

# La Importancia de Internet 2 en Proyectos Genómicos

CUDI. Reunión de Primavera 2006

Oaxaca, Oax., del 3 al 5 de abril de 2006

César Bonavides Martínez

Victor del Moral Chávez

Romualdo Zayas Lagunas



# UNAM, Campus Morelos

Recientemente se formalizó el Campus Morelos de la UNAM

- IBT: Instituto de Biotecnología
- CCF: Centro de Ciencias Físicas
- CRIM: Centro Regional de Investigaciones Multidisciplinarias
- MAT: Unidad de Matemáticas
- CIE: Centro de Investigación en Energía
- CCG: Centro de Ciencias Genómicas

El Campus Morelos es el Campus foráneo más grande en investigación.

# Centro de Ciencias Genómicas

- 7 Programas de Investigación
  - Genómica Computacional
  - Ecología Genómica
  - Genómica Evolutiva
  - Dinámica Genómica
  - Ingeniería Genómica
  - Genómica Funcional de Eucariotes
  - Genómica Funcional de Procaríotes
- Licenciatura en Ciencias Genómicas  
(en conjunto con el Instituto de Biotecnología)

# Desarrollo de las Ciencias Genómicas

En 1999 se firmó el convenio del Proyecto “Desarrollo de las Ciencias Genómicas en México: el Genoma de *Rhizobium etli* como Sistema Modelo”

- Secuenciación del Genoma de *Rhizobium etli* (6.5 Mbp)
- Nodo Nacional de Bioinformática, EMBnet México (2000)
- Fundación de la Sociedad Mexicana de Ciencias Genómicas (2001)
- Cambio de enlace E1 (1.4 Mbps) a E3 (34 Mbps) (beneficio para Dependencias del Campus Morelos, 2002)
- Creación de la Licenciatura en Ciencias Genómicas

# *Rhizobium etli*. El primer genoma de un organismo completo secuenciado en México



# *Rhizobium etli*. El primer genoma de un organismo completo secuenciado en México



# *Rhizobium etli*. El primer genoma de un organismo completo secuenciado en México

**Rhizobium etli Data Base - Microsoft Internet Explorer**

File Edit View Favorites Tools Help

Address http://www.cifn.unam.mx/retdb/

Main Page Genome Project View Citing RetlIDB Documentation About RetlIDB

**R. etli DataBase**

CCG Centro de Ciencias Genómicas

Search: Chromosome Gene In By Object Go!

RFTLIDB is free for non-commercial use  
For commercial use it has to be licensed  
Read the disclaimer

The roots of leguminous plants associate with bacteria of the Rhizobium genus to generate a highly specialized structure, the nitrogen-fixing nodule. Bacterial cells in the nodule fix the atmospheric nitrogen and produce ammonium that is assimilated by the plant, in return, the plant supplies carbon compounds derived from photosynthesis, for the bacteria maintenance.

Many genes from both organisms are required for the establishment and optimal function of this symbiosis. The Rhizobium genome is structured in two types of replicons, the chromosome and several large plasmids. As a general rule, most of the genes involved in the symbiotic processes are plasmid borne and this plasmid is therefore referred to as the symbiotic plasmid or pSym.

6,530,234 bp of the complete Rhizobium etli CFN42 genome has been determined  
Please contact [RetlIDb-Team](#) for additional information

Please, send your comments to [Webmaster](#)

