

ゲノムアノテーション参照システム
Genome Annotation System
ご紹介

株式会社メイズ

Tel: 042-673-3020

<http://www.maze.co.jp/>

〒193-0835 東京都八王子市千人町1-2-17 604

遺伝子情報が僅かな生物種のための機能予測支援DB搭載

既知蛋白質アミノ酸配列への全Blast結果を利用したリッチなアノテーション

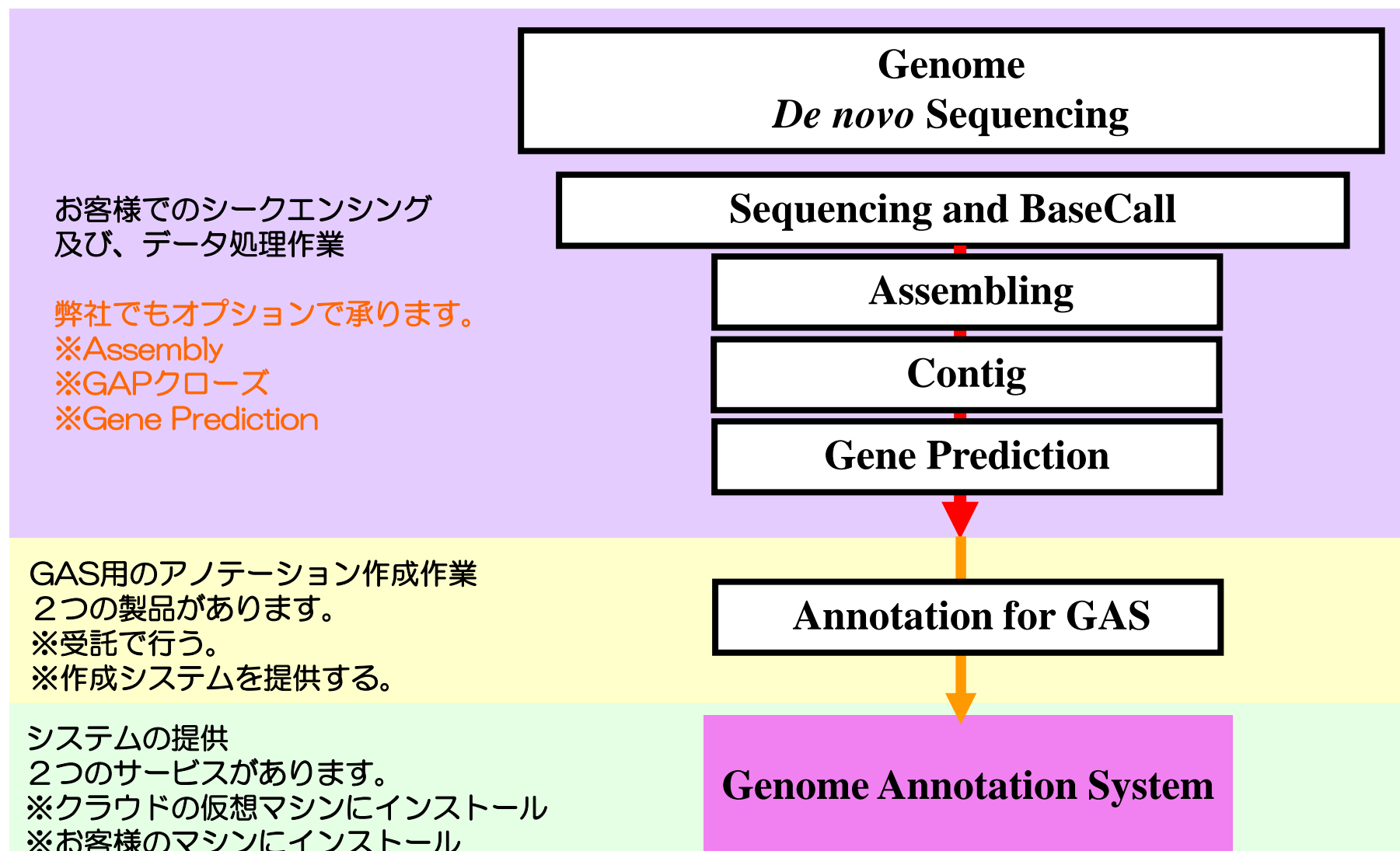
1. はじめに
2. シークエンシングからアノテーション作成及びGASご提供の概要
3. GASの主な特徴（その1）
4. GASの主な特徴（その2）
5. GASの構成図とアノテーション作成の関係図
6. GASのWebインターフェース
7. データ更新と年間サポート
8. GASの機能一覧
9. 公開サーバとして活用することができます

1. はじめに

De novo 解析の結果を格納した実用的なゲノムデータ解析システムとして、“Genome Annotation System”（以降、“GAS”と略します）をご提案いたします。

