



Search datasets to begin... (e.g. Nematoda)

- Use the search box above to find datasets - matching datasets and associated metadata will be displayed in a sortable table.
- If you are unsure what to search for, browse available datasets below or type 'all' to show all available datasets.

Browse datasets:

superkingdom	kingdom
Eukaryota (10774/13554)	Fungi (4562/6697)
	Metazoa (4098/4254)
	Vrirdiplantae (1222/1623)
	Other Eukaryota (892/980)

- show counts for assemblies or species.
- Click a taxon name to list all assemblies in that taxon.
- Numbers indicate available assemblies / total INSDC registered assemblies, click a number to expand taxonomy.

Click on the taxon name to see all datasets for that taxon

Type a taxon name in the search box, will auto-complete

Click on the numbers to expand the taxonomy tree

When you first open BTK Viewer you will only see the **Datasets**, **Help**, and **About** Top menus. When you choose an assembly to view under **Datasets**, you will see the rest of the Top Menus

Top menu	Description	Notes
Datasets	To search for all datasets available in BlobToolKit	Search/browse by species name, or browse by higher level taxon. Summary tables of extra information for each genome assembly that you can sort/filter by
Filters	Change WHAT is shown in a BTK plot	IMPORTANT: Change axes, categories, and taxonomic colouring
Lists	To create lists of sequences (for saving, subsequent filtering)	
Settings	Change HOW a BTK plot looks	IMPORTANT: Type of BTK plots (square or hexagonal grid, sequences as circles, kite plots, line plots, grid plots), Font sizes, Axis limits, Static/Interactive/Display settings. PNG resolution for downloading plots.
Summary	Summary information for each category	Expands the contents of the BTK plot legend in detail. Also allows for click and drag re-ordering of categories.
Help	Frequently Asked Questions	
About	Back to main site homepage	Links to installation, components, publication

Identifiers			Assembly statistics			BUSCO		Coverage
Dataset ID	Taxon	Accession	Sequences	Span (bp)	N50 (bp)	Closest lineage	Closest string	Read sets
AAGJ05	Strongylocentrotus purpuratus	GCA_000002235.3	31,896	990,859,639	419,550	metazoa_odb9	C.93.8%(S:85.7%,D:8.1%),F:1.9%,M:4.3%,n:978	13
APCI01	Anopheles funestus	GCA_000349085.1	1,392	225,223,604	671,960	diptera_odb9	C.98.5%(S:97.9%,D:0.5%),F:1.0%,M:0.6%,n:2799	13
CAAAFJ01.1	Cottoperca gobio	GCA_900634415.1	322	609,391,784	25,156,145	actinopterygii_odb9	C.92.8%(S:90.2%,D:2.7%),F:2.2%,M:4.9%,n:4584	13
CAAAFJ01.1	Cottoperca gobio	GCA_900634435.1	11,555	671,265,480	107,591	actinopterygii_odb9	C.80.6%(S:65.3%,D:1.53%),F:4.2%,M:15.2%,n:4584	13
CABFWM01	Aquila chrysaetos chrysaetos	GCA_902153765.1	3,860	823,060,160	659,735	aves_odb9	C.62.2%(S:60.3%,D:1.0%),F:4.0%,M:33.0%,n:4915	13
JIRV01	Calypte anna	GCA_000699085.1	54,736	1,105,676,412	4,652,191	aves_odb9	C.93.0%(S:92.1%,D:1.0%),F:4.0%,M:2.9%,n:4915	13
RCWQ01	Anopheles funestus	GCA_003951495.1	9,175	444,543,955	239,261	diptera_odb9	C.98.7%(S:60.3%,D:38.4%),F:0.7%,M:0.6%,n:2799	13
mLuLut1_1	Lutra lutra	submitted	43	2,438,442,342	149,004,807	euarchoctogires_odb9	C.92.7%(S:91.3%,D:1.4%),F:3.5%,M:3.8%,n:1192	13
mSciCar1_1	Sciurus carolinensis	submitted	752	2,815,397,268	148,229,995	euarchoctogires_odb9	C.91.8%(S:90.5%,D:1.2%),F:3.7%,M:4.5%,n:1192	13
MRSW03.1	Taenia multiceps	GCA_001923025.3	738	240,610,560	44,815,576	metazoa_odb9	C.61.2%(S:59.8%,D:1.4%),F:9.8%,M:28.9%,n:978	12