Done Internet

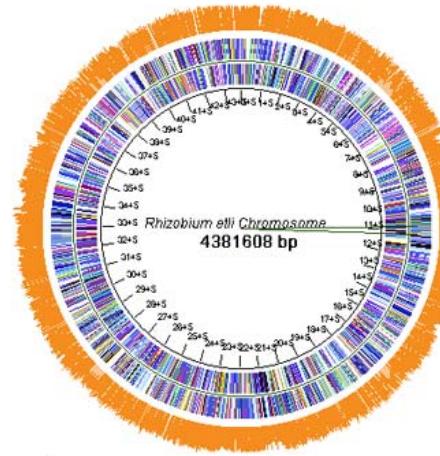
This screenshot shows the homepage of the Rhizobium etli Data Base. The page is framed by a Microsoft Internet Explorer window. At the top, there's a navigation bar with links for Main Page, Genome Project, View, Citing RetlIDB, Documentation, and About RetlIDB. Below the navigation is a search bar with dropdown menus for 'Chromosome' and 'Gene', and a 'Go!' button. To the right of the search bar is the 'R. etli DataBase' logo. Further down, there's a banner featuring a stylized illustration of plant roots and bacteria. Text on the page explains the symbiotic relationship between leguminous plants and Rhizobium bacteria, mentioning the nitrogen-fixing nodule and the symbiotic plasmid pSym. It also states that 6,530,234 bp of the genome has been determined and provides contact information for the RetlIDb-Team. The bottom of the page includes a 'Please, send your comments to Webmaster' link and standard browser footer links for 'Done' and 'Internet'.

# *Rhizobium etli*. El primer genoma de un organismo completo secuenciado en México

Main Page | Genome Project | View | Citing RetiDB | Documentation | About RetiDB

Search:



Genome : Rhizobium etli Chromosome  
Length : 4381608 base pairs  
Genes : 4126

## OBJECT DESCRIPTION

Gene :		Predicted Operon:	
Known Promoter :		Known Operon:	
Predicted Promoter :		Terminator :	
Regulatory Binding Site :			

# *Rhizobium etli*. El primer genoma de un organismo completo secuenciado en México

Rhizobium etli Data Base - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites Home Links Go

Address: http://www.cifn.unam.mx/retlidb/

Main Page Genome Project View Citing RetliDB Documentation About RetliDB

R. etli DataBase

CCG Centro de Ciencias Genómicas

Search: Chromosome Gene crp In By Object Go

350268 353602

Genes ypch00132 RE1CHH007980 dfp

TU RE1CHH007979 RE1CHH007981 356936 356938

353602 ypch00134 ypch00135 ypch00136

Genes ypch00133 RE1CHH007983 RE1CHH007984

TU RE1CHH007981 RE1CHH007982 360270

356936 RE1CHH007984

Genes hutUch yhch0084 yhch0085

TU RE1CHH007985 RE1CHH007986

Legend:

- Not classified regulator
- electron transport
- Transport of small molecules
- Hypothetical/Partial homology
- pantothenate
- Small Molecule Metabolism
- histidine

OBJECT DESCRIPTION

Gene :		Predicted Operon:	
Known Promoter :		Known Operon:	
Predicted Promoter :		Terminator :	
Regulatory Binding Site :			

Operon Name RE1CHH007986 Begin: 359077 End: 363157

Internet

# Nodo Nacional de Bioinformática, EMBnet México

- El Nodo EMBnet México ofrece sitios “espejo” de las bases de datos más importantes para análisis de secuencias y herramientas de bioinformática.
- Actualmente contamos con herramientas como:
  - BIOBASE
  - RSA-Tools
  - GETools

# Nodo Nacional de Bioinformática, EMBnet México

Infraestructura computacional disponible en el CCG:

- Cluster de 27 nodos duales de 2.4 GHz.
- 2 SunFire V880: 8GB RAM, 4 proc. Sparc @900 MHz
- SunFire V890: 16GB RAM, 4 proc. Sparc @1.2 GHz
- SunFire V890: 24GB RAM, 4 proc. Sparc @1.2 GHz (Genoma Humano)
- 4 SunFire V40Z: 8GB RAM, 4 AMD Opteron @2.4 GHz

# Nodo Nacional de Bioinformática, EMBnet México



# Nodo Nacional de Bioinformática, EMBnet México



# Licenciatura en Ciencias Genómicas

- Este año inicia la 4ta generación, aprox. 130 estudiantes (cap. de 160)
- Clases presenciales y por videoconferencia
- Los estudiantes de la LCG utilizan los recursos disponibles del Nodo EMBnet
- De igual forma requieren acceso a grandes bancos de información y a herramientas bioinformáticas que demandan poder de cómputo tanto en servidores locales como remotos

