

Promoting Cross-Species Comparative Approaches to Environmental Health Research: The Comparative Toxicogenomics Database (CTD)

Carolyn J. Mattingly¹, Glenn T. Colby¹, Michael C. Rosenstein¹,
John N. Forrest, Jr.^{1,2}, James L. Boyer^{1,2}

¹Mount Desert Island Biological Laboratory, Salisbury Cove, ME 04672; ²Department of
Medicine, Yale University School of Medicine, New Haven, CT 06520

The etiology of most chronic diseases involves interactions between environmental factors and genes that modulate important physiological processes (Olden and Wilson, 2000).⁴ We are developing a publicly available database, the Comparative Toxicogenomics DatabaseTM (CTDTM; <http://ctd.mdibl.org/>), to promote understanding about the effects of environmental chemicals on human health.^{2,3} CTD identifies interactions between chemicals and genes and facilitates cross-species comparative studies of these genes. The use of diverse animal models and cross-species comparative sequence studies is critical for understanding basic physiological mechanisms and gene and protein functions. These approaches are also being used to explore the molecular mechanisms of action of environmental chemicals and the genetic basis of differential susceptibility.⁴

ctd — The Comparative Toxicogenomics Database™

Home | Sequences | References | Chemicals | Gene Sets | Help | Feedback

Welcome | Overview | Publications | Legal Notices | Privacy Policy

Welcome

The Comparative Toxicogenomics Database (CTD) prototype identifies interactions between chemicals and genes in diverse organisms to advance understanding of how environmental chemicals affect human health.

```

graph TD
    Chemicals --> Reference
    Chemicals --> Sequence
    Chemicals --> GO[Gene Ontology for proteins]
    GeneSets[Gene Sets and Genes] --> Reference
    GeneSets --> Sequence
    GeneSets --> GO
    SpeciesTax[Species Taxonomy] --> Reference
    SpeciesTax --> Sequence
    SpeciesTax --> GO
    Reference --> Sequence
    Reference --> GO
    Sequence --> GO
  
```

CTD integrates and curates gene, sequence, chemical, reference, taxonomic and Gene Ontology data to support your hypotheses about gene-chemical interactions. [more >](#)

Get Answers

1. Which genes are affected by this chemical?
2. Which chemicals interact with this gene?
3. Which references report this gene-chemical interaction?
4. In which organisms has this gene-chemical interaction been studied?
5. Which regions of this toxicologically important protein are conserved in vertebrates and invertebrates?
6. Which cellular functions are affected by this chemical?

Send Feedback | Legal Notices | Privacy Policy | Back to Top

CTD is funded by NIEHS grants ES11267-02 and ES03828-17.
Copyright © 2004 Mount Desert Island Biological Laboratory. All Rights Reserved.

News: October 13, 2004

- ▶ Help us make CTD better: please send feedback as you use the site.
- ▶ Newly curated **Gene Sets** are available.
- ▶ **Gene Sets** now include multiple alignments and phylogenetic trees for all included sequences.
- ▶ Explore our **Chemicals** area, including new chemical detail pages with direct links to TOXNET resources.
- ▶ **Sequences** are now searchable by hierarchical **Gene Ontology** terms and IDs.
- ▶ View current **data integration** status.

Email List

Subscribe to CTD's email news list:

Figure 1. CTD Home Page.

for toxicologically important genes), 4) Gene Sets (sets of curated genes), 5) chemicals (hierarchical vocabulary of chemicals or xenobiotic agents), 6) Gene Ontology terms (GO; hierarchical vocabulary of biological processes, cellular components, and molecular functions), and 7) taxonomy (hierarchical vocabulary of taxa representing taxonomic groups (Figure 2). Nucleotide sequences and annotations are acquired from the National Center for Biotechnology Information. We include only Reference Sequences (RefSeqs) for human (*H. sapiens*), mouse (*M. musculus*), rat (*R. norvegicus*), fruit fly (*D. melanogaster*), and nematode (*C. elegans*). Amino acid sequences and annotations are acquired from

A prototype version of CTD is available via the World Wide Web (Figure 1). Although publicly accessible, improvements are in progress for data curation, data integration, and web site usability. Here, we summarize the major features (entities) of CTD. There are seven *primary* entities that have been integrated in CTD: 1) sequences (nucleotide and protein sequences from vertebrates and invertebrates), 2) references (reference publications), 3) genes (curated, cross-species groups of nucleotide and protein sequences

the European Bioinformatics Institute's Swiss-Prot and TrEMBL databases. References are acquired from PubMed®.

