The *Taenia solium* Genome Project

Universidad Nacional Autónoma de México
The Consortium

Institute of Biotechnology:
E Morett, X Soberón, A García-Rubio, P. Gaytan, J. Yañez

Center of Genomic Sciences:
MA Cevallos, VM González,

School of Medicine:
A. Landa, L Jiménez

School of Sciences:
V. Valdés

Institute of Biomedical Research:
Advisory Board

- Virginia Walbot, Stanford University, USA
- Bruce Roe, Oklahoma University, USA
- Luis Herrera-Estrella, CINVESTAV-Irapuato, MEX
- Charles, B. Shoemaker, Tufts University, USA
- Klaus Brehm, University of Wurzburg, GER
1. *Taenia solium* is the causal agent of human and porcine cysticercosis; a disease that still is **a public health problem** of considerable relevance in México and in several other countries.

2. This parasite/disease **has been studied by multiple groups in Mexico** during at least three decades. A considerable number of contributions on the understanding of the parasite and disease have been made by Mexican scientists. *T. solium* is an organism that the Mexican scientific community can justifiably appropriate.

3. A genomic project of this magnitude (estimated genome size 120 ~ 270 Mb) will promote the **organization of a human team** able to approach this and other projects in genomic sciences, by networking current capabilities in several research centers at UNAM. The project requires a considerable capability on DNA sequencing and a parallel capability on bioinformatics.

4. The project will contribute to the knowledge of an organism with an **interesting phylogenetic position** for studies of comparative genomics, etc.

5. A project of this magnitude will unite **groups with diverse disciplinary backgrounds**: immunology, molecular biology, cell biology, bioinformatics, among others.
6. The complete sequence of the *T. solium* genome will provide us with a much better understanding of the parasite’s physiology, life cycle and metabolism. Also, the knowledge of the complete enzymatic repertoire will help us to identify essential genes, which will be good candidates to be targeted by newly designed drugs.
Life cycle of *Taenia solium*

- Man is the definitive host and harbors the adult tapeworm in the upper small intestine.

- Three stages: larval, embryo, and adult. Can infect man both as larva or as adult.

- Infection with adult occurs through eating uncooked pork. (Embryo eaten and turns into adult in the intestine).

- Infection with the larva causes **cysticercosis**. Occurs when animals or humans become the intermediate host of the **larval** form. Occurs by ingesting eggs. The eggs hatch in the intestine and the larvae burrow through the intestinal walls and disseminate into the soft tissues.

**MAN (adult tapeworm in upper small intestine)**
Eggs intermittently eliminated in feces
Consumed by **PIGS and MAN**
Ova develop into larvae which infect the tissues
**MAN consumes pork**
The adult tapeworm
Cysticercosis

Definition
• A parasitic infection involving the CNS caused by the larval stage of the pork tapeworm, *taenia solium*, which has a marked predilection for neural tissue.
• Infection caused by Cysticercus cellulose which is the larval form.

Epidemiology
• Cysticercosis is the most common parasitic infection involving the CNS.
• Endemic in areas of Mexico, eastern Europe, Asia, Central and South America, and Africa.
• Incidence of Neurocysticercosis (CNS involvement) may reach 4% in some areas.
• 80% of infected individuals show symptoms within 7 years of infection.

• Man is the **definitive host** and harbors the adult tapeworm in the **upper small intestine**. The worm attaches to the wall of the jejunum by 4 suckers and 2 hooklets, where it absorbs food directly. Man is the only permanent host of the adult form. Proglottids (mature segments, each containing reproductive organs) laden with eggs are eliminated intermittently in the feces. If consumed by the intermediate host, the hog, the ova will develop into larvae or onchospheres that penetrate the intestinal wall, then invade the lymphatics and veins, and then disseminates into skeletal muscles and other tissues. Man becomes infected when he ingests poorly cooked, infected pork. The consumed cisticercus develops an **evaginated scolex** that attaches to the jejunal mucosa and develops into an **adult worm** in the human intestine, completing its life cycle.

• Acquisition of **larval forms** producing cisticercosis occurs after ingestion of *T. solium* eggs in food or water contaminated with infected feces. The ova can also be acquired by oral transmission from unclean hands of carriers of the adult tapeworm. Internal autoinfection is also possible by regurgitation of eggs from the jejunum into the stomach through reverse peristalsis. Within 2 months, the larvae develop a **cyst wall**. Within 4 months, they mature into embryos. Embryos can survive 5-7 yrs. Those which die, calcify. When the pigs are eaten, the embryos remain viable and are ingested with the life cycle being repeated.

• **Onchospheres** once outside the GI tract invade other tissues and become **mature cisticerci** which are oval, translucent cysts, containing a single scolex bearing four suckers. Within 12 weeks, the cisticerci mature and are primarily observed in skeletal muscle; other infected sites include the **brain, eyes, liver, lung and subcutaneous tissues**.
Pathophysiology

Cisticercosis (infection with the larva) involves the following common sites and symptoms are referrable to these sites.

i) Brain (60-90%)

ii) Skeletal muscle, eye, subcutaneous tissue

Gross

• Two types of cysts develop in the brain:
  1. **Cysticercus cysts.** Thin walled cysts 3-20mm in diameter forming in the parenchyma or subarachnoid spaces. They contain the scolex (head), and are usually static, producing little inflammation.
  2. **Racemose cysts.** Large (4-12cm), grow actively and produce grape-like clusters in the basal subarachnoid space with intense inflammation. They do not contain larvae.

Microscopic

• Should identify scolex with four suckers and anterior hooklets. Live cysticerci evoke little inflammatory response. Dead cysticerci evoke intense responses, monocytes, macrophages.

CNS Manifestations of Cysts

1. **Diffuse parenchymatous disease.** Without focal mass; disseminated larval death; inflammatory reaction, toxic encephalopathy and meningitis. This probably represents the acute encephalitic phase described in some reports (*J of Child Neurology*, 10:177, 1995). This form is uncommon but more frequently reported in children. A subacute or chronic cysticercotic encephalitis also exists, caused by degeneration of multiple cysticerci.

