tRNADB-CE：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)、斉藤英司1)、後藤大起1)、池村淑道2)、山田優子1)、武藤昱3)、井口八郎1)
1) 新潟大, 2) 長浜バイオ大, 3) 弘前大

http://trna.ie.niigata-u.ac.jp

概要

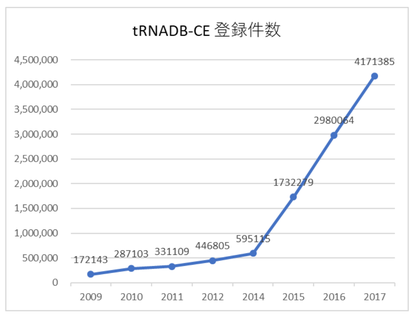
tRNA遺伝子の網羅的特徴把握を目標に、公開されている原核生物の6,024の完全長ゲノム、67,053のドラフトゲノム、876のウイルス・ファージ、121の葉緑体、12の真核生物、ならびに221のメタゲノム配列を対象に網羅的な探索を行い、約417万件のtRNA遺伝を公開している。tRNA遺伝子検索を可能な限り完全にするため、3種の予測プログラム（tRNAscan-SE, Aragorn, tRNAfinder）を併用して検索を行い、予測領域が一致しない場合、シニア世代のtRNA研究者がマニュアルにて精査を行い、その判定結果も登録している。ゲノムの登録数の増加に伴い、蓄積したtRNAの精査情報やtRNA研究者のマニュアル精査の際のチェックポイントを組み込んだtRNA判定システムを構築し、更新作業の効率化も図っている。

本DBは世界最大規模の登録件数で、信頼性の高いDBであり、tRNAの分子生物学的な研究や微生物の進化を研究するための新規情報と解析手法を提供している。

Nucleic Acids Res. 2009;37, D163-D168. doi: 10.1093/nar/gkn692
Nucleic Acids Res. 2011;39, D210-D213, doi: 10.093/nar/gkq1007
Front. Genet. 2014;5:114. doi: 10.3389/fgene.2014.00114

我々がtRNA遺伝子の網羅的把握を目標として開発・公開しているDB．約417万件登録されている
(Data Version 11.0. 2018/1/9 更新. 現在更新中)

