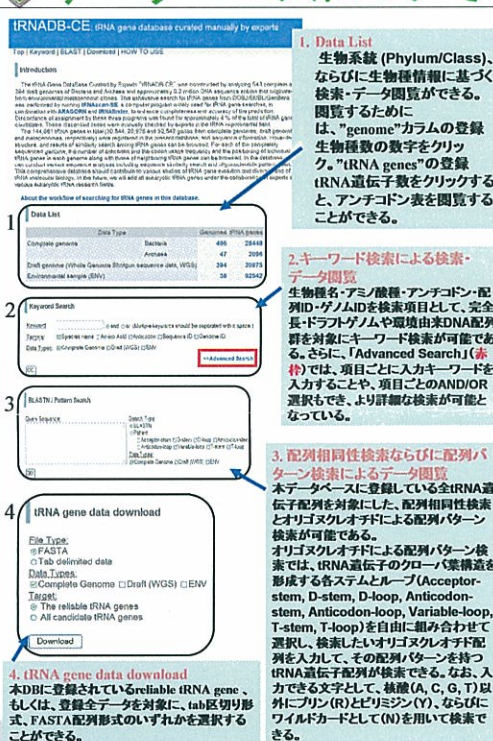
■ Bacteriaのアンチコドン表

[illegible]

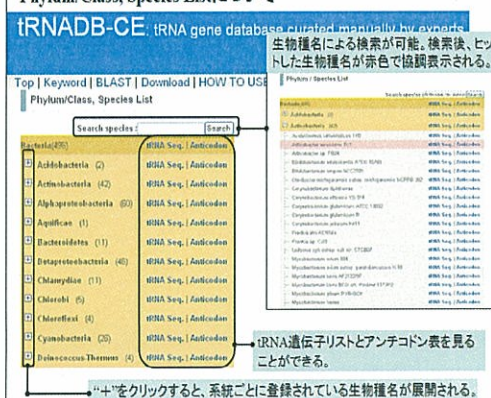
■ Environmental genome



データベースについて



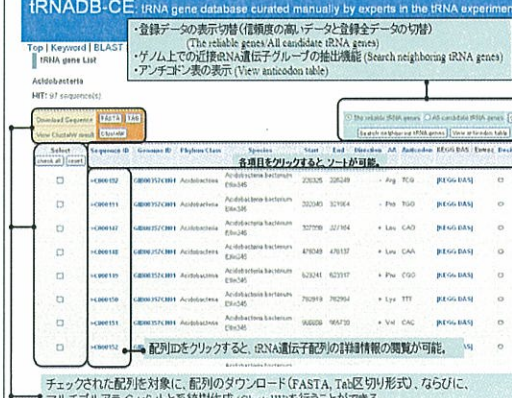
Phylum/Class, Species List(について)



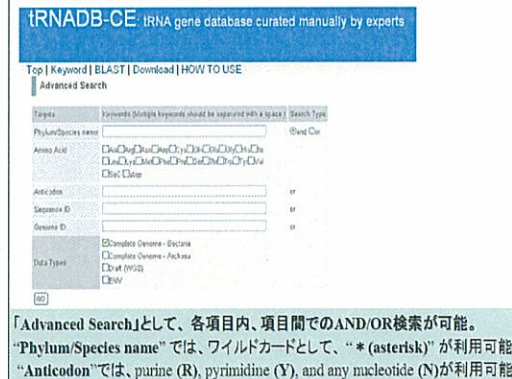
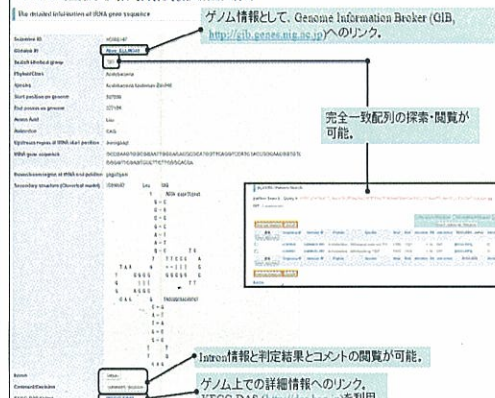
Anticodon table について



tRNA gene List の使い方



Advanced Searchの使い方

[tRNA遺伝子詳細情報閲覧画面](#)

今後

- 年1回、登録データの更新を行う。
- メタゲノムから見出される新規(奇)な tRNA 遺伝子を精査し、新しい発見につなげる。
- Virus (主にPhage)や真核生物 (主に植物ゲノム(イネ))について、tRNA 遺伝子に関するデータの拡充を試みる。

