Bovine Annotation in Ensembl

Tim Hubbard

Wellcome Trust Sanger Institute

Bovine Genome Project Workshop

29-31st March 2005, Houston, Texas

Ensembl Genome Browser

Search Ensembl

Search all species for Anything with Lookup

About Ensembl

Ensembl is a joint project between EMBL - EBI and the Sanger Institute to develop a software system which produces and maintains automatic annotation on metazoan genomes. Ensembl is primarily funded by the Wellcome Trust.

This site provides free access to all the data and software from the Ensembl project. Click on the species buttons to the right to browse the data.

Access to all the data produced by the project, and to the software used to analyse and present it, is provided free and without constraints. Some data and software may be subject to third-party constraints [details].

For all enquiries, please contact the Ensembl *HelpDesk* (helpdesk@ensembl.org).

Help and documentation

- Take the Ensembl tour, go through a step-by-step worked example, or read these papers.
- For help on any web page click: Help
- There is also an index of help pages, and a set of guided How do I....? trails.

Display your own data in Ensembl — DAS

Questions or suggestions? Try the — Help Desk

Documentation (includes tutorial on direct data access & instructions for installing Ensembl on your own site) — Documentation

Try the site map as a good starting point for exploring what Ensembl has to offer

Species - Ensembl v29

Human		NCBI 35	Mar 05
Mouse		NCBI m33	Feb 05
Zebrafish		WTSI Zv4	Sep 04
Rat	pre!	RGSC 3.1	Jul 04
Chicken		WASHUC1	Jul 04
Mosquito		MOZ 2	Feb 05
Fugu		Fugu v2.0	May 04
Fruitfly		BDGP 3.2.1	Feb 05
Chimp		CHIMP1	Mar 05
Honeybee		Amel1.1	Sep 04
Tetraodon		TETRAODON7	Sep 04
Dog		BROADD1	Mar 05
C. elegans		WS 130	Dec 04
X. tropicalis		JGI3	Feb 05
S. cerevisiae		S228C	Mar 05
Cow	pre!	Btau_1.0	
Opossum	pre!	BROAD0.5	

Data

Sequence similarity searches — BLAST/SSAHA

Batch data/sequence retrieval — EnsMart

Ensembl Archive sites — Archive!

Vertebrate Genome Annotation (VEGA) — Vega

Access to whole genome shotgun data (includes additional species) — Trace Server

Download Ensembl data via FTP — Download

Pre! Ensembl Cow

The Wellcome Trust Sanger Institute

EBI

Cow Genome Browser

Ensembl Entry Points

Show scaffold | From 1 | To 10000 | Lookup

BLAST your sequence BLAST

For fast identity search try SSAHA

Cow Genome Project

Btau_1.0 is a preliminary 3x assembly of the draft genome sequence of cow (Bos taurus), Hereford breed, using whole genome shotgun (WGS) reads from small insert clones. The project coordination and genome sequencing and assembly is provided by the Human Genome Sequencing Center at Baylor College of Medicine.

The N50 size is the length such that 50% of the assembled genome lies in blocks of the N50 size or longer. The N50 of the contigs is 4.2 kb. The N50 of the scaffolds is 13.5 kb. The total length of all contigs is 2.26 Gb. When the gaps between contigs in scaffolds are included, the total span of the assembly is 2.34 Gb.

As this is a pre-release, the database does not contain any genes. Subsequent annotations including the ensembl genebuild are ongoing and will be added as soon as they are completed. Future assemblies will include WGS sequences with a larger insert sizes, BAC end sequences, BAC sequences, and marker information for more contiguous assembly, better scaffolding, and chromosome assignment.

Please refer to more details, conditions of use and credits.

Example Data Points

This release of cow data is unassembled, so there are no chromosomes available to browse. Use the BLAST, SSAHA in the Entry Points section above to locate data.

A few example data points :

- Scaffold: Scaffold 90001
- Scaffold: Scaffold 42

The assembly comprises 795212 contigs (which are named by Genbank accession) organised into 449727 scaffolds.

Documentation & Help

About Ensembl — Home

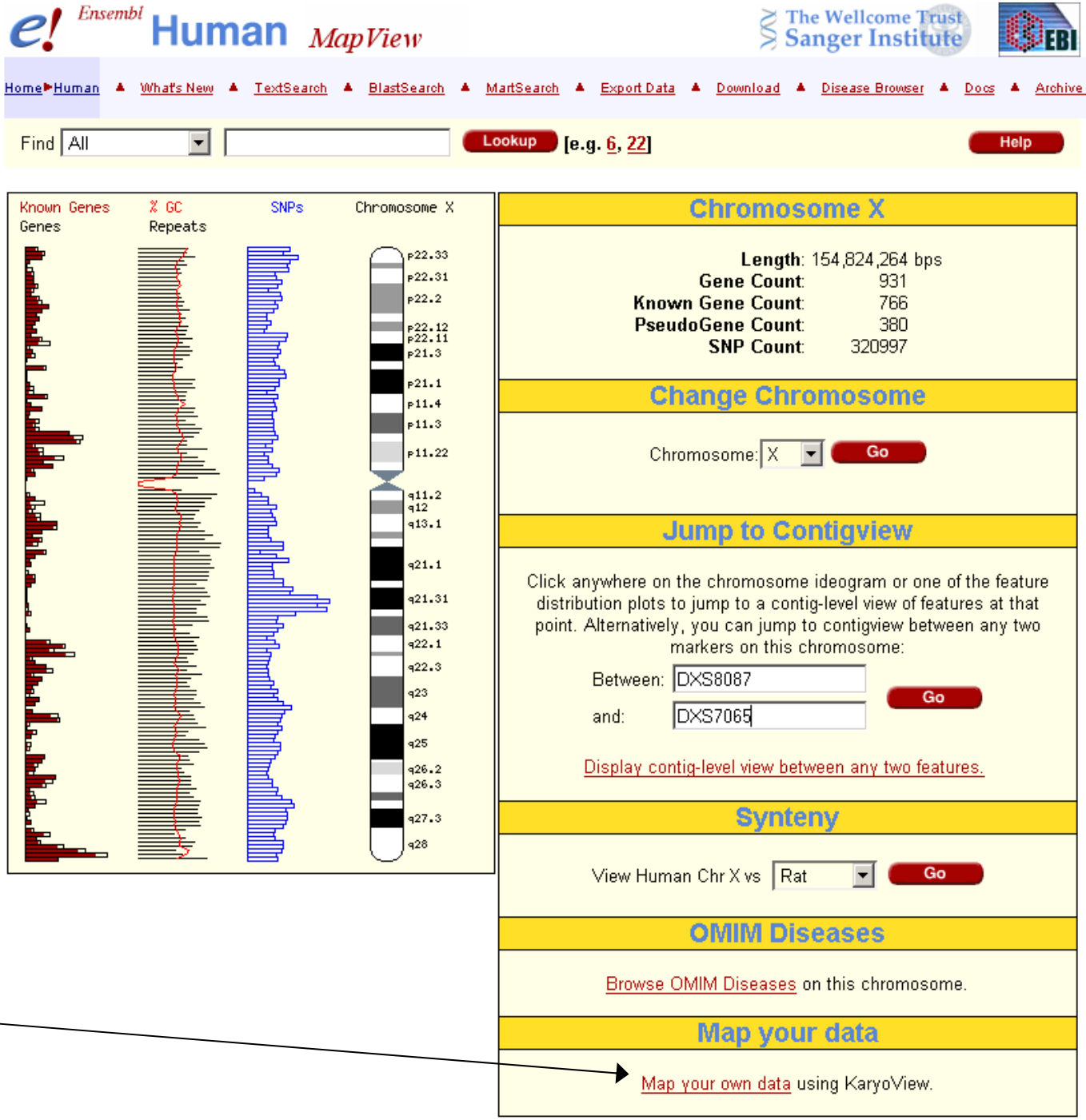
For context-sensitive help on any web page click — Help

Questions or suggestions? Try — Help Desk

Pre-Ensembl

- **Provides rapid access to new genome assemblies before a full gene build is complete**
  - Shows raw alignments to mRNA and Proteins
  - Provides BLAST and SSAHA search services

