



Discovery can't wait.



The Celera Genome Browser:

A Tool for Visualizing and Annotating the Human Genome

*Russell Turner, Ph.D.
Celera Genomics
May, 2002*

† Purpose

- To provide a tool for visualization and structural annotation of large eukaryotic genomes
- To support Celera's internal scientists and customers

† History

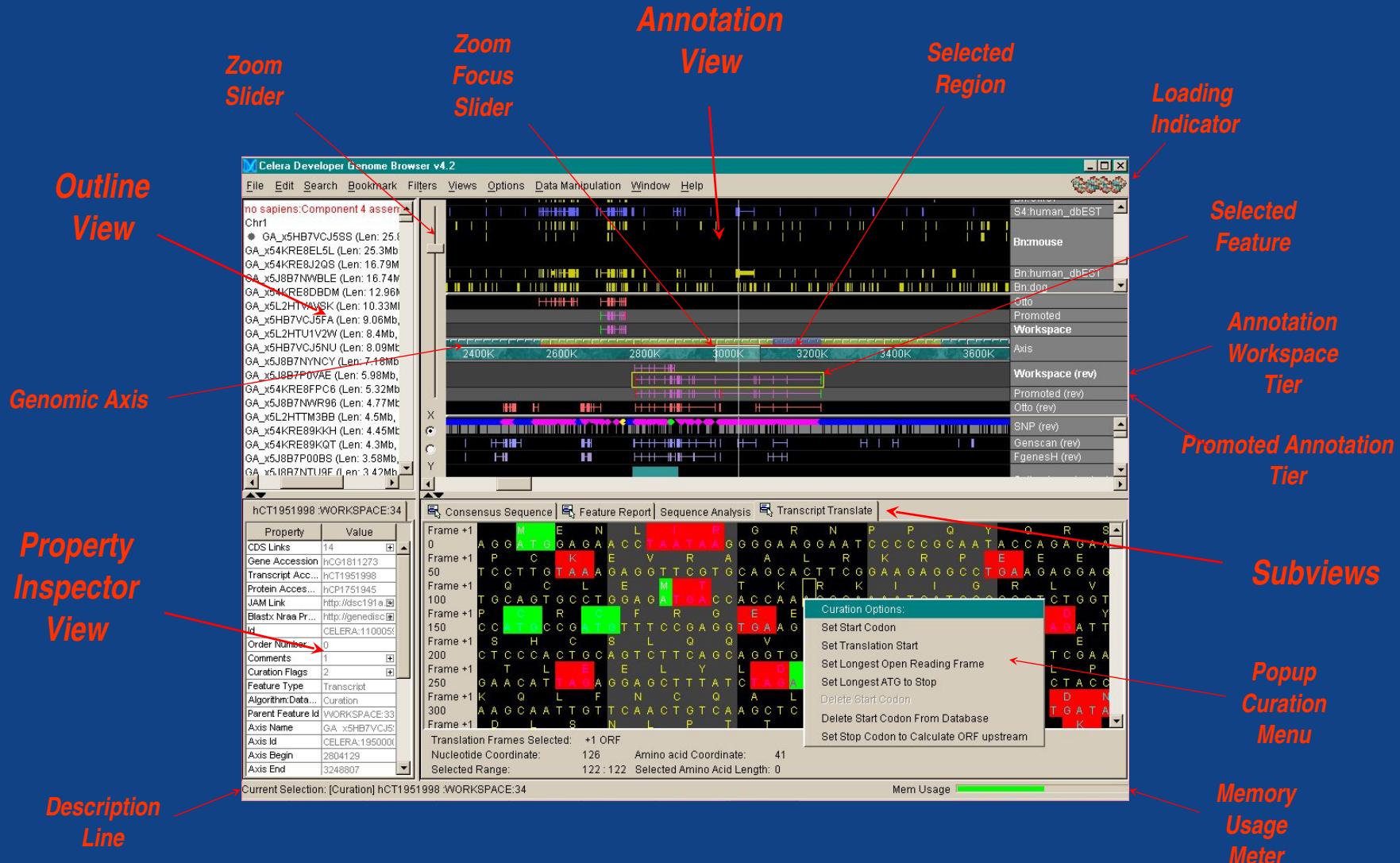
- Version 1 (1999) – Annotation of first assembly of *Drosophila Melanogaster* genome
- Version 2 (2000) – Annotation of Celera's first assembly of the human genome
- Subsequent versions support ongoing internal annotation (including the mouse genome) as well as the needs of Celera's customers
- Version 4.2 currently in development

+ Visualization

- Navigation and display of selected regions of a large genomic database
- Overlay of XML data via local files or HTTP service
- Multiple views of data, including high-performance graphical display
- Continuous pan and zoom from chromosome to nucleotide level
- Interactive Sequence Analysis
- Direct hyperlink to Celera Discovery System

+ Annotation

- Interactive “drag and drop” creation and editing of human-curated features
- Infinite undo capability
- Export of local annotations as XML file



Search Known Features

Search:
 Component 4 assembly from 2001-03-08
 1 loaded Genome Version(s)
 5 available Genome Version(s)

Type: Feature ID
 Find: CELERA:50000090730108
 Search

Status: Search Complete

Results:

Feature 50000090730108

Open New Browser

Panel @@

Options Data Manipulation Window Help

Legend: S4:update_human_F Bn:RIKEN_cDNA Bn:dog S4:Ensembl S4:Invitr_EST GW:human_IPI GW:nraa S4:CHGI GRAIL Bn:CHGI S4:human_dbEST Otto Promoted Workspace Axis Workspace (rev) Promoted (rev) Otto (rev) SNP (rev)

4200K 4400K 4600K 4800K 5000K 5200K 5400K 5600K

Consensus Sequence Feature Report Query Sequence Alignments Sequence Analysis Subject Sequence Alignments

Sum E Value	Number Iden...	Alignment ...	Bit Score	Number ...	Number Gaps	Accession Number	... Percent Iden...
3E-22	67	200	113	107	5	131860	33.5
8E-21	65	181	109	101	2	4204833	35.91
2E-20	64	181	108	101	2	539995	35.36
2E-18	63	200	101	107	3	4150910	31.5
1E-17	59	190	98.7	102	2	5902050	31.05
2E-17	59	190	97.9	104	2	6677755	31.05
1E-20	6E-17	58	96.3	92	2	6321591	36.25
1E-20	6E-17	55	96.3	95	0	1575028	31.25
2.8E-18	2E-14	56	88.2	95	0	131880	29.02
2E-81	9E-78	154	171	298	0	13652240	90.06
1.3E-79	5E-76	157	192	292	1	13928774	81.77
1.2E-12	8E-9	43	162	69.5	4	7292581	26.54
1.2E-8	8E-5	39	166	56.2	6	6323642	23.49
1.5E-9	1E-10	51	208	54.7	4	7522108	24.52