# Licenciatura en Ciencias Genómicas



# Licenciatura en Ciencias Genómicas



# RegulonDB. Base de Datos de Regulación Transcripcional y organización de operones en *Escherichia coli*

- Contiene la colección electrónica más completa de redes de regulación de un organismo.
- Los usuarios pueden obtener información relacionada con: Genes, Unidades de Transcripción, Promotores, Interacciones Reguladoras, etc.
- Actualmente cuenta con más de 3000 visitas mensuales provenientes de más de 30 países

# RegulonDB. Base de Datos de Regulación Transcripcional y organización de operones en *Escherichia coli*



*E. coli* (K12) Transcriptional Network

Main Page | Using RegulonDB | Tools | Downloads | About RegulonDB

Currently the **major** electronically-encoded **regulatory network** of any free-living **organism**

- The complex biology of gene regulation simplified into 3 major navigation streams: genes, operons and regulons.
- Continuously updated curated knowledge of original scientific literature complemented with comprehensive computational predictions.
- Graphic and text-integrated environment with friendly navigation where regulatory information is always at hand.

Search:  
Gene

News:  
Discover all the new features RegulonDB offers in this new update. (more)

Get started RegulonDB  
Learn what you can do with this powerful, knowledgebase (more)

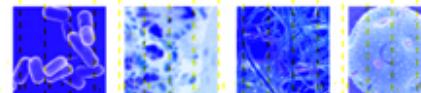
Tools:  
1. Nebulon Tool  
2. Genome Browser

Contact information:  
1. Contact us  
2. Suggestions

Funding  
The development of RegulonDB is funded by grant R01-GM71962 from the NIH (more)



CCG  
Centro de Ciencias Genómicas



# RegulonDB. Base de Datos de Regulación Transcripcional y organización de operones en *Escherichia coli*

RegulonDB 5.0 – Web Browser

http://regulondb.ccg.unam.mx/RegulonListControllerServlet?tf\_fa

Search | Print

Home Bookmarks Java Desktop System Sun Microsystems

## RegulonDB

*E. coli* (K12) Transcriptional Network

Main Page Using RegulonDB Tools Downloads About RegulonDB Gene

### REGULON AraC,CRP, in *Escherichia coli* K12 genome

**TRANSCRIPTION FACTOR:** [info](#)

Name: AraC  
Synonym(s): AraC transcriptional dual regulator  
Gene name(s): araC  
Conformation(s): AraC-Arabinose, AraC,  
Coregulator(s): AraC, CRP,