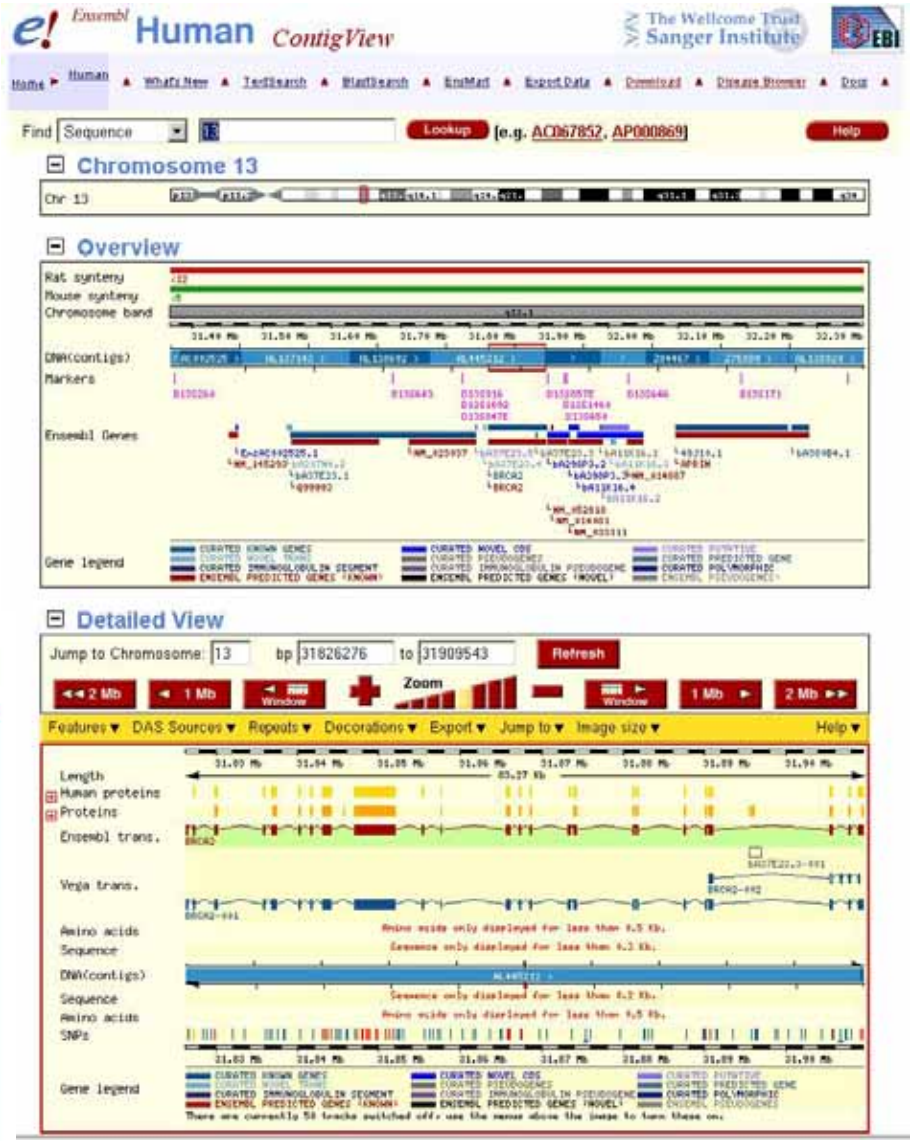
MapView

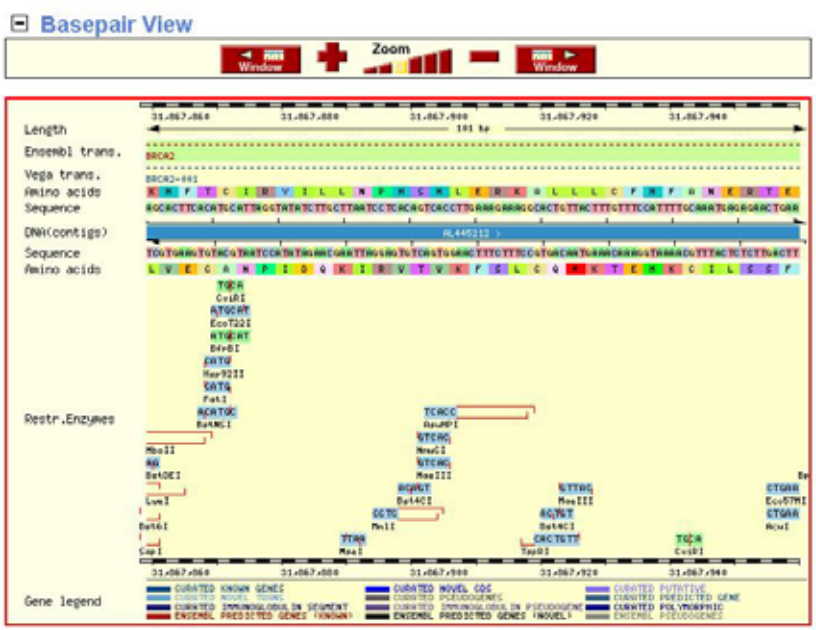


Display your own features

Ensembl

ContigView





new | SETUP | CONFIG | RESULTS | DISPLAY

refresh | Online Help

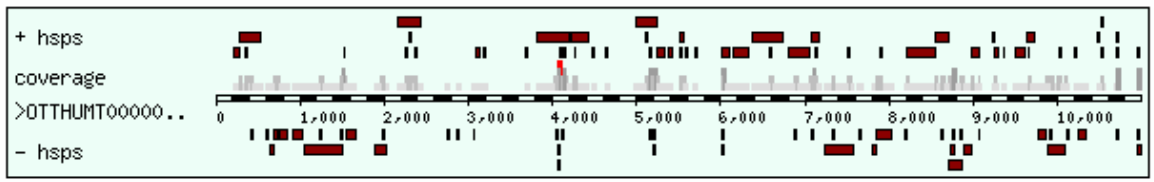
Displaying OTTHUMT00000046000 sequence alignments vs Bos_taurus LATESTGP database

Showing top 100 alignments of 194, sorted by Raw Score refresh

Alignment Locations vs. Karyotype (click arrow to hide)

Karyotype unavailable for Bos_taurus

Alignment Locations vs. Query (click arrow to hide)



Key (%ID): 0 - 20 | 20 - 40 | 40 - 60 | 60 - 80 | 80 - 100

Alignment Summary (click arrow to hide)

Select rows to include in table, and type of sort
(Use the 'ctrl' key to select multiples)

refresh

Query	Subject	Scaffold	Contig	Stats	Sort By
Name	_off_	Name	_off_	E-val	>Score
Start	Name	Start	Name	P-val	<E-val
End	Start	End	Start	%ID	>E-val
Ori	End	Ori	End	Length	<P-val

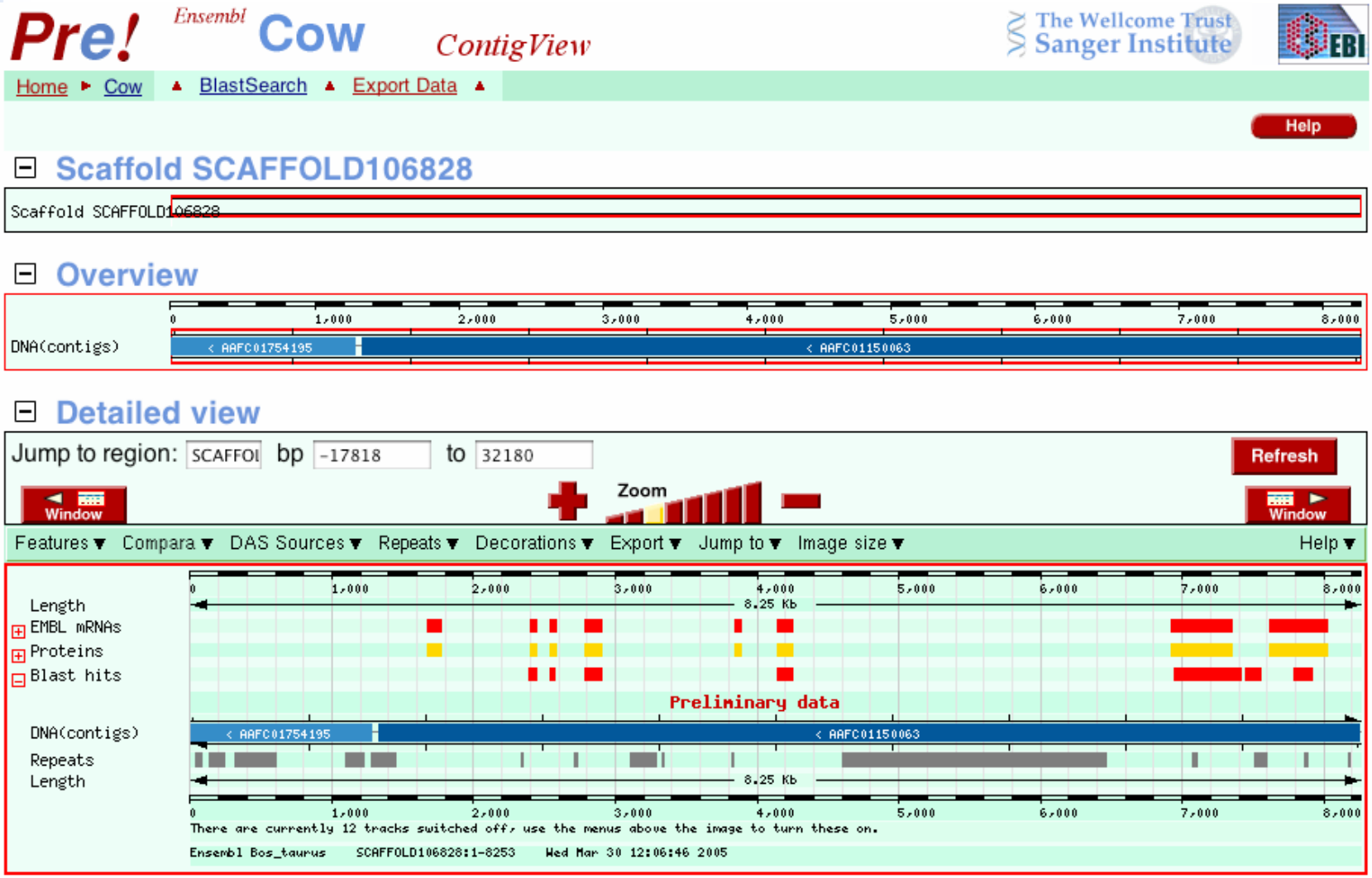
Links	Query Start	Query End	Query Ori	Scaffold Name	Scaffold Start	Scaffold End	Scaffold Ori	Stats Score	Stats E-val	Stats %ID	Stats Length
[A] [S] [G] [C]	8202	8558	+	SCAFFOLD256786	2755	3111	+	202	4.3e-146	89.11	358
[A] [S] [G] [C]	1049	1510	-	SCAFFOLD106828	6949	7413	-	163	6.0e-214	83.73	467
[A] [S] [G] [C]	7842	8033	-	SCAFFOLD256786	767	958	-	152	6.7e-85	94.79	192
[A] [S] [G] [C]	6381	6744	+	SCAFFOLD78058	1429	1798	+	149	0.	84.88	377
[A] [S] [G] [C]	3823	4197	+	SCAFFOLD388566	98	475	+	135	1.9e-99	83.81	383
[A] [S] [G] [C]	6796	7051	+	SCAFFOLD78058	1850	2102	+	125	0.	87.16	257
[A] [S] [G] [C]	292	531	+	SCAFFOLD207125	2446	2682	+	124	4.6e-61	87.92	240
[A] [S] [G] [C]	9877	10101	-	SCAFFOLD96316	9	236	+	120	7.4e-75	87.93	232
[A] [S] [G] [C]	7238	7566	-	SCAFFOLD1352	3494	3826	-	106	3.7e-54	82.84	338

Summary

- **setup**
  - Bos_taurus
  - Genomic sequence
  - BLASTN
  - Low sensitivity
- **configure**
  - -E: 10
  - -B: 100
  - -filter: dust
  - -RepeatMasker
  - -W: 15
  - -M: 1
  - -N: -3
  - -Q: 3
  - -R: 3
- **results**
- **display**