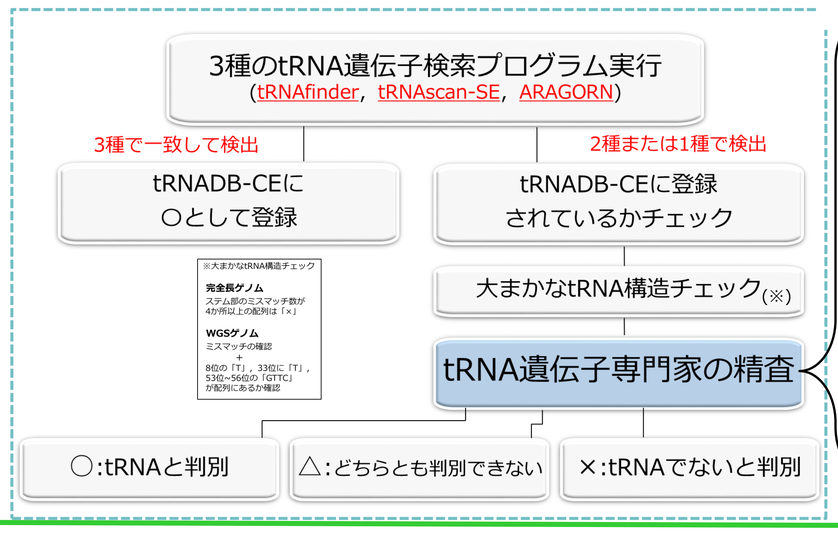
←tRNADB-CE トップページ



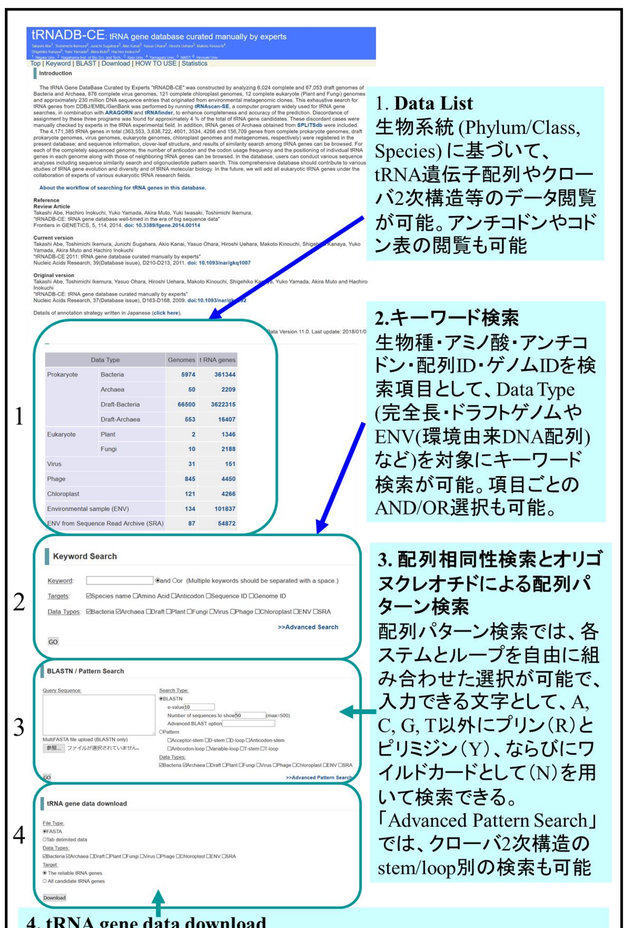
tRNADB-CE→に登録されているデータ件数の推移

年（2008年9月から運用開始）

tRNADB-CEのデータ構築方法更新



➢データベース機能概要

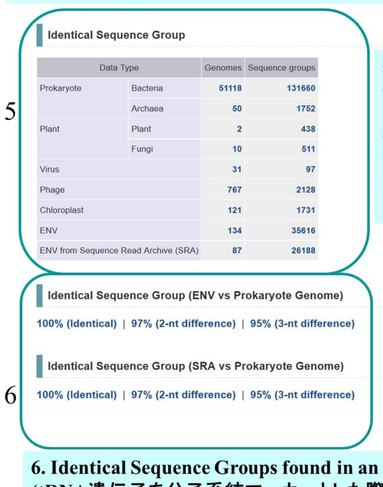


1. **Data List**
生物系統 (Phylum/Class, Species) に基づいて、tRNA遺伝子配列やクローバ2次構造等のデータ閲覧が可能。アンチコドンやコドン表の閲覧も可能

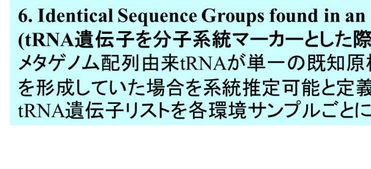
2. **キーワード検索**
生物種・アミノ酸・アンチコドン・配列ID・ゲノムIDを検索項目として、Data Type (完全長・ドラフトゲノムやENV(環境由来DNA配列)など)を対象にキーワード検索が可能。項目ごとのAND/OR選択も可能。

3. **配列相同性検索とオリゴヌクレオチドによる配列パターン検索**
配列パターン検索では、各ステムとループを自由に組み合わせた選択が可能で、入力できる文字として、A, C, G, T以外にプリン(R)とピリミジン(Y)、ならびにワイルドカードとして(N)を用いて検索できる。
「Advanced Pattern Search」では、クローバ2次構造のstem/loop別の検索も可能