第31回日本分子生物学会年会 合同大会
第81回日本生化学会大会

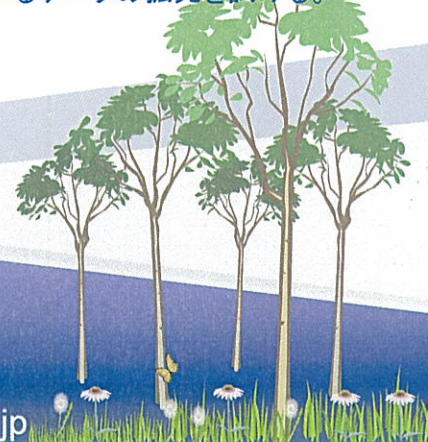
BMB2008 神戸ポートアイランド 2008年12月9日(火)~12月12日(金)

n-bio 長浜バイオ大学 生体分子情報学研究室(池村研究室)

〒526-0829 滋賀県長浜市田村町1266 E-mail: trna@nagahama-i-bio.ac.jp



**Nagahama Institute of
Bio-Science and Technology**



持続可能型社会への貢献遺伝子データベース

“Database for Genes Contributing to Sustainable World”

～ 膨大な 環境由来 メタゲノム配列 からの 有用遺伝子探索 ～



○上原 啓史、棚橋 佳世、阿部 貴志、池村 淑道
長浜バイオ大学 バイオサイエンス学部

○Hiroshi Uehara, Kayo Tanahashi,
Takashi Abe, Toshimichi Ikemura,
Nagahama Inst. of Bio-Sci. and Tech.,

< Abstract >

文部科学省の「ライフサイエンス分野の統合データベース整備事業」の人材養成プログラム活動の一環として行なっている長浜バイオ大学の学生実習等をもとに、「持続可能型社会への貢献遺伝子データベース」の構築を2007年6月から開始している。自然環境の浄化や保全に役立つ、広い意味では「持続可能型社会の実現に貢献できる可能性を持つ遺伝子」を、各学生が独自に想定して、世界で進行している環境生物の混合試料に関する大規模シーケンシング(メタゲノム解析)で得られた配列から発掘を試みている。具体的には、国際塩基配列データベース(DDBJ/EMBL/GenBank)において未だ機能がほとんど特定されていない、メタゲノム解析由来の大量塩基配列を対象に、想定した遺伝子の探索を行なっている。想定した有用遺伝子の候補が発掘できた場合、「持続可能型社会への貢献遺伝子データベース」(プロトタイプ版: <http://dbcls.nagahama-i-bio.ac.jp>)に登録し、公開している(図1)。

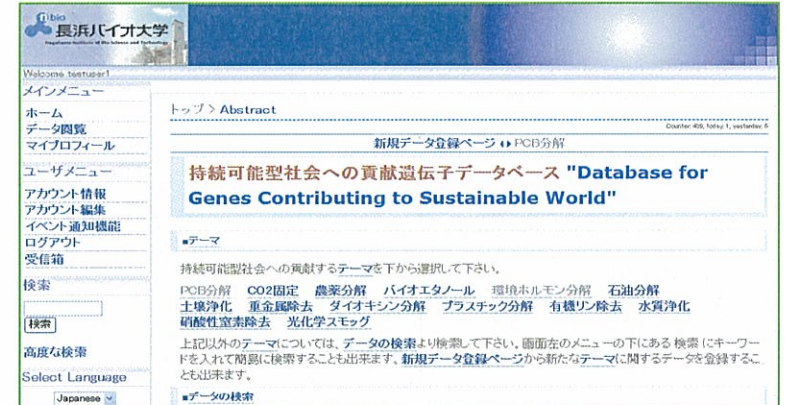
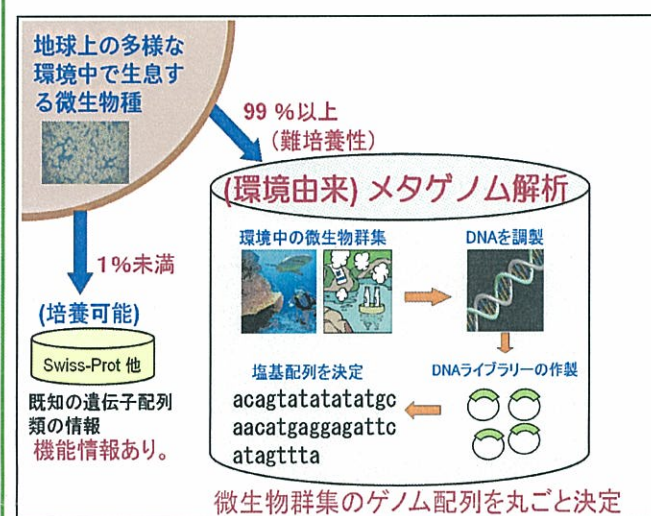