Not yet initialised

Blast hit shown on scaffold



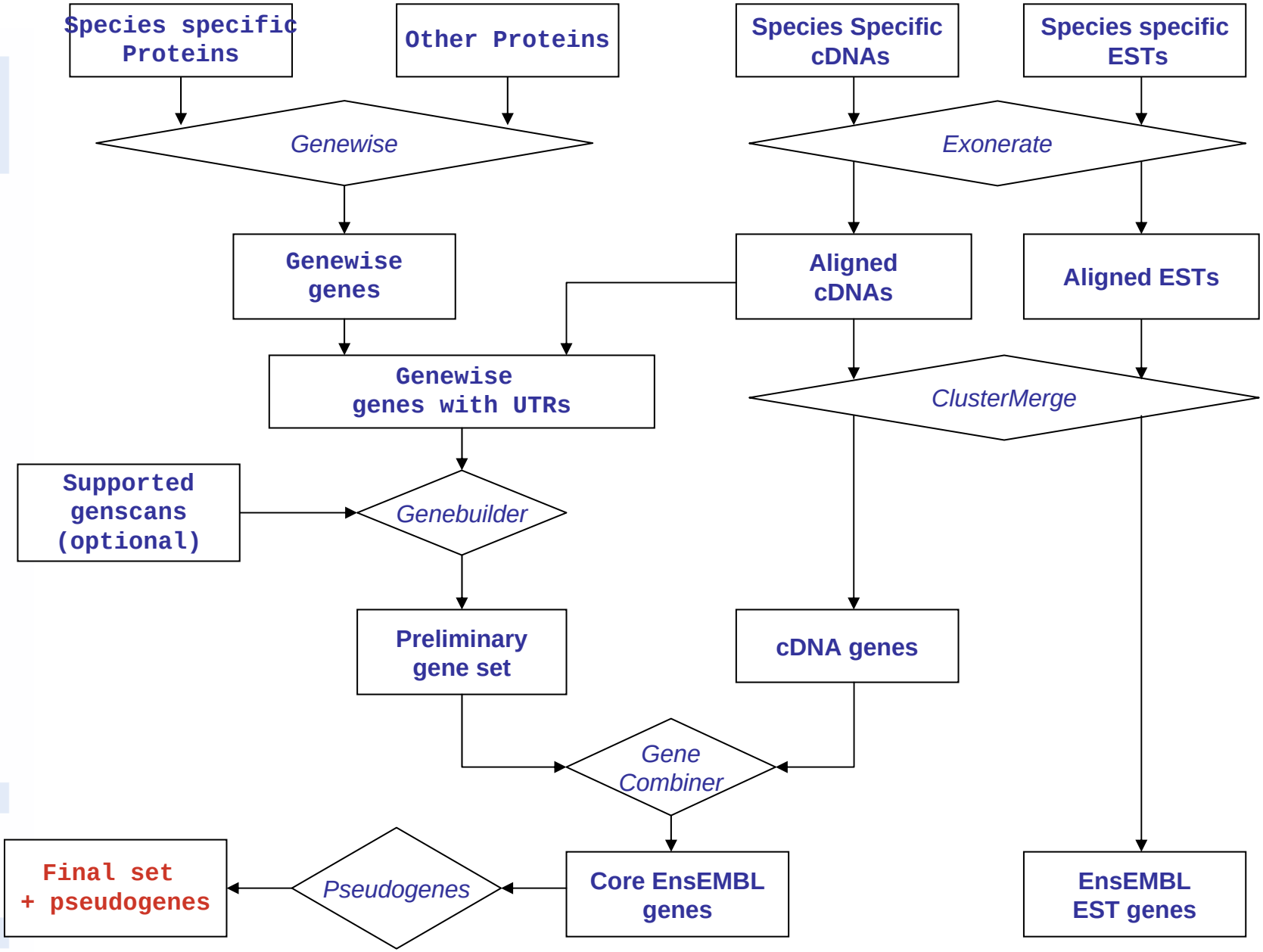
Builds released since last SAB

Species	# Coding Genes	# Transcript	# Builds released since Nov 2003
Human	22218	33869	3 full, 4 patch, 1 pre
Mouse	24461	31535	2 full, 3 patch, 2 pre
Rat	22159	28545	1 full in progress, 2 pre
Dog	18201	30308	1 full, 1 pre
Chicken	17709	28416	2 full (1 release), 1 patch (ncRNA), 1 pre
Fugu	20796	33003	1 full (1st in house)
Honey Bee	9671	16948	1 full (+ 1 in progress), 2 extended pre
Mosquito	14364	15802	1 patch (vectorbase team)
X. tropicalis	24405	52786	1 full, 1 pre (zfish group)
Zebrafish	23524	32062	2 full, 2 pre (zfish group)
Cow	-	-	1 pre, low coverage build development
Chimp	22475	43000	3 transfers from human (API team), 1 pre
Opossum	18936	32270	1 full in progress, 1 extended pre
C. intestinalis	-	-	1 full in progress
Tetraodon	28005	28005	1 import from Genoscope
D. melanogaster	13792	19178	1 import from Flybase
C. elegans	19765	24278	2 imports from Wormbase
S. cerevisiae	6680	6680	1 import from SGD
C. briggsae			1 full, REMOVED

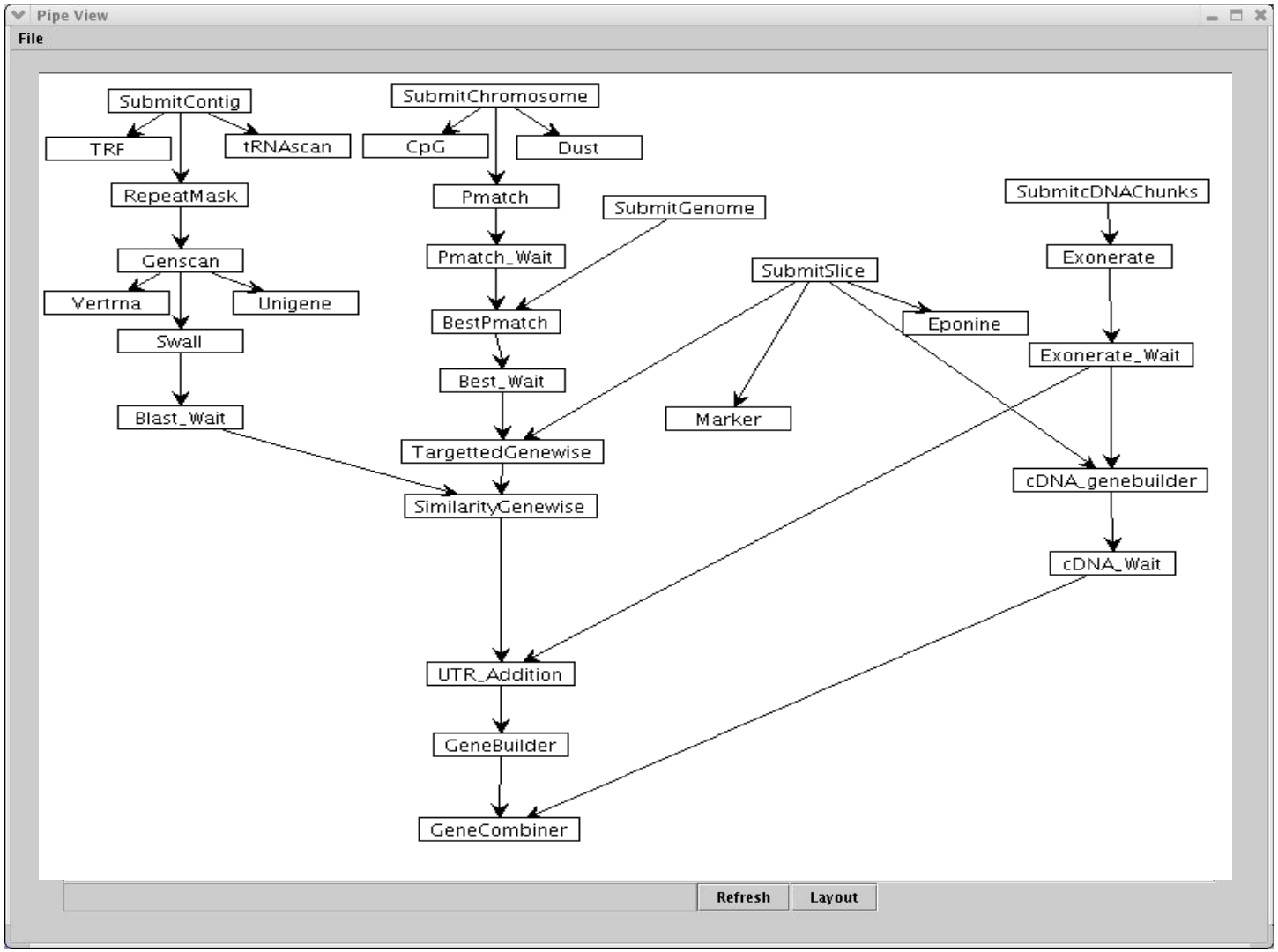
Stages in Genebuild Pipeline

- **the 'raw compute' (eg. blasts, repeat masking)**
  - All dependencies added at the start and run fully automated
- **A set of somewhat species specific steps requiring experimentation and assessment which we call the 'genebuild'**
  - Stages added individually (genewise steps) or in small groups (cDNA build, targetted build) to the pipeline
  - Allows us to modify the procedure, but gives us the pipeline control benefits (automatic job submission, retries, flagging failures, batching jobs)
- **A set of post processing steps**
  - Pseudogene labelling
    - Pipelined during the last year. Can be customised.
  - Protein annotation
    - Fully automated like the 'raw compute' stage
  - Cross reference generation
    - Recently automated by the API team (separate system)