Bn:RIKEN_cDNA Bx:human_IPI Bn:fugu Bx:nraa Bn:rodent_dbEST Bn:CHGI Bn:curated_hCT Bn:CMGI Bn:CRGI
 Bn:hCT Bn:dog Bn:Ensembl Bn:Invitr_index Bn:mouse Bn:human_dbEST Bn:NCBIanno_mRNA Bn:human_RefSeq

Current Selection: [Bn:CHGI] Bn:CHGI:CELERA:50000090730108

Mem Usage

Celera Developer Genome Browser v4.2

File Edit Search Bookmark Filters Views Options Data Manipulation Window Help

GW:nraa
ePCR:allSTS
S4:CHGI
GRAIL
Bn:CHGI
S4:human_dbEST
Bn:mouse
Bx:nraa
Bn:CMGI
Bn:human_dbEST
FgenesH
S4:NCBIanno_mRNA
Otto
Promoted
Workspace
Axis
Workspace (rev)
Promoted (rev)
Otto (rev)

X 2620K 2640K 2660K 2680K 2700K 2720K 2740K 2760K 2780K Y

Consensus Sequence Feature Report Sequence Analysis Transcript Translate

Property Value

CDS Links 14
Gene Accession hCG22840
Transcript Accession hCT13938
Protein Accession hCP40279
JAM Link http://dsc191
Blastx Nraa ... http://genedis...
Id CELERA:1100C
Order Number 0
Comments 2
Curation Flags 3
Feature Type Transcript
Algorithm:Data Curation
Parent Feature WORKSPACE:
Axis Name GA x5HB7VC
Axis Id CELERA:1950C
Axis Begin 2720992
Axis End 2781778

Selected: 2721051 2721051 Select

Mem Usage [progress bar]

Current Selection: [Curation] hCT13938 :WORKSPACE:2

Property	Value
CDS Links	14
Gene Accession	hCG22840
Transcript Accession	hCT13938
Protein Accession	hCP40279
JAM Link	http://dsc191
Blastx Nraa ...	http://genedis...
Id	CELERA:1100C
Order Number	0
Comments	2
Curation Flags	3
Feature Type	Transcript
Algorithm:Data	Curation
Parent Feature	WORKSPACE:
Axis Name	GA x5HB7VC
Axis Id	CELERA:1950C
Axis Begin	2720992
Axis End	2781778

Celera Developer Genome Browser v4.2

Search Known Features

Search:
 Component 4 assembly from 2001-03-08
 1 loaded Genome Version(s)
 5 available Genome Version(s)

Type: Gene (Celera Accession)
Find: hCG1811273
Search

Status: Search Complete

Results:
Gene hCG1811273

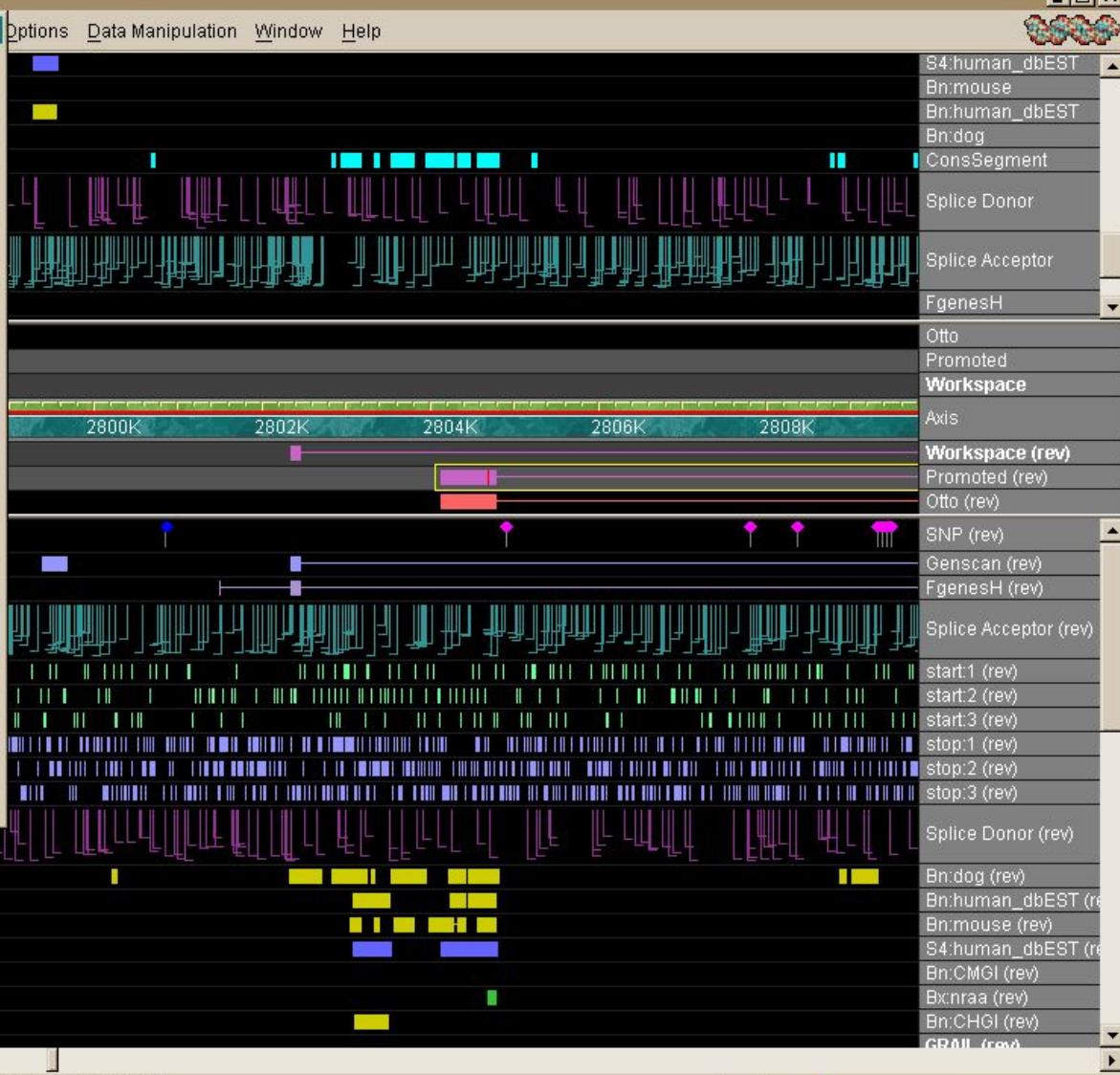
Open New Browser

Close

Order Number: 0
Comments: 1
Curation Flags: 2
Feature Type: Transcript
Algorithm:Dataset: Promoted
Parent Feature Id: CELERA:1100059
Axis Name: GA_x5HB7VCJ5S
Axis Id: CELERA:1950000
Axis Begin: 2804129
Axis End: 3248807

Current Selection: [Promoted] hCT1951998 : CELERA:11000595808682

Mem Usage: 



The genome browser interface displays a genomic track for the gene hCG1811273. The main panel shows multiple tracks of genomic data, including exons (represented by colored bars) and introns (represented by vertical lines). A scale at the bottom indicates positions from 2800K to 2808K. The right side of the interface lists various genomic datasets and their descriptions, such as S4:human_dbEST, Bn:mouse, and various splice donor and acceptor tracks. The status bar at the bottom indicates the current selection is 'Promoted' hCT1951998 : CELERA:11000595808682, and the memory usage is shown as a green progress bar.

