

STN アラート No. 7

DGENE ホモロジー検索のアラート

特許の配列を網羅的に検索できます

本稿は STN アラートに関するシリーズ記事の 7 回目です。今回のテーマは DGENE ファイルのホモロジー検索のアラートです。

STN にはバイオシーケンス(核酸・タンパク質の配列データ)を配列コードから検索できる配列データベースが 4 つあります。いずれのファイルにも特許由来の配列が収録されており、特許の配列調査には最適の検索システムと言えます。

STN の配列データベース

ファイル名	収録されている配列	検索タイプ
DGENE	WPI に収録された世界中の特許(ベーシック特許)の配列	<ul style="list-style-type: none"> 完全配列検索 完全配列ファミリー検索
PCTGEN	PCT 出願の配列	
USGENE	米国の公開特許・登録特許の配列	
REGISTRY	CAplus ファイルに収録された雑誌論文と世界中の特許(ベーシック特許)の配列	<ul style="list-style-type: none"> 部分配列検索 部分配列ファミリー検索 ホモロジー検索

また、調査テーマに応じてさまざまな検索タイプのアラートを登録できます。とりわけ配列の特許性を調査する場合は、多少コードの並びが異なる配列も含めて網羅的に検索できるホモロジー検索がお勧めです。

DGENE ファイルのホモロジー検索

配列データベースのうち、DGENE / PCTGEN / USGENE の 3 ファイルでは、配列検索やアラート登録などの利用方法が同じです。ホモロジー検索においては、以下の 2 種類のプログラムを利用することができます。

BLAST

最もよく利用されています。他のプログラムに比べて高速処理できます。ギャップを考慮しないため、検出感度や選択性が低いと考えられがちですが、実際には他と比べてそれほど遜色はありません。

GETSIM

FASTA 系列のプログラムです。FASTA のように近似を行うことなく、データベース中のすべての配列との間で忠実にアラインメントを行ってホモロジスコアを算定します。

また、核酸配列を検索するのか、タンパク質配列を検索するのかを指定するため、以下の検索フィールドを利用します。核酸をホモロジー検索する場合は、配列質問式として塩基配列のほかにアミノ酸配列も使用できます。

核酸ホモロジー検索の検索フィールド

検索フィールド	検索機能	配列質問式
/SQN	塩基配列の質問式に類似した核酸を検索	塩基配列
/TSQN	データベース中の塩基配列をアミノ酸に翻訳した配列を対象に、アミノ酸配列の質問式に類似した配列を検索	アミノ酸配列

タンパク質ホモロジー検索の検索タイプ

検索フィールド	検索機能	配列質問式
/SQP	アミノ酸配列の質問式に類似したタンパク質・ペプチドを検索	アミノ酸配列

DGENE ホモロジー検索のアラート

ホモロジー検索のアラートでは、回答はオンラインで入手し類似度から適当な配列を選択します。回答はシステム側に自動的に保存されますが、一定期間を過ぎると削除されます。

DGENE/PCTGEN/USGENE ホモロジー検索のアラートの概要

項目	内容
登録方法	・ RUN コマンド行の最後に ALERT と入力する
配列質問式の長さ	・ BLAST : 10,000 コードまで ・ GETSIM : 2,000 コードまで
実行頻度	・ DGENE : 隔週 ・ PCTGEN, USGENE : 毎週
登録可能件数	・ 16 件
回答の保存	<ul style="list-style-type: none"> ・ 回答を呼び出した場合、削除しない限り次の検索まで自動的に保存される。さらに保存しておきたい場合は、SAVE コマンドで保存する ・ 回答を呼び出さなかった場合は、3 ヶ月後にシステムより自動的に削除される
保存可能件数	・ 合計 96 セットまで自動的に保存される。97 番目の回答セットが作成される場合は、最も古い回答セットが削除される

料金体系もユニークです。アラート実行時には課金されず、回答を呼び出す（回収する）際に「SDI 検索結果回収料」が課金されます。回答が 0 件の場合は、回答を回収する前に 0 件であることがわかります。つまり、1 件以上の回答がありそれを回収した場合のみに課金が発生するしくみです。

アラートを登録する場合は、RUN コマンドで以下のよう
に指定してください。

アラート登録方法

=> **RUN** プログラム名 配列質問式 / 検索フィールド
オプション パラメータ **ALERT**
PLEASE ENTER ALERT IDENTIFIER (MAX. 8 CHARS): 登録名
PLEASE ENTER ALERT TITLE (MAX. 40 CHARS): タイトル