Gene build Summary



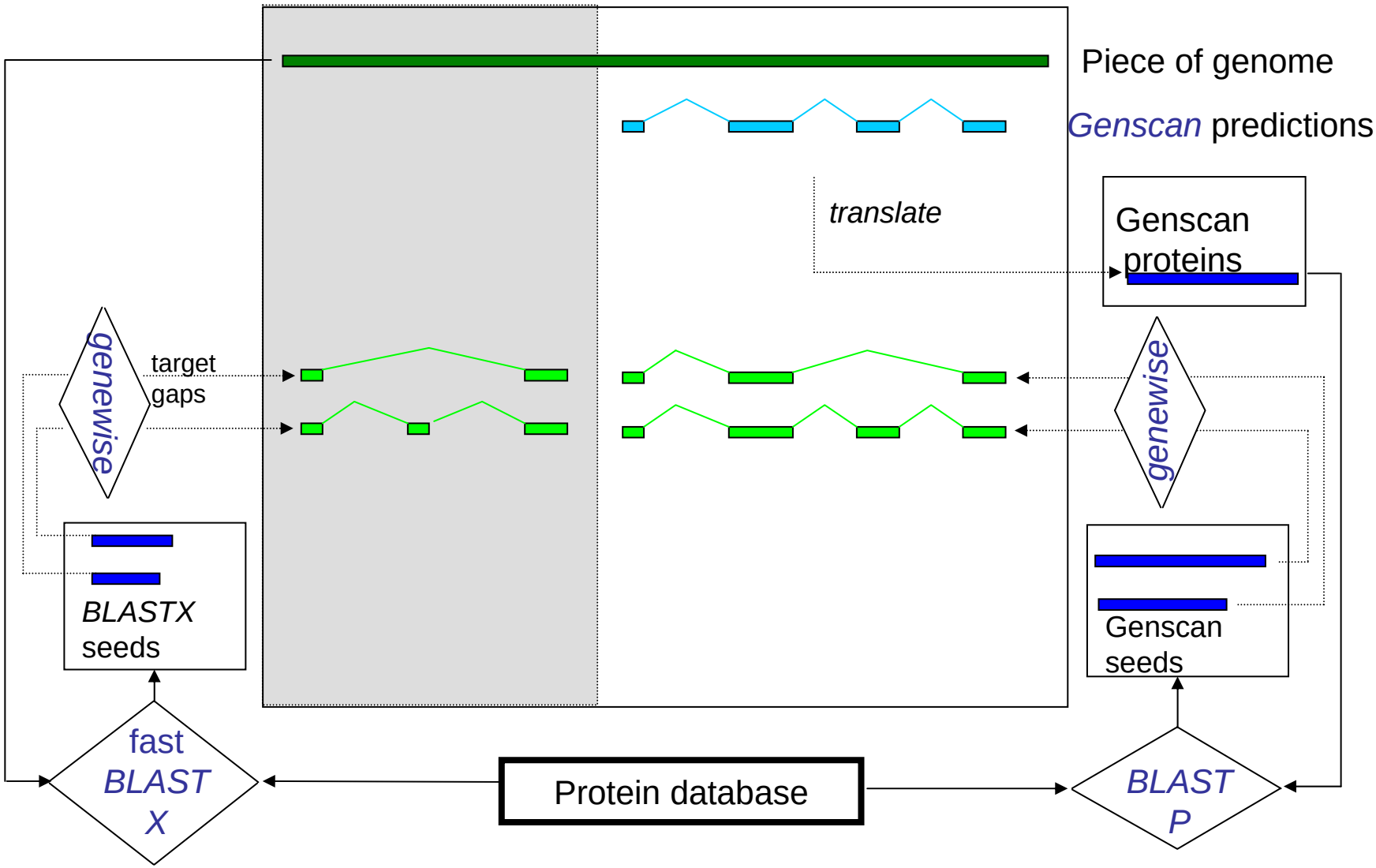
A Genebuild Pipeline



Each organism has different characteristics and resources which affect genebuild design

- Evolutionary distance to closest well characterised species
- Number of species specific protein and cDNA sequences
- Number and quality of ESTs
- Genome size
- Coverage
- Availability of manual annotation data

Fugu Genebuild



Low Coverage Genomes

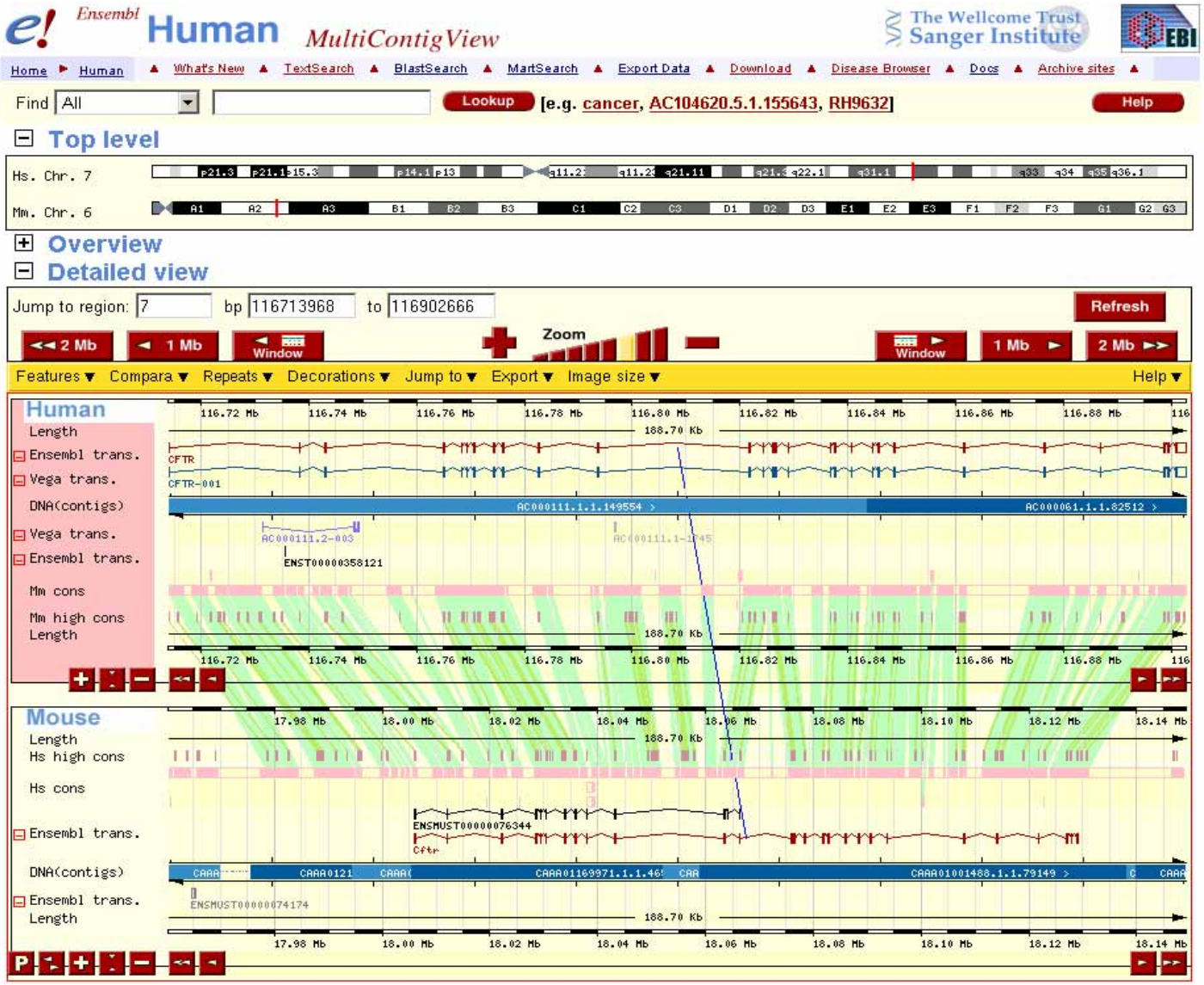
- **8 mammalian genomes sequenced to 2x are expected over the next 2-3 years.**
- **Ensembl is aiming to provide gene sets for these based on alignments to human, building predictions on scaffolds which align to genomic locations of human genes**
- **Test case**
  - Cow preliminary 3x assembly 449727 scaffolds, 795212 contigs
  - Good test case because 6x assembly will be available soon

Cow 3x gene build

- **Define order of cow scaffolds from genome alignment to human (virtual assembly)**
- **Project human genes across alignment onto virtual cow assembly**
- **Identify Cow specific genes using standard pipeline**
  - coverage depends on coverage of Cow cDNAs)
  - Initial results: 80/2500 cDNA alignments gave matches that did not overlap with human projections. Some appear to be junk; some likely to be real.

- **Many issues resolving Cow-Human genome matches, assembly issues etc.**
- **Will investigate projection from Dog as Dog gene build/assembly improves**