Celera Developer Genome Browser v4.2

File Edit Search Bookmark Filters Views Options Data Manipulation Window Help

no sapiens:Component 4 assert

Chr1

- GA_x5HB7VCJ588 (Len: 25. GA_x54KRE8EL5L (Len: 25.3Mb GA_x54KRE8J2QS (Len: 16.79M GA_x5JB87NWBLE (Len: 16.74M GA_x54KRE8DBDM (Len: 12.96I GA_x5L2HTVAVSK (Len: 10.33M GA_x5HB7VCJ5FA (Len: 9.06Mb GA_x5L2HTU1V2W (Len: 8.4Mb, GA_x5HB7VCJ5NU (Len: 8.09Mb GA_x5JB87NYNCY (Len: 7.18Mb GA_x5JB87P0VAE (Len: 5.98Mb, GA_x54KRE8FPC6 (Len: 5.32Mb GA_x5JB87NWR96 (Len: 4.77Mb GA_x5L2HTTM3BB (Len: 4.5Mb, GA_x54KRE89KKH (Len: 4.45Mb GA_x54KRE89KQT (Len: 4.3Mb, GA_x5JB87P00BS (Len: 3.58Mb GA_x5IRR7NTU9F (Len: 3.42Mb

GRAIL

- Bn:CHGI
- Bx:nraa
- Bn:CMGI
- S4:human_dbEST

Bn:mouse

- Bn:human_dbEST
- Bn:dog

Otto

- Promoted
- Workspace

Axis

X

Y

Workspace (rev)

- Promoted (rev)
- Otto (rev)

SNP (rev)

Bxnraa:CELERA:5000009072895

Property	Value
Id	CELERA:500000:
Aliases (Number)	0
Order Number	1
Comments	0
Feature Type	High Scoring Pair
Algorithm:Data...	Bxnraa
Parent Feature...	CELERA:500000:
Axis Name	GA_x5HB7VCJ5
Axis Id	CELERA:1950001
Axis Begin	5261006
Axis End	5261183
Entity Length	177
Entity Orientation	Forward
Is Child	true
Is Composite	false
Relative Asse...	985037830
Display Priority	Low

Consensus Sequence | Feature Report | Query Sequence Alignments | Sequence Analysis | Subject Sequence Alignments

K K E L T Q I K Q K +3
K K E L T Q I K Q K +2
S G Q R G S S K L K G D D L Q A I K K +1
S G Q R G S S E L K G D D L Q A I R R DNA +
gtggcatgggatcttc aatt gaaagg gat gaccc t cagggcat aa aaggaaat gac c agat aaaac aaaaa
gtggcatgggatcttc aatt gaaagg gat gaccc t cagggcat aa aaggaaat gac c agat aaaac aaaaa

GTGGGCA GTGGGGATCTTCCGAATTGAAAGGAGATGACCTTCAGGCCATAAAGAAGGAATTGACCCAGATAAAAACAAAAA Query Sequence

5260960 5260970 5260980 5260990 5261000 5261010 5261020 5261030 Axis

DNA -

-1
-2
-3

Current Selection: [Bxnraa]Bxnraa:CELERA:50000090728954

Mem Usage

Celera Developer Genome Browser v4.2

File Edit Search Bookmark Filters Views Options Data Manipulation Window Help

Chr1

- GA_x5HB7VCJ5SS (Len: 25.8)
- GA_x54KRE8EL5L (Len: 25.3Mb)
- GA_x54KRE8J2QS (Len: 16.79M)
- GA_x5J8B7NWBLE (Len: 16.74M)
- GA_x54KRE8DBDM (Len: 12.96M)
- GA_x5L2HTVAVSK (Len: 10.33M)
- GA_x5HB7VCJ5FA (Len: 9.06Mb,
- GA_x5L2HTU1V2W (Len: 8.4Mb,
- GA_x5HB7VCJ5NU (Len: 8.09Mb)
- GA_x5J8B7NYNCY (Len: 7.18Mb)
- GA_x5J8B7P0VAE (Len: 5.98Mb,
- GA_x54KRE8FPC6 (Len: 5.32Mb)
- GA_x5J8B7NWR96 (Len: 4.77Mb)
- GA_x5L2HTTM3BB (Len: 4.5Mb,
- GA_x54KRE89KKH (Len: 4.45Mb)
- GA_x54KRE89KQT (Len: 4.3Mb,
- GA_x5J8B7P00BS (Len: 3.58Mb,
- GA_x5J8B7NTLUF (Len: 3.42Mb)

84:human_dbEST
Bn:mouse
Bn:human_dbEST
Bn:dog
Otto
Promoted
Workspace
Axis
Workspace (rev)
Promoted (rev)
Otto (rev)
SNP (rev)
Genscan (rev)
FgenesH (rev)

2400K 2600K 2800K 3000K 3200K 3400K 3600K

X Y

hCT1951998 :WORKSPACE:34

Property	Value
CDS Links	14
Gene Accession	hCG1811273
Transcript Acc...	hCT1951998
Protein Acces...	hCP1751945
JAM Link	http://dsc191a
Blastx Nraa Pr...	http://genedisc
Id	CELERA:110005
Order Number	0
Comments	1
Curation Flags	2
Feature Type	Transcript
Algorithm:Data...	Curation
Parent Feature Id	WORKSPACE:33
Axis Name	GA_x5HB7VCJ5:
Axis Id	CELERA:1950000
Axis Begin	2804129
Axis End	3248807

Consensus Sequence | Feature Report | Sequence Analysis | Transcript Translate

Frame +1 M E N L I R G R N P P Q Y Q R S
 0 A G G A T G G A G A A C C T A A T A A G G G G A A G G A T C C C C C C G C A A T A C C A G A G A A
 Frame +1 P C K E V R A A L R K R P E E E
 50 T C C T T G T A A A G A G G G T C G T G C A G C A C T T C G G A A G A G G G C T G A A G A G G A G
 Frame +1 Q C L E M T T K R K I I G R L V
 100 T G C A G T G C C T G G A G A T G A C C A C C A A C C A A A C C A A A C C A A A C C A A A C C A A
 Frame +1 P C R C F R G E E
 150 C C A T C C C G A T T T C C G A G G T G A A G
 Frame +1 S H C S L Q Q V
 200 C T C C C A C T G C A G T C T T C A G C A G G T G
 Frame +1 T L E E L Y L D
 250 G A A C A T T A G A G G A G C T T T A T C T A G A
 Frame +1 K Q L F N C Q A L
 300 A A G C A A T T G T T C A A C T G T C A A G G C T C
 Frame +1 D L S N L P T T