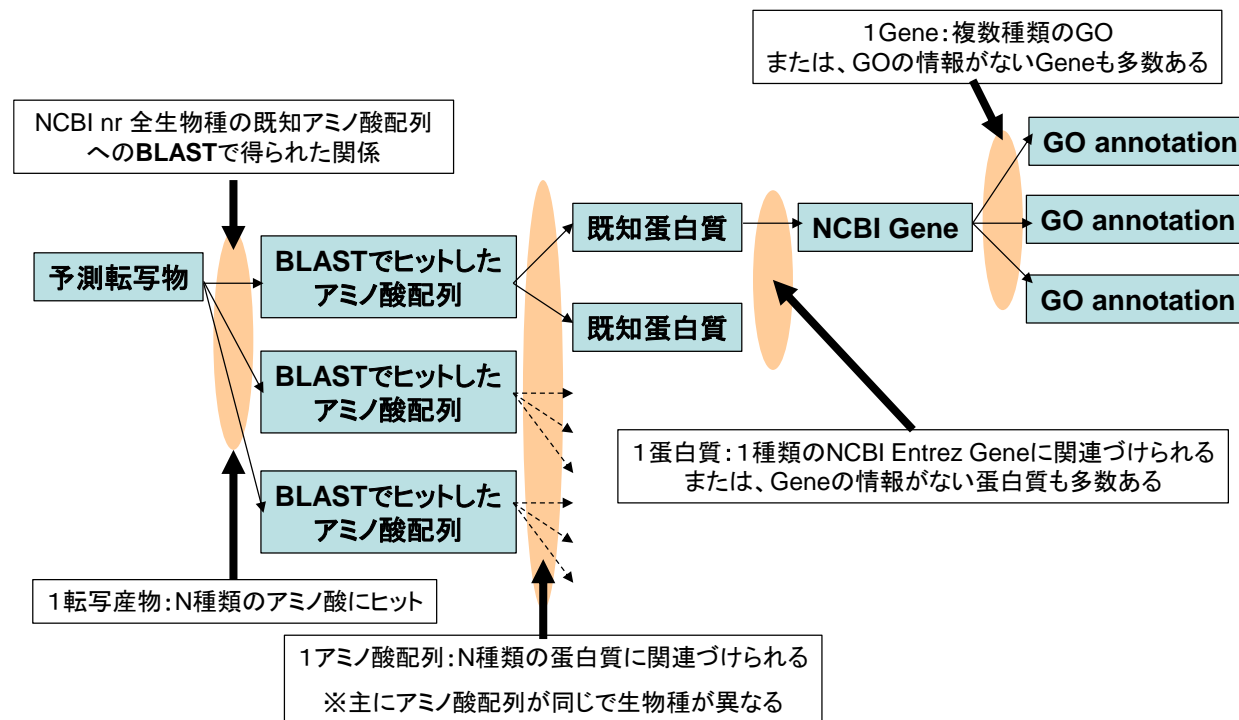
- ★ ゲノム配列から予測された転写配列にリッチなアノテーションを付与したデータベースをWebのユーザインタフェースで簡単に検索・参照できるシステムです。
- ★ コンピュータの専門家に頼らず、簡単にアノテーション情報を引き出すことができるツールです。
- ★ 共同研究者同士でのデータを共有システムとして、また、そのまま公開用システムとして利用可能です。