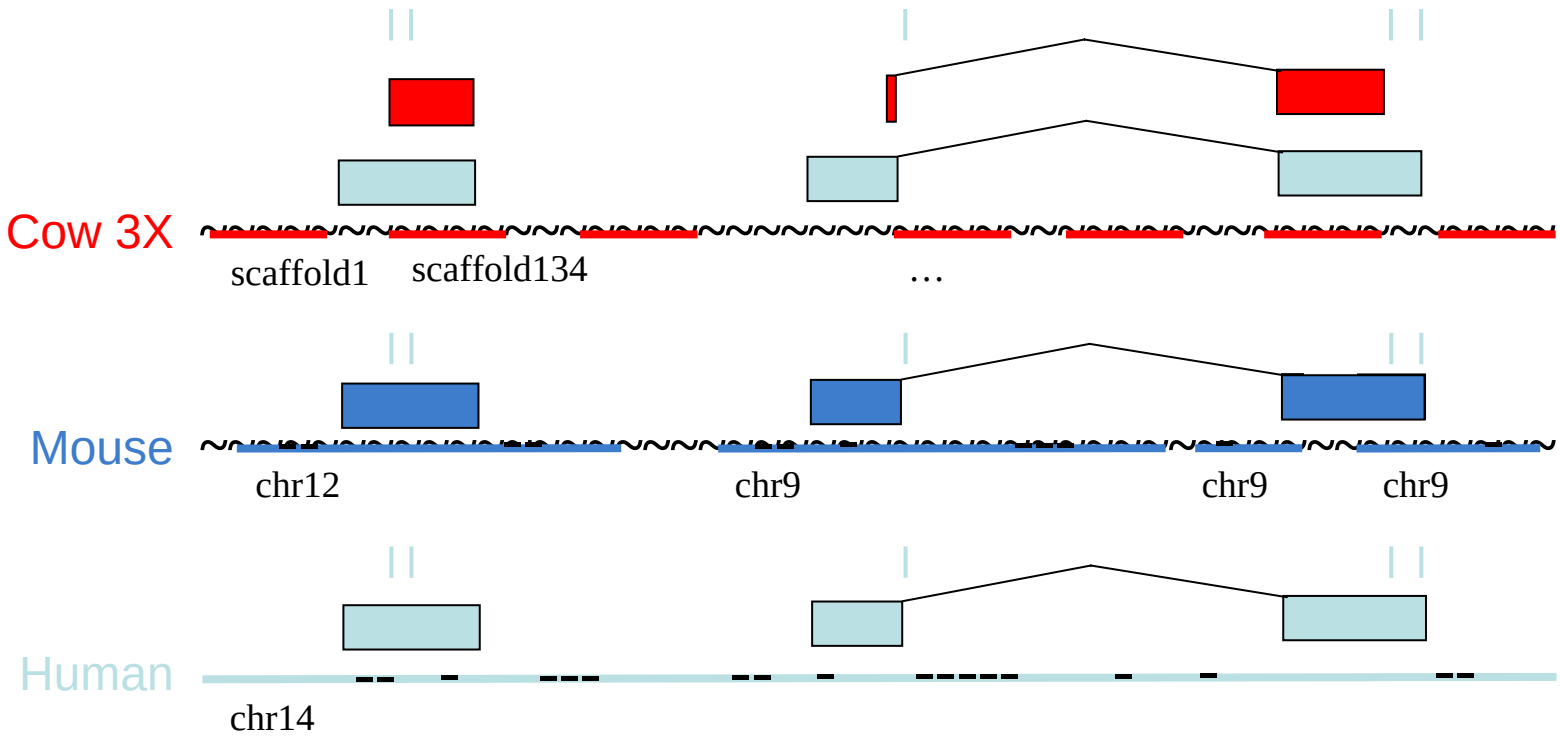
MultiContigView



AlignSlice API

- Using whole genome pairwise/multiple alignment data, to generate a reference coordinate system common to the aligned species in the genomic region of interest.
- Being able to project a transcript from one species to another through the alignment data
- Give gene context information across species, and more generally give annotation context information.
- Needed as a significant number of genomes are going to be 2X/3X where no sensible gene building is likely to give good quality gene set.

AlignSliceView mock-up



Acknowledgements

EBI
Ensembl

Comparison
Abel Ureta-Vidal
Jessica Severin
Cara Woodwark
Javier Herrero

Mart
Arek Kasprzyk
Craig Melsopp
Glenn Proctor
Damian Smedley

Core API
Arne Stabenau
Ian Longden

Helpdesk
Xose Fernandez-S.
Damien Keefe

Guy Slater
Yuan Chen
Darin London

Ewan Birney

Sanger
Ensembl

Genebuild
Steve Searle
Val Curwen
Dan Andrews
Laura Clarke
Kevin Howe
Vivek Iyer
Felix Kokocinsci
Jan Vogel
Simon White

Web
James Stalker
James Smith
Fiona Cunningham
Paul Bevan

Zebrafish
Kerstin Jekosch
Mario Caccaro

Vega
Patrick Meidl
Steve Trevanion

Systems
Guy Coates
Tim Cutts
Mark Rae
Simon Kelley

Tim Hubbard
Tony Cox
Richard Durbin

NCBI and the Bovine Genome

- Bovine Genome specific public resources at NCBI
- Bovine assembly and annotation activities
- Ensuring the Bovine Genome information is included in a broad information context

NCBI

110101

Bovine Specific Resources

UniSTS - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address http://www.ncbi.nlm.nih.gov/genome/sts/sts.cgi?uid=278327

NCBI UniSTS Integrating Markers and Maps

PubMed Entrez BLAST OMIM Taxonomy Structure

Search UniSTS for Go

Entrez UniSTS
- Help
- Query tips
- Submit
- Submit map
- FTP site
- Statistics

Related sites
- e-PCR
- Map Viewer
- Gene
- UniGene
- dbSNP
- GeneMap'99
- RHdb
- GDB
- MGD
- ZFIN

Genomic biology
- *Bos taurus*
- *Canis familiaris*
- *Danio rerio*
- *Homo sapiens*
- *Mus musculus*
- *Rattus novegicus*
- *Sus scrofa*

UniSTS:278327 Links

EDG1

Bos taurus chromosome 3
Homo sapiens locus EDG1
Pan troglodytes chromosome 1, locus LOC457063

Found by e-PCR in sequences from *Homo sapiens*, *Pan troglodytes* and *Sus scrofa*.

Primer Information

Forward primer:	TGGCCCTCTCAGACCTGTTG
Reverse primer:	TGGCGAGGAGACTGAACACG
PCR product size:	not available
GenBank Accession:	M31210

Bos taurus

Name: EDG1

Mapping Information

EDG1	ILTX-2004 Map:	Chr 3|LG1	Map Viewer
	Position:	138.21 (cR5000)	

Homo sapiens

Cross References

Gene	GeneID:	1901
	Symbol:	EDG1
	Description:	endothelial differentiation, sphingolipid G-protein-coupled receptor, 1
	Position:	1p21

Internet

Map Viewer - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address http://www.ncbi.nlm.nih.gov/mapview/maps.cgi?TAXID=9913&CHR=3&BEG=0.00&END=312.8

NCBI NCBI Map Viewer

PubMed Entrez BLAST OMIM Taxonomy Structure

Search Find Find in This View Advanced Search

Map Viewer Home
- Map Viewer Help
- Cow Maps Help
- Data As Table View
- Maps & Options
- Compress Map
- Region Shown: Go

out zoom in

You are here:

default
master

Bos taurus (cow) **Build 1.1** BLAST The Cow Genome

Chromosome: 1 2 [3] 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 X Y MT

Master Map: ILTX Summary of Maps Maps & Options

Region Displayed: **0.00-312.81 cR**

MARC ILTX LG1 LG2

Marker	cR	LG	LinkOut	MARC	ILTX	Polymorphism
UAP1	1.6	LG1	STS acc	·	●	N
MUC1	55.6	LG1	STS acc	·	●	N
S100A6	57	LG1	STS acc	·	●	N
TUFT1	63	LG1	STS acc	·	●	N
RME23	75.1	LG1	STS acc	●	●	Y
PSMD4	90	LG1	STS acc	·	●	N
NRAS	96.8	LG1	STS acc	·	●	N
OVGP1	124.9	LG1	STS acc	·	●	N
FLJ10330	128.4	LG1	STS acc	·	●	N
BE217498	138.2	LG1	STS acc	·	●	N
EDG1	138.2	LG1	STS acc	·	●	N
CDC7L1	142.1	LG1	STS acc	·	●	N
TGFBR3	142.1	LG1	STS acc	·	●	N
ILSTS044	8.3	LG2	STS acc	·	●	N
AW267062	23.3	LG2	STS acc	·	●	N
NEDD5	63.5	LG2	STS acc	·	●	N
MUF1	88.3	LG2	STS acc	·	●	N
ACF7	93.6	LG2	STS acc	·	●	N
MIP-T3	121.5	LG2	STS acc	·	●	N
SFPQ	125.3	LG2	STS acc	·	●	N

Internet

NCBI

110101

Bovine Specific Resources

Gene - Microsoft Internet Explorer

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene&cmd=retrieve&dopt=graphics&list_uids=280758

NCBI Entrez Gene

My NCBI [Sign In] [Register]

All Databases PubMed Nucleotide

Search Gene for

Limits Preview/Index

Display Graphics

All: 1 Genes Genomes:

Entrez
SITE MAP
Entrez Help

Gene
About
Search
Gene Help
NCBI Handbook

My NCBI
My NCBI help

FAQ

FTP site

Related sites
Entrez Genome
Genome Project
Genomic Biology
GEO
HomoloGene
Map Viewer
OMIM
RefSeq
UniGene
UniSTS

Feedback
Help Desk
Corrections
About GeneRIFs

Subscriptions
RefSeq

1: **DBH dopamine monooxygenase)** [*Bc*

GeneID: 280758

Genomic context: chromo:

Gene type: protein coding

Gene description: dopami

RefSeq status: Provisional

Organism: *Bos taurus*

Lineage: *Eukaryota; Metc Eutheria; Laurasiatheria;*

Gene aliases: DBH

Bibliography:

PubMed links

GeneRIFs:

1. dioxygen and substrate ac
2. The present study present beta-hydroxylase (DBH; do independently by pH and by

General gene informati

Markers (Sequence Tagg

DBH (e-PCR)

General protein informa

Name: dopamine beta-hy

NCBI Reference Seque

mRNA Sequence NM

Entrez PubMed - Microsoft Internet Explorer

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=12966

NCBI PubMed National Library of Medicine NLM

My NCBI [Sign In] [Register]

All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books

Search PubMed for Go Clear

Limits Preview/Index History Clipboard Details

Display Abstract Show: 20 Sort Send to Text

All: 1 Review: 0

About Entrez

Text Version

Entrez PubMed
Overview
Help | FAQ
Tutorial
New/Noteworthy
E-Utilities

PubMed Services
Journals Database
MeSH Database
Single Citation Matcher
Batch Citation Matcher
Clinical Queries
LinkOut
My NCBI (Cubby)

Related Resources
Order Documents
NLM Catalog
NLM Gateway
TOXNET
Consumer Health
Clinical Alerts
ClinicalTrials.gov
PubMed Central

1: J Biol Chem. 2003 Dec 12;278(50):49691-8. Epub 2003 Sep 9. Related Articles, Links

FREE full text article at www.jbc.org

Evidence that dioxygen and substrate activation are tightly coupled in dopamine beta-monooxygenase. Implications for the reactive oxygen species.

Evans JP, Ahn K, Klinman JP.

Department of Chemistry, University of California, Berkeley, California 94720-1460, USA.

