



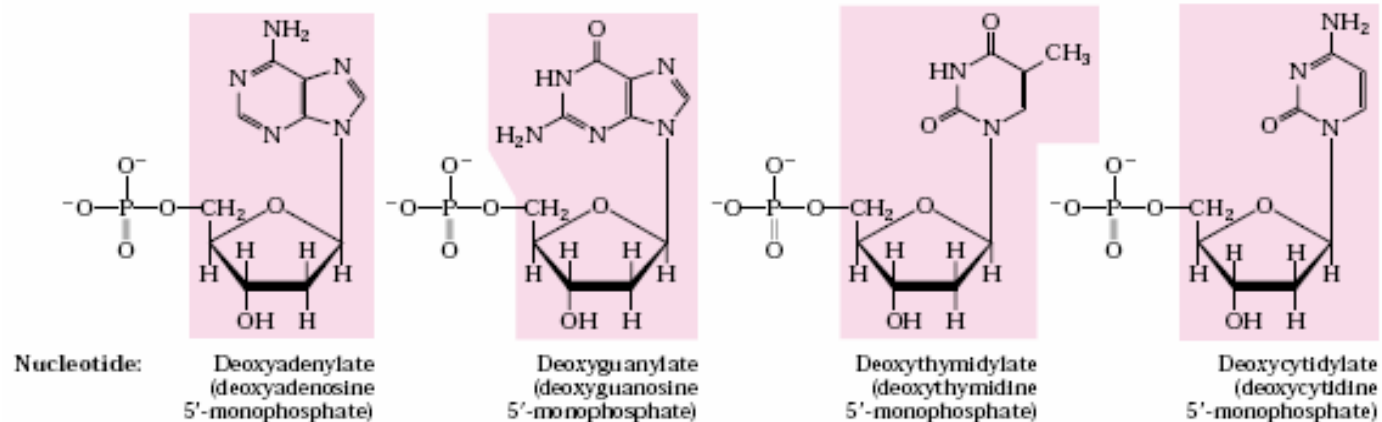
GenBank

Jenny Consuelo Núñez
Ximena Carolina Pérez

POSTGRADO INTERFACULTADES DE MICROBIOLOGÍA
BIOINFORMÁTICA

CONCEPTOS BÁSICOS

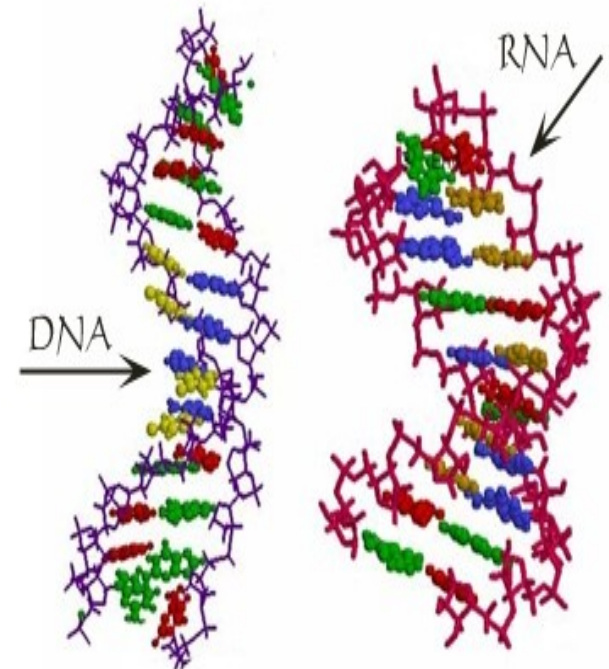
NUCLEÓTIDOS



- Pentosa: Ribosa o desoxirribosa
- Base nitrogenada: Purinas (A-G)
Pirimidinas (C-T-U)
- Fosfato

CONCEPTOS BÁSICOS

- **Ácidos nucleicos: polímeros de ribonucleótidos o desoxirribonucleótidos unidos por enlaces fosfodiéster. (DNA y RNA)**
- **Gen: es una secuencia lineal de nucleótidos que contiene la información necesaria para la síntesis de una macromolécula con función celular específica**



Samanburður á sameindaröðun DNA og RNA

Que es GenBank?