2. シークエンシングからアノテーション作成及びGASご提供の概要

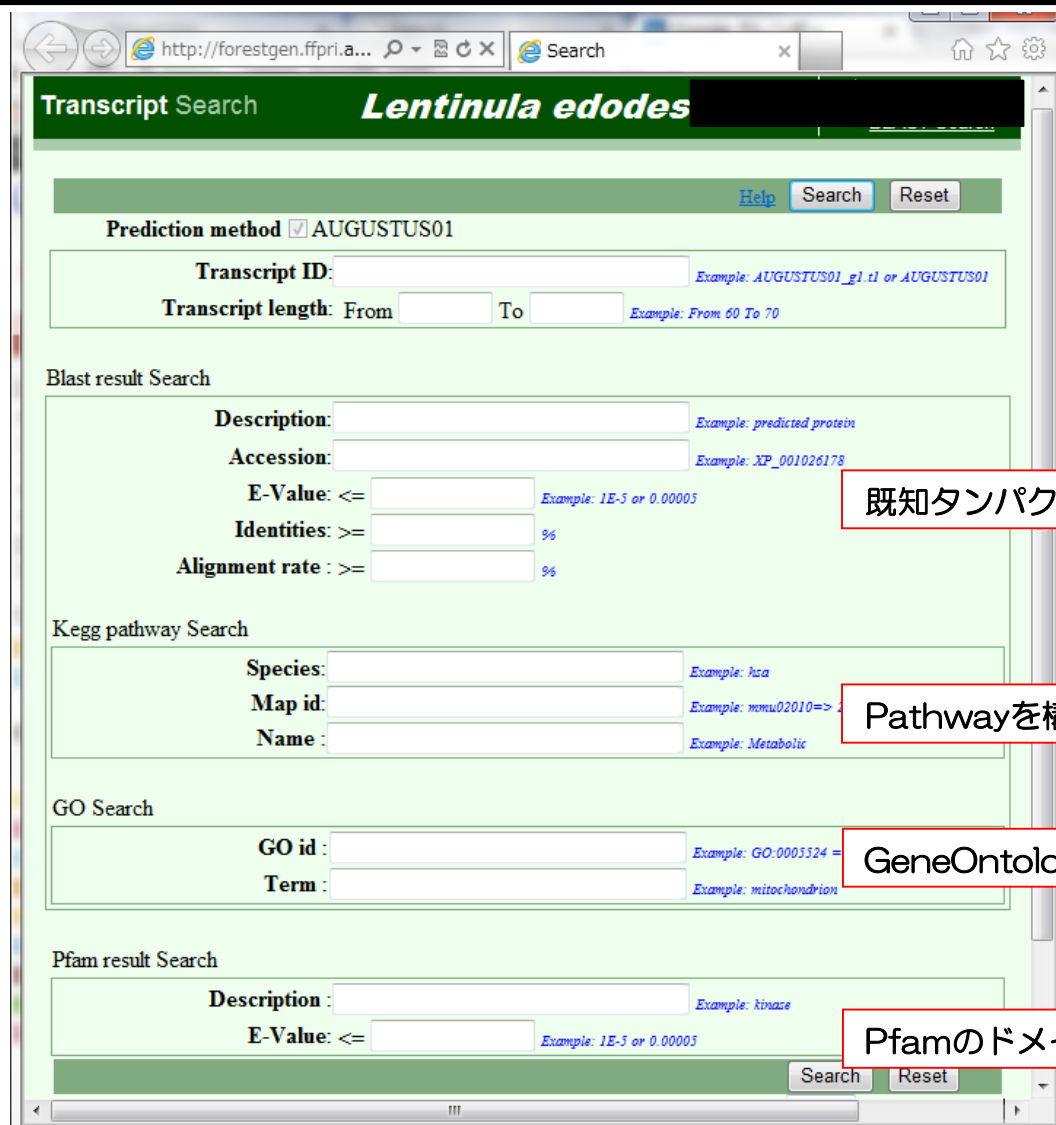


3. GASの主な特徴 NCBInrを利用したリッチアノテーション

- 予測遺伝子の核酸配列を全生物種の登録蛋白質（NCBI Nr）へのBlastxを行い、類似性を示したすべての蛋白質情報と類似度を検索対象のアノテーションとして格納します。
- 類似性を示したすべての蛋白質ごとに、NCBIの”Gene”データベースからKEGGとGeneOntologyのアノテーションを抽出して、蛋白質と関連づけて格納することにより、予測遺伝子とKEGG・GOの関係を作成します。
- 予測遺伝子のアミノ酸配列から、Pfam検索を行い、すべての結果を格納します。



リッチアノテーション 検索条件入力画面



Transcript Search **Lentinula edodes**

Help Search Reset

Prediction method AUGUSTUS01

Transcript ID: Example: AUGUSTUS01_g1.t1 or AUGUSTUS01

Transcript length: From To Example: From 60 To 70

Blast result Search

Description: Example: predicted protein

Accession: Example: XP_001026178

E-Value: <= Example: 1E-5 or 0.00005

Identities: >= %

Alignment rate: >= %

Kegg pathway Search

Species: Example: hsa

Map id: Example: mmu02010=>

Name: Example: Metabolic

GO Search

GO id: Example: GO:0005524 =

Term: Example: mitochondrion

Pfam result Search

Description: Example: kinase

E-Value: <= Example: 1E-5 or 0.00005

Search Reset

既知タンパク質を指定して、類似性のある予測遺伝子を検索

Pathwayを構成するタンパク質に類似性のある予測遺伝子を検索

GeneOntologyで予測遺伝子を検索

Pfamのドメインを持つ予測遺伝子を検索

リッチアノテーション 検索結果一覧画面

Transcript Listing Build: 1
2011/12/15 Main Menu
BLAST Search

Items 1-20 of 13089 Page: 1 / 655 >>

Transcript ID	Best-Hit protein		Best-Hit of KEGG-annotated protein(s)		Best-Hit of GO-annotated protein(s)			Pfam Hit(s)			
	Protein	E-Value /Score	Protein	Pathway/Name	Protein	GO/Term	E-Value /Score	Model	E-Value /Score	Description	
AUGUSTUS01_g1.tl GB	>gi254254100ref ZP_04947417.1 Phospholipase C [Burkholderia dolosa AUO158] gi124898745 gb EAY70588.1 Phospholipase C [Burkholderia dolosa AUO158]	0.16 41	>gi78778800ref YP_396912.1 malate:quinone oxidoreductase [Prochlorococcus marinus str. MIT9312] gi123554568 sp Q31CB9.1 MQO_PROM9 RecName: Full=Probable malate:quinone oxidoreductase; AltName:Full=MQO; AltName: Full=Malate dehydrogenase [quinone] gi78712299 gb ABB49476.1 Malate dehydrogenase (acceptor) [Prochlorococcus marinus str. MIT9312]	pmi00620 Pyruvate metabolism	2.3 37						
AUGUSTUS01_g2.tl GB	>gi170100753ref XP_001881594.1 predicted protein [Laccaria bicolor S238N-H82] gi164643553 gb EDR07805.1 predicted protein [Laccaria bicolor S238N-H82]	2E-7 61	>gi68535782ref YP_250487.1 dihydroipoamide acetyltransferase [Corynebacterium jeikeium K411] gi68263381 emb CAI36869.1 dihydroipoamide succinyltransferase [Corynebacterium jeikeium K411]	cjk01110 Biosynthesis of secondary metabolites cjk00020 Citrate cycle (TCA cycle) cjk01120 Microbial metabolism in diverse environments cjk01100 Metabolic pathways cjk00310 Lysine degradation	0.17 41		Hemocyanin_M	0.057 12	Hemocyanin, copper containing domain		
AUGUSTUS01_g3.tl GB	>gi170112037ref XP_001887221.1 predicted protein [Laccaria bicolor S238N-H82] gi164637782 gb EDR02064.1 predicted protein [Laccaria bicolor S238N-H82]	1E-101 372	>gi170112037ref XP_001887221.1 predicted protein [Laccaria bicolor S238N-H82] gi164637782 gb EDR02064.1 predicted protein [Laccaria bicolor S238N-H82]	lbc04111 Cell cycle - yeast lbc04113 Meiosis - yeast	1E-101 372		MIT	5.2 7	MIT (microtubule interacting and transport) do		
AUGUSTUS01_g4.tl GB	>gi336369299 gb EGN97641.1 hypothetical protein SERLA73DRAFT_161591 [Serpula lacrymans var.lacrymans S7.3] gi336382084 gb EGO23235.1 hypothetical protein	3E-70 270	>gi326912605ref XP_003202639.1 PREDICTED: LOW QUALITY PROTEIN: eukaryotic translation initiationfactor 5B-like [Meleagris gallopavo]	mcp03013 RNA transport	0.009 46		ORC2	2.6E-57 194	Origin recognition complex subunit 2		