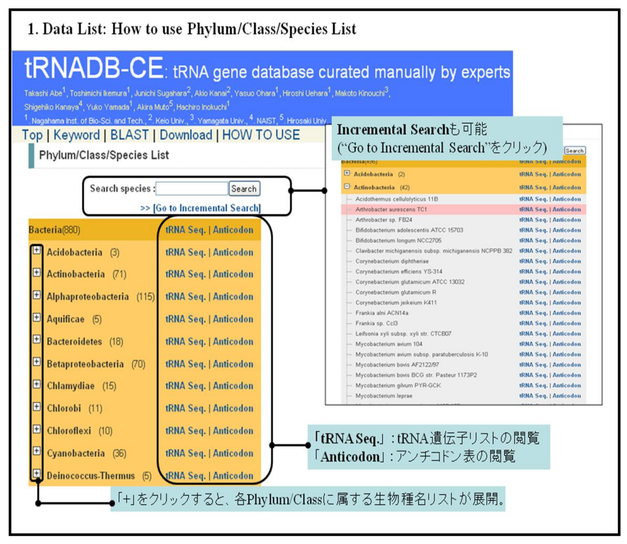
4. **tRNA gene data download**
登録されているreliable tRNA gene 、もしくは、登録tRNA遺伝子候補データ(All candidate tRNA gene)を対象に、tab区切り形式、FASTA配列形式でダウンロード可能。

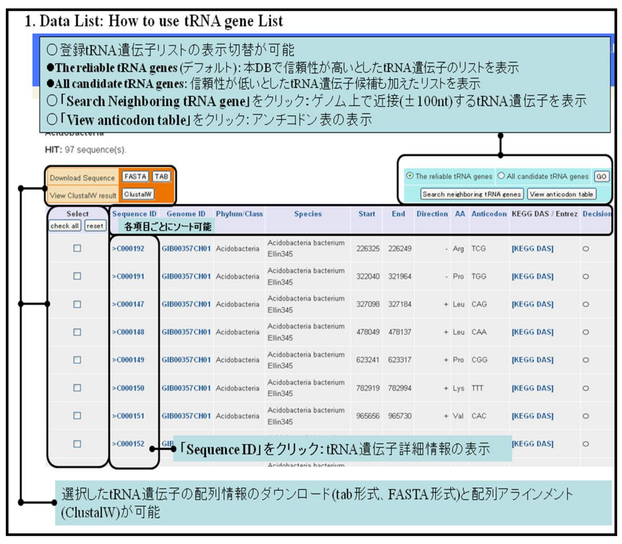


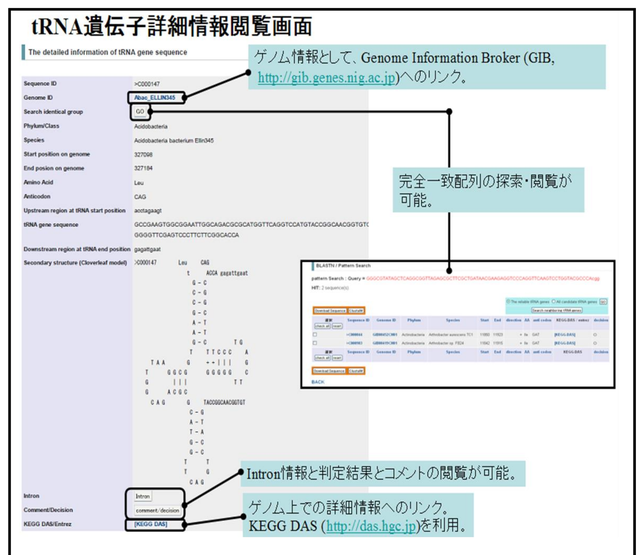
5. **Identical Sequence Group (ISG)**
各Data Type別の配列クラスタリング結果を生物系統 (Phylum/Class, Species) に基づく閲覧が可能。