- Click headers to sort results.
- Click a row to view an assembly.
- Type in the box at the top of each column to filter assemblies.
- Customise table

Click to add columns

If Read sets = 0, blob view will not be available

Identifiers: Dataset ID Taxon Accession

Assembly statistics: GC (%) AT (%) N (%) Sequences Span (bp) N50 (bp) | Taxon Value

BUSCO: Closest lineage Closest string | Lineage Value

Taxonomy: Lineage Taxon ID Target Rank Target taxon

Summary statistics: No hit (%) Hits matching target (%) N50:Span ratio

Coverage: Read sets

The **Views** submenu shows up when you are viewing an assembly, and when you select **Filters**, **Lists**, **Settings**, or **Summary** from the Top menu

Views submenu	Description	Notes
blob	Taxon-annotated GC-coverage plots (default)	IMPORTANT: Axes can be changed using Filters: numeric variable attributes, and the colours can be changed using Filters: categorical attributes
busco	BUSCO completeness plots for each relevant BUSCO lineage for that species	BTK also shows BUSCO completeness plots for bacteria and archaea.
cumulative	X axis - number of contigs or sequences in assembly sorted by decreasing order of length; Y axis - cumulative length of all contigs	Steep slopes at the start indicate better assembled sequences. Colours of taxonomic categories can be set using Filters: category attributes
detail	Details of the assembly + BTK pipeline used	Can be downloaded as a json file
report	Single page with all views in one page	
snail	Snail plot for assembly contiguity	Smooth snails in the middle indicate many contigs - ie a poor assembly. Highly stepped spiral indicates long well assembled sequences.
table	Table view of all contigs or sequences, columns are all attributes active in the Filters menu	IMPORTANT: Useful for looking at individual sequences and their taxonomic assignments. To add an extra column, activate it in the Filters top menu. Can be downloaded as a csv file

Datasets	Filters	Lists	Settings	Summary	Help	About
FILTERS	Description	Notes				
selection	Activating this shows only the selected contigs. Side buttons: Invert Selection, Select None, Select All, Show selection	This is the only Filter set that is not stored in the browser URL, so if you want to save the selected contigs, use Lists to save and download the current selection as a json file.				
length	Numerical variable attribute: Filter sequences by length	Use sliders or text fields to set min and max length				
position	For plotting values along sequence position windows (eg 1kb, 100kb, 1Mb sized windows)	Very useful for chromosomal assemblies when combined with Settings : shape → plot grid. Use the side button to set this variable on the X axis, Then set other variables on the Y axis like gc, coverage, busco_odb10_count. Use Settings to change window size				
proportion	For plotting values by sequence proportion windows (eg 1% or 10% windows)	Very useful for chromosomal assemblies when combined with Settings : shape → plot grid. Use the side button to set this variable on the X axis, Then set other variables on the Y axis like gc, coverage, busco_odb10_count. Use Settings to change window size				
gc	Numerical variable attribute: Filter sequences by gc content	Use sliders or text fields to set min and max gc.				
ncount	Numerical variable attribute: Count of "N" (unknown) bases in sequence	Useful for identifying scaffolds made up of large numbers of Ns (seen in old assembly algorithms)				
masked	Numerical variable attribute: Count of masked (repetitive) bases in sequence	If a contig or scaffold is made up of mostly masked repetitive bases, the taxonomic assignment is usually not reliable.				
taxon_odb10_count	Numerical variable attribute: Count of BUSCO genes on each sequence	(New BTK pipeline runs) The "Benchmarking Universal Single Copy Orthologs" (BUSCO) tool was run using the named taxon lineage. This numerical attribute shows the number of BUSCO genes found on that sequence. eg: ascomycota_odb10_count fungi_odb10_count eukaryota_odb10_count				

