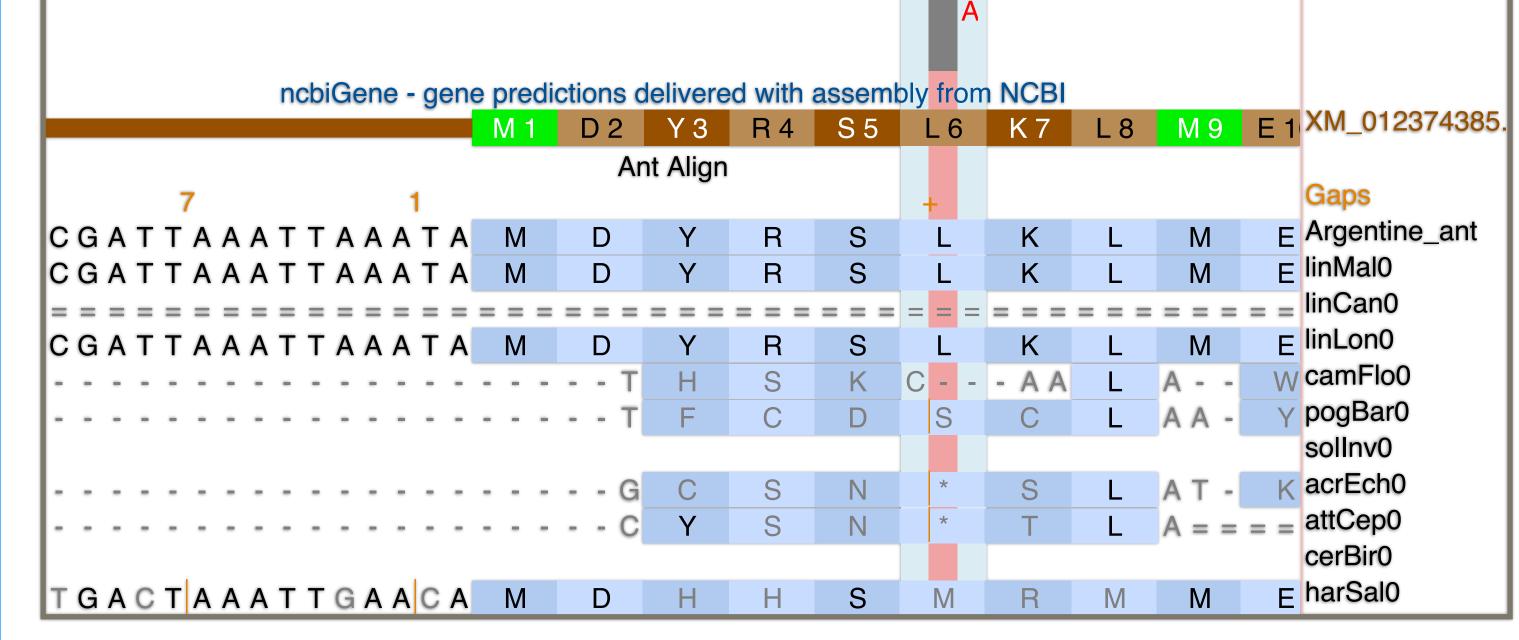
# Annotating Variants and Displaying Multiple Alignments on Your **Assemblies Using the UCSC Genome Browser**

Brian T. Lee<sup>1</sup>, Brian Raney<sup>1</sup>, Matthew L. Speir<sup>1</sup>, Galt P. Barber<sup>1</sup>, Christopher M. Lee<sup>1</sup>, Kate R. Rosenbloom<sup>1</sup>, Angie S. Hinrichs<sup>1</sup>, Jonathan Casper<sup>1</sup>, Hiram Clawson<sup>1</sup>, Ann S. Zweig<sup>1</sup>, Max Haeussler<sup>1</sup>, Robert M. Kuhn<sup>1</sup>, W. James Kent<sup>1</sup>

<sup>1</sup> Univ. of California Santa Cruz Genomics Institute Santa Cruz, CA, US