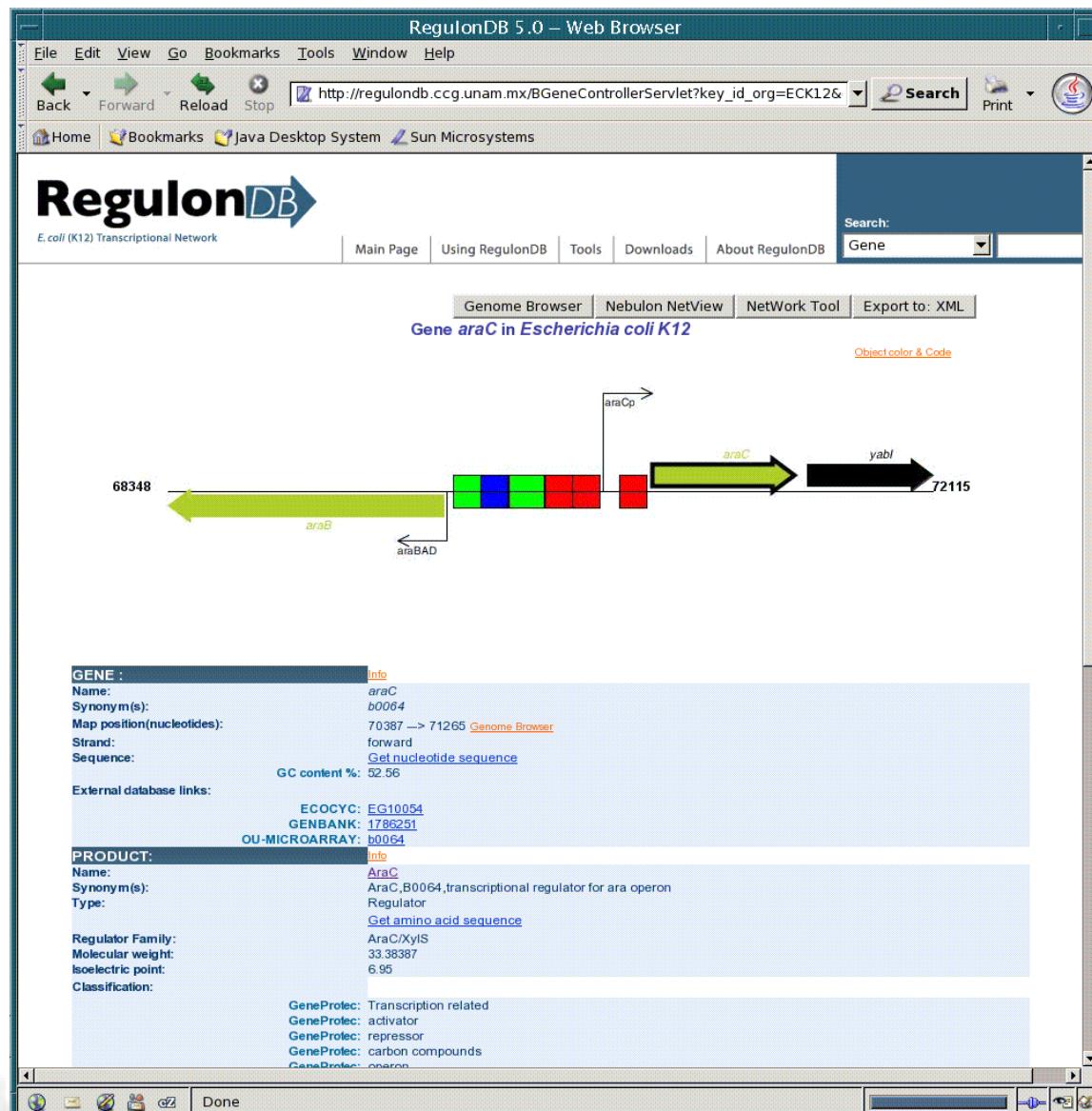
**TRANSCRIPTION FACTOR:** [info](#)

Name: CRP  
Synonym(s): CRP transcriptional dual regulator  
Gene name(s): crp  
Conformation(s): CRP-cAMP,  
Coregulator(s): AraG, AraC, AraA, AraR, CRP, CaiF, ChbR, CsiR, CspA, CytR, DcuR, DeoR, DnaA, DsdC, EbaR, ExuR, Fnr, FadR, FhlA, Fis, FlhD, FruR, Fur, Gade, GadW, GalX, GalS, GlpR, GntR, GutM, GutR, H-Ns, HU, HdrR, HvrR, IHF, IdnR, LacI, LrhA, Lrp, MalI, Malt, McR, MhpR, Mlc, MtrR, Nag, NagC, NanR, NarL, NtcR, OmpR, OxyR, PaaX, PdhR, PhoB, PurR, QseB, RbsR, RcsB, RhaR, RhsA, SlyA, TdcR, TdcR, TorR, TreR, UhpA, UidR, UlaR, UvuR, XviR, YiaJ.