2. **Cysts** (classified by location).
   i) **Intraparenchymatous cysts** as mass lesions. Solitary or part of multifocal disease. Tend to present with seizures.
   ii) **Subarachnoid and cisternal cysts.** The commonest location for extraventricular, nonparenchymatous CNS involvement is the basal cisterns (found in almost all cisterns). Often forms a basal meningitis in the cisterns. These are usually of the racemose variety. Dosolateral subarachnoid space can be involved by the cysticerci type and usually cause minimal symptoms.
   iii) **Intraventricular cysts.** May be solitary or multiple.
   iv) **Spinal cysts**
   v) **Mixed lesions**

3. **Basilar adhesive and racemose form.** Obliterative arachnoiditis; hydrocephalus; mixed types with cisternal cysts or spinal disease.
Clinical Manifestations

Clinical
• Manifestations are the consequence of the organism lodging in the meninges, parenchyma, subarachnoid space, etc. This can cause blockage of CSF flow, a mass lesion, chronic meningitis, cranial neuropathy, and/or seizures.

Laboratory Investigations
• Diagnosis is by:
  1. Identification of probable epidemiologic exposure.
  2. Origin of patient from endemic region
  3. History and physical
  4. Radiological data
  5. Serology
     i) Eosinophilia. Suggestive of parasitic infection but is inconsistent and unreliable.
     ii) Serology of serum and CSF. Cysticercosis antibody titers determined by ELISA. Indirect hemagglutinins or indirect immuofluorescence.
     iii) CSF may be normal. Pleocytosis (eosinophils) seen in 15%.
        (NB. T. solium ova only present in the stool in 1/3 of cases.

Imaging
• Soft tissue X-rays may show calcified nodules.
• On CT scans, ring-enhancing lesions represent living cysticerci. Little edema as long as larva is alive. Central punctate high density probably represents scolex. Calcified (dead) forms of larvae are identified best by CT. Usually without surrounding edema.
• Contrast enhancement generally does not identify the walls of live cysts in the CSF spaces. Water soluble positive contrast ventriculography or cisternography historically was the best method for detecting lesions in the CSF spaces. MRI is the preferred imaging modality for these lesions. A T2 weighted image (or FLAIR) will often reveal multiple areas of signal prolongation. Following gadolinium, the cysts with the scolex can be seen
Treatment & Results

Medical
1. Steroids: indicated during acute neurological deterioration or while on other therapy.
2. Antihelminthics. A single dose will kill the tapeworm.
i) **Praziquantel.** May be used in basilar adhesive arachnoiditis, cisternal cystic lesions, and active parenchymatous disease. Given as 50 mg/kg in three divided doses for 15 days. Drug of choice for intestinal infestation.
   ii) **Albendazole.** A newer agent which is superior to praziquantel. 15 mg/kg divided into 2-3 doses for 3 months.
   iii) **Niclosamide.** May be used to treat GI tapeworms.

Surgical
- Indications:
  1. Establishing the diagnosis (by biopsy, stereotactic or open)
  2. Treatment
     i) Palliation of hydrocephalus.
     ii) Excision of intracisternal or intraparenchymatous cysts because of mass effect.
     iii) Excision of intraventricular cysts.
     iv) Excision of rare spinal masses.

- **Total surgical excision is necessary**; the wall of the larval form may continue to cause difficulties even in the absence of a persistent scolex.
- Little danger of disseminating cysts through CSF diversion.
- Intraventricular cysts should be removed surgically whenever possible especially when they cause obstruction of the IVth ventricle. IVth ventricular cysts are the commonest
- Treat patients with **steroids** perioperatively
Cysticercosis in the world
Sequencing the *Taenia solium* genome

May 2007 report
Sequences are produced at 3 sequencing centers.

CINVESTAV (Irapuato).
Centro de Ciencias Genómicas (Cuernavaca).
Instituto de Biotecnología (Cuernavaca).

All sequences are collected and analyzed at IBT.
There are 2 kinds of projects:

1 ESTs
2 Genomic shotgun

Two technologies are used:

Sanger:
Pros: long reads, higher quality, easier assembly
Cons: more expensive, some genome regions could be poorly represented

454 pyrosequencing:
Pros: cheaper, high throughput, no cloning required, unbiased representation.
Cons: short reads, lower quality, homo-nucleotide compression, difficult to assemble.
DNA library in Sephadex beads

Water-oil emulsion

microreactors

Clonal amplification by PCR

Break of microreactors

a) PCR emulsion and clonal PCR

b) Pyrosequencing

loading in chamber

DNA synthesis coupled to light emission

CCD detection

454 Pyrosequencing
The libraries

Up to now, 8 "libraries" have been built:

An adult cDNA library (cd1)
Two larva cDNA libraries (cd2 and cd3)
A genomic library with 2-5 Kb inserts (sg1)
A genomic library with 1-3 Kb inserts (sg2)
A genomic library with 7-10 Kb inserts (sg3, starting sequencing)
A genomic fosmid library (sg4)
Several anonymous 454 libraries (454)
Sanger production by library (reads):

24039 Adult EST (cd1)
14870 Larva EST (cd2 + cd3)

------
38909 EST total

105219 Shotgun 2-5Kb (sg1)
16128 Shotgun 1-3Kb (sg2)

------
121347 Shotgun total (sum of both ends)

------
160,256 Grand total
Sanger sequence processing

Centers deposit 3 kinds of files: Chromatograms, sequences and quality files produced by the sequencing software.

New sequences and quality files are produced with an alternative basecaller (Phred).

Lucy takes the sequences from both basecallers and the Phred quality file, and extends the "good quality" regions. The average gain is 60 bases per read.