プログラム名	・GETSIM ・BLAST
検索フィールド	・/SQN ・/TSQN ・/SQP
オプション (核酸ホモロジー検索の場合)	<ul style="list-style-type: none"> ・SIN: 入力した配列コードについてホモロジー検索します (GETSIM 利用時のデフォルト) ・COM: 入力した配列コードの相補鎖についてホモロジー検索します ・BOTH: 入力した配列コードとその相補鎖についてホモロジー検索します (BLAST 利用時のデフォルト)
パラメータ (BLAST 利用の場合)	<ul style="list-style-type: none"> ・フィルタや期待値など、9 種類のパラメータを指定できます ・詳しくは以下をご参照ください www.jaici.or.jp/stn/pdf/para.pdf

また、アラートの登録内容を確認したり、回答を回収する場合も、以下のように RUN コマンドを用います。

アラート登録の確認、または回答の回収方法

=> **RUN** **ALERT**
Enter "R" to process alert results
or "Q" to process alert queries
or alert id to retrieve results
or enter . for (end)
ALERT REQUEST: 項目

項目	目的
Q	アラートの登録内容を確認する
R	アラートの回答を確認または回収する
END	指定を中断し => プロンプトに戻る

アラート活用例

それでは、DGENE ファイルのホモロジー検索のアラート活用例をご紹介します。今回は、ある塩基配列の相補鎖について、類似性の高い配列を GETSIM ホモロジー検索でモニタリングします。インターフェースは STN Express V8.3 を用います。

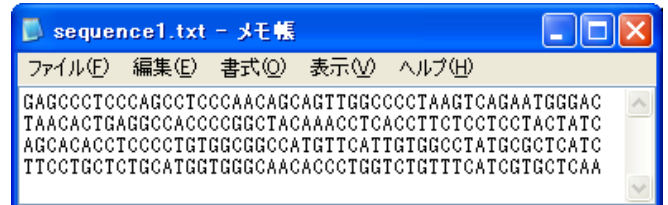
調査テーマ：以下の塩基配列の相補鎖について、類似性の高い配列をモニタリングする

```
GAGCCCTCCCAGCCTCCCAACAGCAGTTGGCCCTAAGTCAGAATGGGAC
TAACACTGAGGCCACCCCGGCTACAAACCTCACCTTCTCCTCTACTATC
AGCACACCTCCCCTGTGGCGGCCATGTTTCATTGTGGCCTATGCGCTCATC
TTCTGCTCTGCATGGTGGGCAACACCCCTGGTCTGTTTCATCGTGTCTCAA
```

1. アラートを登録します

以下のルールに従い、配列質問式のテキストファイルを作成します。


- ・配列コードのみを入力します。
- ・曖昧コードは利用できませんが、ギャップ記号・特殊記号は利用できません。
- ・スペースは含まれていてもかまいません。
- ・一行 300 文字以内に分割します。

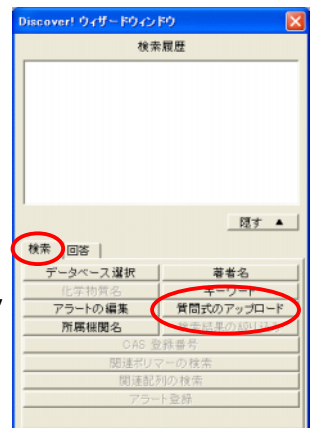


DGENE ファイルで配列質問式をアップロードします。

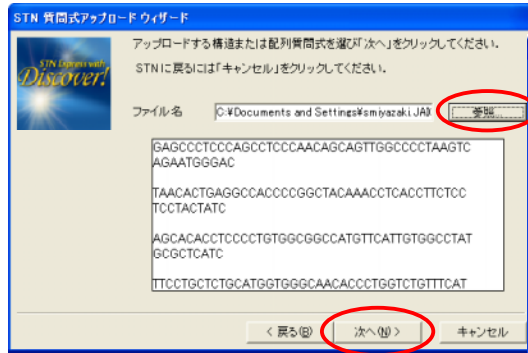
=> **FILE** **DGENE** DGENE ファイルに入ります

Discover! ウィザードウィンドウの「検索」タブを選択し「質問式のアップロード」をクリックします。

Discover! ウィザードウィンドウが開いていない場合は、セッション画面上の  アイコンをクリックしてください。



「参照」をクリックして配列質問式のファイルを選択し「次へ」をクリックします。



「DGENE」を選択し「終了」をクリックすると配列質問式がアップロードされます。



```
=> UPL R BLAST
UPLOAD SUCCESSFULLY COMPLETED
L1 GENERATED
```