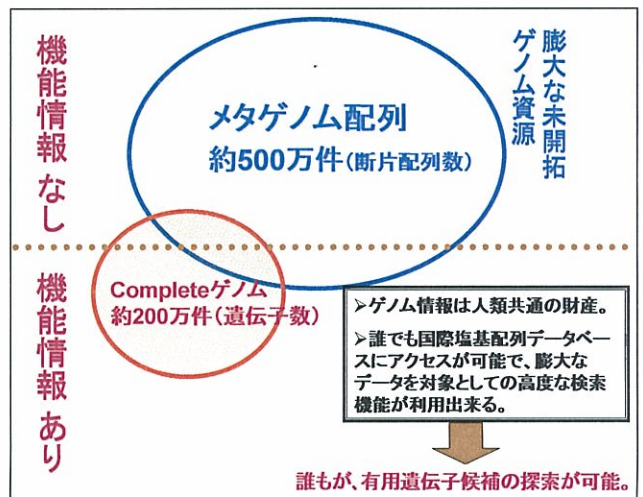
図1. 「持続可能型社会への貢献遺伝子データベース」トップページ
URL: <http://dbcls.nagahama-i-bio.ac.jp>

< Introduction >

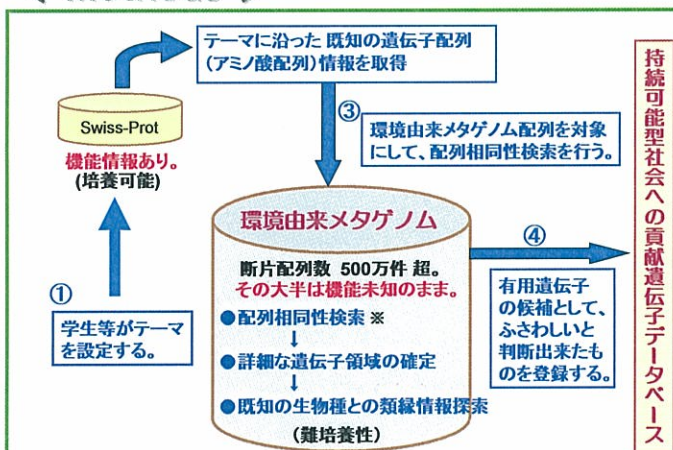


環境中で生息する微生物種の99%以上は、実験室では培養が困難(難培養性)と言われており、膨大なゲノム資源が未開拓のまま残されてきた。近年、これらの難培養性微生物類のゲノムDNA混合物を、環境試料から培養過程を経ずに直接に抽出し、混合ゲノムを対象にしてショットガンシーケンスする「メタゲノム解析法」が普及している。このメタゲノム解析によって得られた多様な環境からの塩基配列データは、断片配列数で500万件を超えている。注目すべきことに、その大半が機能に関する記述なしに、国際塩基配列データベースに収録されている。

私達は、これらの機能に関する記述なしに、(従って利用価値が低いままに)国際塩基配列データベースに収録されている、500万件を超える膨大な環境由来の遺伝子塩基配列(メタゲノム配列)に着目し、この中から環境の浄化や保全に役立つ可能性のある遺伝子の候補を探索し、探索出来たものを「持続可能型社会への貢献遺伝子データベース」に登録し公開を行っている。



< Methods >



※ ● E-value : $1e-10$ 以下 ● アミノ酸配列の一致率: 30%以上 ● カバー率: 30%以上

持続可能型社会の実現に貢献が期待できる、具体的には環境浄化・環境ホルモン分解・重金属除去・PCB分解・微生物によるCO₂固定・バイオエタノール生産等に能力を発揮できる遺伝子やタンパク質の候補を学生等が自由に設定する。設定した遺伝子やタンパク質について、既知のアミノ酸配列をSwiss-Protから取得し、取得したアミノ酸配列を問合せ配列とし、環境生物試料から得られた膨大なメタゲノム配列を対象に、配列相同性検索を行う。“E-valueが $1e-10$ 以下”、“アミノ酸配列の一致率30%以上”、並びに“問い合わせ配列の配列長に対するカバー率が30%以上”の配列を有用遺伝子の候補とする。詳細な遺伝子領域を確定し、配列情報・位置情報並びに既知の生物種との類縁関係等の情報を付加して、「持続可能型社会への貢献遺伝子データベース」に登録を行う。探索した人の所属・名前入りで公開を行っている。探索方法を説明したテキストの配布や講習会も行っている。

