

Numbers of putatively functional full-length *CYP* genes in 54 whole eukaryotic genomes

August 9, 2004 D. Nelson

[Vertebrates \(deuterostomes\)](#)

<i>Homo sapiens</i> (human)	57	Apr. 2003
<i>Pan troglodytes</i> (chimp)	? (probably 57 or 58)	
<i>Mus musculus</i> (mouse)	102	Feb. 2003 build 30
<i>Rattus norvegicus</i> (rat)	?	
<i>Canis familiaris</i> (dog)	54	Sep. 2003
<i>Bos taurus</i> (cow)	at least 53	
<i>Gallus gallus</i> (chicken)	at least 41	
<i>Takifugu rubripes</i> (pufferfish)	54	Aug. 2002 v.3.0
<i>Tetraodon nigroviridis</i> (freshwater puffer)	?	
<i>Danio rerio</i> (zebrafish)	at least 81	

[Urochordates \(deuterostomes\)](#)

<i>Ciona intestinalis</i> (sea squirt)	80	Dec. 2002
<i>Ciona savignyi</i> (sea squirt)	97	Apr. 2003 release 1

[Echinoderms \(deuterostomes\)](#)

Strongylocentrotus purpuratus (purple sea urchin) ?

Genome size = 870Mb, based on haploid genome of 0.89pg X 0.978 pg/Gb

White paper states genome size is about 800Mb.

At Caltech:

Database: purpuratus-complete

Posted date: June 11, 2004

Number of letters in database: 67,056,103 about 8%

Number of sequences in database: 89,094

Database: BAC-end sequence tag connectors (STCs)

Number of letters in database: 46,412,324 about 5% of the genome

Number of sequences in database: 76,020

Database: purpuratus-pmc-ests (ESTs of primary mesenchyme cells)

10,206 sequences; 7,908,988 total letters

at Baylor: (BLASTN only)
Posted date: Sep 25, 2003
Spurpuratus/genome/Spur20030922-genome
90,680 sequences; 269,576,112 total letters, about 30% of genome

at NCBI:

over 8.1 million traces in shotgun and WGS (at about 500bp/read)

This may be about 4Gb or about 4-5X coverage

Insects (protostomes)

Drosophila melanogaster (fruit fly) 84 Mar. 2000

Drosophila pseudoobscura (a second fly) 79

Drosophila simulans <http://www.genome.wustl.edu/blast/client.pl>

of letters in database: 155,439,225

of sequences in database: 98,443

Drosophila yakuba <http://www.genome.wustl.edu/blast/client.pl>

of letters in database: 169,400,961

of sequences in database: 23,959

Anopheles gambiae (mosquito) 105 Oct. 2002

Bombyx mori (silkworm) ~79

Apis mellifera (honeybee) ~60

Nematodes (protostomes)

Caenorhabditis elegans (nematode) 74 Dec. 1998

Caenorhabditis briggsae (a second nematode) ?

Mycetozoa

Dictyostelium discoideum (slime mold) 42 Apr. 2003

Plants

Arabidopsis thaliana (thale cress) 249 Dec. 2000

Oryza sativa (rice) 323 Apr. 2002

Chlamydomonas reinhardtii (green algae) at least 33

Cyanidioschyzon merolae (red algae) 5

Fungi

Neurospora crassa (bread mold) 38 Apr. 2003

Saccharomyces cerevisiae (baker's yeast) 3 Oct. 1996

Schizosaccharomyces pombe (fission yeast) 2 Feb. 2002

Fusarium graminearum (plant pathogen, fungi) 110

Magnaporthe grisea (rice blast fungus) 120

Aspergillus nidulans (plant pathogen, fungi) 111

Aspergillus fumigatus (filamentous fungi) ?

Candida albicans ?

<http://sequence-www.stanford.edu/group/candida/index.html>

Coprinus cinereus 10X coverage

http://www.broad.mit.edu/annotation/fungi/coprinus_cinereus/whatsnew.html

Ustilago maydis 10X coverage

http://www.broad.mit.edu/annotation/fungi/ustilago_maydis/whatsnew.html

Cryptococcus neoformans 11X coverage 24Mb

http://www.broad.mit.edu/annotation/fungi/cryptococcus_neoformans/index.html

Encephalitozoon cuniculi 0

Botrytis cinerea = *Botryotinia fuckeliana* (fungi) 5X ?

sequenced at Syngenta (private), cDNA project at Genoscope

Natalie L. Catlett, Olen C. Yoder, and B. Gillian Turgeon*

Whole-Genome Analysis of Two-Component Signal Transduction Genes in Fungal Pathogens

Eukaryotic Cell, December 2003, p. 1151-1161, Vol. 2, No. 6

Torrey Mesa Research Institute/Syngenta Research and Technology, San Diego, California 92121

Gibberella moniliformis = *Fusarium verticillioides* 5X ?

sequenced at Syngenta (private)

Natalie L. Catlett, Olen C. Yoder, and B. Gillian Turgeon*

Whole-Genome Analysis of Two-Component Signal Transduction Genes in Fungal Pathogens

Eukaryotic Cell, December 2003, p. 1151-1161, Vol. 2, No. 6

Torrey Mesa Research Institute/Syngenta Research and Technology, San Diego, California 92121

Cochliobolus heterostrophus = *Bipolaris maydis* 5X ?

sequenced at Syngenta (private)

Natalie L. Catlett, Olen C. Yoder, and B. Gillian Turgeon*

Whole-Genome Analysis of Two-Component Signal Transduction Genes in Fungal Pathogens

Eukaryotic Cell, December 2003, p. 1151-1161, Vol. 2, No. 6

Torrey Mesa Research Institute/Syngenta Research and Technology, San Diego, California 92121

Alveolates/Ciliates (free living)

Tetrahymena thermophila (ciliate) 48

Paramecium tetrauralia (ciliate) 22

Alveolates/Apicomplexan parasites

Plasmodium falciparum (malaria) 0

Stramenopiles, heterokonts, Chromista

Phytophthora ramorum (Sudden Oak Death) 24

Phytophthora sojae (stramenopile) 29

Thalassiosira pseudonana (centric marine diatom) 10

Euglenozoans

Trypanosoma cruzi (Chagas disease, euglenozoan) 3

Trypanosoma brucei (African sleeping sickness) 2

Trypanosoma vivax (euglenozoan) 2

Trypanosoma congolense (euglenozoan) 2

Leishmania major (Leishmaniasis organism, euglenozoan) 4

Leishmania infantum (euglenozoan) 4

Parabasala, Archaezoa

Giardia lamblia (hiker's diarrhea) 0