=> D LQUE L1 配列質問式の内容は表示確認できます

```
L1 ANSWER 1 DGENE COPYRIGHT 2008 THOMSON REUTERS on STN
LQUE GAGCCCTCCAGCCTCCCAACAGCAGTTGGCCCTAAGTCAGAATGGGACTAACACTGA
GGCCACCCCGGCTACAAACCTCACCTTCTCCTCTACTATCAGCACACCTCCCTGTGG
CGGCCATGTTCAATTGTGGCTATGCGCTCATCTTCTGCTCTGCATGGTGGCAACACC
CTGGTCTGTTTCATCGTGTCTCAA
```

RUN コマンドでアラートを登録します。

```
=> RUN GETSIM
PLEASE ENTER SEQUENCE QUERY OR ? : L1 /SQN COM ALERT
PLEASE ENTER ALERT IDENTIFIER (MAX. 8 CHARS) : ALERTN1
PLEASE ENTER ALERT TITLE (MAX. 40 CHARS) : Similarity Search of sequence 1
NEW ALERT CREATED
```

GETSIM ホモロジー検索の指示
COM で相補鎖検索, ALERT でアラート登録を指示
登録名
タイトル
アラート登録の完了

2. アラートの回答を受け取ります

```
=> RUN ALERT
Enter "R" to process alert results
or "Q" to process alert queries
or alert id to retrieve results
or enter . for (end)
ALERT REQUEST: R
```

R を入力するとアラートの回答リストが表示されます

```
CURRENT RESULTS AVAILABLE
NAME RUN DATE
1) ALERTN1 20081022 (getsim)
2) ALERTN2 20080808 (getsim)
3) ALERTN2 20080822 (getsim)
4) ALERTN2 20080905 (getsim)
5) ALERTN2 20080920 (getsim)
6) ALERTN2 20081003 (getsim)
7) ALERTN2 20081022 (getsim)
8) ALERTN3 20080808 (getsim)
:
13) ALERTN3 20081022 (No answers - getsim)
:
35) DGGETSIM 20081022 (getsim)
-----
```

ALERTN1 の 2008 年 10 月 22 日
検索実行分

ALERTN2 のように回答を回収しない
場合は 3 ヶ月後に自動的に削除され
ます

回答がない場合は
リストに表示されません

アラートを登録した後は、隔週ごとに GETSIM ホモロジー検索が自動検索され、システム側に回答が保存されます。

回答は一度呼び出す(回収する)と次の自動検索時に削除されます。回収しなかった場合も 3 ヶ月後に自動的に削除されます。このため必要な回答は自動削除される前に SAVE コマンドで保存することをお勧めします。

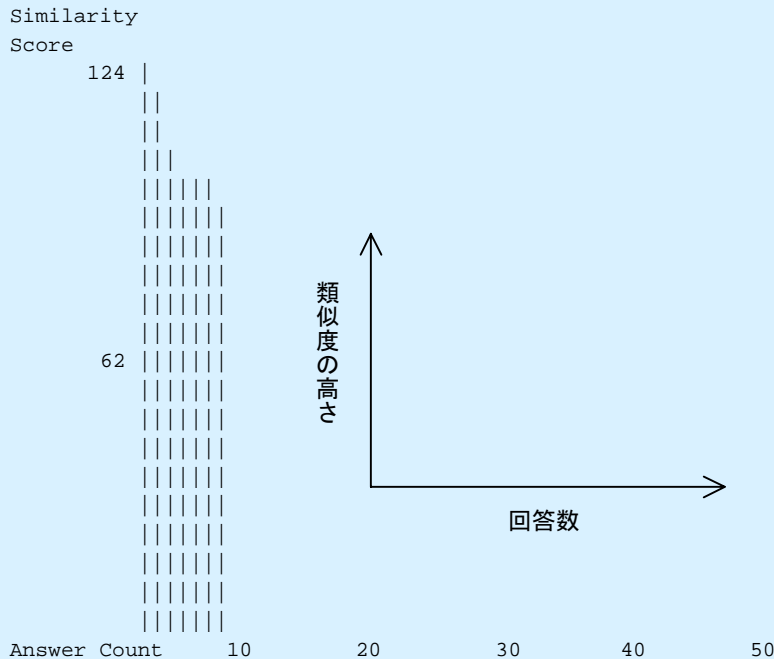
回答を回収する場合は RUN ALERT コマンドを実行します。R と入力すると回答リストが表示されます。