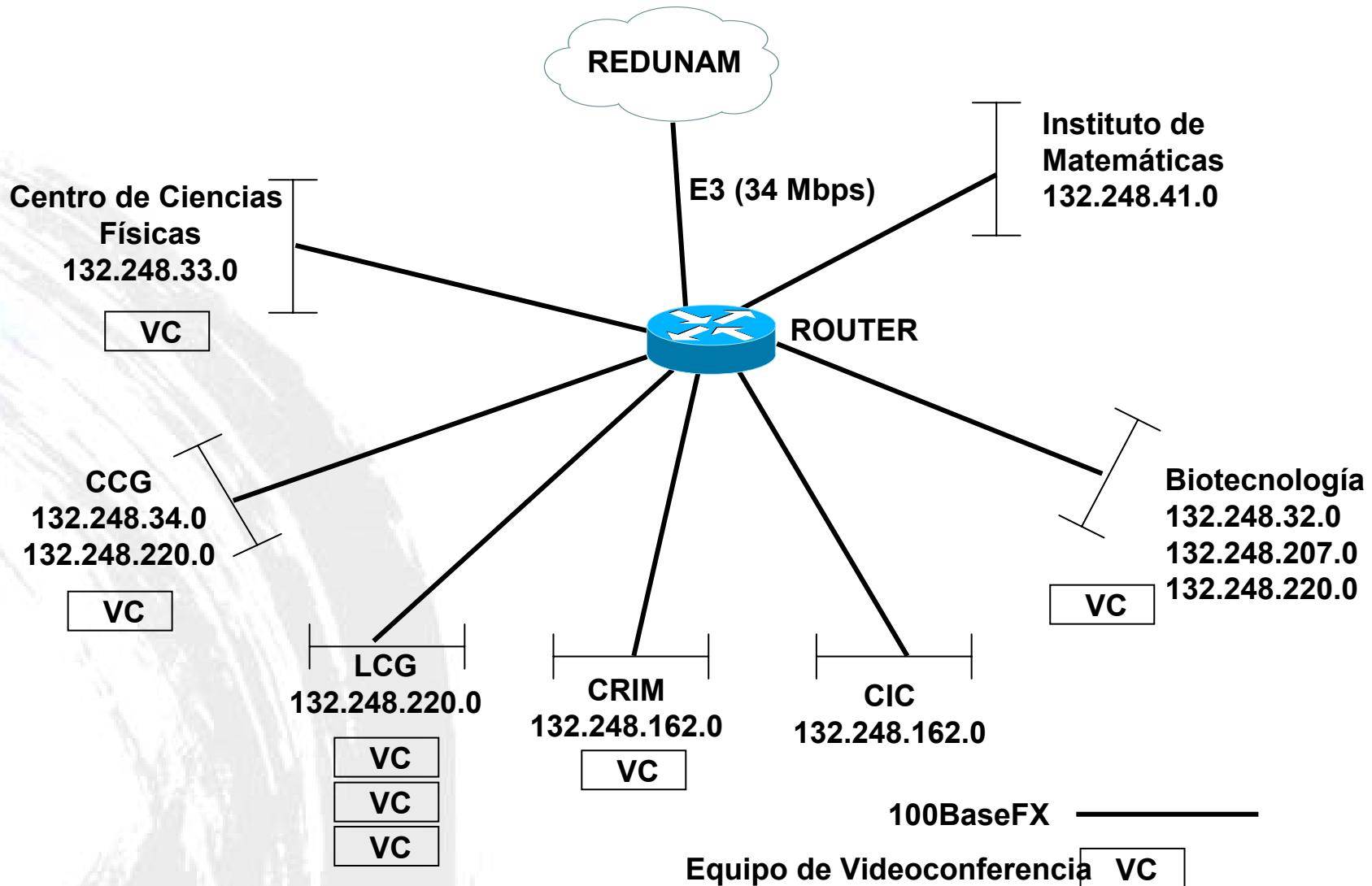
**REGULATION EXERTED BY AraC,CRP,** [info](#)