リッチアノテーション BLASTX結果画面

e99.t1 - Mozilla Firefox

Searching.....done

Best-Hit of KEGG-annotated protein [KEGG](#)

[gi|219118732|ref|XP_002180133.1](#) predicted protein [Phaeodactylum tricornutum CCAP 1055/1]
[gi|217408390|gb|EEC48324.1](#) predicted protein [Phaeodactylum tricornutum CCAP 1055/1]

map id	Name
09040	Spliceosome

Best-Hit of GO-annotated protein [GO](#)

[gi|40807485|ref|NP_036601.2](#) pre-mRNA-processing factor 6 [Homo sapiens]
[gi|24212088|sp|O94906.1](#) PRP6 pre-mRNA processing factor 6; AltName: Full=Pre-mRNA-processing factor 6; AltName: F
Short=U5-102 kDa protein
[gi|7658291|gb|AAAF66128.1](#) AF221842.1 U5 snRNP-associated 102 kDa protein [Homo sapiens]
[gi|4164166|db|BAA37140.1](#) unnamed protein product [Homo sapiens]
[gi|1290451|gb|AAH01666.1](#) PRP6 pre-mRNA processing factor 6 homolog (S. cerevisiae) [Homo sapiens]
[gi|55664585|emb|CAH71582.1](#) PRP6 pre-mRNA processing factor 6 homolog (S. cerevisiae) [Homo sapiens]
[gi|55968583|emb|CAC16610.2](#) PRP6 pre-mRNA processing factor 6 homolog (S. cerevisiae) [Homo sapiens]
[gi|56203967|emb|CA21906.1](#) PRP6 pre-mRNA processing factor 6 homolog (S. cerevisiae) [Homo sapiens]
[gi|119595580|gb|EAW75174.1](#) chromosome 20 open reading frame 14, isoform CRA_a [Homo sapiens]
[gi|261858980|db|BA146012.1](#) PRP6 pre-mRNA processing factor 6 homolog [synthetic construct]

GO	Term	aspect
GO:0005681	spliceosomal complex	Component
GO:0005622	intracellular	Component

Hit list by Species:

gi	Species	Score	E Value
gi 217408390 gb EEC48324.1	predicted protein	1212	0
gi 224002959 ref XP_002291151.1	RNA splicing fact...	1162	0
gi 298705024 emb CBJ28499.1	conserved unknown pro...	815	0
gi 301106837 ref XP_002902501.1	pre-mRNA-processi...	812	0
gi 255553813 ref XP_002517947.1	pre-mRNA splicing...	805	0
gi 40807485 ref NP_036601.2	pre-mRNA-processing f...	798	0
gi 261858980 db BA146012.1	PRP6 pre-mRNA process...	798	0
gi 242083614 ref XP_002442232.1	hypothetical prot...	797	0
gi 73992767 ref XP_534481.2	PREDICTED: similar to...	797	0
gi 149638068 ref XP_001507859.1	PREDICTED: simila...	796	0
gi 118100678 ref XP_417426.2	PREDICTED: similar t...	795	0
gi 148222496 ref NP_001087374.1	PRP6 pre-mRNA pro...	795	0
gi 119331084 ref NP_001073234.1	pre-mRNA-processi...	795	0
gi 71895939 ref NP_001025642.1	PRP6 pre-mRNA proc...	794	0
gi 21539655 ref NP_598462.1	pre-mRNA-processing f...	794	0
gi 114052114 ref NP_001039840.1	pre-mRNA-processi...	792	0
gi 168004776 ref XP_001755087.1	predicted protein...	791	0
gi 114683093 ref XP_001152053.1	PREDICTED: hypoth...	791	0
gi 225446942 ref XP_002267416.1	PREDICTED: hypoth...	790	0
gi 224078337 ref XP_002197944.1	PREDICTED: simila...	790	0
gi 18071414 gb AA158273.1	AC068923_15 putative pre...	790	0
gi 125532522 gb EAY79087.1	hypothetical protein O...	790	0
gi 16326551 gb AA488558.1	C20orf14 homolog (H. sa...	789	0

Hit list by Species:

