

Genome Statistics

Genome Statistics under Compare Genomes is consist of 2 parts:

- *Summary Statistics*, which shows the count of all genomes in IMG;
- *General Statistics*, which shows statistics per genome.

Summary Statistics

Summary statistics shows the total count of coding bases, scaffolds and genes in the IMG database. The statistics in IMG/M ER shows the total counts of all public and private genomes. Therefore, the numbers of genes and scaffolds a general user can access are less than the total counts shown on this page.

The statistics also shows the breakdown of protein coding genes based on various functional annotations such as COG and Pfam (see Figure 1).

Summary Statistics		
	Number	% of Total
DNA, total number of bases	12395631088558	100.00%
DNA coding number of bases	299839437626	2.42%
DNA G+C number of bases	6881106923813	55.51% ¹
DNA scaffolds	47946850868	
Genes total number	49983988345	
Protein coding genes	49629769139	
Pseudo Genes	457936	
Uncharacterized Genes	20695	
RNA genes	354166167	
rRNA genes	204178861	
5S rRNA	5348438	
16S rRNA	75816559	
18S rRNA	332	
23S rRNA	123011224	
28S rRNA	495	
tRNA genes	124443293	
Other RNA genes	25523122	
Protein coding genes with function prediction	21874478647	
without function prediction	27755290492	
Protein coding genes with enzymes	8610261964	
w/o enzymes but with candidate KO based enzymes	9080490	
Protein coding genes connected to KEGG pathways ³	8485138386	
not connected to KEGG pathways	41078268923	82.18%
Protein coding genes connected to KEGG Orthology (KO)	13976774406	27.96%
not connected to KEGG Orthology (KO)	35597859593	71.22%
Protein coding genes connected to MetaCyc pathways	5579367710	11.16%
not connected to MetaCyc pathways	43970375497	87.97%
Protein coding genes with COGs ³	16458835310	32.92%
with KOGs ³	38668179	0.08%
with Pfam ³	16753268611	33.52%
with TIGRfam ³	70819115	0.14%
with InterPro	114731583	0.23%
with IMG Terms	42682585	0.09%
with IMG Pathways	15508321	0.03%
with IMG Parts List	15313025	0.03%
with MyIMG Annotation	49024	0.00%
in internal clusters	128082993	0.26%
Fused Protein coding genes	9784008	0.02%
Protein coding genes coding signal peptides	47100206	0.09%
Protein coding genes coding transmembrane proteins	67234617	0.13%
Obsolete Protein coding genes	3193	0.00%
Revised Genes	7000	0.00%

Figure 1. Summary Statistics: The left hand side shows the first part of the summary statistics, and the right hand side shows the second part of the statistics (scroll down the window to view it).

General Statistics

General statistics shows statistics per genome. Only the genomes that a user has access to will be displayed. The default display only shows domain, sequencing status, genome name, total gene count, GC percentage and total number of bases per genome (as shown in Figure 2(i)). Users can scroll down to find the **Configuration** section to add more statistics fields to the table display (Figure 2(ii)). To add additional fields to the display, first select fields of interest, and then click the "Display Genomes Again" button below the field selection. It is also possible for users to export the statistics table (see Figure 2(ii)) or to select genomes of interest to be added to the Genome Cart for further analysis.

(i)

Select	Domain	Status	Genome Name	Total gene count	GC (percentage)	Total number of bases
<input type="checkbox"/>	A	F	Acidianus hospitalis W1	2404	34%	2137654
<input type="checkbox"/>	A	D	Acidianus sp. DS-80	2713	34%	2233448
<input type="checkbox"/>	A	D	Acidianus sp. DS-80 - draft draft	12271	56%	10699091
<input type="checkbox"/>	A	F	Acidilobus saccharovorans 345-15	1547	57%	1496453
<input type="checkbox"/>	A	D	Acidilobus sulfurireducens sp. nov.	1925	58%	1648899
<input type="checkbox"/>	A	D	Acidilobus-DT1-04 (from MC-3)	1755	59%	1548506
<input type="checkbox"/>	A	D	Acidiplasma ae		36%	1959402

Export Genome Table

Export Tab Delimited To Excel

(ii)

Configuration

Show	Column Name
<input type="checkbox"/>	Taxon Object ID
<input type="checkbox"/>	NCBI Taxon ID
<input type="checkbox"/>	Phylum
<input type="checkbox"/>	Class
<input type="checkbox"/>	Order
<input type="checkbox"/>	Family
<input type="checkbox"/>	Genus
<input type="checkbox"/>	Proposal GOLD ID
<input checked="" type="checkbox"/>	Total gene count
<input type="checkbox"/>	Number of CDS genes
<input type="checkbox"/>	CDS genes (percentage)
<input type="checkbox"/>	Number of RNA genes

Figure 2. General Statistics.