- Es la base de datos mas conocida de secuencias DNA públicamente disponibles. hay aproximadamente 65,369,091,950 bases.
- organizaciones intercambian datos diariamente.
- GenBank y sus colaboradores reciben secuencias genéticas producidas en laboratorios de todo el mundo, procedentes de más de 100.000 organismos distintos.

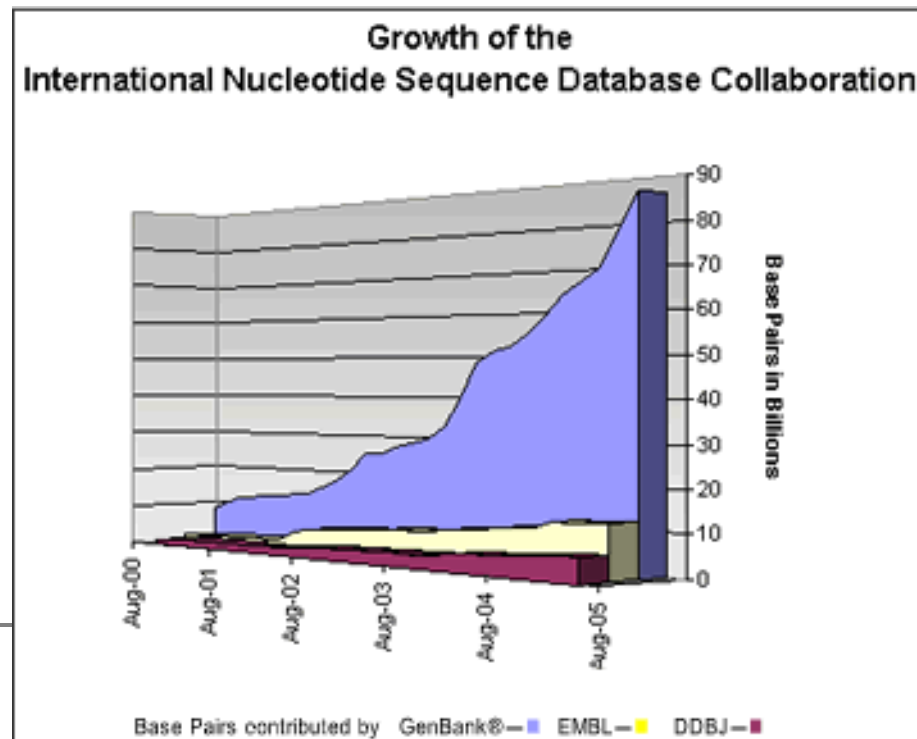
Aceso NCBI

<http://www.ncbi.nlm.gov>

GenBank	➡	E.U.U
EMBL	➡	Europea
DDBJ	➡	Japon

Bases datos GenBank

- Crecimiento exponencial.
- Base datos internacional excede 100 gigabases.



Revisiones secuencias

- Las puede hacer actualizaciones, revisiones entradas al GenBank.
 - Utilidad sirve buscar secuencias de nucleótidos.
 - a la vez comparar alineamientos a través del BLAST
-

Actualizaciones

- GenBank recibe a menudo las actualizaciones a las sumisiones en un formato.
 - La información siguiente proporciona los métodos diferentes para someter las actualizaciones a GenBank que asegurará que su actualización se procesa rápidamente y correctamente.
 - Las actualizaciones deben ser directamente emailed a gb-admin@ncbi.nlm.nih.gov
Bankit Update
SequinMacroSend.
-

Acceso al GenBank

<http://www.ncbi.nlm.gov>

- Esta disponible para investigar NCBI por varios métodos.
 - Los GenBank nucleotide archivos están disponibles en las divisiones CoreNucleotide, dbEST, o dbGSS y pueden investigarse juntos o independientemente en Entrez.
-