(No answers)の表記のある回答は 0 件の場合です。この回答は回収できませんし、課金も発生しません。

Enter Number of result to be selected
or "-" before Number to delete result
or "Q" to process alert queries
or enter . for (end)

RESULT REQUEST: 6 回答番号を指定すると回答が呼び出されます

7 ANSWERS FOUND ABOVE A THRESHOLD OF 93
QUERY SELF SCORE VALUE IS 1000
BEST ANSWER SCORE VALUE IS 124



ENTER EITHER THE NUMBER OF ANSWERS YOU WISH TO KEEP
OR ENTER MINIMUM PERCENT OF SELF SCORE FOLLOWED BY %
(BEST ANSWER PERCENTAGE IS 12%)

ENTER (ALL) OR ? : ALL ALL と入力すると今回のアラートで得られた
配列全件の回答セットが作成されます

L1 RUN STATEMENT CREATED
L1 7 GAGCCCTCCCAGCCTCCCAACAGCAGTTGGCCCCTAAGTCAGAATGGGAC
 TAACACTGAGGCCACCCCGGCTACAAACCTCACCTTCTCCTCTACTATC
 AGCACACCTCCCCTGTGGCGCCATGTTTCATTGTGGCCTATGCGCTCATC
 TTCTGCTCTGCATGGTGGCAACACCCTGGTCTGTTTCATCGTCTCAA
 /SQN.COM

Answer set arranged by accession number; to sort by descending
similarity score, enter at an arrow prompt (=>) "sor score d".

CURRENT RESULTS AVAILABLE
NAME RUN DATE
1) ALERTN1 20080808 (getsim)
2) ALERTN1 20080822 (getsim)
:

Enter Number of result to be selected
or "-" before Number to delete result
or "Q" to process alert queries
or enter . for (end)

RESULT REQUEST: END END を入力すると表示モードが終了します

回答リストの中から目的の回答の番号を選択して RESULT REQUEST: の後に入力すると、グラフが表示されます。

これは、類似度別に配列の分布を表しています。回答には類似度のかなり低い配列も含まれますので、調査目的に応じて類似度の高い配列に限定する必要があります。

グラフ表示の後に、呼び出す（回収する）配列件数の入力（ALL と入力すれば全ての配列が回収されます）が促されます。

また、5 などのように数字を入力すると類似度順に上位 5 位までの配列が回収されます。このほか、最低スコア値（%）で限定することもできます。

回答が回収されると回答セットの L 番号（L1）が作成されます。この時点で「SDI 検索結果回収料」が課金されます。

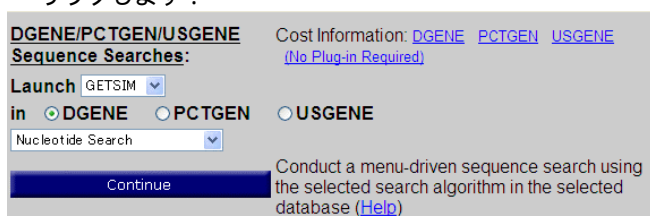
STN on the Web の Sequence Search Assistant

STN on the Web では、配列検索が簡単に行えるようユーザーをサポートする Sequence Search Assistant 機能がご利用いただけます。特に DGENE / PCTGEN / USGENE の 3 ファイルについては、すべての検索タイプの配列検索をサポートしており、メニュー方式で簡単に検索していただくことができます。また、アラートの登録や回答表示についてもホモロジー検索の場合は Sequence Search Assistant 機能を利用できます（ホモロジー検索以外の検索タイプのアラートを登録する場合は Alert Assistant 機能がご利用いただけます）。

STN on the Web の Sequence Search Assistant 機能を利用してホモロジー検索をアラート登録する手順を簡単にご紹介します。詳しくは「STN on the web 配列検索ガイド」（www.jaici.or.jp/stn_web/webseq.pdf）をご覧ください。

STN on the Web に接続後、画面左のツールバーから「Search Assistants」をクリックし「Sequence Asst.」をクリックします。