Curation Options:

- Set Start Codon
- Set Translation Start
- Set Longest Open Reading Frame
- Set Longest ATG to Stop
- Delete Start Codon
- Delete Start Codon From Database
- Set Stop Codon to Calculate ORF upstream

Translation Frames Selected: +1 ORF
 Nucleotide Coordinate: 126 Amino acid Coordinate: 41
 Selected Range: 122:122 Selected Amino Acid Length: 0

Current Selection: [Curation] hCT1951998 :WORKSPACE:34

Mem Usage

Celera Developer Genome Browser v4.2

File Edit Search Bookmark Filters Views Options Data Manipulation Window Help

Conssegment EgenesH
Bn:dog TRIEST:MPSS_Tag
Genscan RepeatMasker
Contig Otto
Promoted Workspace
Axis
Workspace (rev)
Promoted (rev)
Otto (rev)
SNP (rev)
Genscan (rev)
TRIEST:MPSS_Tag (rev)
Bn:dog (rev)

000K 2500K 3000K 3500K 4000K 4500K 5000K 5500K Axis

hCT1951998 :WORKSPACE:34 | Consensus Sequence Feature Report Sequence Analysis Transcript Translate

Query Sequence: User-entered Sequence External DB: blastx Tier Name (Optional): DB To Search: Celera Genomic Axis

Subject Sequence:

```
TTATACCATATGTACAGTTACTTCATAAGGCATTATCATGTAATTATTGAAAAAAGATTCAATATAGTAAAGTCTTA
TTTTAAGATATATTTCTCTATCTGAGATGTAGTTATGAAAATTATTGTTTCTCTATATTCTTTGTATATAT
GCCAACATATACTTGCTAGAAATTATTGGAAATTACAATTCTTTATATTGTATTCTATATAAAGGCATATA
CAACTTGAGGAGTGTGATGGAGAGCAGAATTCTAAAGAGCTTGTGTCTCCGTGCTCCGAAATTATATGA
AATCCATTGAGAAGAAACTTATTGATTAAAAACAAAAACAGAAAACAAAAACTGCACCAAT
```

Inputs Results Parameters Begin Algorithm Execution Clear Subject Sequence

Current Selection: [Curation] hCT1951998 :WORKSPACE:34 Mem Usage:

Celera Developer Genome Browser v4.2

File Edit Search Bookmark Filters Views Options Data Manipulation Window Help

Homo sapiens:Chr1

Chr1

GA_x5HB7VC

GA_x54KRE8EL

GA_x54KRE8J2C

GA_x5J8B7NWB

GA_x54KRE8DB

GA_x5L2HTVAV8

GA_x5HB7VCJ5I

GA_x5L2HTU1V:

GA_x5HB7VCJ5I

GA_x5J8B7NYN1

GA_x5J8B7P0VA

GA_x54KRE8FP

GA_x5J8B7NWR

GA_x5L2HTTM3I

GA_x54KRE89K

GA_x54KRE89K

GA_x5J8B7P00E

GA_x518R7NTU1

hCT1951998 :WORKSPACE:

Property	Value
CDS Links	14
Gene Acc...	hCG111273
Transcript ...	hCT1951998
Protein Acc...	hCP1751945
JAM Link	http://dsc19
Blastx Nraa ...	http://gened
Id	CELERA:1100
Order Number	0
Comments	1
Curation Fla...	2
Feature Type	Transcript
Algorithm:D...	Curation
Parent Feat...	WORKSPACE
Axis Name	GA_x5HB7VC
Axis Id	CELERA:1950
Axis Begin	2804129
Axis End	3248807

Current Selection: [Curation] hCT1951998 :WORKSPACE:34

Mem Usage

GW:human_IPI

Preferences: Genomic Axis Annotation View

Edit Colors | Edit Features | Edit Tiers | View Preference File |

Add Tier | Delete Tier | Make All Visible

Tier Name	Position	Visible State	Hide Empty	Is Docked	Background
Axis	0	Collapsed	false	true	Gray
Consensus	1	Collapsed	true	true	Gray
Contig	2	Collapsed	true	false	Gray
Workspace	3	Expanded	false	true	Dark Gray
Promoted	4	Collapsed	true	true	Soft Gray
Otto	5	Collapsed	true	true	Black
BlastN dbST...	6	Collapsed	true	false	Black
Clustered Bla...	7	Collapsed	true	false	Black
clustered Bla...	8	Collapsed	true	false	Black
RepeatMasker	9	Collapsed	true	false	Black
pombePep	10	Collapsed	true	false	Black
SNP	11	Collapsed	true	false	Black
Genscan	12	Collapsed	true	false	Black
TmaScan	13	Collapsed	true	false	Black
yeast	14	Collapsed	true	false	Black
CandidaAlbi...	15	Collapsed	true	false	Black
gw:yeast	16	Collapsed	true	false	Black
gw:pombeP...	17	Collapsed	true	false	Black
gw:Candida...	18	Collapsed	true	false	Black
S4_Bn_GenB...	19	Collapsed	true	false	Black
GlimmerM	20	Collapsed	true	false	Black
Bn_Asprgillu...	21	Collapsed	true	false	Black
Bn_Fusarium...	22	Collapsed	true	false	Black
Bn_Neurosp...	23	Collapsed	true	false	Black
Bn_FungiEst	24	Collapsed	true	false	Black
Bn_GenBank...	25	Collapsed	true	false	Black
S4_human_d	26	Collapsed	true	false	Black

OK Cancel Apply

Celera Developer Genome Browser v4.2

File Edit Search Bookmark Filters Views Options Data Manipulation Window Help

Homo sapiens:Com Chr1

2485K 2490K 2495K 2500K 2505K 2510K 2515K Axis

Preferences: System

Application Settings Data Source Settings Splice Profiles Workspace Backup

CDS Login Information

User Name: joe_user Password: ***** Save Login Information

XML Directory

C:\WINNT\Profiles\safforta\Desktop

Add to Current Directories Remove Selected Directory

Validation Options

None

XML Service

Current URLs:

Add to Current URLs Remove Selected URL

New URL:

OK Cancel Apply

hCT195 Proper CDS Links Gene Acc Transcript Protein Ac JAM Link Blastx Nra Id Order Num Comments Curation F Feature Ty Algorithm Parent Fee Axis Name Axis Id Axis Begin Axis End 3248807