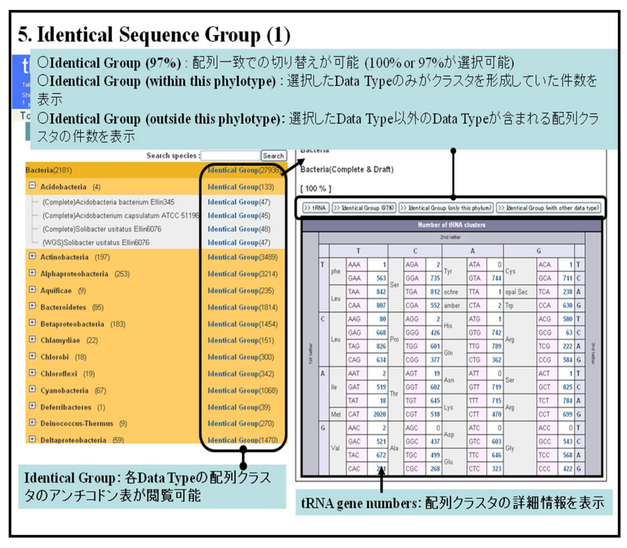


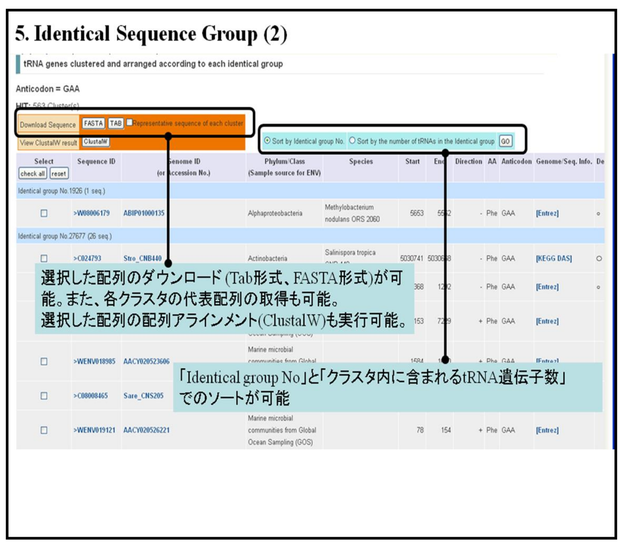
6. **Identical Sequence Groups found in an environmental sample (tRNA遺伝子を分子系統マーカーとした際の系統推定結果閲覧)**
メタゲノム配列由来tRNAが単一の既知原核生物由来tRNAとISGを形成していた場合を系統推定可能と定義し、系統推定できたtRNA遺伝子リストを各環境サンプルごとに閲覧が可能。