Zapping-awk trims poly-A, sequencing vector and poor quality regions from both ends.
Sequence number and average length after trimming
(initial/ final /bases)

<table>
<thead>
<tr>
<th></th>
<th>initial</th>
<th>final</th>
<th>bases</th>
</tr>
</thead>
<tbody>
<tr>
<td>cd1</td>
<td>24039</td>
<td>17529</td>
<td>568</td>
</tr>
<tr>
<td>cd2</td>
<td>12470</td>
<td>9132</td>
<td>582</td>
</tr>
<tr>
<td>cd3</td>
<td>2400</td>
<td>1409</td>
<td>531</td>
</tr>
<tr>
<td>sg1</td>
<td>55219</td>
<td>50405</td>
<td>665</td>
</tr>
<tr>
<td>sg2</td>
<td>16128</td>
<td>14130</td>
<td>640</td>
</tr>
</tbody>
</table>
Efficiency after trimming (reads)

- cd1
- cd2
- cd3
- sa1
- sa2
- Row
- ibt
- ira
- cca
- total

- initial
- final
Sequence cleaning

Trimmed sequences (zap.seq) are masked against the cloning vector, the sequencing vector, the NCBI UniVect data base, Lambda phage, Eschericia coli genome, and the T. solium mitochondrion.

Chimerism is detected by "wrong place" poly-A presence and by internal vector-insert splice sites.

Chimeras are cleaved judiciously and all masked regions are removed. The resulting fragments longer than 150bp are kept (split.seq).

Chimeras are frequent in the EST libraries.
Size distributions of clean (split.seq) ESTs

EST frequency vs length
Size distributions of clean (split.seq) shotgun traces
454 sequencing

There have been 56 runs, each producing ~20Mb.

Reads are assembled at the flowgram level by the 454 “de novo” assembler (Newbler).

28 runs is the maximum that Newbler can handle, so the reads have been assembled in two batches, with ~560 Mb each.

Both batches behaved very similarly: they produced ~79 Mb of assembled sequence, incorporating >95% of the reads. Assemblies included 362 and 382 thousand contigs, with average contig lengths of 217 and 203 bases.
454 sequence reassembly

Contigs from both batches were assembled at the sequence level with PCAP. The number of contigs was reduced from 744 K to 215 K, while average length almost doubled (from 210 to 382 bases).

N50 (the smallest size in the collection of the longest contigs which hold 50% of the assembled sequence) changed from 239 to 448 bases.

There are 130 contigs longer than 5Kb. The longest one has 16336 bases.
Whole genes in the 454 contigs

NCBI has 95 complete CDS *T. solium* sequences. After eliminating very similar sequences (nr90) we made a collection of 45 whole genes.

Using MegaBlast, four genes did not match the 454 contigs, and only 2 genes were entirely contained in a single contig. Aprox. 65% of the genes were covered by contigs.

Most genes matched 2 or more contigs, which indicates that gap sizes in the 454 assembly are within the range of gene sizes.

If the sequences of the genes are taken as correct, the 454 sequences have an error rate slightly over 1% (1.08%)
What is the amount of repeated sequences in the genome?

ReapeatMasker suggests it is \(~6.8\%\)

The naïve approach of comparing all sequences vs all sequences, and taking as repeated any segment with at least 4 hits, gives a similar value: \(7.0\%\)

The following small (53 bp) tandem repeat represents \(0.5\%\) of the genome:

CGCTCTCACTGAATGCGATTTCGTATGAGTGATCTTCACCAGACTGCAGATTT \(\text{(Pst I)}\)

Different tetranucleotide repeats of the form \((Txxx)n\) amount to \(4.5\%\) of the genome.
The genome size

The “footprint” calculation

Presently, the sum of 454 contigs is 82Mb.

Those 82Mb should be 65% of the genome, because 65% of NCBI complete CDS sequences find a match in those contigs, and , similarly, 65% of ‘unique’ (non repeated) sequences from the sanger shotgun (10.2 out of 15.9 Mb) find a match in those contigs.

Thus, the genome size must be:

\[
82\,\text{Mb} / 0.65 = 126\,\text{Mb}
\]
The “coverage” calculation.

At 560 Mb of 454 sequence (half the amount we have now) we estimated that the distribution of reads per contig was consistent with a 4-5X coverage. The calculation did not require assuming a genome size.

From this data we inferred that the genome size should be in the range:

\[
560\text{Mb}/5 = 112 \text{Mb} \quad \text{to} \quad 560\text{Mb}/4 = 140\text{Mb}
\]

So this and the footprint calculation are in good agreement.
The Lander-Waterman corollary

According to the Lander-Waterman formula of random nucleotide sampling, for any genome size, there should be a 5% decrease in gaps (a 5% increase in sequence) when going from 3X to 6X coverage.

That is nearly what happened when we assembled the two 560Mb batches; so this is consistent with a 560MB batch being 3X the genome.

However, that same formula shows that at 6X, 99.7% of the sequence should be contained. Thus, the genome size must be 82Mb.

This is consistent with the saturation of the curves we have observed, but contradicts the fact that the contigs are missing ~35% of all searched query sequences, as shown in the footprint calculation.
Saturation of the 454 assembly
From ESTs to genes

Unique genes are identified by clustering all EST-fragments with an assembler (minimus).

Out of the initial 23290 ESTs, 19067 were incorporated into 2564 genes (contigs).

We have ~6800 “genes”, including the 2564 contigs, plus 2592 larva and 1611 adult ESTs that remained as solitons.
Highly expressed genes.

There are 349 “genes” with 10 or more sequences. They account for 50% of all transcripts. Many are differentially expressed. (Note: because we have more adult than larva sequences, a 1.5X normalization factor should be applied for proper comparison).
Larva/adult ESTs in 3 categories of high expression contigs
Homology annotation

With a Blast cutoff (expect < 1e-10), aprox $\frac{1}{3}$ (2038) of the 6800 genes have a significant match in SwissProt, so some annotation can be inferred.

Gene Ontology (GO) and Gene Ontology broad categories (GoSlim) can be assigned to 36% (2448 genes) of our genes.