Current Selection: [Curation] hCT1951998 :WORKSPACE:34

Mem Usage

The screenshot displays the Celera Developer Genome Browser interface. A modal dialog box titled "Preferences: System" is open, specifically showing the "Data Source Settings" tab. This tab contains sections for "CDS Login Information" (User Name: "joe_user", Password: masked, "Save Login Information" checked), "XML Directory" (set to "C:\WINNT\Profiles\safforta\Desktop"), "Validation Options" (set to "None"), and "XML Service" (Current URLs section with "Add to Current URLs" and "Remove Selected URL" buttons, and a "New URL:" input field). At the bottom of the dialog are "OK", "Cancel", and "Apply" buttons. The main workspace shows a genomic track for "Chr1" with multiple tracks of data. On the left, a sidebar lists various genomic features like "hCT195", "Proper", "CDS Links", etc. The top menu bar includes "File", "Edit", "Search", "Bookmark", "Filters", "Views", "Options", "Data Manipulation", "Window", and "Help". The status bar at the bottom indicates the current selection as "[Curation] hCT1951998 :WORKSPACE:34".

Genome Browser – Intensity Mapping

Discovery can't wait.

Celera Developer Genome Browser v4.2

File Edit Search Bookmark Filters Views Options Data Manipulation Window Help

Otto
Promoted
Workspace
Axis
Workspace (rev)
Promoted (rev)
Otto (rev)

520K 525K 530K 535K 540K 545K 550K 555K 560

542.0K 542.5K 543.0K

clustered BlastX nr.aa precomputes

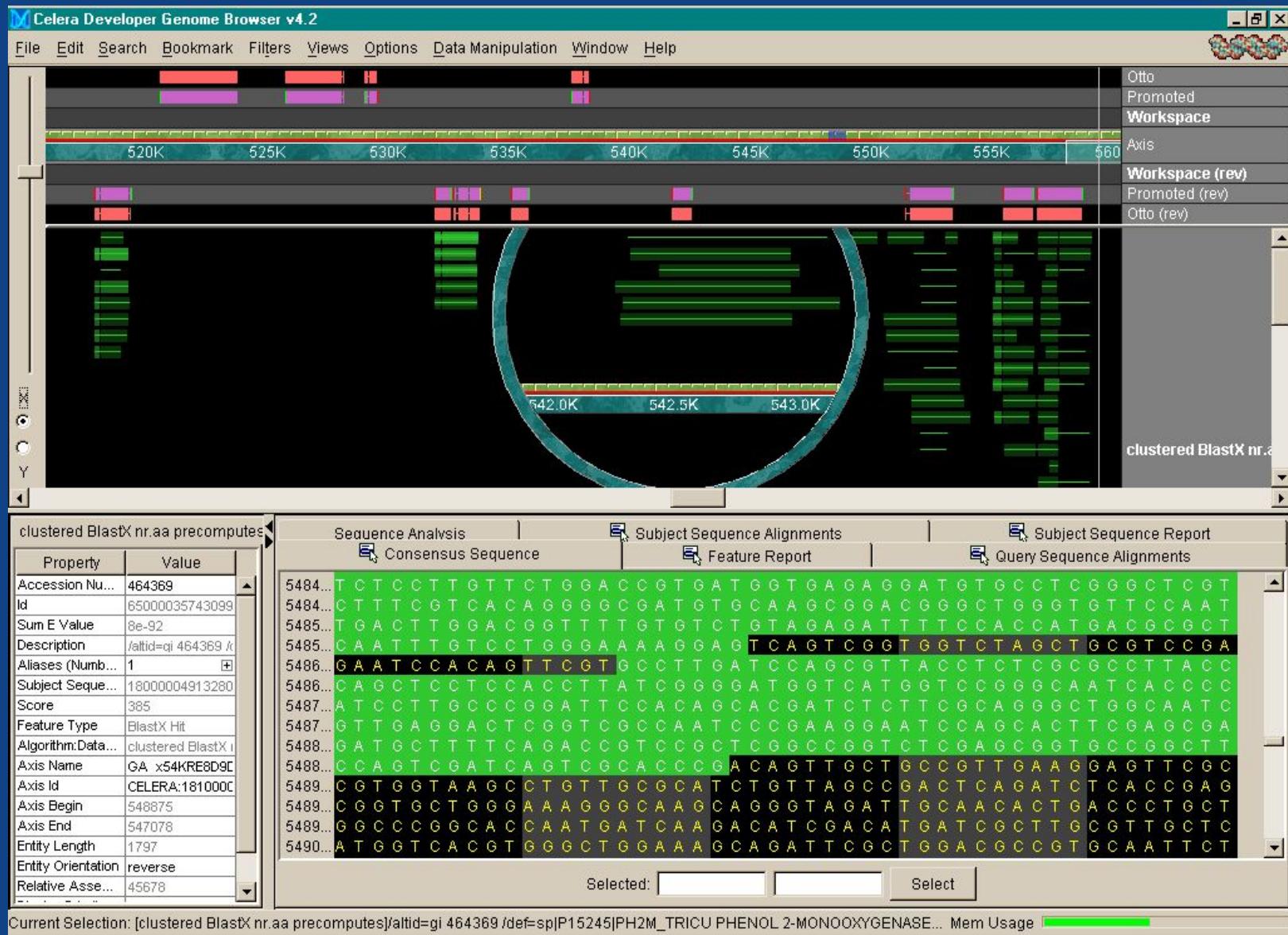
Sequence Analysis | Subject Sequence Alignments | Subject Sequence Report

Consensus Sequence | Feature Report | Query Sequence Alignments

Property	Value
Accession Nu...	464369
Id	65000035743099
Sum E Value	8e-92
Description	/altid=gi 464369 /k
Aliases (Numb...	1
Subject Seque...	18000004913280
Score	385
Feature Type	BlastX Hit
Algorithm:Data...	clustered BlastX I
Axis Name	GA_x54KRE8D9C
Axis Id	CELERA:181000C
Axis Begin	548875
Axis End	547078
Entity Length	1797
Entity Orientation	reverse
Relative Asse...	45678

Selected: [] [] Select

Current Selection: [clustered BlastX nr.aa precomputes]/altid=gi 464369 /def=sp|P15245|PH2M_TRICU PHENOL 2-MONOXYGENASE... Mem Usage [██████████]