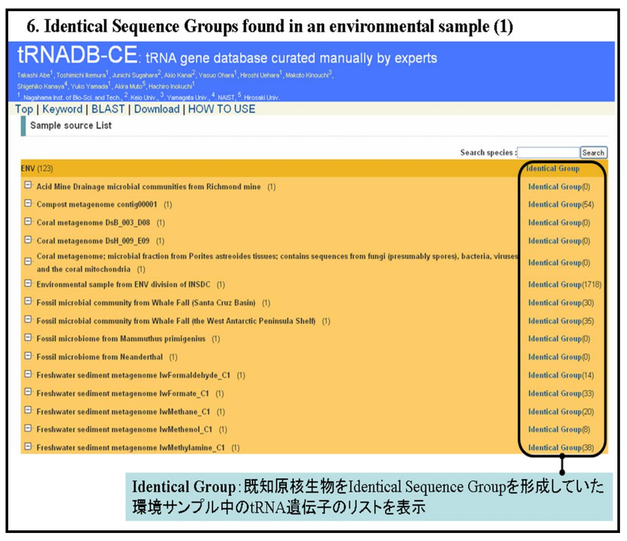


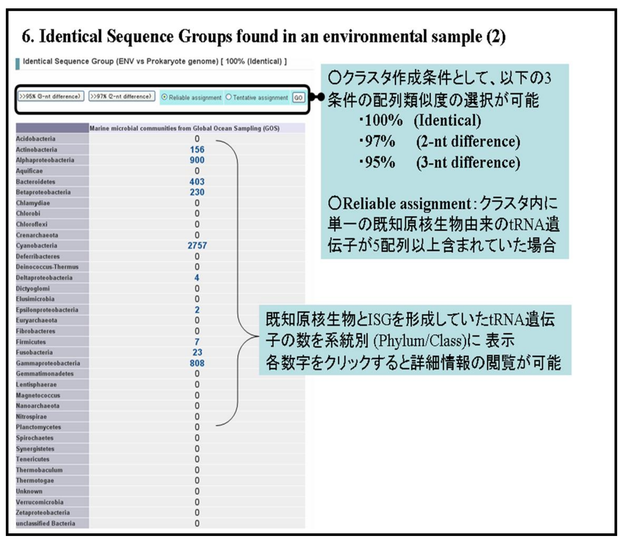
新機能：コメント入力機能

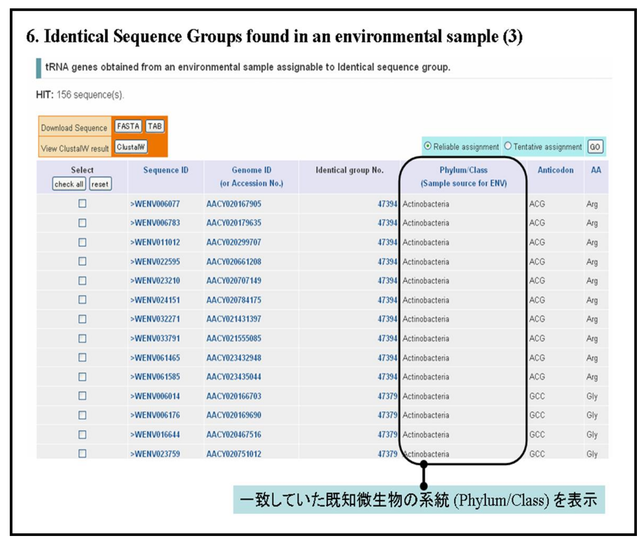
各tRNA遺伝子に対するコメント入力機能を追加
(英語 or 日本語)