Comparison of GoSlim annotations to the GoSlim annotations of C. elegans leads to contradictions, indicating that very little real information is being transferred.

Aprox. 27% of the genes do not have a match (expect< 1e-3) in SProt + TREMBL, so they could be new genes.
High expression genes with good SProt homologues
Polymorphism

Polymorphism in the contigs can hint to assembly errors, sequencing errors or poor alignments. The 3 possibilities can frequently be distinguished by close examination of the alignments.

We are analyzing correlated polymorphism: sequences that share non-consensus nucleotides at different positions. These can hint to heterozygosis, gene duplications, and alternative splicing.

Around 20% of the contigs that can be analyzed (those with 4 sequences or more) show correlated polymorphism. In general, this is not associated to a larva/adult division.
Strategies for the classification of polymorphisms

**Heterozygosis** can only explain *dimorphism*, and the divergent sequences tend to be aligned along their full lengths.

If the alignment is *tri-morphic*, or the divergent sequences have in common end positions that are distinct from the other sequences, *duplication* can be assumed.

Shared **indels**, close to highly divergent regions, can indicate **alternative splicing**.

The work is in progress, but a relevant data is that genes with **paralogues** should be less than 20%, probably ~10%. 
Different length paralogues
Tri-morphic contig
Summary

Mysteries aside, all bioinformatics measures indicate that the genome size should be between 82 and 130 Mb.

Taking the upper limit as the correct figure, the project has achieved:

- **9X** in 454 sequences,
- **0.5X** in sanger sequences, and
- **25000 ESTs**

Genome size was estimated at about 270 Mb by cytofluorometry on isolated cyton nuclei (Diploid?)
Other cousins (flatworms) being sequenced:

Schmidtea mediterranea (480Mb); WU; (3 million reads) 7X coverage + 50000 ESTs

Schistosoma mansoni (270Mb/8Chr); Sanger + TIGR; (370K reads)

Isodiametra pulchra (Acoela); WU; 10,000-20,000 shotgun, planed

Clonorchis sinensis; Seoul National Univ.

Fasciola hepatica; Sanger; 15000 ESTs, finished

Echinococcus granulosus (150Mb); Sanger 10000 EST

Echinococcus multilocularis (150Mb); Sanger 10000 EST

Schistosoma haematobium (270 Mb); Sanger; 15000 ESTs
Analysis of *Taenia solium* adult and larvae cDNA libraries

- 14,000 adult clones
- 9,000 larvae clones

- 2,564 contigs
- 1,611 adult singleton clones
- 2,592 larvae singleton clones

Larvae has ca 60% fewer sequences but ca 60% more singletons
Adult has many genes with high level of expression.

Blast analysis with SwissProt to identify cDNAs with putative function.
About half of the contigs showed identity with SwissProt entries
From higher to lower expression

**Fatty acid-binding protein**

<table>
<thead>
<tr>
<th>Exp_rank</th>
<th>contig_id</th>
<th>category</th>
<th>exp_total</th>
<th>exp_larva</th>
<th>exp_adult</th>
<th>Swiss_id</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>878</td>
<td>HE</td>
<td>422</td>
<td>30</td>
<td>392</td>
<td>Q9BMK3:</td>
</tr>
</tbody>
</table>

**Swiss_description**
Fatty acid-binding protein.

**SIMILARITY:** Belongs to the fatty-acid binding protein (FABP)
Are fatty acids the main food source of *T. solium*?

Biochim Biophys Acta. 2000

The fatty acid transport function of fatty acid-binding proteins.

Storch J, Thumser AE.

The intracellular fatty acid-binding proteins (FABPs) comprise a family of 14-15 kDa proteins which bind long-chain fatty acids... Collectively, data from these studies have provided strong support for defining the FABPs as fatty acid transport proteins.
Q10442: Putative mitochondrial carrier
Mitochondrial aspartate-glutamate transporter
SUBCELLULAR LOCATION: Mitochondrion; mitochondrial inner membrane;

Transport of glutamate in mitochondria is required for mitochondrial transamination reactions and ornithine synthesis. Plays also a role in malate-aspartate NADH shuttle, which is critical for growth on acetate and fatty acids!!!
Is the Fatty Acid Binding Protein of *T. solium* a good candidate for a vaccine?

*Proc Natl Acad Sci U S A.* 1996

A *Schistosoma mansoni* fatty acid-binding protein, Sm14, is the potential basis of a dual-purpose anti-helminth vaccine

Miriam Tendler et al

Molecular cloning of components of protective antigenic preparations has suggested that related parasite fatty acid-binding proteins could form the basis of the protective immune crossreactivity between the parasitic trematode worms *Fasciola hepatica* and *Schistosoma mansoni*... A recombinant form of the *S. mansoni* antigen, rSm14, protected outbred Swiss mice by up to 67% against challenge with *S. mansoni* cercariae... The same antigen also provided complete protection against challenge with *F. hepatica* metacercariae in the same animal model. The results suggest that it may be possible to produce a single vaccine that would be effective against at least two parasites, *F. hepatica* and *S. mansoni*, of veterinary and human importance, respectively.
Actin and actin related functions

FUNCTION: Actins are highly conserved proteins that are involved in various types of cell motility and are ubiquitously expressed in all eukaryotic cells.

FUNCTION: Severin blocks the ends of F-actin and causes the fragmentation and depolymerization of actin filaments. This severin binds stably with actin both in a Ca(2+) dependent and a Ca(2+) independent manner.

FUNCTION: May play a role in the organization of actin filament arrays within muscle cells.
Tubulin

FUNCTION: Tubulin is the major constituent of microtubules. It binds two moles of GTP, one at an exchangeable site on the beta chain and one at a non-exchangeable site on the alpha-chain.

SUBUNIT: Dimer of alpha and beta chains.