Transcription Factor	Regulated	Binding Sites	Evidences	References					
Conformation	Function	Promoter	Gene(s)	LeftPos	RighPos	CenterPos	Sequence		
KNOWN BINDING SITES (The centerpos is relative to the promoter +1)									
<b>AraC dual CRP activator</b>									
CRP-cAMP	activator	araCp	araC	70157	70178	-72.5	atggcatagcAAAGTGTGACGCCGTGCAAATAtacaatgg	[BPP] [GEA] [HIBSCS] [SM]	[1] [2] [3] [4] [5] [6]
CRP-cAMP	activator	araBAD	araA,araB,araD	70157	70178	-93.5	ccacatgtatTTTGACGGCGTCACATTtgtatgcatt	[BPP] [GEA] [HIBSCS] [SM]	[1] [2] [3] [4] [5] [6]
AraC	repressor	araCp	araC	70341	70358	109.5	gtctgtatcaATATGGACAATTGGTTCTtctctgaat	[BCE] [BPP] [HIBSCS] [SM]	[7] [2] [8] [9] [10] [11]
AraC	repressor	araBAD	araA,araB,araD	70341	70358	-275.5	attcagagaaGAAACCAATTGCCATATtgcatcagac	[BCE] [BPP] [HIBSCS] [SM]	[7] [2] [8] [9] [10] [11]
AraC-Arabinose	dual	araCp	araC	70129	70146	-102.5	atcogctaatCTTATGGATAAAAATGCTatggcatagc	[BPP] [GEA] [HIBSCS] [SM]	[12] [7] [8] [9] [10] [13]
AraC-Arabinose	dual	araBAD	araA,araB,araD	70129	70146	-63.5	gctatgcatAGCATTTTATCCATAAGtagatggat	[BPP] [GEA] [HIBSCS] [SM]	[12] [7] [8] [9] [10] [13]
AraC	activator	araCp	araC	70109	70126	-122.5	ataaaaaaggcGTCAGGTAGGATCCGCTAatcttatgta	[AIBSCS] [BPP] [SM]	[7] [2] [10]
AraC-Arabinose	activator	araBAD	araA,araB,araD	70109	70126	-43.5	tccataagatTAGCGGATCCTACCTGACgttttatac	[BPP] [GEA] [HIBSCS] [SM]	[12] [7] [2] [3] [4] [10] [13] [11]

# RegulonDB. Base de Datos de Regulación Transcripcional y organización de operones en *Escherichia coli*



# Infraestructura de telecomunicaciones actual del Campus Morelos

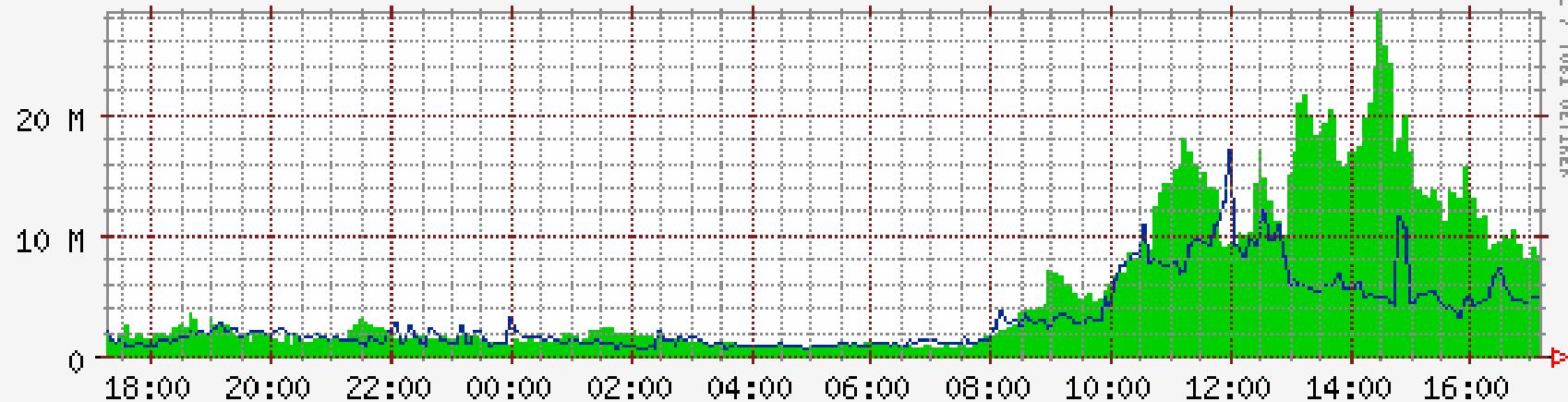


# Infraestructura de telecomunicaciones actual del Campus Morelos

- 6 Dependencias con más de 1500 usuarios y 1300 computadoras
- Actualmente nuestra conexión principal es una E3 (34 Mb), sin embargo no contamos con Internet 2.
- Contamos con 7 equipos de videoconferencia (prox. 10)