-
- **NCBI** está desarrollando las nuevas herramientas continuamente y existentes reforzando acceso a **GenBank**.
 - NCBI Anuncie la lista del correo electrónico, lea las Noticias de NCBI, Web.
 - Los Datos de la Sucesión sometiendo a GenBank.
 - Los datos intercambian entre GenBank, EMBL y DDBJ ocurre diariamente. Pueden guardarse datos de la sucesión sometidos de antemano de publicaciones confidencial.
 - Anualmente se reúnen las tres bases de datos con el fin de corregir errores técnicos.
-

Herramienta de búsqueda

Alineación local BLAST

- La Herramienta de Búsqueda de Alineación Local Básica (BLAST) encuentra regiones de similitud local entre las secuencias.
 - El programa compara nucleótido o secuencias de la proteína a las bases de datos de la sucesión.
 - **BLAST** puede usarse para inferir las relaciones funcionales y evolutivas entre las secuencias así como ayuda identifica a los miembros de familias del gen.
-

-
- National Center for Biotechnology Information (NCBI)
 - Molecular Biology Laboratory (EMBL)
 - DNA Data Bank of Japan (DDBJ)
-

NCBI HomePage - Windows Internet Explorer

http://www.ncbi.nlm.nih.gov/ 1

Archivo Edición Ver Favoritos Herramientas Ayuda

Buscar en el Web

NCBI HomePage

Página principal

NCBI National Center for Biotechnology Information

National Library of Medicine National Institutes of Health

PubMed All Databases BLAST OMIM Books TaxBrowser Structure

Search All Databases for Go

SITE MAP
Alphabetical List
Resource Guide

About NCBI
An introduction to NCBI

GenBank
Sequence submission support and software

Literature databases
PubMed, OMIM, Books, and PubMed Central

Molecular databases
Sequences, structures, and taxonomy

Genomic

What does NCBI do?

Established in 1988 as a national resource for molecular biology information, NCBI creates public databases, conducts research in computational biology, develops software tools for analyzing genome data, and disseminates biomedical information - all for the better understanding of molecular processes affecting human health and disease. [More...](#)

Hot Spots

- Assembly Archive
- Clusters of orthologous groups
- Coffee Break, Genes & Disease, NCBI Handbook
- Electronic PCR
- Entrez Home
- Entrez Tools
- Gene expression omnibus (GEO)
- Human genome resources
- Influenza Virus Resource

New dbGaP
NCBI's dbGaP Genome Wide Association Database

NCBI's **dbGaP** (database of Genotype and Phenotype) provides data from Genome Wide Association (GWA) studies. The resource is intended to help elucidate the link between genes and disease. For each study, users have access to detailed information about the phenotypic variables measured and pre-computed associations between subjects' phenotypes and genotypes. Click here to read the [press release](#). To read more about GWA projects, see NCBI's [GWA resource page](#)

Listo

Inicio NCBI HomePage - Wi...

1. Acceso a la página principal de NCBI:
<http://www.ncbi.nlm.nih.gov>

2. Acceso al Link GenBank

GenBank Overview

NCBI GenBank Overview

PubMed Entrez BLAST OMIM Books Taxonomy

Search Entrez for Go

NCBI
SITE MAP

Submit to GenBank 3

Updates 4

Search GenBank
[Entrez Nucleotide](#)
[BLAST](#)

International sequence databases exceed 100 gi

In August 2005, the INSDC announced the DNA sequence 100 gigabases. GenBank is proud of its contributions toward thank all the scientists who have worked through the submit GenBank and made their sequence data available to the world [press release](#).