124: 136 HE 24 7 17: O17449: Tubulin beta-1 chain

156: 1614 AD 19 0 19: Q68FR8: Tubulin alpha-3 chain

216: 1886 HE 14 2 12: Q9BQE3: Tubulin alpha-6 chain

237: 593 HE 13 4 9: Q6P9T8: Tubulin beta-2C chain
Dynein

Dynein is a motor protein (also called molecular motor or motor molecule) in cells which converts the chemical energy contained in ATP into the mechanical energy of movement. Dynein transports various cellular cargo by "walking" along cytoskeletal microtubules towards the minus-end of the microtubule, which is usually oriented towards the cell center. Thus, they are called "minus-end directed motors," while kinesins, motor proteins that move toward the microtubules' plus end, are called plus-end directed motors.

FUNCTION: May be involved in some aspects of dynein-related intracellular transport and motility. May play a role in changing or maintaining the spatial distribution of cytoskeletal structures.
Myosin and related proteins

25: 375 HE 92 24 68: P02612: Myosin regulatory light chain 2
FUNCTION: Plays an important role in regulation of both smooth muscle and nonmuscle cell contractile activity. ENZYME REGULATION: Phosphorylation of MLC-2 by the enzyme MLC kinase in the presence of calcium and calmodulin increases the actin-activated myosin ATPase activity and thereby regulates the contractile activity.

128: 1515 HE 24 11 13: Q24756: Myosin light chain (also P27166 Calmodulin)

112: 1356 HE 27 2 25: Q95PU1: Tropomyosin
FUNCTION: Tropomyosin, in association with the troponin complex, plays a central role in the calcium dependent regulation of muscle contraction.


255: 766 HE 12 9 3: Q08093: Calponin-2
FUNCTION: Thin filament-associated protein that is implicated in the regulation and modulation of smooth muscle contraction. It is capable of binding to actin, calmodulin, troponin C and tropomyosin. The interaction of calponin with actin inhibits the actomyosin Mg-ATPase activity.

297: 919 HE 11 6 5: P46150: Moesin/ezrin/radixin homolog
FUNCTION: Involved in connections of major cytoskeletal structures to the plasma membrane.
305: 1639 HE 11 2 9: Q8T305: Paramyosin
Calmodulin

FUNCTION: Calmodulin mediates the control of a large number of enzymes and other proteins by Ca(2+). Among the enzymes to bestimulated by the calmodulin-Ca(2+) complex are a number of protein kinases and phosphatases.

Collagen

FUNCTION: Type I collagen is a member of group I collagen (fibrillar forming collagen).

FUNCTION: Type I collagen is a member of group I collagen (fibrillar forming collagen).
Sperm protein

23: 1455     AD  95  0  95: Q8K450: Sperm-associated antigen
FUNCTION: Necessary for sperm flagellar function.
SUBUNIT: Interacts with SPAG6.

Does Taenia has sperm flagella?
Protein degradation

30: 2240 AD 71 0 71: P23398: Ubiquitin
FUNCTION: Protein modifier which can be covalently attached to target lysines either as a monomer or as a lysine-linked polymer. Attachment to proteins as a Lys-48-linked polymer usually leads to their degradation by proteasome. Attachment to proteins as a monomer or as an alternatively linked polymer does not lead to proteasomal degradation and may be required for numerous functions, including maintenance of chromatin structure, regulation of gene expression, stress response, ribosome biogenesis and DNA repair (By similarity).

180: 42 HE 17 3 14: P23398: Ubiquitin
269: 2475 HE 12 3 9: P23398: Ubiquitin
66: 1452 AD 41 1 40: Q41365: 26S protease regulatory subunit(ATPase 6)
FUNCTION: The 26S protease is involved in the ATP-dependent degradation of ubiquitinated proteins. The regulatory (or ATPase) complex confers ATP dependency and substrate specificity to the 26S complex

191: 1307 LV 16 16 0: Q965X6: E3 ubiquitin-protein ligase
FUNCTION: E3 ubiquitin-protein ligase that mediates ubiquitination and subsequent proteasomal degradation of target proteins. E3 ubiquitin ligases accept ubiquitin from an E2 ubiquitin-conjugating enzyme in the form of a thioester and then directly transfers the ubiquitin to targeted substrates. It probably triggers the ubiquitin-mediated degradation of different substrates

244: 2026 AD 13 0 13: P21670: Proteasome subunit alpha type 4
FUNCTION: The proteasome is a multicatalytic proteinase complex which is characterized by its ability to cleave peptides with Arg, Phe, Tyr, Leu, and Glu adjacent to the leaving group at neutral or slightly basic pH. The proteasome has an ATP-dependent proteolytic activity.

303: 627 HE 11 9 2: Q06AA9: Ubiquitin-conjugating enzyme
338: 1508 HE 10 3 7: O73817: Proteasome subunit beta
Nucleotidase

33: 1809 AD 69 1 68: Q9D020: Cytosolic 5'-nucleotidase III
CATALYTIC ACTIVITY: A 5'-ribonucleotide + H(2)O = a ribonucleoside+ phosphate.
Annexines

The annexins are a family of proteins that bind calcium-dependently to phospholipid membranes. They are found in all kingdoms (animal, plant and fungi) with the exception of the bacteria.

40: 1527   HE   62   5   57: P51901: Annexin A
FUNCTION: May associate with CD21. May regulate the release of Ca(2+) from intracellular stores.

267: 303   HE   12   1   11: P20073: Annexin
FUNCTION: Calcium/phospholipid-binding protein which promotes membrane fusion and is involved in exocytosis.

FUNCTION: Binds specifically to calcyclin in a calcium-dependent manner.

281: 403   HE   12   10   2: Q29471: Annexin A13
FUNCTION: Involved in vesicular traffic to the apical plasma membrane.
Transcription factors

49: 1077  AD  51  1  50: Q06543: Eukaryotic Transcription factor 5
SIMILARITY: Belongs to the FET5 family.