< Results >

■ 主なテーマにおけるメタゲノム配列からの新規探索 遺伝子候補数

環境由来メタゲノム配列から、既に環境ホルモン分解、重金属除去、バイオエタノール生産等に能力を発揮する可能性のある遺伝子候補を、7,000件以上見出し、その情報をデータベースに登録し公開している。

テーマ	(原生物由来) CO ₂ 固定	窒素除去	PCB分解	農薬分解	リン除去	有機物分解	バイオエタノール生産	(合計)
サルガッソ海	2,388	99	1,331	115	1,110	222	185	5,450
ハワイ沖海水	70	8	84	8	23	20	16	229
ミネソタ土壌	60	26	37	3	37	3	11	177
リッヂモンド湖底泥	28	0	2	0	2	4	1	37
鯉魚微生物 (深海)	47	20	38	6	30	16	16	173
ヒト腸内	300	4	66	4	144	15	40	573
シロアリ腸内	83	1	1	1	17	0	49	152
その他	175	51	30	1	88	28	15	388
(合計)	3,151	209	1,589	138	1,451	308	333	7,179

■ バイオエタノール生産

テーマ	cellulase	Lac	Lip	(合計)
サルガッソ海	63	101	21	185
ハワイ沖海水	2	9	5	16
ミネソタ土壌	5	2	4	11
鯉魚微生物 (深海)	1	13	2	16
ヒト腸内	40	0	0	40
シロアリ腸内	49	0	0	49
その他	8	8	0	16
(合計)	168	133	32	333

Lac: ラッカーゼ
Lip: リグニンペルオキシダーゼ

■ PCB分解

	bph A	bph B	bph C	bph D	bph E	bph F	bph G	合計
サルガッソ海	301	338	236	44	11	296	105	1,331
ハワイ沖海水	22	24	4	5	7	18	4	84
ミネソタ土壌	8	6	2	5	2	4	10	37
鯉魚微生物 (深海)	8	4	6	5	3	5	7	38
ヒト腸内	4	25	0	4	2	27	4	66
その他	2	12	0	2	0	7	10	33
合計	345	409	248	65	25	357	140	1,589

bph A: ビフェニル2,3-ジオキシゲナーゼ、
bph B: ジヒドロジオールデヒドロゲナーゼ
bph C: ジヒドロキシビフェニル1,2-ジオキシゲナーゼ
bph D: ヒドロキシオキシフェニルヘキサジエン酸ヒドラーゼ
bph E: ヒドロキシベンタジエン酸ヒドラーゼ
bph F: ヒドロキシオキシベンザルドラーゼ
bph G: アセトアルデヒドデヒドロゲナーゼ

■ データベース画面例

➤ 既知遺伝子 画面の一例 (PCB分解)

テーマ	PCB分解
技術情報(注1)	PCB分解: Spingonases
遺伝子配列(注2)	ヒンメル2,3-ジオキシゲナーゼ
遺伝子配列(注3)	ヒンメル2,3-ジオキシゲナーゼ (EC 1.12.1.28) (2,3-dioxygenase) (2,3-dihydroxybiphenyl dioxygenase) (dbbA)
遺伝子の略号	hbc
遺伝子配列(注4)	芳香族化合物分解の分解代謝経路において、芳香環を開裂する酵素。Extradiol 型芳香環分解酵素(注5)の一種。
Function and Characteristics	
UniProt/GenBank等のデータベースへの登録番号	P11122
所属	長浜バイオ大学
登録者	金 野村 貴志 阿部 貴志 池村 淑道 上原 啓史
登録日	2007/06/01
更新日	2007/06/01

➤ 新規探索遺伝子候補 画面の一例 (PCB分解)

テーマ	PCB分解
遺伝子配列(注1)	ヒンメル2,3-ジオキシゲナーゼ (EC 1.12.1.28) (2,3-dioxygenase) (2,3-dihydroxybiphenyl dioxygenase) (dbbA)
遺伝子の略号	hbc
遺伝子配列(注2)	芳香族化合物分解の分解代謝経路において、芳香環を開裂する酵素。Extradiol 型芳香環分解酵素(注3)の一種。
Function and Characteristics	
UniProt/GenBank等のデータベースへの登録番号	P11122
所属	長浜バイオ大学
登録者	金 野村 貴志 阿部 貴志 池村 淑道 上原 啓史
登録日	2007/06/01
更新日	2007/06/01