Celera Developer Genome Browser v4.2

File Edit Search Bookmark Filters Views Options Data Manipulation Window Help

no sapiens:Component 4 asserted

Chr1

- GA_x5HB7VCJ5SS (Len: 25.3Mb)
- GA_x54KRE8EL5L (Len: 25.3Mb)
- GA_x54KRE8J2QS (Len: 16.79Mb)
- GA_x5J8B7NWBLE (Len: 16.74Mb)
- GA_x54KRE8DBDM (Len: 12.96Mb)
- GA_x5L2HTVAVSK (Len: 10.33Mb)
- GA_x5HB7VCJ5FA (Len: 9.06Mb)
- GA_x5L2HTU1V2W (Len: 8.4Mb)
- GA_x5HB7VCJ5NU (Len: 8.09Mb)
- GA_x5J8B7NYNCY (Len: 7.18Mb)
- GA_x5J8B7P0VAE (Len: 5.98Mb)
- GA_x54KRE8FPC6 (Len: 5.32Mb)
- GA_x5J8B7NWR96 (Len: 4.77Mb)
- GA_x5L2HTTM3BB (Len: 4.5Mb)
- GA_x54KRE89KKH (Len: 4.45Mb)
- GA_x54KRE89KQT (Len: 4.3Mb)
- GA_x5J8B7P00BS (Len: 3.58Mb)
- GA_y5.IRR7NTU9F (Len: 3.42Mb)

X

2720K 2730K 2740K 2750K

Y

GRAIL

Bn:CHGI

Bx:nraa

Bn:CMGI

S4:human_dbEST

Bn:mouse

Bn:human_dbEST

Bn:dog

ConsSegment

EgenesH

Genscan

RepeatMasker

Otto

Promoted

Workspace

Axis

80K

Workspace (rev)

Otto (rev)

Transcript Options:

- Delete Gene from WorkSpace
- Split Transcript
- Merge with Selected Transcript
- Delete Start Codon From Database
- Create New Gene
- Attach Transcript To Selected Gene
- Detach Transcript From its Gene
- GeneOptions:

hCG22840 :WORKSPACE:1

Property	Value
CDS Links	4
Gene Accession	hCG22840
Id	CELERA:110005
Comments	0
Feature Type	Gene
Algorithm:Data...	Curation
Axis Name	GA_x5HB7VCJ5
Axis Id	CELERA:195000
Axis Begin	2720992
Axis End	2781778
Entity Length	60786
Entity Orientation	Forward
Is Alternate Splice	false
Is Child	false
Assigned By	otto
Assigned Date	05/23/2001 22:31
Modified By	safforta

Consensus Sequence | Feature Report | Sequence Analysis

Selected: [] Select []

Mem Usage: []

Current Selection: [Curation] hCG22840 :WORKSPACE:1

Genome Browser – Evidence Highlighting

Discovery can't wait.

Celera Developer Genome Browser v4.2

File Edit Search Bookmark Filters Views Options Data Manipulation Window Help

no sapiens:Component 4 assembly

Chr1

- GA_x5HB7VCJ5SS (Len: 25.3Mb)
- GA_x54KRE8EL5L (Len: 25.3Mb)
- GA_x54KRE8J2QS (Len: 16.79Mb)
- GA_x5JB7NWBLE (Len: 16.74Mb)
- GA_x54KRE8DBDM (Len: 12.96Mb)
- GA_x5L2HTVAVSK (Len: 10.33Mb)
- GA_x5HB7VCJ5FA (Len: 9.06Mb)
- GA_x5L2HTU1V2W (Len: 8.4Mb)
- GA_x5HB7VCJ5NU (Len: 8.09Mb)
- GA_x5JB7NYNCY (Len: 7.18Mb)
- GA_x5JB7P0VAE (Len: 5.98Mb)
- GA_x54KRE8FPC6 (Len: 5.32Mb)
- GA_x5JB87NWR96 (Len: 4.77Mb)
- GA_x5L2HTTM3BB (Len: 4.5Mb)
- GA_x54KRE89KKH (Len: 4.45Mb)
- GA_x54KRE89KQT (Len: 4.3Mb)
- GA_x5JB87P00BS (Len: 3.58Mb)
- GA_x5JB87NTU9F (Len: 3.42Mb)

X 2720K 2730K 2740K 2750K Y

GRAIL Bn:CHGI Bx:nraa Bn:CMGI 84:human_dbEST Bn:mouse Bn:human_dbEST Bn:dog ConsSegment FgenesH Genscan RepeatMasker

Evidence Options: Use to set left edges of all Exons in the selected Transcript Use to set right edges of all Exons in the selected Transcript

Workspace (rev) Otto (rev)

hCT13938 :WORKSPACE:2

Consensus Sequence Feature Report Sequence Analysis Transcript Translate

Property	Value
CDS Links	14
Gene Accession	hCG22840
Transcript Accession	hCT13938
Protein Accession	hCP40279
JAM Link	http://dsc191a
Blastx Nraa Protein	http://genedisc
Id	CELERA:110005
Order Number	0
Comments	2
Curation Flags	3
Feature Type	Transcript
Algorithm:Data...	Curation
Parent Feature...	WORKSPACE:1
Axis Name	GA_x5HB7VCJ5
Axis Id	CELERA:195000
Axis Begin	2720992
Axis End	2781778

Frame +3
0 C A C C T T T A A G G C G T C A C G G G T G G G G C T G C A G C T T T C T G G A C C T T A G G A C T T T
Frame +3 M S R L K R I A G Q D L R A G F
52 C T G T C G G C C T G A A G C G G A T A G C G G G G C A G G A T C T C C G C G C T G G T T T C A
Frame +3 A G G R D C O T S V P Q O L L K A
104 G C A G G T G G A A G A G A C T G C G G T A C C T C G G T A C C C C A A G G G C T G T T G A A G G
Frame +3 R K S G Q L N L S G R N L S E V
156 C G A G G A A G A G C G G G C C A G T T A A A C C T G T C G G G T A G A A A C C T C A G T G A A G T
Frame +3 Q C V W R I N V D I P E E A N Q I
208 G C A G T G T G T C T G G A G A T A A A T G T G G A T A T C C C T G A G G A A G C T A A T C A G A
Frame +3 L S F G A T E R W W E Q T D L T K
260 C T T T C G T T T G G T G C T A C T G O A A A G A T G G T G G G G A G C A G A C A G A T T T G A C C A
Frame +3 I I S N N K L Q S L T D D L R L
312 T A A T A A T A T C A A A C A A T A A A C T T C A G T C A C T T A C A G A T G A C C T G C G A C T
Frame +3 P A L T V L D I H D N Q L T S L

Translation Frames Selected: +3 ORF
Nucleotide Coordinate: 23 Amino acid Coordinate:
Selected Range: Selected Amino Acid Length:

Current Selection: [Curation] hCT13938 :WORKSPACE:2 Mem Usage:

† 3-Tiered Architecture

- Interactive graphical client front end
- J2EE application server middle tier
- Oracle relational database back end