### Cuernavaca - A Telecom 1 - Se4/0

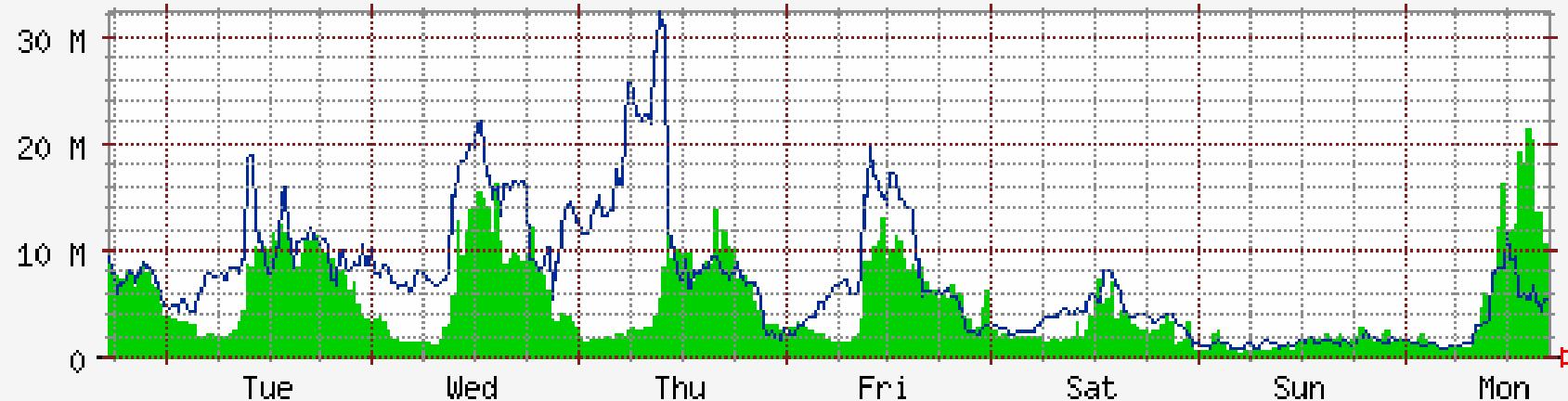
bits per second



Inbound Current: 8.20 M Average: 5.45 M Maximum: 28.45 M  
Outbound Current: 5.09 M Average: 3.23 M Maximum: 16.97 M

### Cuernavaca - A Telecom 1 - Se4/0

bits per second



Inbound Current: 9.26 M Average: 4.93 M Maximum: 21.21 M  
Outbound Current: 4.97 M Average: 7.26 M Maximum: 32.36 M

# Beneficios de una mejor conexión del Campus Morelos

- El contar con Internet 2 y un ancho de banda mayor tanto para Campus Morelos como para el Backbone de la UNAM traería grandes beneficios para la enseñanza y la investigación colaborativa.
- Nuestras necesidades de servicios de videoconferencia están creciendo.
- Los proyectos de investigación están cada vez más enfocados a la colaboración a distancia.
- Conferencias con asesores internacionales para discusión de diferentes proyectos (SRI-

# Beneficios de una mejor conexión del Campus Morelos

Con respecto a las Ciencias Genómicas:

- Las bases de datos bioinformáticas se duplican cada seis meses, por lo tanto también las necesidades de descarga de información (prom. de 40 GigaBytes c/2 semanas)
- Las herramientas bioinformáticas requieren de transmisión y recepción de grandes volúmenes de información en línea

# Beneficios de una mejor conexión del Campus Morelos

...continuación.

- El Nodo EMBnet podría contar con más herramientas y bases de datos para poder ofrecer más y mejores servicios a la comunidad científica (mayor velocidad de respuesta en interfaces gráficas), sin embargo, no basta con tener poder de cómputo suficiente, **si no contamos con infraestructura para poder entregar resultados rápido.**
- Un Grid bioinformático se vería ampliamente beneficiado.
- Mejoraría el tiempo de sincronización de la base de datos RegulonDB EcoCyc.