E-mailとパスワードを入力いただければどなたでもコメントが入力可能です。

ご意見をお待ちいたしております。



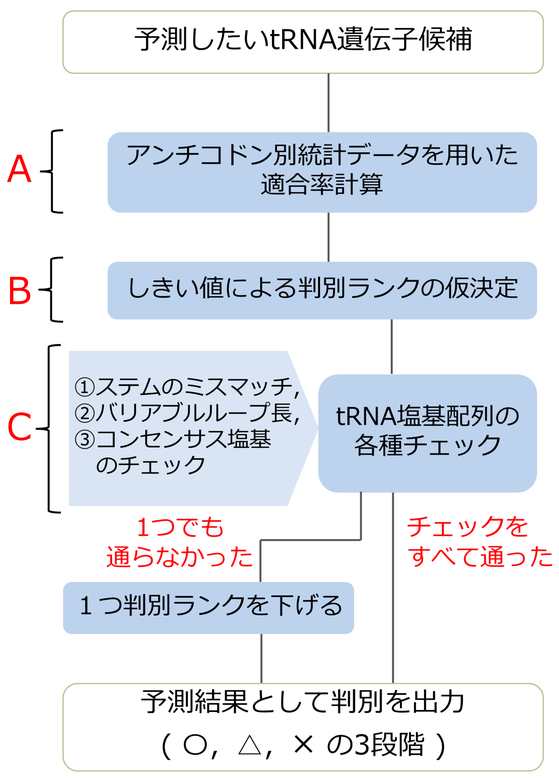




➢専門家の知識を組み込んだ予測システムの開発

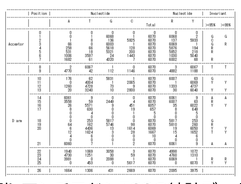
登録データを活用したtRNA遺伝子予測システム

開発したtRNA予測システムのフローを以下に示す



A アンチコドン別統計データを用いた適合率計算

tRNADB-CE 統計データ
- 配列の位置毎に塩基が集計
- 95%に達している塩基をコンセンサス塩基 (CS塩基)



例) アンチコドンTCCの統計データ

予測配列と統計データを比較（全位置計算）
- CS塩基と一致： 2 点
- 一致しない or CS塩基がない： $\frac{\text{各塩基数}}{\text{その位置の総塩基数}}$ 点

統計データで求められる最高得点で除算
＝ 予測したい配列の適合率とする

B しきい値による判別ランクの仮決定

A にて適合率を求めた後,
- 70%以上 ：○
- 60～70% ：△
- 60%以下 ：×

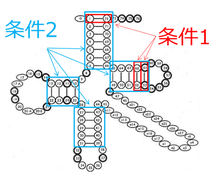
として1回目の判別を行う

C-① ステムのミスマッチチェック

A-T, G-C (G+T)以外のミスマッチ(mm) をチェック

条件1. 1-72, 53-61, 52-62にmm

条件2. G+T対を0.3点, mmを1点として3点以上

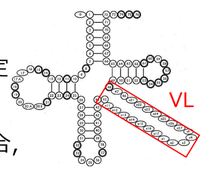


以上の場合，判別を1つ下げる

C-② バリアブルループ長のチェック

バリアブルループ部の塩基数 (VL)をチェック

- 最低長 ： 4
- 最大長 ： アンチコドン別に決定（例 CGA：20）



予測配列のVLがこの範囲でない場合，判別を1つ下げる

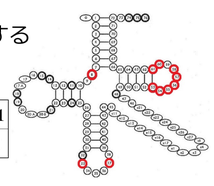
C-③ コンセンサス塩基のチェック

tRNA塩基配列全体に見られるCS塩基をチェック

⇒ 11か所 (赤丸部)の塩基を確認する

表：チェックする配列の位置とコンセンサス塩基

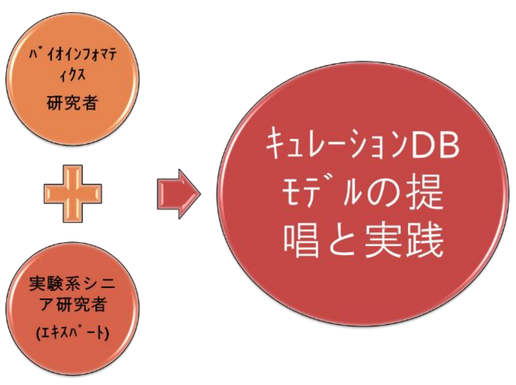
配列の位置（ポジション）	8	33	37	53	54	55	56	57	58	60	61
コンセンサス塩基	T	T	R	G	T	T	C	R	A	Y	C



3か所以上外れていた場合，判別を1つ下げる

・大学院生と共同で，tRNADB-CEとtRNA専門家の知見を取り入れ組み合わせることで，マニュアルチェック結果と80%以上一致する精度の高いシステムの構築を行った.

・tRNADB-CEの機能の一つとして公開することでより網羅性の高いデータベースの実現を目指す



精査の結果の発表と議論の過程を，院生を含む若手研究者が参加して行うことで，知識継承の場となっている．シニア世代も，若手と議論を深め，計算機やDBを使う技術が向上し，自らの関心事項を最新のデータを活用した研究も可能になる．

Licensed under a Creative Commons表示4.0国際ライセンス
(c)2018 阿部貴志（新潟大学）