Growth of the International Nucleotide Sequence Database Collaboration

Time Period	Approximate Bases (Millions)
1	65
2	68
3	72
4	75
5	80
6	85
7	90

3. Envío de datos a GenBank: Como someter secuencias.

4. Actualizaciones: Como enviar actualizaciones de la información publicada



NCBI

GenBank Overview

[PubMed](#)[Entrez](#)[BLAST](#)[OMIM](#)[Books](#)[Taxonomy](#)

Search

Entrez



for

Go

[NCBI](#)[SITE MAP](#)[Submit to
GenBank](#)[Updates](#)[Search GenBank](#)[Entrez Nucleotide](#)[BLAST](#)

International sequence databases exceed 100 gi

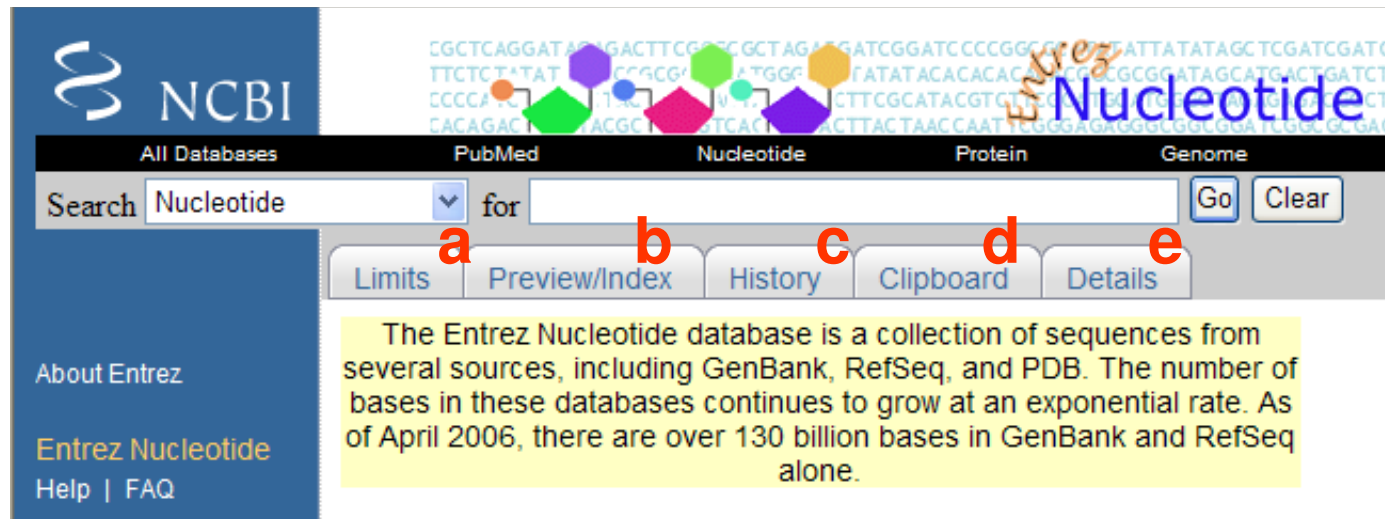
In August 2005, the INSDC announced the DNA sequence 100 gigabases. GenBank is proud of its contributions toward thank all the scientists who have worked through the submit: GenBank and made their sequence data available to the world [press release](#).