296: 762  HE  11  8  3: O43474: Krueppel-like factor 4
FUNCTION: May act as a transcriptional activator. Binds the CACCC core sequence. May be involved in the differentiation of epithelial cells and may also function in the development of the skeleton and kidney.
Carbohydrate degradation; glycolysis

51: 821  HE  48  19  29:  P51469: Glyceraldehyde-3-phosphate dehydrogenase
CATALYTIC ACTIVITY: D-glyceraldehyde 3-phosphate + phosphate + NAD(+) = 3-phospho-
D-glyceroyl phosphate + NADH. PATHWAY: Carbohydrate degradation; glycolysis; pyruvate
from D-glyceraldehyde 3-phosphate: step 1.

CATALYTIC ACTIVITY: D-fructose 1,6-bisphosphate = glycerone phosphate + D-
glyceraldehyde 3-phosphate. PATHWAY: Carbohydrate degradation; glycolysis; D-
glyceraldehyde 3-phosphate and glycerone phosphate from D-glucose: step 4.

82: 959  HE  36  23  13:  Q27655: Enolase
CATALYTIC ACTIVITY: 2-phospho-D-glycerate = phosphoenolpyruvate + H(2)O.

99: 2111  AD  30  0  30:  Q7VS43: 2,3-bisphosphoglycerate-dependent phosphoglycerate
mutase
FUNCTION: Catalyzes the interconversion of 2-phosphoglycerate and 3-phosphoglycerate.

151: 1394  HE  20  7  13:  P27443: NAD-dependent malic enzyme
CATALYTIC ACTIVITY: (S)-malate + NAD(+) = pyruvate + CO(2) + NADH.

164: 711  HE  18  12  6:  P53442: Fructose-bisphosphate aldolase
CATALYTIC ACTIVITY: D-fructose 1,6-bisphosphate = glycerone phosphate + D-
glyceraldehyde 3-phosphate.

188: 1954  AD  16  0  16:  P06745: Glucose-6-phosphate isomerase
CATALYTIC ACTIVITY: D-glucose 6-phosphate = D-fructose 6- phosphate.

252: 413  HE  12  3  9:  Q589R5: Triosephosphate isomerase

263: 243  HE  12  8  4:  Q69HD8: Phosphoglucomutase
Phosphoenolpyruvate carboxykinase

**CATALYTIC ACTIVITY:** GTP + oxaloacetate = GDP + phosphoenolpyruvate + CO(2).

**ENZYME REGULATION:** Activity is affected by a number of hormones regulating this metabolic process (such as glucagon, insulin, or glucocorticoids).

**PATHWAY:** Carbohydrate biosynthesis; gluconeogenesis.
Protein folding

54: 1328   HE  48   14   34:  P14088: Peptidyl-prolyl cis-trans isomerase
FUNCTION: PPIases accelerate the folding of proteins. It catalyzes the cis-trans isomerization of proline imidic peptide bonds in oligopeptides.
Cytochromes

62: 1454  AD  42  0  42: P75925: Cytochrome b561
COFACTOR: Binds 2 heme B (iron-protoporphyrin IX) groups permolecule.
<table>
<thead>
<tr>
<th>Queue</th>
<th>Gene ID</th>
<th>Accession No.</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>50:</td>
<td>1451</td>
<td>AD</td>
<td>51: P49154: 40S ribosomal protein S2</td>
</tr>
<tr>
<td>64:</td>
<td>2228</td>
<td>HE</td>
<td>0: P55935: 40S ribosomal protein S9</td>
</tr>
<tr>
<td>75:</td>
<td>1516</td>
<td>HE</td>
<td>8: Q5E973: 60S ribosomal protein L18</td>
</tr>
<tr>
<td>97:</td>
<td>2056</td>
<td>HE</td>
<td>5: Q4R5P3: 60S ribosomal protein L10a</td>
</tr>
<tr>
<td>103:</td>
<td>750</td>
<td>HE</td>
<td>8: Q3SYR7: 60S ribosomal protein L9</td>
</tr>
<tr>
<td>130:</td>
<td>329</td>
<td>HE</td>
<td>7: Q29361: 60S ribosomal protein L35</td>
</tr>
<tr>
<td>147:</td>
<td>1161</td>
<td>HE</td>
<td>6: Q2I0I6: 60S ribosomal protein L26</td>
</tr>
<tr>
<td>154:</td>
<td>721</td>
<td>HE</td>
<td>6: P48162: 60S ribosomal protein</td>
</tr>
<tr>
<td>223:</td>
<td>345</td>
<td>HE</td>
<td>10: P39018: 40S ribosomal protein S19</td>
</tr>
<tr>
<td>224:</td>
<td>22</td>
<td>HE</td>
<td>9: P23403: 40S ribosomal protein S20</td>
</tr>
<tr>
<td>225:</td>
<td>894</td>
<td>HE</td>
<td>11: O61231: 60S ribosomal protein L10</td>
</tr>
<tr>
<td>226:</td>
<td>1166</td>
<td>HE</td>
<td>9: Q9NB34: 60S ribosomal protein L34</td>
</tr>
<tr>
<td>230:</td>
<td>217</td>
<td>HE</td>
<td>4: Q2I0I6: 60S ribosomal protein S19</td>
</tr>
<tr>
<td>235:</td>
<td>595</td>
<td>LV</td>
<td>12: Q3T0B7: 40S ribosomal protein S27</td>
</tr>
<tr>
<td>250:</td>
<td>1790</td>
<td>HE</td>
<td>12: Q90YR6: 40S ribosomal protein S8</td>
</tr>
<tr>
<td>258:</td>
<td>653</td>
<td>HE</td>
<td>7: Q17445: 60S ribosomal protein L15</td>
</tr>
<tr>
<td>264:</td>
<td>454</td>
<td>HE</td>
<td>7: Q4R7Y8: 60S ribosomal protein L8</td>
</tr>
<tr>
<td>278:</td>
<td>16</td>
<td>HE</td>
<td>8: Q4R7Y8: 60S ribosomal protein L8</td>
</tr>
<tr>
<td>279:</td>
<td>849</td>
<td>HE</td>
<td>6: Q5R7Y8: 60S ribosomal protein L8</td>
</tr>
<tr>
<td>282:</td>
<td>536</td>
<td>HE</td>
<td>8: P20280: 60S ribosomal protein</td>
</tr>
<tr>
<td>284:</td>
<td>432</td>
<td>HE</td>
<td>9: P20280: 60S ribosomal protein</td>
</tr>
<tr>
<td>289:</td>
<td>1246</td>
<td>HE</td>
<td>3: Q5R7Y8: 60S ribosomal protein L8</td>
</tr>
<tr>
<td>306:</td>
<td>2521</td>
<td>HE</td>
<td>11: P47840: 40S ribosomal protein S12</td>
</tr>
<tr>
<td>307:</td>
<td>312</td>
<td>HE</td>
<td>6: O57592: 60S ribosomal protein L17a</td>
</tr>
</tbody>
</table>