遺伝子配列(注1)	遺伝子配列(注2)	遺伝子配列(注3)	遺伝子配列(注4)
ヒンメル2,3-ジオキシゲナーゼ (EC 1.12.1.28) (2,3-dioxygenase) (2,3-dihydroxybiphenyl dioxygenase) (dbbA)	ヒンメル2,3-ジオキシゲナーゼ (EC 1.12.1.28) (2,3-dioxygenase) (2,3-dihydroxybiphenyl dioxygenase) (dbbA)	ヒンメル2,3-ジオキシゲナーゼ (EC 1.12.1.28) (2,3-dioxygenase) (2,3-dihydroxybiphenyl dioxygenase) (dbbA)	ヒンメル2,3-ジオキシゲナーゼ (EC 1.12.1.28) (2,3-dioxygenase) (2,3-dihydroxybiphenyl dioxygenase) (dbbA)
ヒンメル2,3-ジオキシゲナーゼ (EC 1.12.1.28) (2,3-dioxygenase) (2,3-dihydroxybiphenyl dioxygenase) (dbbA)	ヒンメル2,3-ジオキシゲナーゼ (EC 1.12.1.28) (2,3-dioxygenase) (2,3-dihydroxybiphenyl dioxygenase) (dbbA)	ヒンメル2,3-ジオキシゲナーゼ (EC 1.12.1.28) (2,3-dioxygenase) (2,3-dihydroxybiphenyl dioxygenase) (dbbA)	ヒンメル2,3-ジオキシゲナーゼ (EC 1.12.1.28) (2,3-dioxygenase) (2,3-dihydroxybiphenyl dioxygenase) (dbbA)
ヒンメル2,3-ジオキシゲナーゼ (EC 1.12.1.28) (2,3-dioxygenase) (2,3-dihydroxybiphenyl dioxygenase) (dbbA)	ヒンメル2,3-ジオキシゲナーゼ (EC 1.12.1.28) (2,3-dioxygenase) (2,3-dihydroxybiphenyl dioxygenase) (dbbA)	ヒンメル2,3-ジオキシゲナーゼ (EC 1.12.1.28) (2,3-dioxygenase) (2,3-dihydroxybiphenyl dioxygenase) (dbbA)	ヒンメル2,3-ジオキシゲナーゼ (EC 1.12.1.28) (2,3-dioxygenase) (2,3-dihydroxybiphenyl dioxygenase) (dbbA)

各データに、コメントの入力出来るデータベース

Comment

Edit Card

名前

本文

➤本データベースでは、登録している各データに対して、コメントの書き込みを可能としている。学生が発掘した遺伝子候補や、興味を持った他大学の学生や若手研究者、企業の研究者並びに一般の方が発掘した遺伝子候補に関して、相互に情報交換を持ちながら新たな知識発見を行うことの出来る「集合知の形成」が可能となるデータベースを目指している。複数の方が同じテーマで同じメタゲノム配列データ特定したものについては、高い専門知識をお持ちのシニア世代の先生方にキュレーションをして頂くことも予定している。産業界が現在利用している工業的に有用な酵素についても、今までの常識とは違った環境中のメタゲノム配列の中から更に高い活性を示すものや、同程度の活性であっても培養スピードが速い、もしくは培養が容易な生物種の中から見出せれば、それは産業界においてコストダウンにも繋がる。

➤今後、本データベースに配列相同性検索によらないBLSOMによる生物系統の推定情報も追加していくことを予定している。新規性の高いメタゲノム配列に対して正確な生物系統が推定できれば、現時点で難培養性の生物種でも培養法の開発に指針を提供できる。新規な遺伝子類を豊富に保有するとみられる環境由来のメタゲノム配列データは、加速度的な勢いで登録数を増大している。本データベースが、新たな発見の手がかりとなることを目指している。広く多様な皆様から利用されるデータベースとなるよう、新規情報を組み入れ、利用者との双方向性を持つデータベースとして維持・更新していく。ぜひ、利用いただいた皆様からのご意見をお寄せいただきたい。

※ テキストご希望、講習会の開催依頼、ご意見 等につきましては、h_uehara@nagahama-i-bio.ac.jp 長浜バイオ大学 上原 迄ご連絡下さい。

健康への貢献ならびに持続可能型社会への貢献遺伝子データベース
～ 膨大な 環境由来メタゲノム配列からの 有用遺伝子探索 ～