Growth of the
International Nucleotide Sequence Database Collaboration



5

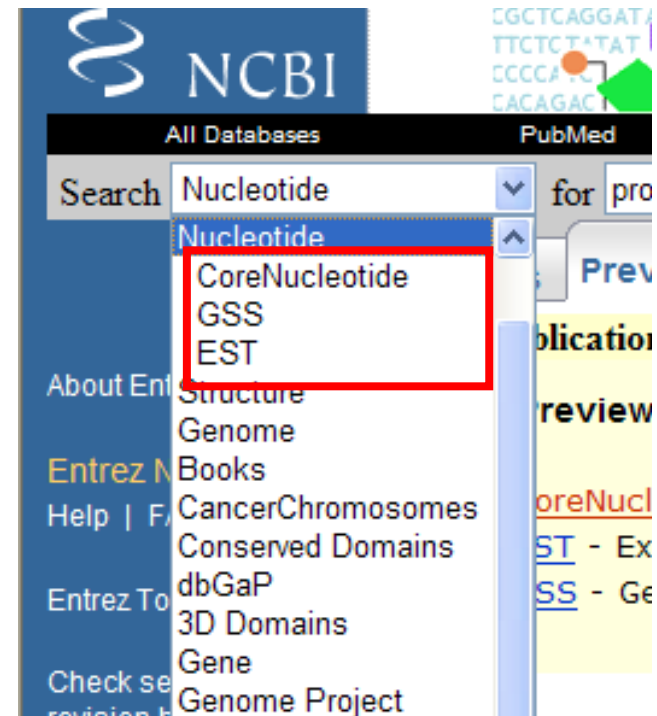
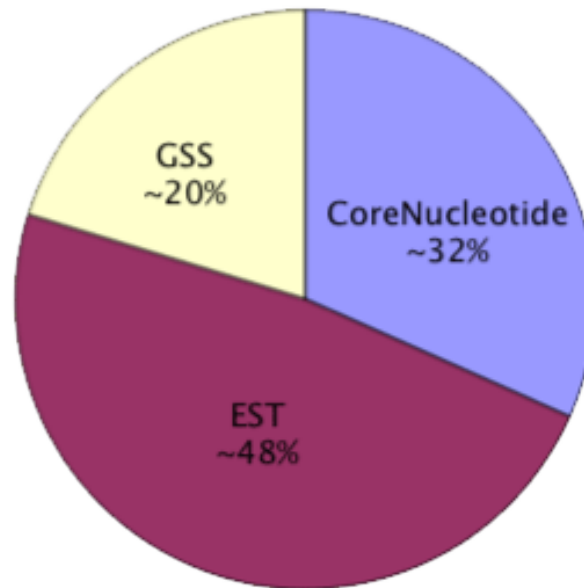
Entrez Nucleotide

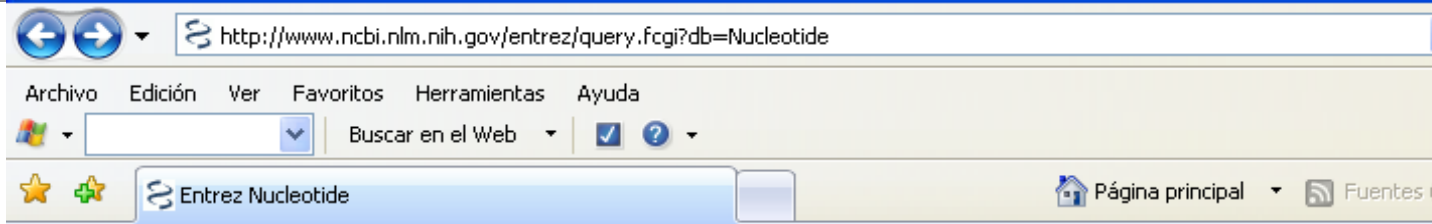


- a. Límites: Restricción de búsqueda
- b. Índice: Lista alfabética de términos de búsqueda dentro de la base de datos
- c. Historia: Expediente de búsquedas durante una sesión
- d. "Clipboard": lugar temporal para guardar resultados.
- e. Detalles: Muestra las estrategias usadas para la búsqueda

Subconjuntos de GenBank

- CoreNucleotide
- EST
- GSS







CGCTCAGGATAGACTTCGCGCTAGATCGGATCCCCGGCGATTATTATAGCTCGATCGATCT
TTCTCTATATCCGCGATGGGATATACACACACACCGCGGATAGCATGACTGATCTA
CCCCATCTTCTCGCATACGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
CACAGACTACGCT

All Databases PubMed Nucleotide Protein Genome Structure

Search Nucleotide for **propanediol AND Clostridium**

Limits Preview/Index History Clipboard Details

About Entrez

Entrez Nucleotide
Help | FAQ

Entrez Tools

Check sequence
revision history

LinkOut

My NCBI (Cubby)

Related resources
BLAST

Reference sequence
project

Search for Genes

Submit to GenBank

The Entrez Nucleotide database is a collection of sequences from several sources, including GenBank, RefSeq, and PDB. The number of bases in these databases continues to grow at an exponential rate. As of April 2006, there are over 130 billion bases in GenBank and RefSeq alone.