CTD data are linked to 26 other sequence, protein domain, and toxicology databases (e.g., Swiss-Prot, Pfam, and TOXNET®, respectively). We provide two-dimensional chemical drawings and links to regulatory and toxicology data for approximately 53,000 of the chemical terms in our vocabulary. We created "vocabulary browsers" with detail pages in CTD that allow users to navigate the hierarchical structures of controlled vocabularies (chemicals, taxonomy, and GO) to formulate queries or quickly access data (Figure 3).

Genes and Gene Sets are manually curated in CTD to promote cross-species comparisons of toxicologically important genes and proteins. In CTD, genes are defined by their constituent nucleotide and protein sequences from vertebrates and invertebrates and are presented in a cross-species context. We use sequence analysis methods in combination with literature review to curate genes. We developed the concept of a Gene Set to group closely related, curated genes, such as those that have undergone duplication events in specific species

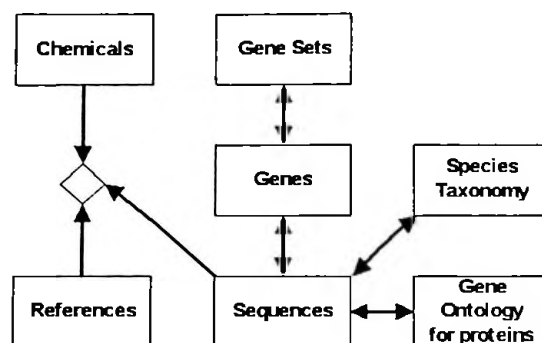


Figure 2. High-Level View of the Primary Entities in CTD. Lines indicate a relationship between two entities. Single- and double-headed arrows indicate one-to-many and many-to-many relationships, respectively. Lines that meet at a diamond indicate a multi-way association among several entities.

(e.g., CYP1A4, CYP1A5) or are members of large families (e.g., ABC transporters). Gene Sets provide the user with a broad perspective about their gene of interest. For example, the CYP1A Gene Set includes the curated genes CYP1A, CYP1A1, CYP1A2, CYP1A3, CYP1A4, and CYP1A5. By combining these genes, a user familiar only with mammalian CYP1A1 and CYP1A2 genes is introduced to the avian CYP1A4 and CYP1A5 genes and associated supplementary information (sequences, references, associated chemicals, and GO terms). Multiple alignments and phylogenetic trees are constructed from sequences of curated Gene Sets and may be downloaded from Gene Set detail pages (Figure 4). These files assist users in evaluating conservation and divergence in toxicologically important sequences among diverse organisms and in developing hypotheses about the

ctd - The Comparative Toxicogenomics Database

Home | Sequences | References | Chemicals | Gene Sets | Help | Feedback

Chemicals: arsenic trioxide

Chemicals | contains | arsenic trioxide | Search

[Jump to: General Info | Other Data Resources | Hierarchies]

General Information

Name: arsenic trioxide

Chemical Synonyms: arsenic(III) oxide, arsenous anhydride, diarsenic trioxide, Trisenox, Trisenox

CAS Registry No.: 1327-53-3

Chemical Drawing: $\left[\text{As}^{3+} \right]_2 \left[\text{O}^{2-} \right]_3$

Related CTD Data: 36 sequences, 8 references

Other Data Resources

ChemIDplus: 1327-53-3 - Chemical Report [ChemIDplus]

Chemical Carcinogenesis Research Information System: 1327-53-3 - Chemical Report [CCRIS]

GENE-TOX: 1327-53-3 - Chemical Report [GENE-TOX]

Hazardous Substances Data Bank: 1327-53-3 - Chemical Report [HSDB]

NCBI MeSH: C006632 - MeSH term [MeSH]

Substance Registry System: 1327-53-3 - Substance List [SRS]

TOXLINE Core: arsenic trioxide / 1327-53-3 - [Search TOXLINE Core]

TOXLINE Special: arsenic trioxide / 1327-53-3 - [Search TOXLINE Special]

Hierarchies

[Jump to: Path 1 | Path 2 | Path 3 | Path 4 | ...]