Creating Assembly Hubs	Annotating Variants in Assembly Hubs
The UCSC Genome Browser (https://genome.ucsc.edu/) is a free, web-based tool that allows researchers to visualize and explore genomic data, including new sequences. Using Assembly Hubs users can visualize any sequence and annotations upon it. An underlying binary indexed "2bit" file hosted remotely represents the ACTG of the new sequence. These assembly hubs can be further annotated by researchers with most of UCSC's track formats including Gene Prediction Tracks (https://genome.ucsc.edu/goldenpath/help/bigGenePred.html), which can be used in turn to predict the effects of novel variants. Further additional remotely hosted binary indexed data (" <b>big</b> " files such as <b>bigPsI</b> , <b>bigChain</b> , <b>bigMaf</b> , ect.) can annotate alignments or other genomic data.	The Variant Annotation Integrator (VAI) tool can be used with bigGenePred annotations to predict functional effects of variants on transcripts. For example, a variant might be located in the coding sequence of one transcript, but in the intron of an alternatively spliced transcript and VAI will predict the functional impact for each gene transcript. A TGGATTATCGTAGTTTAAAAATTGATGGA change of UUA to UAA (Note: gene and display is of the - strand) A predictions delivered with assembly from NCBI M 1 D 2 Y 3 R 4 S 5 L 6 K 7 L 8 M 9 E
C G A T T A A A T T A A A T A A T G G A T T A T C G T A G T T T A A A A T T G A T G G A> Variant Change A to T and codon change of UUA to UAA (Note: gene and display is of the - strand)	Genomes Genome Browser Tools Mirrors Downloads My Data



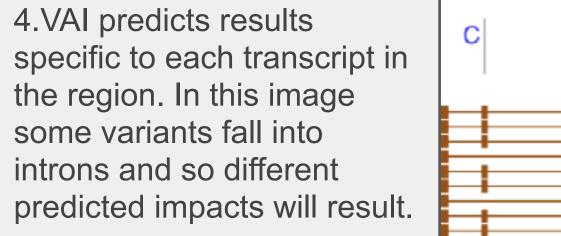
This image shows an assembly hub of Argentine Ant (GCA\_000217595.1, limHum0) where the display is flipped to view the reverse strand. The region has been annotated with a gene track (showing a gene XM\_012374385) above several Multiple Alignment Files for 10 other aligned ant species (using **bigMaf** format). A red highlight emphasizes a variant track (T>A) resulting in a change in a region highlighted in blue from coding for a Leucine to a Stop Codon.

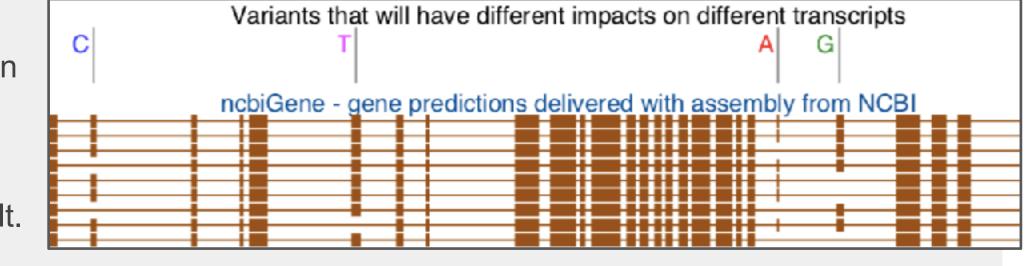
2. With a variant track and a gene track, the VAI tool can **annotate impacts** of up to 10,000 variants. Filters allow screening out synonymous, intronic and intergenic variants so only desired impacts are returned.

Feature	Feature type	Consequence	Position in cDNA	Position in CDS	Position in protein	Amino acid change
XM_012374385.1	Transcript	stop_gained	125	17	6	L/*

Select Genome Assemb	ly and Rec	Table Browser	
clade genome	assembly	Variant Annotation Integrator	
Ant Hub CArgentine ant C region to annotate	GCA_000217595	Data Integrator	
genome C		Gene Interactions	
		Gene Sorter	
Select Variants		Genome Graphs	
variants: Variant Change A to T and maximum number of variants		In-Silico PCR	of the - strand)
manage custom tracks track hubs	To reset all u	LiftOver	ustom tracks), click

3. The VAI output predicts the variant's impact. In this case a change from a Leucine to a Stop Codon. VAI output is in Ensembl VEP format.