Oxygen activation occurs at a wide variety of enzyme active sites. Mechanisms previously proposed for the copper monooxygenase, dopamine beta-monooxygenase (DbetaM), involve the accumulation of an activated oxygen intermediate with the properties of a copper-peroxo or copper-oxo species before substrate activation. These are reminiscent of the mechanism of cytochrome P-450, where a heme iron stabilizes the activated O2 species. Herein, we report two experimental probes of the activated oxygen species in DbetaM. First, we have synthesized the substrate analog, beta,beta-difluorophenethylamine, and examined its capacity to induce reoxidation of the prereduced copper sites of DbetaM upon mixing with O2 under rapid freeze-quench conditions. This experiment fails to give rise to an EPR-detectable copper species, in contrast to a substrate with a C-H active bond. This indicates either that the reoxidation of the enzyme-bound copper sites in the presence of O2 is tightly linked to C-H activation or that a diamagnetic species Cu(II)-O2* has been formed. In the context of the open and fully solvent-accessible active site for

http://www.ncbi.nlm.nih.gov/entrez/utils/lofref.fcgi?PrId=3051&uid=12966104&db=pubmed&url=http://www.jbc.org/

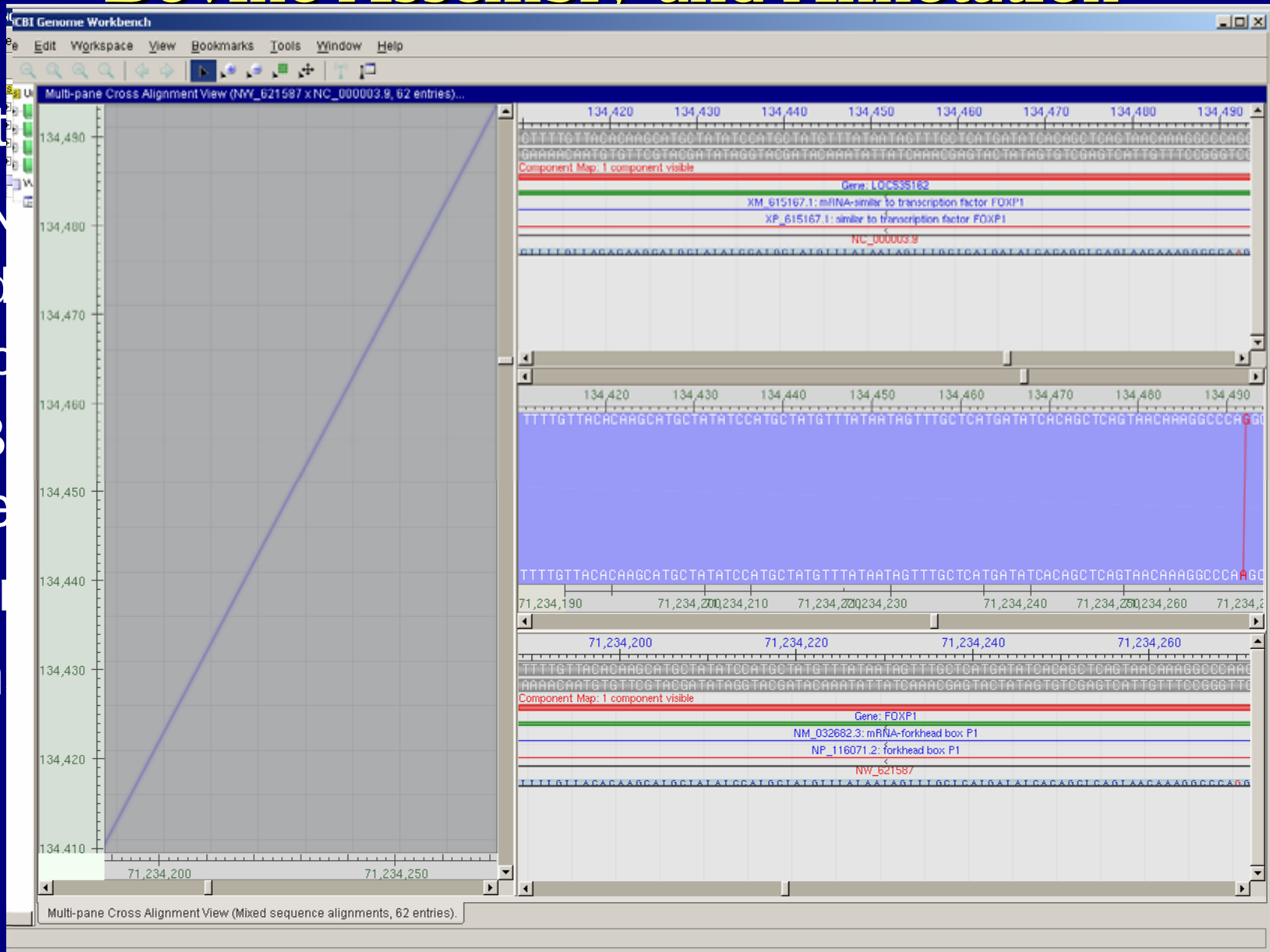
Internet

NCBI

110101

Bovine Assembly and Annotation

- Cont
- mRN
  - Ad
  - inc
- NCB gene
- Bovi
- Tran

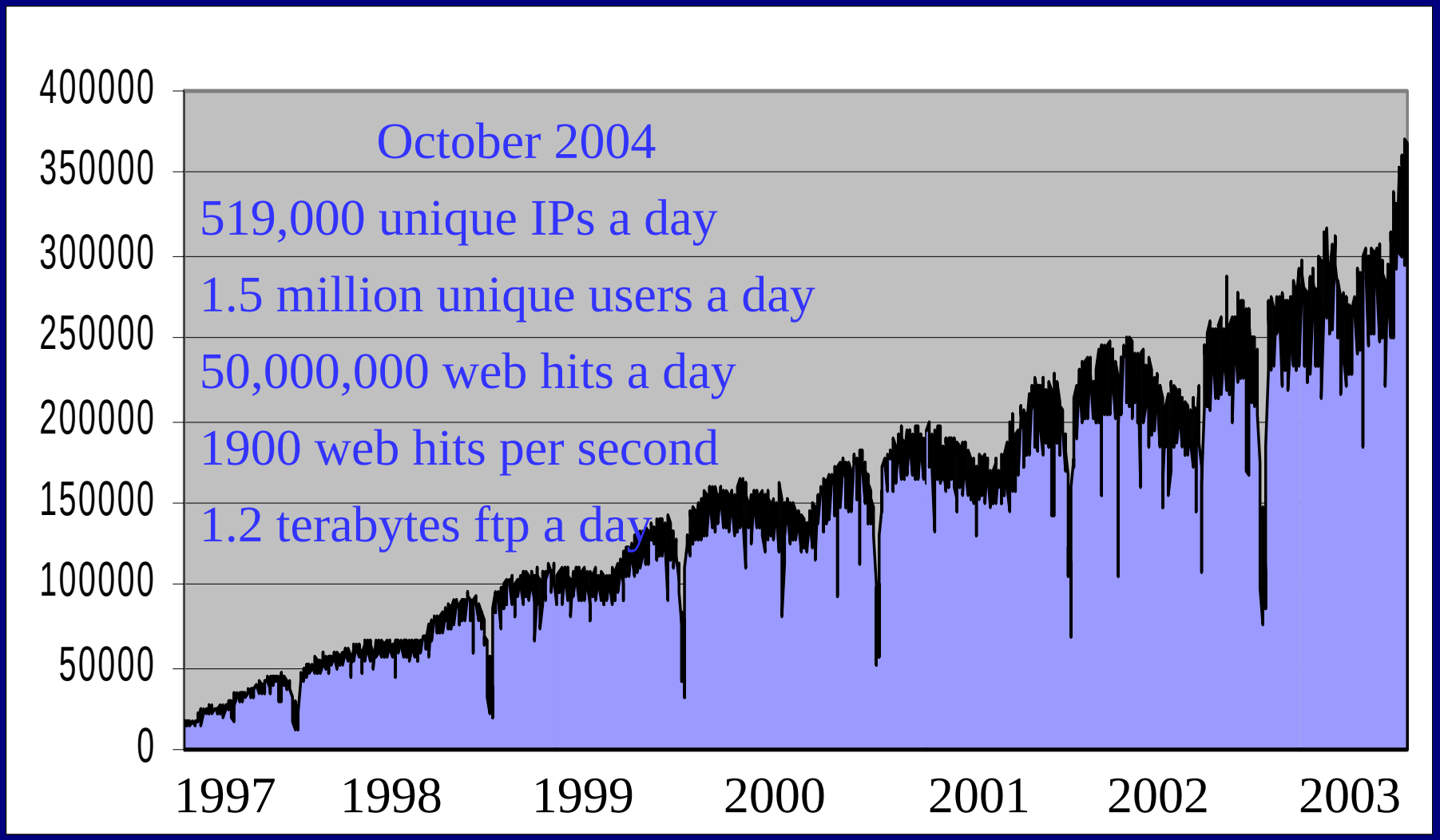


Bovine Assembly and Annotation

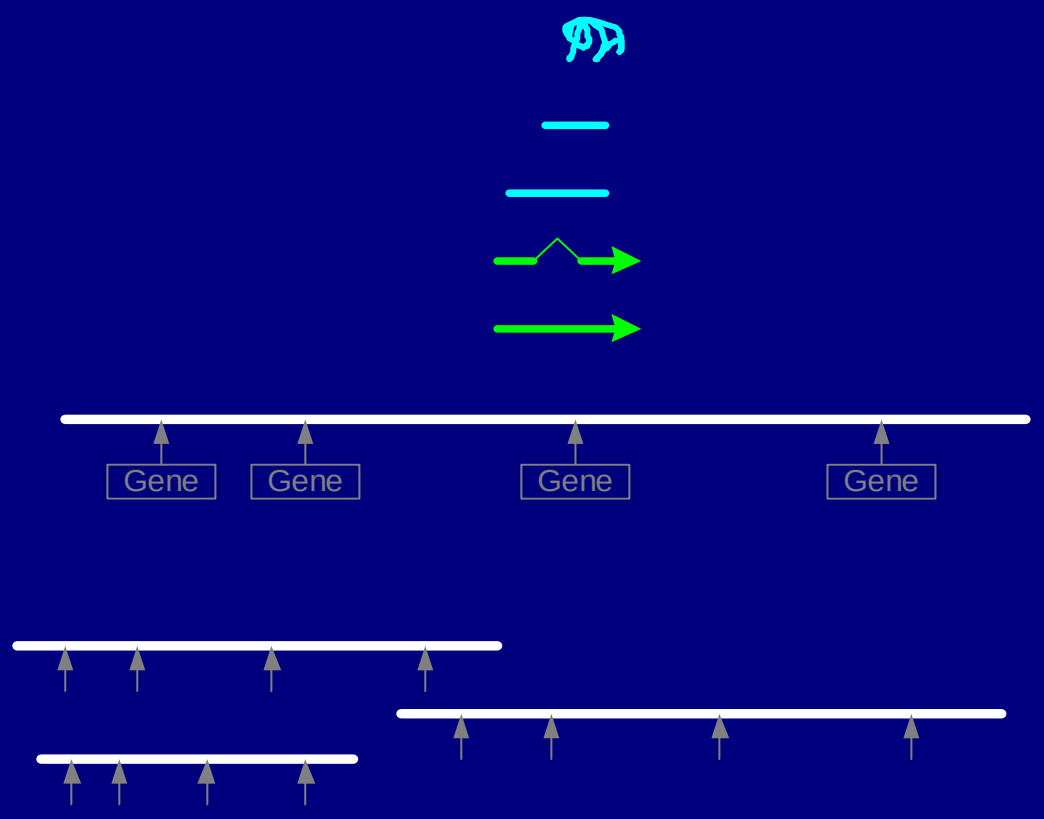
- Contamination Screening – 167 contigs in 3x
- mRNA Based Super-scaffolds
  - Added 6297 aligned human cDNAs, 169 bovine
  - increased max contig length 2X
- NCBI genome pipeline placed 2340 known bovine genes (created 35,483 models)
- Bovine MGC picking, sequence evaluation
- Transcript anchored genomic alignment
- Traces, Assembly Database, SNPs