Human Genome

Explore [human genome resources](#) or browse the human genome sequence using the [Map Viewer](#).

Building the human genome

The Human Genome Reference DNA Sequence was completed in April 2003. The current version is listed as a build number on the [Genome View](#) page and includes an accompanying set of [statistics](#) and [release notes](#).

***Homo sapiens (human)* genome view** BLAST search the human genome

[Build 36.2 statistics](#) [Switch to previous build](#)



Entrez Nucleotide

 100% ▼

Información general

Search for

Limits Preview/Index History Clipbo

Display Show Send to Hide: ☐ sequence ☐ all but gene, CDS and mRN

Range: from to ☐ Reverse complemented strand Features:

☐ 1: [DQ901409](#). Reports *Clostridium* sp. I...[gi:113171052]

[Features](#) [Sequence](#)

LOCUS DQ901409 781 bp DNA linear BCT 28-AUG-2006

DEFINITION *Clostridium* sp. IBUN 158B 1,3-propanediol dehydrogenase (dhaT) gene, partial cds.

ACCESSION DQ901409

VERSION DQ901409.1 GI:113171052

KEYWORDS .

SOURCE *Clostridium* sp. IBUN 158B

ORGANISM [Clostridium sp. IBUN 158B](#)
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae; Clostridium.

REFERENCE 1 (bases 1 to 781)

AUTHORS Quilaguy,D.M., Suarez,Z.R., Bernal,M., Aristizabal,F.A. and Montoya,D.

TITLE Direct Submission

JOURNAL Submitted (15-AUG-2006) Grupo de Bioprocesos y Bioprospeccion, Universidad Nacional de Colombia, Cra 30 No 45-03-Edificio Manuel Ancizar, Bogota D.C., Cundinamarca 14490, Colombia

FEATURES

Características

Universidad Nacional de Colombia, Cra 30 No 45-03-Edificio Manuel Ancizar, Bogota D.C., Cundinamarca 14490, Colombia

FEATURES

source

1..781

/organism="Clostridium sp. IBUN 158B"

/mol_type="genomic DNA"

/strain="IBUN 158B"

/isolation_source="tomato crops"

/db_xref="taxon:400088"

/country="Colombia"

/note="similar to Clostridium butyricum"

[gene](#)

<1..>781

/gene="dhaT"

[CDS](#)

<1..>781

/gene="dhaT"

/EC_number="1.1.1.202"

/function="3-hydroxypropionaldehyde + NADH = 1,3-propanediol + NAD(+)"

/note="oxidoreductase activity; iron ion binding; 1,3-propanediol oxidoreductase; 3-hydroxypropionaldehyde reductase"

/codon_start=1

/transl_table=11

/product="1,3-propanediol dehydrogenase"

/protein_id="ABI30641.1"

/db_xref="GI:113171053"

/translation="DYAGIETLVNALPPIVAVNTSAGTASELTRHCVLPNTKKKIKFV
IVSWRNLPVLSINDAMLMVKKPAGLTAATGMDVVTHAIEAYVSKDANPVTDAIAIQAI
KLISQNLQRSVALGENLEARENMAYASLLAGMAFNANLGYVHAMAHQLGGLYDMAHG
VANAMLLPHVERYNMISNPKKFADIAEFMGENISGLSVMEAAEKAINAMFRLSEDVGI
PKSLKEMGVKQEDFEHMAELALLDGNFASNPRKGNAKDIINI"

