

UCSC Genome Browser GenArk (Genome Archive) Assembly Hubs

Brian T. Lee¹, Hiram Clawson¹, Brian J. Raney¹, Mark Diekhans¹, Galt P. Barber¹, Gerardo Perez¹, Luis R. Nassar¹, Christopher M. Lee¹, Angie S. Hinrichs¹, Jonathan Casper¹, Daniel Schmelter¹, Jairo Navarro Gonzalez¹, Brittney Wick¹, Matthew L. Speir¹, Clay Fischer¹, Ann S. Zweig¹, Robert M. Kuhn¹, W. James Kent¹, Max Haussler¹
¹Univ. of California Santa Cruz Genomics Institute Santa Cruz, CA, US

Discover even more genomes in the UCSC Genome Browser

- Search species on Browser Gateway
- Use GCF ID to launch Browsers
- Visit taxonomic grouping pages

Access the New GenArk Assembly Hub Collections

The UCSC Genome Browser has a new collection of over 1,500 Genome Archive (GenArk) assembly hubs. The new GenArk assemblies come ready-for-use with several pre-computed annotation tracks and can be loaded via searching our Gateway, using direct links, or visiting pages of taxonomic groupings.

On the Browser Gateway, <https://genome.ucsc.edu/cgi-bin/hgGateway>, in the top left box you can search a term such as “dog” and find all the genomes hosted in our internal databases and in external Public Hubs that have dog in the name.

In this image a specific dog breed, Labrador, in the GenArk Genome Archive public assembly hub for Mammal assemblies is selected (GCF_014441545.1).

When you know your GCF identifier the species can be directly loaded with short /h/GCF accession links such as this one for Labrador: https://genome.ucsc.edu/h/GCF_014441545.1

The Genomes menu has a link to the “Genome Archive GenArk” homepage, which in turn lists all GenArk hubs by taxonomic groups.

The GenArk homepage links to separate pages for birds, fish, plants, and other taxonomic groups. Each individual page provides a table with links to load Browsers and access related details, a snapshot for birds seen here: <https://hgdownload.soe.ucsc.edu/hubs/birds/>

count	common name and view in browser	scientific name and data download	NCBI assembly
1	Adelie penguin	Pygoscelis adeliae	GCF_00069105.1_ASM6910v1
2	African ostrich	Struthio camelus australis	GCF_00069896.1_ASM6989v1
3	American crow	Corvus brachyrhynchos	GCF_00069197.1_ASM69197v1
4	Anna's hummingbird	Calypte anna	GCF_00395755.1_bCalAnn1_v1.p
5	bald eagle	Haliaeetus leucocephalus	GCF_00073746.1_Haliaeetus_leucocephalus-4.0

Examine GenArk Hubs with BLAT and PCR searches, or External Tools

Once you've loaded a GenArk hub, it is ready for action. BLAT and In-Silico PCR under the Tools menu allow the ability to search the assembly for primer pairs or DNA, protein, or RNA sequences.

In the left image a GenArk Hub for Zebu Cattle with the In-Silico PCR option under the Tools menu is selected, where in the main view a PCR match is highlighted. In the right image PCR results show two matches on different scaffolds of the assembly.

Under the Tools menu you can also access BLAT to search for DNA or protein or RNA sequence on GenArk Assembly Hubs. In the below left image is an example search of a human protein on a GenArk Hub where clicking on the details allows for seeing the side-by-side alignment to both identify the location of the protein in the organism and how certain amino acids differ.

DNA can also be exported to External Tools from UCSC via an option in the View menu, such as for primer design.

In the right image a region of Zebu Cattle DNA was sent to RSBAT (Regulatory Sequence Analysis Tools), which found a potential TFBS in the region. The predicted binding motif is shown in the Short Match track.

GenArk Hubs come with BLAT and PCR Tools

Blog posts about GenArk Hubs



<https://genome-blog.soe.ucsc.edu/blog/?s=GenArk>

Learn the answers to questions such as, “Which annotations do these genome browsers have?” and “Do they come mapped with genes, repeats, and config tracks?” See how to discover if your favorite model organism is already in GenArk, or how to get it added. In three parts, the blog posts review real examples of requesting new GenArk Hubs, such as on zebrafish, and how to go about accessing, using, and learning the technical details behind GenArk Hubs.



Funding and Acknowledgments

- UCSC Browser Funding provided by the National Human Genome Research Institute [U41-HG002371].
- GenArk Data: sourced from new assembly data placed in the National Center for Biotechnology Information (NCBI) and the Vertebrate Genomes Project (VGP).
- We would like to thank the collaborators and data contributors that generate the GenArk Data, our loyal users, and our tireless systems administrators for their support.

Come Ask Questions @ PAG 2022

PAG Workshop: 6:20 pm Sat. Jan. 8th Pacific H-J (2nd Fl)

GenArk Home Page: <https://hgdownload.soe.ucsc.edu/hubs/>

Download this poster: http://bit.ly/PAG_Poster_2022

Follow us:

[@GenomeBrowser](https://twitter.com/GenomeBrowser) [/ucscGenomeBrowser](https://www.facebook.com/ucscGenomeBrowser) [UCSC Genome Browser](https://www.youtube.com/channel/UCSCGenomeBrowser)



References

Lee BT, Barber GP, Benet-Pages A, Casper J, Clawson H, Diekhans M, Fischer C, Gonzalez JN, Hinrichs AS, Lee CM, AS, Pranav M, Nassar LR, Reagan N, Tiana P, Perez G, Raney BJ, Rosenbloom KR, Schmelter D, Speir ML, Wick BD, Zweig AS, Haussler D, Kuhn RM, Kent WJ. The UCSC Genome Browser database: 2022 update. Nucleic Acids Res. 2021 Oct 28; <https://doi.org/10.1093/nar/gkab959>

A future GenArk manuscript is currently in the works to be published in 2022.

Your genome is not in GenArk?

If you can't find the assembly you want in the GenArk hub collection, but you do already have the GCA/GCF identifier you can email us at genome@ucsc.soe.edu