Species	Score	E Value
Phaeodactylum tricornutum CCAP 1055/1	(556484)	
Thalassidaira pseudonana CCMP1335	(298543)	
Ectocarpus siliculosus	(2880)	
Phytophthora infestans T30-4	(403677)	
Rhizinus communis	(3988)	
Homo sapiens	(9606)	
[synthetic construct]	(32630)	
Sorghum bicolor	(4558)	
Canis lupus familiaris	(9615)	
Omithorhynchus anatinus	(9258)	
Gallus gallus	(915)	
Xenopus laevis	(8364)	
Rattus norvegicus	(9913)	
Xenopus (Silurana) tropicalis	(8364)	
Mus musculus	(10090)	
Bos taurus	(9913)	
Physcomitrella patens subsp. patens	(145481)	
Pan troglodytes	(9598)	
Vitis vinifera	(29760)	
Taeniopygia guttata	(59729)	
Oryza sativa Japonica Group	(39947)	
Oryza sativa Indica Group	(39946)	
Danio rerio	(7955)	

1 GOやKEGGのデータが存在する蛋白質のTop Hitを提示します

3 GOとKEGGのデータが存在するマーククリックでGOとKEGGの詳細情報を参照

2 生物種ごとにTop Hitの蛋白質を提示

4. GASのその他の特徴

●ゲノムブラウザの構築

GBrowse (※1) を構築し、ゲノムコンティグと予測遺伝子とアノテーションを参照可能とします。

●予測遺伝子に対するコメント機能

各予測遺伝子について、コメントを登録・編集・検索することができます。

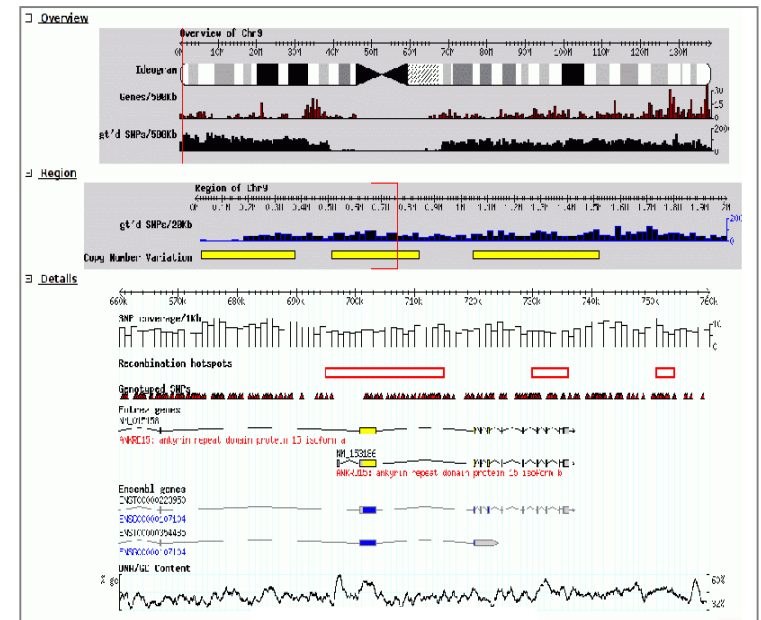
●Blast検索機能

ゲノム配列へのBlast検索

予測遺伝子のアミノ酸配列に対するBlast検索

※1：GBrowseは、フリーのゲノムブラウザです。

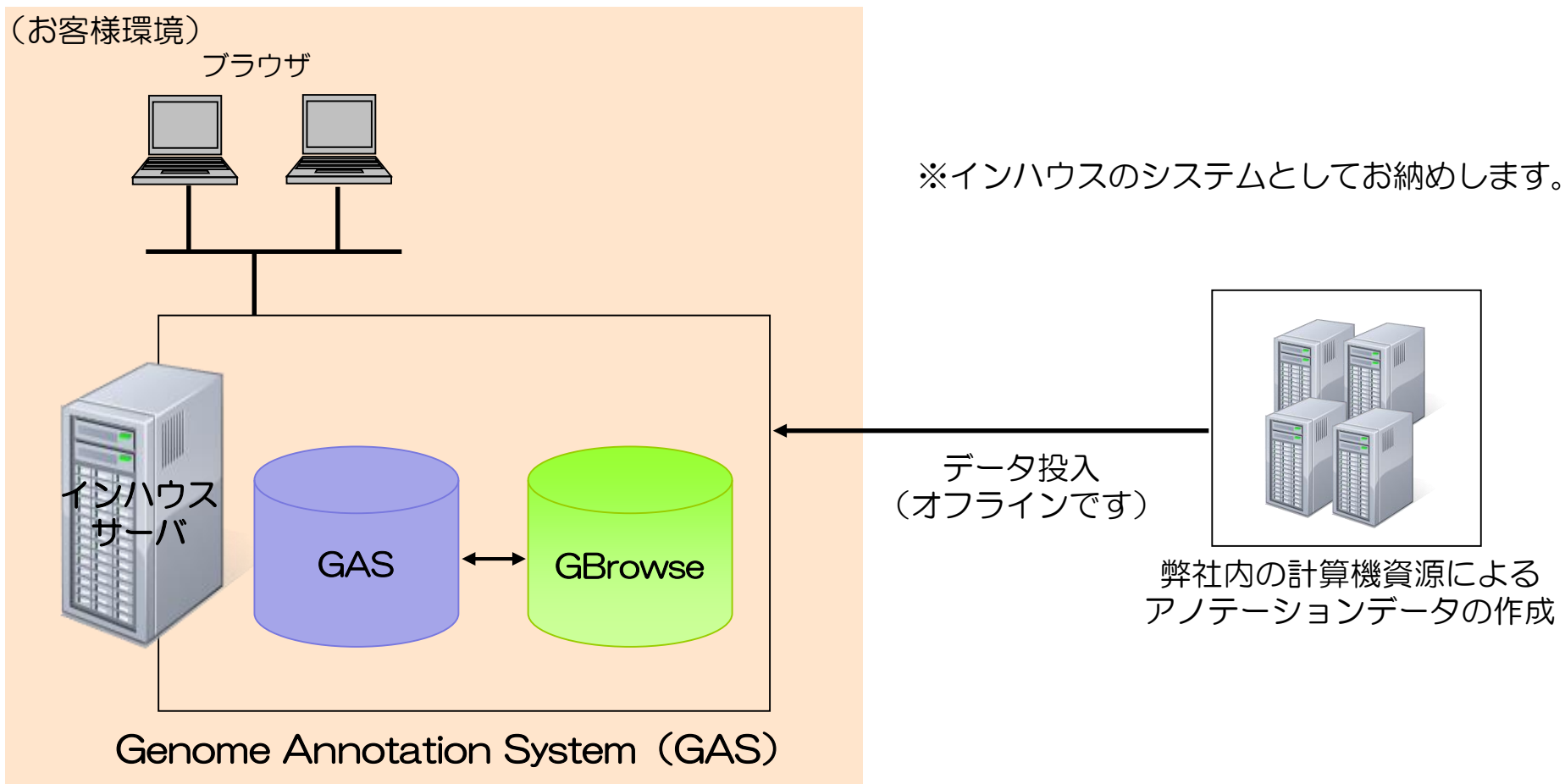
現在、GBrowseには環状に表示する機能はありません。



GBrowse画面

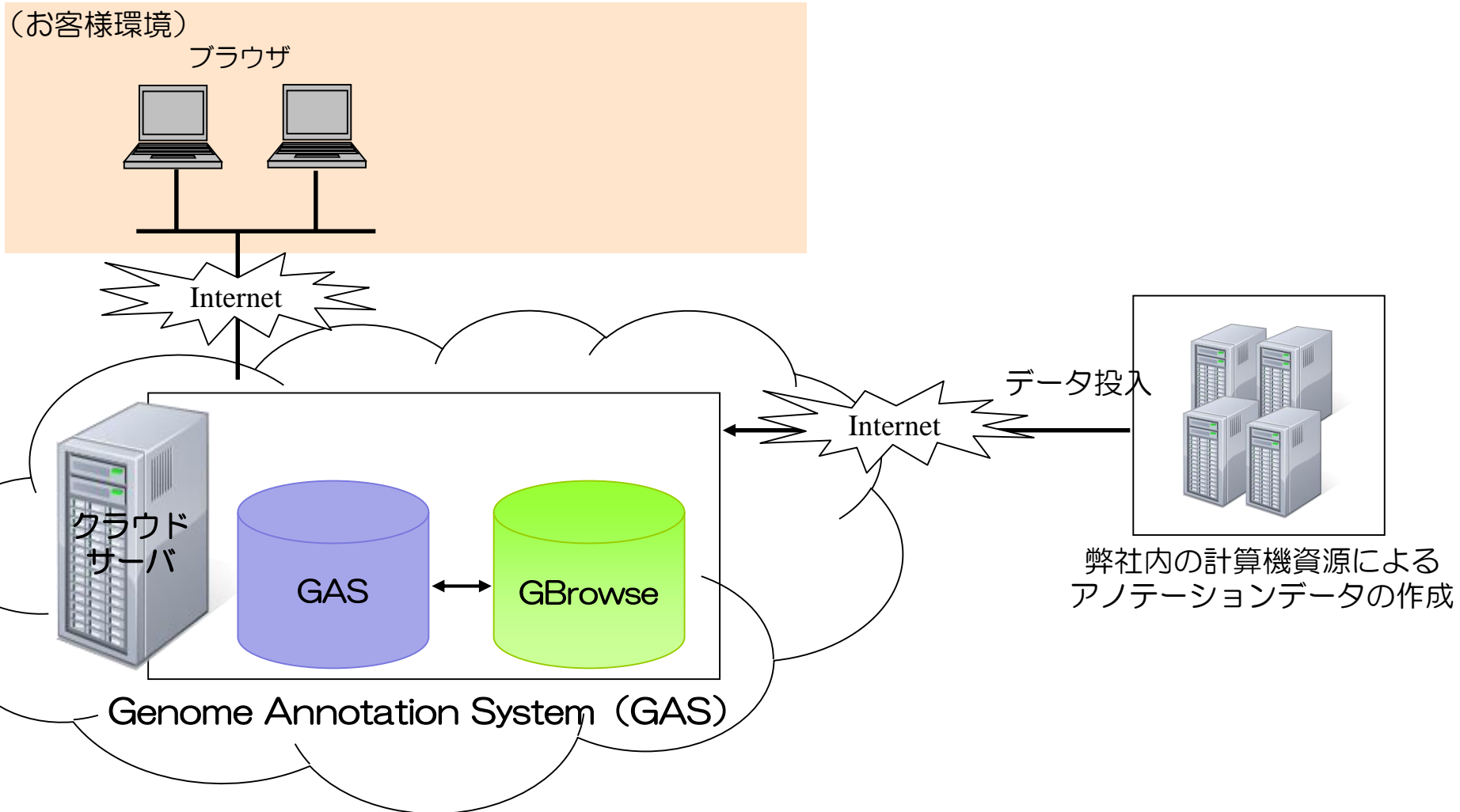
5. GASの構成とアノテーション作成の関係図

- イメージ1 受託でリッチアノテーションを作成し、お客様のマシンにインストールする場合



5. GASの構成とアノテーション作成の関係図

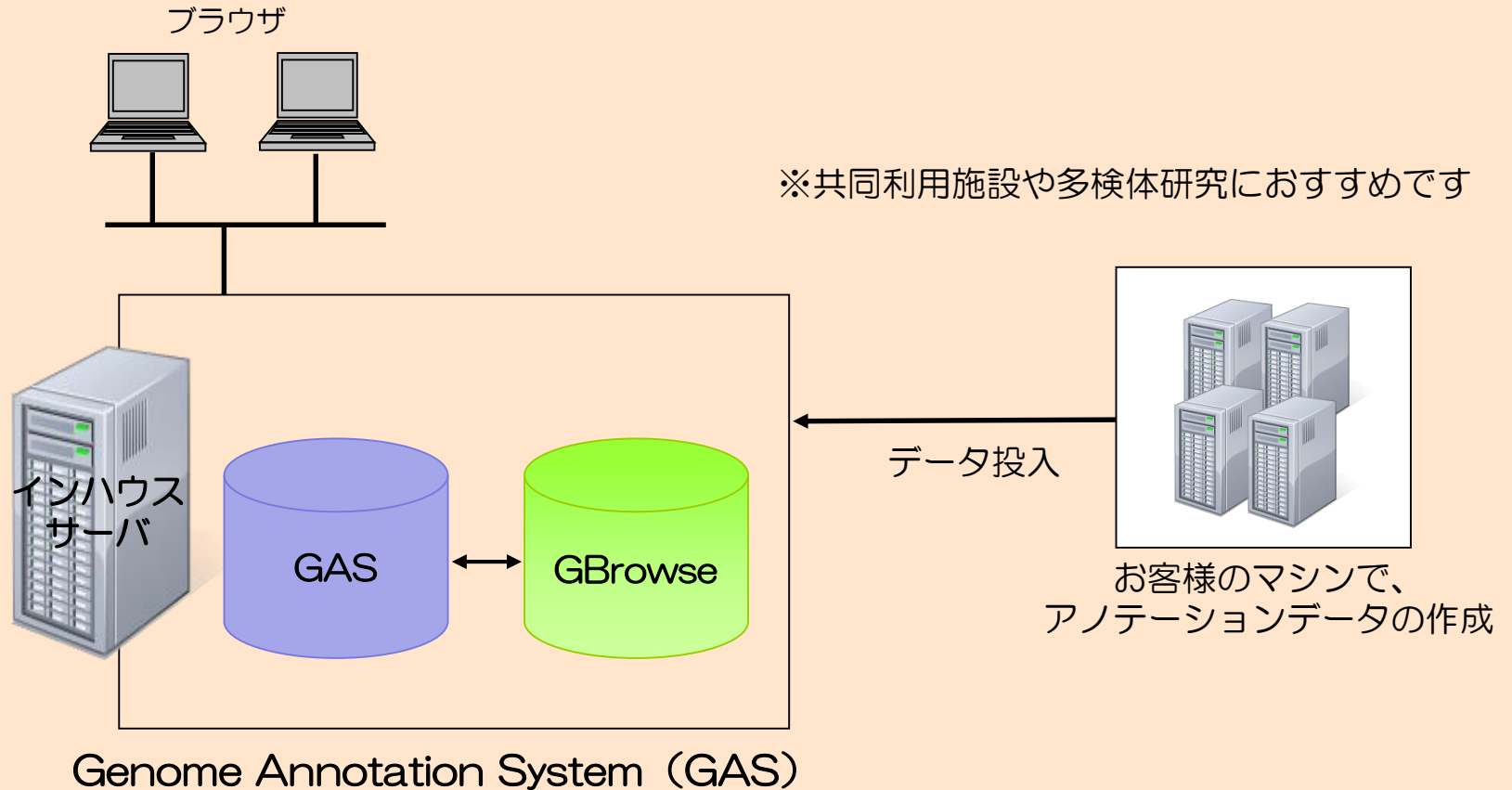
- イメージ2 受託でリッチアノテーションを作成し、クラウドの仮想マシンにインストールする場合



5. GASの構成とアノテーション作成の関係図

- イメージ3 お客様の環境にリッチアノテーション作成システムを提供し、お客様のマシンで参照・検索システムが動作します。

(お客様環境)



6. GASのWebインターフェース



The diagram illustrates the workflow of the ForestGEN web interface. It starts with the ForestGEN homepage, which provides information about the Lentinula edodes genome. A user can search for transcripts, leading to a detailed transcript page for AUGUSTUS01. This page provides information about the transcript, including its description, accession number, and alignment. The transcript page then links to a genome browser showing the Lentinula edodes genome. A separate path shows the BLAST Search Tool interface, which allows users to search for sequences. The BLAST Search Tool interface includes options for the BLAST program, database, and query sequence. The BLAST Search Result page displays the search results, including the sequence and alignment information.