NCBI

Bovine in a Broader Context



Bovine in a Broader Context



Disease

Function

Structure

Mature Peptide

ProPeptide

mRNA

Transcript

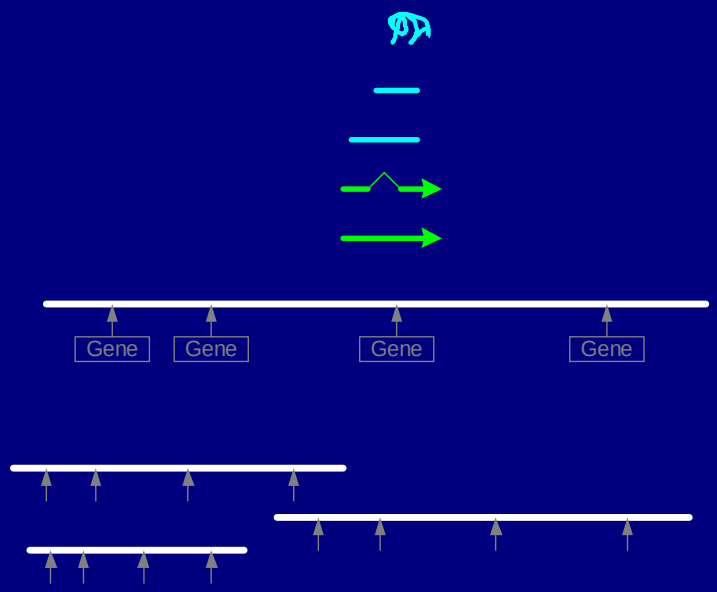
Chromosome

Genetics

Genomes

Organisms

Bovine in a Broader Context



Disease
Function
Structure
Mature Peptide
ProPeptide
mRNA
Transcript
Chromosome
Genetics
Genomes
Organisms

GenBank
RefSeq
Human Genome
Bacterial Genome
Virus Genome
MMDB
PubMed
UniGene(s)
LocusLink
OMIM
Taxonomy
GEO
PopSet
BLAST
Entrez
ePCR
Sequin

NCBI

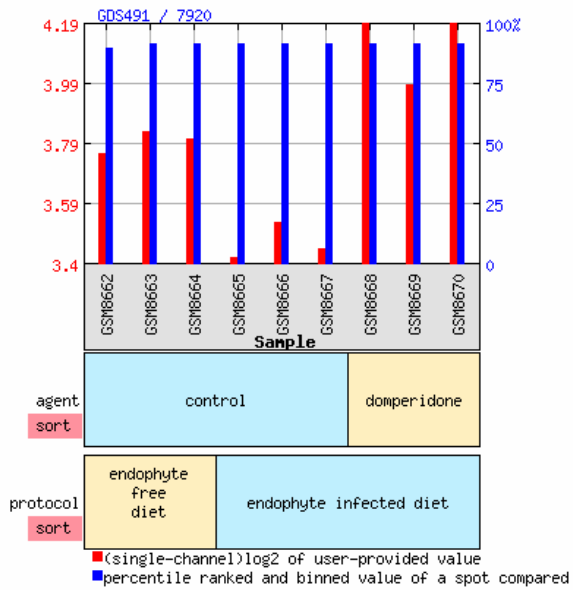
110101

Bovine in a Broader Context

http://www.ncbi.nlm.nih.gov/projects/geo/gds/gdsGraph.cgi?&...

File Edit View Favorites Tools Help

Back Address http://www.ncbi.nlm.nih.gov/p



GSM8662:Endophyte free, rep 1
GSM8663:Endophyte free, rep 2
GSM8664:Endophyte free, rep 3
GSM8665:Endophyte infected, rep 1
GSM8666:Endophyte infected, rep 2
GSM8667:Endophyte infected, rep 3
GSM8668:Endophyte infected + domperidone, rep 1
GSM8669:Endophyte infected + domperidone, rep 2
GSM8670:Endophyte infected + domperidone, rep 3

Done Internet

Entrez Protein - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Address http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=DisplayFiltered&DB=protein

NCBI Entrez Protein [Sign In] [Reg

All Databases PubMed Nucleotide Protein Genome Structure PMC Taxonomy

Search Protein for Go Clear

Limits Preview/Index History Clipboard Details

Display Summary Show 20 Sort by Send to

All: 211 bacteria: 1 RefSeq: 46

Items 1 - 20 of 46 Page 1 of 3 Next

About Entrez

Entrez Protein
Help | FAQ

Entrez Tools

Check sequence revision history

LinkOut

My NCBI (Cubby)

Related resources
BLAST

Reference sequence project

Search for Genes

Clusters of orthologous groups

Protein reviews on the web

Search for full length cDNAs

1: NP_037290 Reports BLink, Domains, Links
dopamine beta hydroxylase [Rattus norvegicus]
gi|25742780|ref|NP_037290.1|[25742780]

2: NP_620392 Reports BLink, Domains, Links
dopamine beta hydroxylase [Mus musculus]
gi|20336728|ref|NP_620392.1|[20336728]

3: NP_851338 Reports BLink, Domains, Links
dopamine beta-hydroxylase (dopamine beta-monooxygenase) [Bos taurus]
gi|30794286|ref|NP_851338.1|[30794286]

4: NP_000778 Reports BLink, Domains, Links
dopamine beta-hydroxylase precursor [Homo sapiens]
gi|18426906|ref|NP_000778.2|[18426906]

5: XP_520341 Reports BLink, Domains, Links
PREDICTED: similar to dopamine beta-hydroxylase precursor; dopamine beta-monooxygenase [Pan troglodytes]

Internet

NCBI

110101

Bovine in a Broader Context

Entrez GEO - Gene Expression Omnibus - Microsoft Internet Explorer

Gene - Microsoft Internet Explorer

1: Gad1 glutamate decarboxylase 1 [*Rattus norvegicus*]

GeneID: 24379 Locus tag: RGD:2652; RATMAP: 33958 updated 16-Mar-2005

Official Symbol: Gad1 **and Name:** glutamate decarboxylase 1 **provided by** Rat Genome Database

Transcripts and products: RefSeq below

NM_017167

Genomic context: chromosome: 3; **Maps:** 3q21

Gene type: protein coding

Gene name: Gad1

Gene description: glutamate decarboxylase 1

RefSeq status: Provisional

Organism: *Rattus norvegicus*

Lineage: *Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; E Glires; Rodentia; Sciurognathi; Muridae; Murinae; Rattus*

Gene aliases: Gad67

Summary: SUMMARY: plays a role in the biosynthesis of gamma-aminobutyric acid [RGD]

Bibliography: Gene References into Function (GeneRIF): Submit help

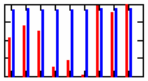
PubMed links

GeneRIFs:

1. palmitoylation of GAD65 regulates the trafficking of the protein from Golgi membranes to an endo in axons that is dependent on Rab5a and is required for the targeting of several synaptic vesicle prot clusters
2. The results revealed a significant decline in GAD-IR cells between middle and old age in CA1 but CA3.
3. With age, overall expression of GAD(67) mRNA decreases in the area surrounding the organum terminalis and in the anteroventral periventricular nucleus. Young rats display a diurnal rhythm in GA regions.
4. Many areas of the brainstem and cerebellum involve double-labeled neurons with GLYT2 and GAD67 mRNAs and suggests that the corelease of glycine and GABA from single neurons is more widespread than has been reported. PubMed
5. The effects of a subthalamic nucleus lesion on 6-OHDA- or repeated D2 antagonist-induced changes in globus pallidus GAD(67) mRNA expression in parvalbumin[PV]+ and PV- neurons was examined in rats PubMed
6. The mRNA expression of GAD65 is up-regulated in the vestibular nuclei bilaterally 50 h after labyrinthectomy. In the flocculus, GAD65 mRNA expression isbilaterally up-regulated 50 h post-operatively. PubMed
7. acute furosemide treatment reduced the expression of GAD67 mRNA, the active holoenzyme predictive of GABA synthesis, within the lamina terminalis PubMed
8. in 6-hydroxydopamine-lesioned rats, GAD67 mRNA levels in striatonigral and striatopallidal pathways were selectively modified, and the modification correlated to dopamine agonist priming PubMed