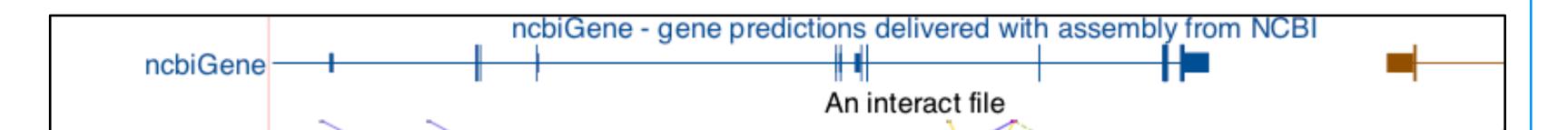




Variant Annotation Integrator

## Displaying Multiple Alignments and other Custom Data Types and Views in Assembly Hubs

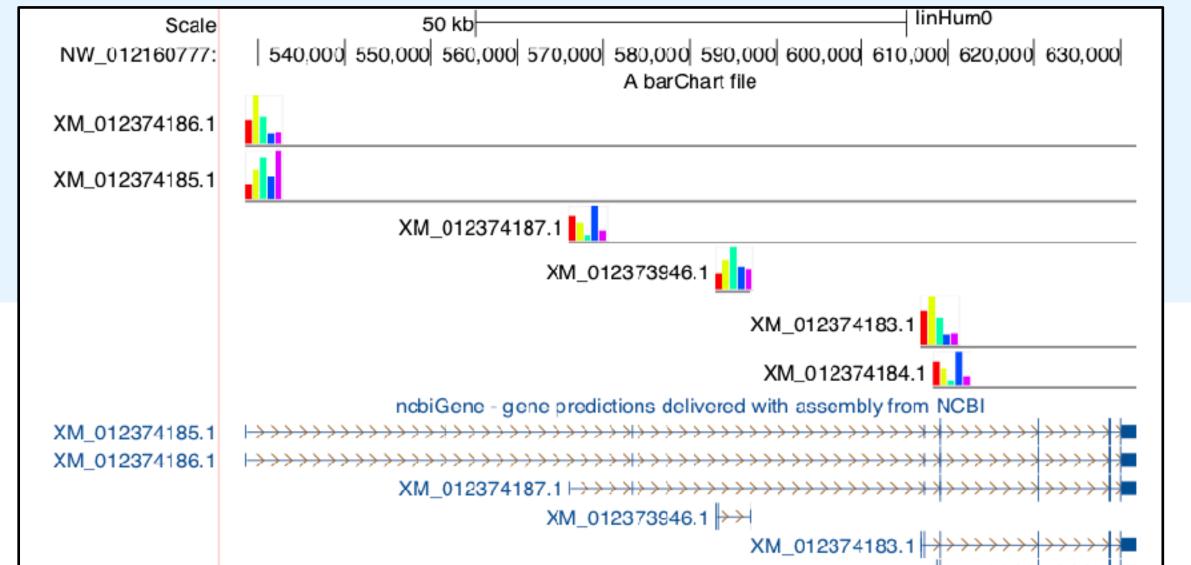
Scale	10 bases			lir	Hum0					
NW_012160777:									-	
>	TCAACCGCAGCAGGCTGACATA	ATG	ACG	тсс	ATG	$\top \top \top$	GAA	CAG	CAG	АТС
	ncbiGene - gene predictions delivered with assembly from NCBI									
XM 012374547.1		M 1	T 2	S 3	M 4	F 5	E 6	07	0.8	19



AM_012374547.1	IVI I	12	33	IVI 4	<b>- 3</b>	EO	Q./	20	19
Ant Align									
Gaps + 3									
Argentine_ant TCAACCGCAGCAGGCTGACATA	м	Т	S	М	F	Е	Q	Q	1
linMal0 T C A A C C G C A G C A G G C T G A C A T A	М	Т	S	М	F	Е	Q	Q	1
linCan0 = = A A C C G C A G C A G G C T G A C A T A	М	Т	TNC	ΝΤG	ΝΤΤ	Е	Q	Q	ANC
linLon0 = = = = = = = = = = = = = = = = = = =									
camFlo0 T C C A C C G C A G C A G A C T G A T A T A	М	Т	S	М	F	Е	Q	Q	1
pogBar0 T C A A C C G C A G C A G C T G A C A T T	М	Т	S	М	F	D	Q	Q	1
solinv0 CCAACCGCAGCAGGCTGACATT	М	Т	S	м	F	D	Q	Q	1
acrEch0 T C A A C C A C A G C A G A C T G T C A T T	М	Т	S	М	F	D	Q	Q	1
attCep0 TCAACCACAGCAGACTGTCATT	М	Т	S	М	F	D	Q	Q	1
cerBir0 T C A A C C G C A G C A G A C G G A T A T A	М	Т	S	М	F	Е	Q	Q	1
harSal0 T C A A G C A C A G C A G C A C C G A C A T A	М	Т	S	М	F	Е	Q	Q	1

**bigMaf:** displays Multiple Alignment Files visualizing other aligned sequences (https://genome.ucsc.edu/goldenpath/help/bigMaf.html).