Datasets	Filters	Lists	Settings	Summary	Help	About
FILTERS (cont)	Description	Notes				
readaccession_cov	Numerical variable attribute: Average sequence coverage or depth for each sequence using reads from this public accession.	Publicly available read data from the read accession shown was mapped to the assembly if it was linked to the assembly in the public databases. The average sequencing coverage or depth is shown for each sequence. If you plot "position" or "proportion" on the X axis in the grid view (Settings : shape → plot grid), this numerical variable attribute shows the average coverage per sequence window instead. eg: ERR6054835_cov SRR1234344_cov				
buscogenes_taxonrank	Categorical attribute: BUSCO genes on each sequence were searched against the Uniprot Refseq Proteome database using diamond blastp.	(New BTK pipeline runs). The best hits at each taxon rank are shown here. eg: buscogenes_phylum buscogenes_order				
buscoregions_taxonrank	Categorical attribute: BUSCO-dense regions (or 100kb random regions if no BUSCO genes were found on a contig) were searched using diamond blastx against the Uniprot Refseq Proteome database.	(New BTK pipeline runs). If no hits were found, contigs were searched against the NCBI nt database using blastn eg: buscoregions_phylum buscoregions_order				
bestsumorder_taxonrank	Categorical attribute: total bitscore from a single database search, with scores taken from successive databases for contigs or scaffolds that failed to identify hits in the first database	(Old BTK pipeline runs) The databases were searched in this order: blastn vs NCBI nt, followed by diamond blastx against Uniprot Refseq Proteome. So any contigs with no hits in the first database were searched in the second database. eg: bestsumorder_phylum bestsumorder_order				
taxrule_taxonrank_cindex	Numerical variable attribute: cindex is a "confusion index" which shows how many extra taxa had hits at that taxon rank	Ideally, the cindex would be 0, i.e., only one taxon at that rank (eg phylum or order or genus etc) was found during similarity searches. But if a contig or sequence hits many different taxa, then the overall single assignment could be wrong				
taxrule_taxonrank_score	Numerical variable attribute: total score of all hits for that contig or sequence	If the score is low, the taxonomic assignment is less reliable, so you can filter out low score values by setting a higher minimum value.				

Datasets	Filters	Lists	Settings	Summary	Help	About
Settings	Description	Notes				
Interactive / Static	Switch between interactive and static views	Static views load faster for assemblies with hundreds of thousands of contigs				
shape	Change the main plot shape: Square bins, Hexagonal bins, circles, connected windows, plot grid, kite plot	<p>IMPORTANT: Default plots for older BTK pipeline runs with more than 2000 contigs show up as square bins. Switch to circle to show each contig as a circle (size of circle proportional to length of contig by default)</p> <p>Square and hexagonal bins are better for visualising many thousands of contigs which would get hidden in the circle view.</p> <p>Plot grid is useful for comparing chromosomes in well assembled genomes.</p> <p>Kite plots show a pair of lines representing two standard deviations about the mean on each axis (weighted to account for scaffold lengths) that intersect at a point representing the weighted median</p>				
window size	Choose between proportional windows 10% or 1%, or position (size) windows: 100k or 1M	Choose 10% or 1M for large scale trends, or 1% or 100k for more resolution				
plot style	Circles or Lines	Lines connect the points and make it easier to see trends				
reducer function	The function to plot square/hex binned points and histograms on the axes: Sum (default), max, min, count, mean	By default, the "z-axis" of the BTK plot is the length of each contig or sequence. Thus the default "reducer" for each bin is the "sum" of the lengths of the contigs in that bin.				
scale function	Scaling function for size of all points: Square root (default), log, linear	Log will reduce the perceived differences between sizes; Linear will increase the perceived differences.				
scale factor	Change the relative size of all points in the BTK plot (by default it makes the circles bigger/smaller)					
x-axis range	Set the min and maximum(default: lower and upper bounds of data)	IMPORTANT: Use this to zoom into an area of the plot				
y-axis range	Set the min and max (default: lower and upper bounds of data)	IMPORTANT: Use this to zoom into an area of the plot				

Datasets	Filters	Lists	Settings	Summary	Help	About
Settings (cont)	Description	Notes				
histogram maximum	Set the highest value on the histogram (default: highest possible value)	Useful if you want to compare two BTK plots and set both histograms to the same scale.				
display total	Show or Hide "Total" as a category in the legend					
palettes	Change the default colours of the categories	Click on each colour to customise it				
png resolution	Downloadable PNG image file max dimension (Default: 2000 pixels)	Change to 3000 or 4000 if you want a high-resolution image for publication				
static threshold	Total number of contigs or sequences above which the BTK plot is shown as a static plot					
nohit threshold	Total number of contigs or sequences above which the no-hit contigs are not shown in the blob view					
category index						
use larger fonts	Increase/Decrease the font size in the BTK plot legend					