Secuencia

```
1,3-propanediol oxidoreductase; 3-hydroxypropionaldehyde
reductase"
/codon_start=1
/transl_table=11
/product="1,3-propanediol dehydrogenase"
/protein_id="ABI30641.1"
/db_xref="GI:113171053"
/translation="DYAGIETLVNALPPIVAVNTSAGTASELTRHCVLPNTKKIKFV
IVSWRNLPLVSINDAMLMVKKPAGLTAATGMDVVTHAIEAYVSKDANPVTDAIQA
KLISQNLRQSVLGENLEARENMAYASLLAGMAFNANLGYVHAMAHQLGGLYDMAHG
VANAMLLPHVERYNMISNPKKFADIAEFMGENISGLSVMEAAEKAINAMFRLSE
PKSLKEMGVKQEDFEHMAELALLDGNAFSNPRKGNAKDIINI"

ORIGIN
1 gattatgcag gaatagaaac acttgtcaat gcattgccac caatagtagc tgtaaatact
61 tctgcaggaa ctgctagtga attaactcgt catttgttat tgcctaatac aaaaaagaaa
121 ataaaaattt ttatagttag ctggagaaat ttgcctctag tatctataaa tgatgcaatg
181 cttatgggtca aaaaacctgc aggattaaca gcagctacag gaatggatgt tgtaacacat
241 gcaatagaag catatgtatc aaaagatgca aatccagtaa cagatgcttc agcaatacaa
301 gctattaaat taatttcaca aaatttaaga caatctgtag ctttaggaga aaatcttgaa
361 gcaagagaaa atatggctta tgcattcata ctagcaggaa tggcatttaa taatgcta
421 ttaggatatg tacatgcaat ggctcatcaa ttagggggac tgtatgatat ggcacatggt
481 gttgctaata caatgctatt accacatggt gaacgttata atatgatata aaatcctaag
541 aagtttgca atatagcaga atttatggga gaaaatatat ctggactttc tgtaatggaa
601 gcagcagaga aagccataaa tgcaatgttc aggccttcag aggatgttgg aattccgaaa
661 agtttaaaag agatgggagt gaaacaagaa gattttgagc atatggcaga actagctcct
721 ttagatggaa atgcatttag caatccaaga aaaggaaatg caaagatat tataaatatt
781 t

//

Disclaimer | Write to the Help Desk
NCBI | NLM | NIH
```

Formato

The screenshot shows the NCBI Nucleotide search interface. The search bar contains the text "Nucleotide" and "for". The "Display" dropdown menu is open, showing options: GenBank, GenBank(Full), INSDSeq XML, **FASTA** (highlighted with a red box), tnySeq XML, ASN.1, XML, Graph, GI List, Brief, and Summary. The search results for "Clostridium sp. I...[gi:113171052]" are displayed, showing a 781 bp DNA linear sequence. The "FEATURES" section is visible at the bottom.

NCBI Nucleotide

Search Nucleotide for [Go] [Clear]

Limits Preview/Index History Clipboard

Display GenBank Show 5 Send to Hide: ☐ sequence ☐ all but gene, CDS and mRNA features

Range: GenBank to end ☐ Reverse complemented strand Features: + Refresh

☐ 1: D Clostridium sp. I...[gi:113171052]

Feature tnySeq XML ASN.1 XML Graph GI List Brief Summary

LOCUS 781 bp DNA linear BCT 28-AUG-2006

DEFINITION Clostridium sp. IBUN 158B 1,3-propanediol dehydrogenase (dhaT)

ACCESSION

VERSION DQ901409.1 GI:113171052

KEYWORDS .

SOURCE Clostridium sp. IBUN 158B

ORGANISM Clostridium sp. IBUN 158B
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
Clostridium.

REFERENCE 1 (bases 1 to 781)

AUTHORS Quilaguy,D.M., Suarez,Z.R., Bernal,M., Aristizabal,F.A. and Montoya,D.

TITLE Direct Submission

JOURNAL Submitted (15-AUG-2006) Grupo de Bioprocesos y Bioprospeccion, Universidad Nacional de Colombia, Cra 30 No 45-03-Edificio Manuel Ancizar, Bogota D.C., Cundinamarca 14490, Colombia

FEATURES Location/Qualifiers

source 1..781

Listo