Path 1: Chemicals > Inorganic Chemicals > Arsenicals

arsenic trioxide
[chemTree:G.D01.075/C006632]

36 sequences
8 references

Figure 3. CTD Chemical Browser Detail Page.

functions of these genes in modulating chemical actions.

ctd The Comparative Toxicogenomics Database

Home | Sequences | References | Chemicals | Gene Sets

AHR (Aryl Hydrocarbon Receptor) Gene Set

(Jump to: Description | Included Genes | Alignments)

Description

Aryl hydrocarbon receptor is a ligand-activated transcription factor involved in the regulation of the response to aromatic hydrocarbons. AHR has been shown to regulate xenobiotic-metabolism and has been shown to be involved in a variety of aromatic hydrocarbons. (from Entrez Gene)

Included Genes

Gene	Synonyms	Related Sequences
1. AHR (Aryl hydrocarbon receptor precursor)	AH RECEPTOR, ARYL HYDROCARBON RECEPTOR, ARYL HYDROCARBON RECEPTOR	59
2. AHR1 (Aryl hydrocarbon receptor 1)	AHR1, AHR3	16
3. AHR1 beta (Aryl hydrocarbon receptor 1 beta)	AHR1BETA	2
4. AHR2 (Aryl hydrocarbon receptor 2)		8
5. AHR2 alpha (Aryl hydrocarbon receptor 2 alpha)	AHR2ALPHA	1
6. AHR2 beta (Aryl hydrocarbon receptor 2 beta)	AHR2BETA	1
7. AHR2 gamma (Aryl hydrocarbon receptor 2 gamma)		2
8. AHR2 delta (Aryl hydrocarbon receptor 2 delta)		2

Alignments and Phylogenetic Trees

Alignment | Aligned Sequences | Max. Parsimony | Max. Likelihood

Nucleotides | Amino Acids | Sequences | Phylogenetic Trees

Sequences

Gene	Organism	Chemicals	Mol Type	Gene Category (GO)	Accession/Description
1. AHR (Aryl hydrocarbon receptor precursor)	<i>Caenorhabditis elegans</i> (nematode)		PROT	None	Q04712 Aryl hydrocarbon receptor precursor
	<i>Canis familiaris</i> (dog)	* Tetrachlorodibenzo-p-dioxin	mRNA	N	
	<i>Canis familiaris</i> (dog)		PROT	* Receptor * Receptor * Receptor	

Figure 4. AHR Gene Set Page. Member genes are indicated, with links to sequence analysis results, sequences, and chemical and GO annotations.

From the CTD home page, users can initiate searches with sequence or reference query forms or by browsing the chemical vocabulary. Queries of varying complexity for the novice and experienced molecular toxicologist are possible. Examples of supported queries include: Which chemicals interact with my gene of interest? Which genes are affected by my chemical of interest? Which regions of my favorite toxicologically important protein are conserved in vertebrates and invertebrates? Which references report information about a particular gene-chemical interaction? In which organisms has a particular gene-chemical interaction been studied? Are the proteins affected by my chemical of interest involved in a particular biological process (e.g., apoptosis)? Priorities for future development include expanding the set of references in CTD, curating specific types of gene-chemical interactions described in the literature (e.g., protein "X" binds chemical "Y"), and continuing to curate genes and Gene Sets. The community is encouraged to participate in CTD

development by providing feedback (ctd@mdibl.org). Please contact us if you are interested in submitting curated data sets for inclusion in CTD.

This project is funded by NIEHS ES11267 and is a component of the MDIBL CMTS.

1. Ballatori N., J.L. Boyer, J.C. Rockett. Exploiting genome data to understand the function, regulation, and evolutionary origins of toxicologically relevant genes. *Environ Health Perspect Toxicogenomics*. 111: 61-5, 2000
2. Mattingly, C.J., G.T. Colby, J.N. Forrest Jr., and J.L. Boyer. The Comparative Toxicogenomics Database (CTD). *Environ Health Perspect*. 111: 793-795, 2003.
3. Mattingly, C.J., G.T. Colby, M.C. Rosenstein, J.N. Forrest Jr., and J.L. Boyer. Promoting comparative molecular studies in environmental health research: an overview of the comparative toxicogenomics database (CTD). *Pharmacogenomics J*. 4: 5-8, 2004.
4. Olden, K. and S. Wilson. Environmental health and genomics: visions and implications. *Nat Rev Genet*. 1: 149-153, 2000.