Sequence Search Assistant 画面が表示されたら使用するファイル（DGENE）とプログラム（GETSIM）、核酸検索する旨を指定し「Continue」をクリックします。



DGENE/PCTGEN/USGENE Sequence Searches: Cost Information: [DGENE](#) [PCTGEN](#) [USGENE](#) (No Plug-in Required)

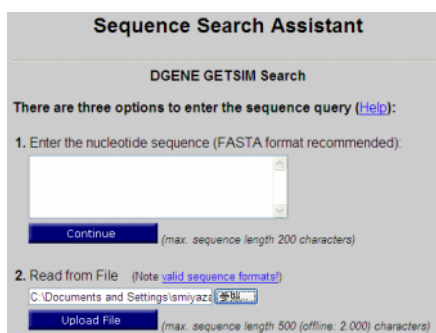
Launch GETSIM

in DGENE PCTGEN USGENE

Nucleotide Search

Continue Conduct a menu-driven sequence search using the selected search algorithm in the selected database ([Help](#))

「参照」をクリックし、予め用意しておいた配列質問式のファイルを選択して「Upload File」をクリックします。



Sequence Search Assistant

DGENE GETSIM Search

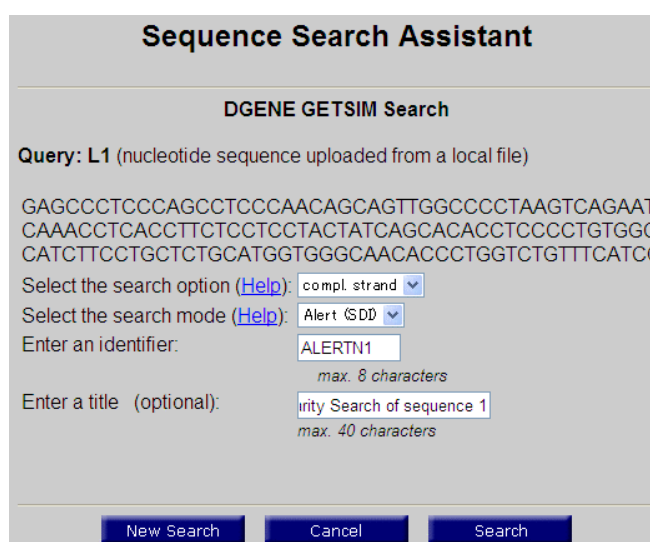
There are three options to enter the sequence query ([Help](#)):

1. Enter the nucleotide sequence (FASTA format recommended):

Continue (max. sequence length 200 characters)
2. Read from File (Note valid sequence formats):
C:\Documents and Settings\simiyaz\参照...
Upload File (max. sequence length 500 (offline: 2,000) characters)

配列質問式がアップロードされ、L 番号が付与されます。ここで、検索オプションと検索モードを以下のように指定して「Search」ボタンをクリックします。これで登録は完了します。

- Select the search option: compl. Strands (入力した配列の相補鎖のみを検索)
- Select search mode: Alert (SDI) (アラートを登録)
- Enter an identifier: アラート登録名
- Enter a title: タイトル



Sequence Search Assistant

DGENE GETSIM Search

Query: L1 (nucleotide sequence uploaded from a local file)

GAGCCCTCCCAGCCTCCCAACAGCAGTGGCCCTAAGTCAGAATCAAACCTCACCTTCTCCTCTACTATCAGCACACCTCCCCTGTGGCCATCTTCTGCTCTGCATGGTGGGCAACACCCTGGTCTGTTTCATCC

Select the search option ([Help](#)): compl. strand

Select the search mode ([Help](#)): Alert (SDI)

Enter an identifier: ALERTN1 (max. 8 characters)

Enter a title (optional): urity Search of sequence 1 (max. 40 characters)

New Search Cancel Search

補足情報

DGENE ファイルホモロジー検索の詳細や各インターフェースの操作方法については以下をご参照ください。また、お気軽にヘルプデスクにお問い合わせください。

DGENE/PCTGEN/USGENE BLAST ホモロジー検索のパラメータ

www.jaici.or.jp/stn/pdf/para.pdf

STN on the web 配列検索ガイド:

www.jaici.or.jp/stn_web/webseq.pdf

STN の配列検索関連資料

www.jaici.or.jp/stn/stn_doc_02.html#06

ヘルプデスク:

www.jaici.or.jp/helpdesk/index.htm