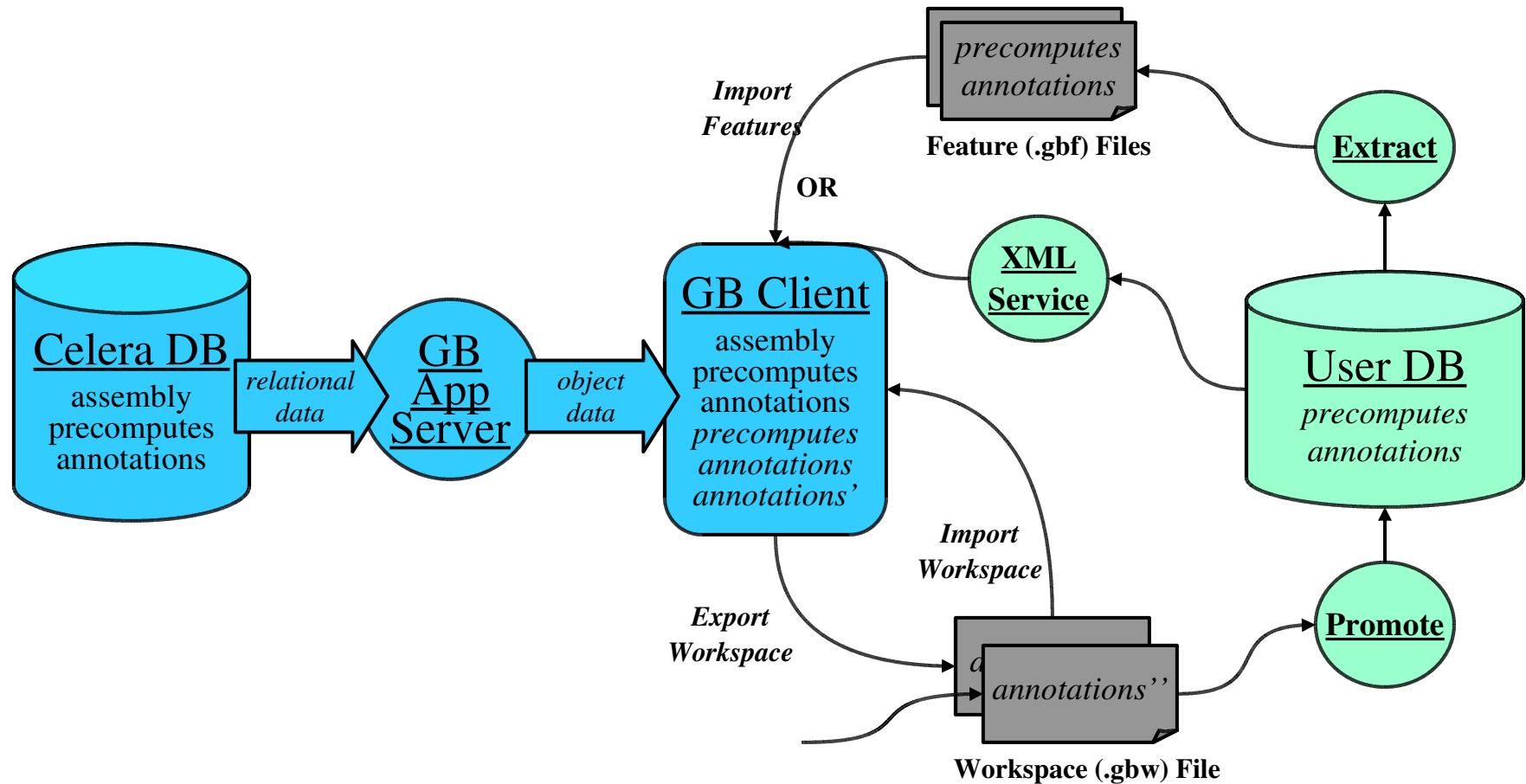
† Pure Java client and server implementation