**Ribosomal proteins**
Translation factors

190: 328  LV  16  16  0:  P61220: translation initiation factor 1b
FUNCTION: Probably involved in translation.

207: 1823  HE  15  4  11:  O55135: Eukaryotic translation initiation factor 6 (eIF-6)
FUNCTION: Binds to the 60S ribosomal subunit and prevents its association with the 40S ribosomal subunit to form the 80S initiation complex

243: 980  HE  13  4  9:  Q0ULD0: Nascent polypeptide-associated complex subunit beta
FUNCTION: Component of the nascent polypeptide-associated complex (NAC), a dynamic component of the ribosomal exit tunnel, protecting the emerging polypeptides from interaction with other cytoplasmic proteins to ensure appropriate nascent protein targeting (By similarity). The NAC complex also promotes mitochondrial protein import by enhancing productive ribosome interactions with the outer mitochondrial membrane and blocks the inappropriate interaction of ribosomes translating non-secretory nascent polypeptides with translocation sites in the membrane of the endoplasmic reticulum (By similarity). EGD1 may act as a transcription factor that exert a negative effect on the expression of several genes that are transcribed by RNA polymerase

261: 1738  HE  12  1  11:  P13549: Elongation factor
FUNCTION: This protein promotes the GTP-dependent binding of aminoacyl-tRNA to the A-site of ribosomes during protein biosynthesis.
ATP synthesis

70: 964  HE  39  4  35: Q5RAP9: ATP synthase lipid-binding protein
FUNCTION: This protein is one of the chains of the nonenzymatic membrane component (F0) of mitochondrial ATPase.

329: 99  LV  10  9  1: P34546: Vacuolar ATP synthase

340: 2217  AD  10  0  10: P05630: ATP synthase delta chain
-S-S- redox proteins

84: 910  HE  35  22  13: O14463: Thioredoxin
FUNCTION: Participates in various redox reactions through the reversible oxidation of its active center dithiol to a disulfide and catalyzes dithiol-disulfide exchange reactions.

115: 2484  HE  27  10  17: Q8T6C4: Thioredoxin peroxidase
FUNCTION: Reduces peroxides with reducing equivalents provided through the thioredoxin system. It is not able to receive electrons from glutaredoxin. May play an important role in eliminating peroxides generated during metabolism. Might participate in the signaling cascades of growth factors and tumor necrosis factor-alpha by regulating the intracellular concentrations of H(2)O(2) (By similarity).
CATALYTIC ACTIVITY: 2 R'-SH + ROOH = R'-S-S-R' + H(2)O + ROH.

249: 582  HE  13  4  9: P35705: Thioredoxin-dependent peroxide reductase (Peroxiredoxin)
FUNCTION: Involved in redox regulation of the cell. Protects radical-sensitive enzymes from oxidative damage by a radical- generating system.

321: 1179  HE  10  4  6: Q17770: Protein disulfide-isomerase
CATALYTIC ACTIVITY: Catalyzes the rearrangement of -S-S- bonds in CC proteins.
Glutathione S-transferase

FUNCTION: Conjugation of reduced glutathione to a wide number of exogenous and endogenous hydrophobic electrophiles

96: 1524  AD  32  2  30: O35660: Glutathione S-transferase

157: 1245  AD  19  1  18: O35660: Glutathione S-transferase

168: 1543  AD  18  0  18: Q9N0V4: Glutathione S-transferase

215: 772  LV  14  14  0: P10299: Glutathione S-transferase P
Heat shock proteins

FUNCTION: Molecular chaperone. Has ATPase activity

146: 71     HE    20    16     4:  P27541: Heat shock protein 70 kDa protein

162: 112    HE    18     5    13:  Q5ZM98: Stress-70 protein, mitochondrial precursor
FUNCTION: Implicated in the control of cell proliferation and cellular aging. May also act as a chaperone

245: 84     HE    13    10     3:  P31689: DnaJ (Chaperone protein)
FUNCTION: Co-chaperone of Hsc70. Seems to play a role in protein import into mitochondria.
**Histones**

116: 943    HE  26    3    23: Q10453: Histone H3.3 type 1

FUNCTION: Variant histone H3 which replaces conventional H3 in a wide range of nucleosomes in active genes. Constitutes the predominant form of histone H3 in non-dividing cells and is incorporated into chromatin independently of DNA synthesis. Deposited at sites of nucleosomal displacement throughout transcribed genes, suggesting that it represents an epigenetic imprint of transcriptionally active chromatin. Nucleosomes wrap and compact DNA into chromatin, limiting DNA accessibility to the cellular machineries which require DNA as a template. Histones thereby play a central role in transcription regulation, DNA repair, DNA replication and chromosomal stability. DNA accessibility is regulated via a complex set of post-translational modifications of histones, also called histone code, and nucleosome remodelling.
Protein trafficking

165: 344 HE 18 2 16: Q5E971: 21 kDa transmembrane-trafficking protein
FUNCTION: Involved in vesicular protein trafficking

260: 960 HE 12 6 6: P51823: ADP-ribosylation factor
FUNCTION: GTP-binding protein that functions as an allosteric activator of the cholera toxin catalytic subunit, an ADP-ribosyltransferase. Involved in protein trafficking; may modulate vesicle budding and uncoating within the Golgi apparatus.