【納入実績】

- （独）森林総合研究所様の「ForestGEN」からシイタケゲノム用のユーザインターフェースが公開されています。BLAST検索画面から、DNA塩基配列やアミノ酸配列情報を入力し検索すると、その配列と相同性の高いシイタケのDNA塩基配列、遺伝子、アミノ酸配列等を調べることができます。また、シイタケ遺伝子の転写産物（transcript）の検索画面から、機能や構造の特徴等について詳細な情報を調べることができます。ご参考にしていただければ幸いです。

http://forestgen.ffpri.affrc.go.jp/ja/info_le.html

- 東京農工大学様、その他研究機関

7. データ更新と年間サポート

●データ更新

弊社の計算機資源で、GASのアノテーション情報を更新しご提供いたします。
アノテーション情報の更新は、別途費用にて承ります。

●GASの納入初年度の年間サポートについて

(納入初年度)

お問合せ窓口の設置

不具合が発生した場合の原因一時切り分け

オフサイトによる障害復旧作業

(瑕疵による障害は、オンサイト或いはオフサイトにて復旧作業にあたります)

※ 瑕疵による障害復旧作業を除くオンサイト作業は、都度お見積のうえ承ります。

●納入次年度以降の年間サポートについて

納入初年度と同様のサポート内容で次年度以降の年間サポートを別途費用にて承ります
(任意契約)。

●リッチアノテーションのデータ作成システムのサポートについて

リモートメンテナンスが可能であれば、オンサイトと同様のリッチなサポートを提供します。

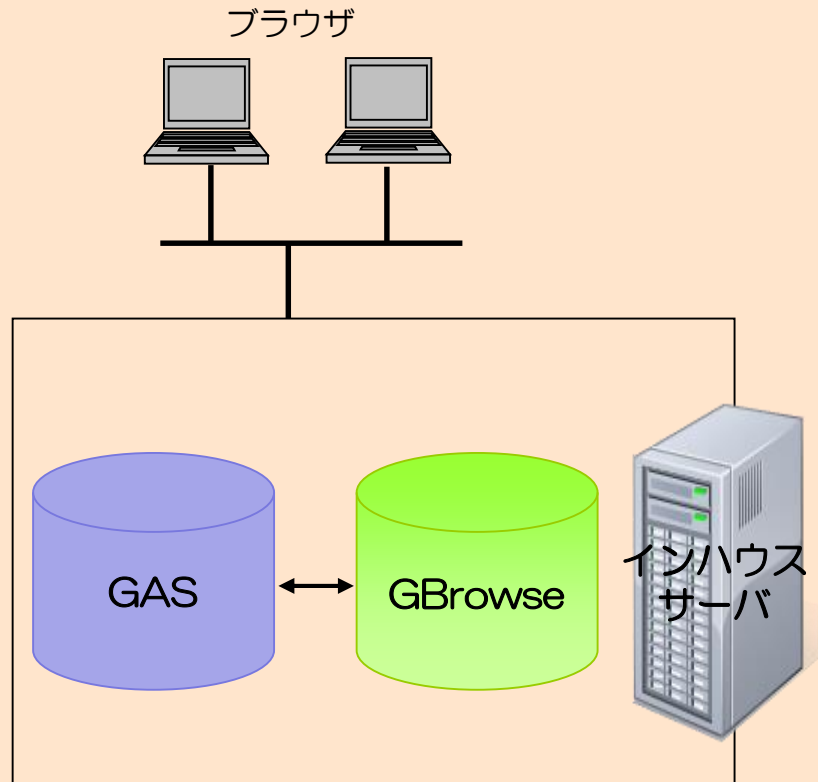
8. GASの機能一覧

項目名	内容	オプション
アノテーション	<ul style="list-style-type: none"> ● 予測遺伝子と全NCBI NrとのBlast結果とDescription ● Best-Hit protein (NCBI Nr) ● Pfam Hit ● Best-Hit of KEGG-annotated protein ● Best-Hit of GO-annotated protein 	(例) Reactome Conserved Domain ゲノム比較
全BlastX結果による リッチなアノテーション		
検索機能	<ul style="list-style-type: none"> ● アノテーションに対する条件検索 ● ユーザコメントに対する条件検索 ● Blast検索 (Contig、予測遺伝子) 	
ビューア	<ul style="list-style-type: none"> ● ゲノムブラウザ <ul style="list-style-type: none"> - GBrowse - Genome Contig と予測遺伝子を表示 ● Transcript Listing (予測遺伝子一覧表示画面) <ul style="list-style-type: none"> - 予測遺伝子と付与されたアノテーションの一覧を表示 ● Transcript Detail View (予測遺伝子の詳細画面) <ul style="list-style-type: none"> - シークエンスとアノテーションの表示 	
その他	<ul style="list-style-type: none"> ● ユーザコメント機能 <ul style="list-style-type: none"> - 登録・追加・削除 ● 配列データダウンロード機能 	

9. 公開サーバとして活用することができます

●ゲノム解析支援システムとして、共同研究先との共有はもとより、公開用のシステムとしてもご利用いただけます。

(お客様環境)



Genome Annotation System (GAS)

※サーバへの不正侵入や改竄などを監視する“サポートサービス”をご用意しています。
詳しくは、担当者までお問合せ下さい。
※サーバを一般公開する場合は、サポートサービスをお薦めいたします。

- 共同研究者の間で共用利用
- 成果の公開としてサーバを公開



u14329500 www.fotosearch.jp