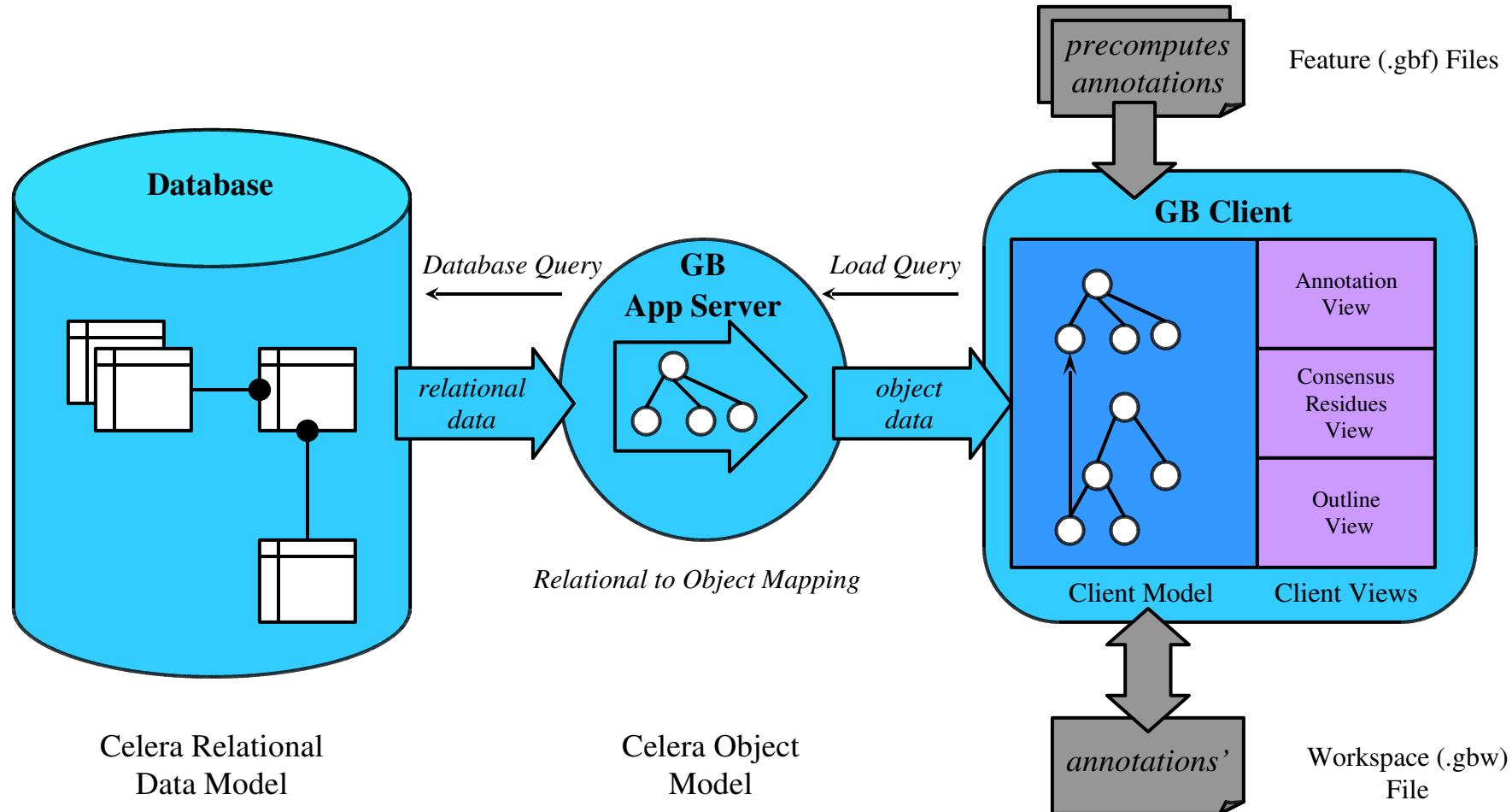
† Standalone Java “thick” client

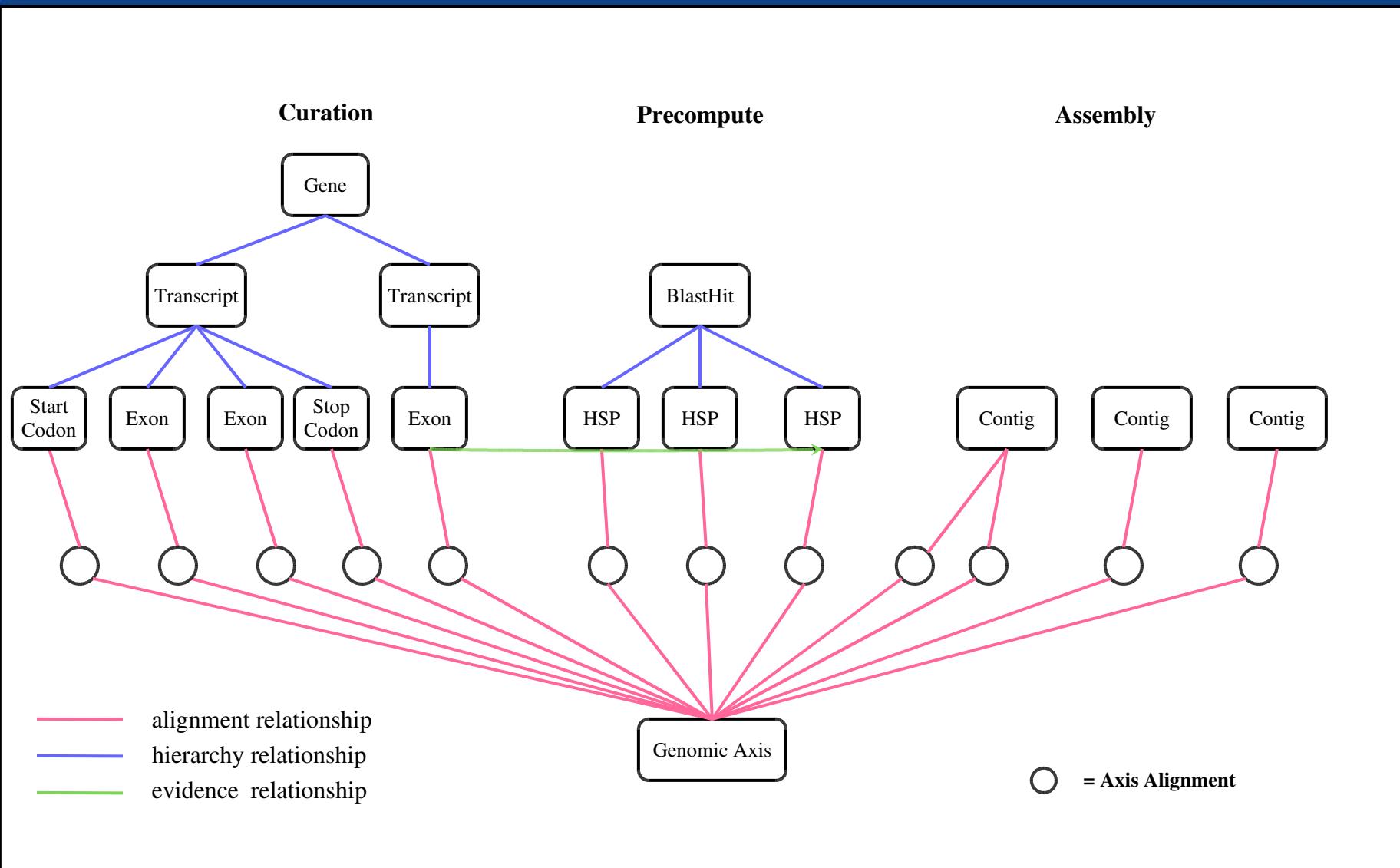
- High-performance, interactive graphics
- Client-side data analysis
- Client-side data integration via XML

† Two-way HTML browser connectivity with search capability

† Downloadable installation with self-updating client







✚ Main View Tiered Feature Display

- High-performance zoom
- Dragable Tiers
- Dockable Tiers
- Color coding of feature types
- Full user preferences for tiers, features, colors, orders etc.
- Color intensity mapping of arbitrary feature properties
- Highlighting of curation evidence, exon edge/splice matching, feature edge matching
- Thick Intron lines for adjacent HSP's in subject space
- Length of splice site glyph denotes score
- SNP glyphs shape and color

✚ Subviews

- Provide simultaneous micro-macro views with linked visual cues
- Nucleotide Sequence
- Spliced Transcript View with start/stop codons and reading frames
- Color-coded comparison of DNA and Protein sequences

✚ Navigation

- Bookmarks
- Searching
- Navigation launched from web browsers, email, (URL's)
- Sequence Analysis on the fly

† Direct Manipulation Operations

- “Drag and Drop” creation of transcripts, evidence
- Dragging exon edges to resize
- Infinite-level undo
- Annotation logging

† Right-click popup actions

- Deleting exons
- Splitting/Merging transcripts

† Reloadable Curation “Workspace”

- † Redraw Rate: 150,000 features/sec
- † Data Load Rate: 600 features/sec
- † Client Memory Footprint: 8 Mbyte + 700 byte/feature
- † Typical Database Size: 1 Terabyte, 150 Million features
- † Development Human Resources: ~20 person-years
- † Number of Classes: 1250 (client and server)
- † Lines of Code: 87,000 (client and server)

† For help preparing this talk

- Todd Safford, Patrick Dunn, Peter Davies, James Baxendale

† Visualization Team

- Deepali Bhandari, Peter Davies, Les Foster, Michael Harris, Olivier Jojic, Mark Lewis, Doug Mason, Rick Wise, Todd Safford, Jay Schira, Eric Sun, David Wu

† Apiary Team

- James Baxendale, Lou Blick, Bob Crandall, Ken Graham, Joe McDaniel, Michael Simpson, Mary Wen

† And many more...

- Mark Adams, Matteo di Tommaso, Patrick Dunn, Scott Henderson, Jackie Keeling, Frank Mayberry, M'Liz Robinson, Marian Skupski, and many others...