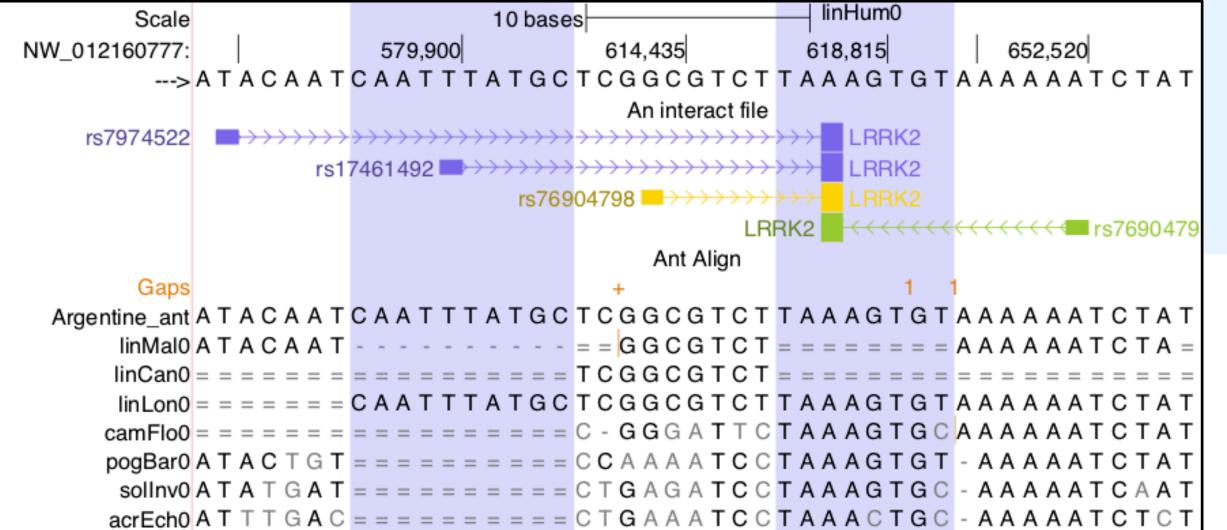
**bigBarChart:** displays a graph of category-specific values over genomic regions (https://genome.ucsc.edu/goldenpath/help/barChart.html).



Ant Align 

**bigInteract:** displays pairwise interactions as arcs or half-rectangles connecting two genomic regions on the same chromosome. For directional interactions such as SNP/gene, the interactions in the reverse direction are dashed (https://genome.ucsc.edu/goldenpath/help/intearact.html).

Multi-Region View: displays a genome sliced into user-defined regions, even across different chromosomes. End-points of SNP/gene interactions are shown (http://genome.ucsc.edu/goldenPath/help/multiRegionHelp.html).



## **Future work**

- Allow for **PCR** searches on Assembly Hubs. The future addition of PCR gfServers will allow verifying primers on novel genomes.

- Allow interaction displays to provide a direct link to a multi-region view.

### More information

PAG XXVII San Diego 2019 Talk: Jan 16th Wed. 10:30am in Royal Palm **Salon 1-2** UCSC Genome Browser - a home for all organisms

Assembly Hub Help: http://genomewiki.ucsc.edu/index.php/ Assembly Hubs

**Download this poster:** http://bit.ly/2CTnmlt



Follow us:

**UCSC** Genome Browser

#### References

Haeussler M, Zweig AS, Tyner C, Speir ML, Rosenbloom KR, Raney BJ, Lee CM, Lee BT, Hinrichs AS, Gonzalez JN, Gibson D, Diekhans M, Clawson H, Casper J, Barber GP, Haussler D, Kuhn RM, Kent WJ. The UCSC Genome Browser database: 2019 update. Nucleic Acids Res. 2018 Nov 8; https://doi.org/10.1093/nar/gky1095

#### Funding and acknowledgements

Funding provided by National Human Genome Research

Institute [U41HG002371].

We would like to thank our loyal users, our collaborators and data contributors, and our tireless systems administrators for their support.