**Signaling**

171: 455  HE  18  12  6:  Q26537: 14-3-3 protein homolog
SIMILARITY: Belongs to the 14-3-3 family.

290: 60   HE  11  7  4:  O49998: 14-3-3-like protein F

347: 857  HE  10  6  4:  Q5ZKC9: 14-3-3 protein zeta
FUNCTION: Adapter protein implicated in the regulation of a large spectrum of both general and specialized signaling pathway. Binds to a large number of partners, usually by recognition of a phosphoserine or phosphothreonine motif. Binding generally results in the modulation of the activity of the binding partner

181: 1969  AD  16  0  16:  P92177: 14-3-3 protein epsilon
FUNCTION: Positively regulates Ras-mediated pathways. Acts downstream or parallel to Raf, but upstream of nuclear factors in Ras signaling. Three mutants have been isolated, that suppress the rough eye phenotype caused by mutated Ras1
Tricarboxylic acid cycle

81: 138  HE  36  19  17: Q04820: Malate dehydrogenase
CATALYTIC ACTIVITY: (S)-malate + NAD(+) = oxaloacetate + NADH.

246: 1425  AD  13  0  13: O87840: Succinyl-CoA synthetase beta chain
CATALYTIC ACTIVITY: ATP + succinate + CoA = ADP + phosphate + succinyl-CoA.

270: 220  HE  12  7  5: P21912: Succinate dehydrogenase [ubiquinone] iron-sulfur protein
CATALYTIC ACTIVITY: Succinate + ubiquinone = fumarate + ubiquinol.

324: 570  HE  10  3  7: P21912: Succinate dehydrogenase
Signal transduction pathways

251: 96 HE 12 9 3: P24406: Transforming protein RhoA precursor (Rho1)
FUNCTION: Regulates a signal transduction pathway linking plasma membrane receptors to the assembly of focal adhesions and actin stress fibers. May be an activator of PLCE1.
Placental proteins!!

268: 576 HE 12 3 9: P21128: Placental protein 11 precursor
FUNCTION: Probable serine protease.
Pentose-phosphate pathway

304: 720  

FUNCTION: Transaldolase is important for the balance of metabolites in the pentose-phosphate pathway.


PATHWAY: Carbohydrate degradation; pentose phosphate pathway; D-fructose 6-phosphate and D-glyceraldehyde 3-phosphate from D-ribose 5-phosphate and D-xylulose 5-phosphate (non-oxidative stage): step 2.
Lipid synthesis

319: 681  HE  10   6   4:  O35547: Long-chain-fatty-acid--CoA ligase
FUNCTION: Activation of long-chain fatty acids for both synthesis of cellular lipids, and degradation via beta-oxidation.
tRNA synthetases

334: 146  HE  10  3  7: P17248: Tryptophanyl-tRNA synthetase
Miscellaneous functions

72: 2334  AD  39  0  39:  Q8CBW7: Cysteine-rich hydrophobic domain 1 protein

78: 1467  AD  36  0  36:  O75828: Carbonyl reductase [NADPH] 3
CATALYTIC ACTIVITY: R-CHOH-R' + NADP(+) = R-CO-R' + NADPH.

92: 215  HE  33  10  23:  P07943: Aldose reductase
FUNCTION: Catalyzes the NADPH-dependent reduction of a wide variety of carbonyl-containing compounds

101: 1775  AD  29  0  29:  Q99LM2: CDK5 regulatory subunit-associated protein 3
FUNCTION: Potential regulator of CDK5 activity. May be involved in cell proliferation.

118: 1972  HE  26  6  20:  P97315: Cysteine and glycine-rich protein
FUNCTION: Could play a role in neuronal development.

121: 2034  AD  25  0  25:  P37111: Aminoacylase-1
FUNCTION: Involved in the hydrolysis of N-acylated or N-acetylated amino acids (except L-aspartate).

136: 1754  HE  21  2  19:  Q5RCU5: Carbonyl reductase
FUNCTION: Catalyzes the reduction of a wide variety of carbonyl compounds.

159: 76  HE  19  11  8:  P18238: ADP,ATP carrier protein
FUNCTION: Catalyzes the exchange of ADP and ATP across the mitochondrial inner membrane.

160: 219  HE  19  12  7:  P37805: Transgelin-3  Abundant and ubiquitous expression in neurons
201: 248 AD 15  1  14: Q9DCV4: Protein FAM82B.

204: 811 HE 15  3  12: Q13310: Polyadenylate-binding protein 4
FUNCTION: Binds the poly(A) tail of mRNA.

205: 244 HE 15  6  9: O46119: Ferritin heavy chain
FUNCTION: Stores iron in a soluble, non-toxic, readily available form.

219: 2318 AD 14  0  14: Q27245: Putative aminopeptidase.

233: 1863 AD 13  0  13: O88986: 2-amino-3-ketobutyrate coenzyme A ligase.

285: 756 HE 11  5  6: Q969X1: Transmembrane BAX inhibitor motif-containing protein

330: 414 HE 10  3  7: O94272: Autophagy-related protein 8
FUNCTION: Involved in cytoplasm to vacuole transport (Cvt) vesicles and autophagosomes formation.

342: 694 HE 10  8  2: Q9DCT1: Aldo-keto reductase family 1 member
FUNCTION: Catalyzes the reduction of various aldehydes and quinones.

344: 104 HE 10  5  5: Q4R596: Adenosylhomocysteinase
FUNCTION: Adenosylhomocysteine is a competitive inhibitor of S-adenosyl-L-methionine-dependent methyl transferase reactions;
CATALYTIC ACTIVITY: S-adenosyl-L-homocysteine + H(2)O = L-homocysteine + adenosine.