LOC529488

9 samples Profile Neighbors, Sequence Neighbors, Homologs, Links

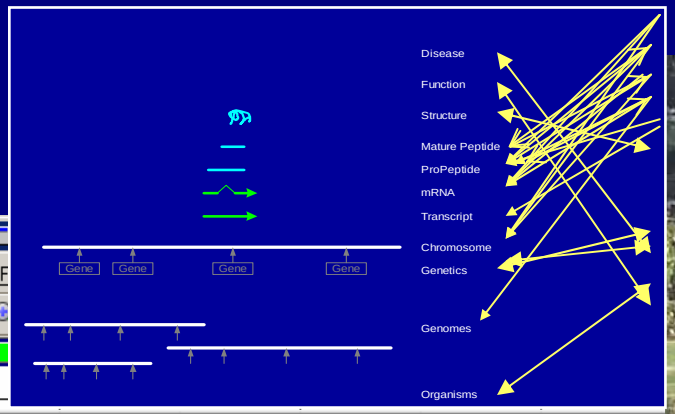


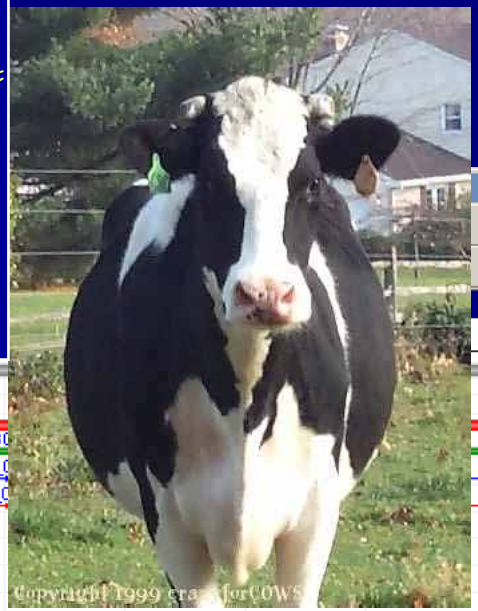
channel nucleotide count

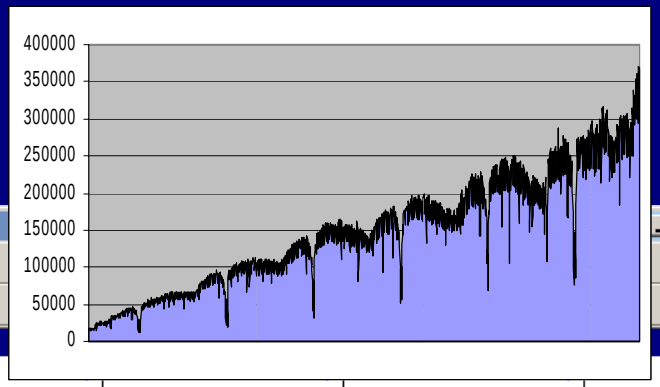
9 samples Profile Neighbors, Links

NCBI and the Bovine Genome

- Bovine Genome specific public resources at NCBI
- Bovine assembly and annotation activities
- Ensuring the Bovine Genome information is included in a broad information context







NCBI

Bovine Genome in Context

ANALYSES FOR THE CATTLE GENOME PROJECT

Chris Ponting

Leo Goodstadt

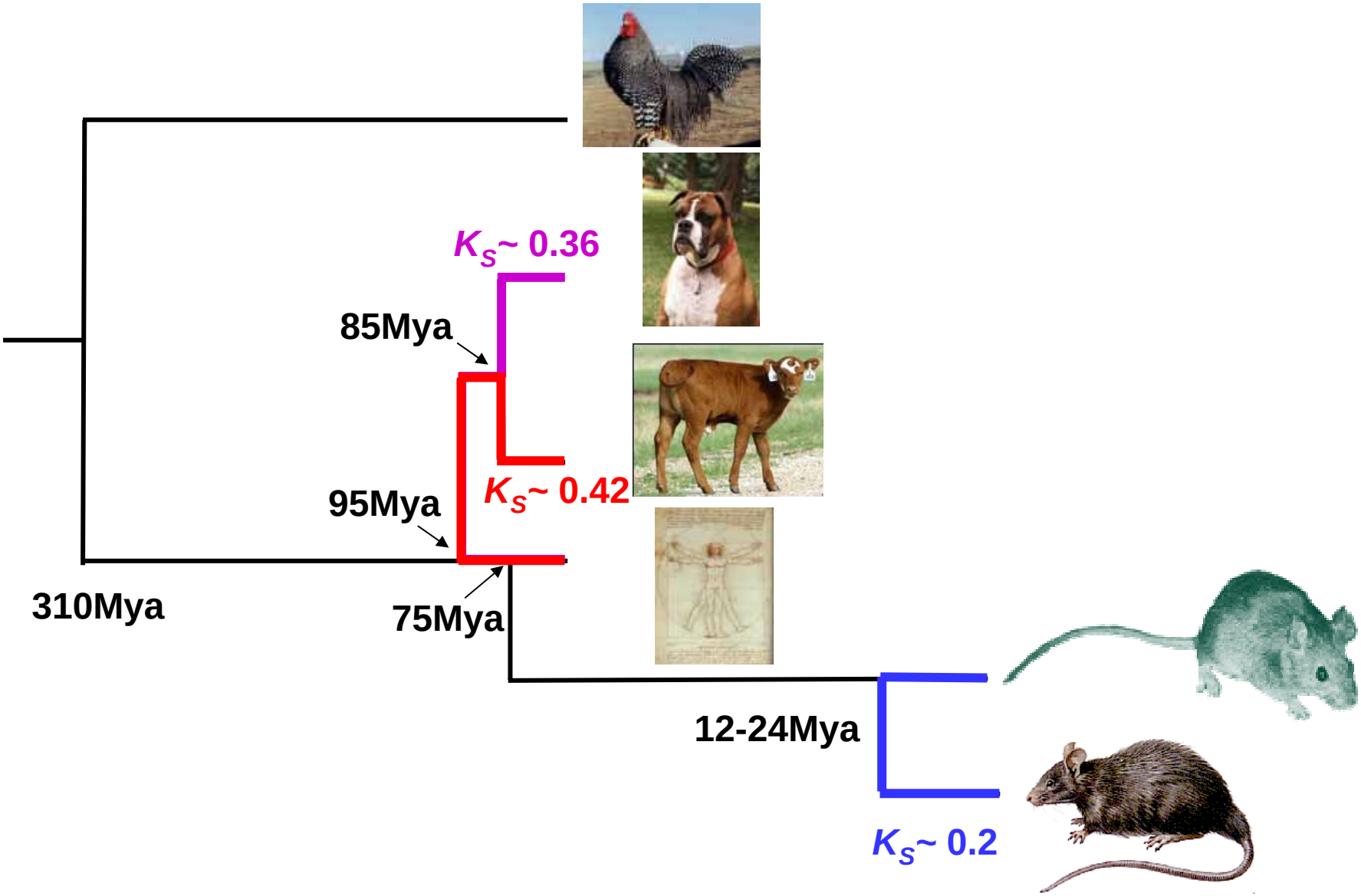
Caleb Webber

Andreas Heger

MRC Functional Genetics Unit

2001	2002	2003	2004	2005
			  	
EVOLUTION				
Domain architectures	**Protein Evolution Gene Duplication**		**Genome Remodelling**	**Gene Loss Gene Gain**
VARIATION				
		SNPs	**K_S elevation**	**Allelic nulls**
DISEASE				
Paralogues		**Conserved sites**	**K_S elevation**	

Amniote Phylogeny



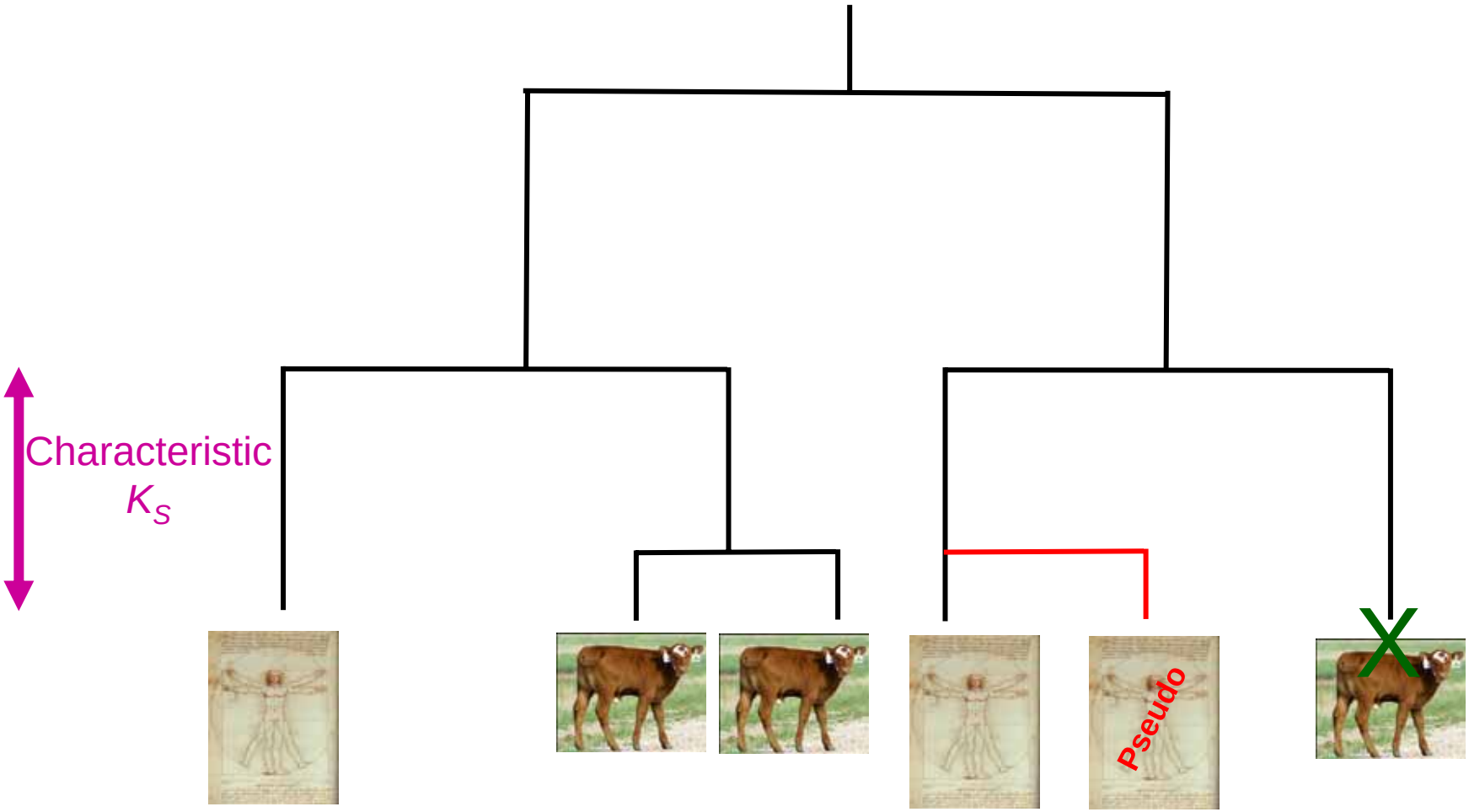
K_S: fraction of synonymous changes at synonymous sites

- Redundant genetic code, e.g.

$$\left.\begin{array}{l}\mathbf{GCA}\\ \mathbf{GCC}\\ \mathbf{GCG}\\ \mathbf{GCT}\end{array}\right\} \rightarrow \textbf{Alanine}$$

- Third base of a codon "wobbles" without changing the translated amino acid
- K_S measures <u>neutral</u> mutation rate in coding regions without selection

Gene Phylogeny

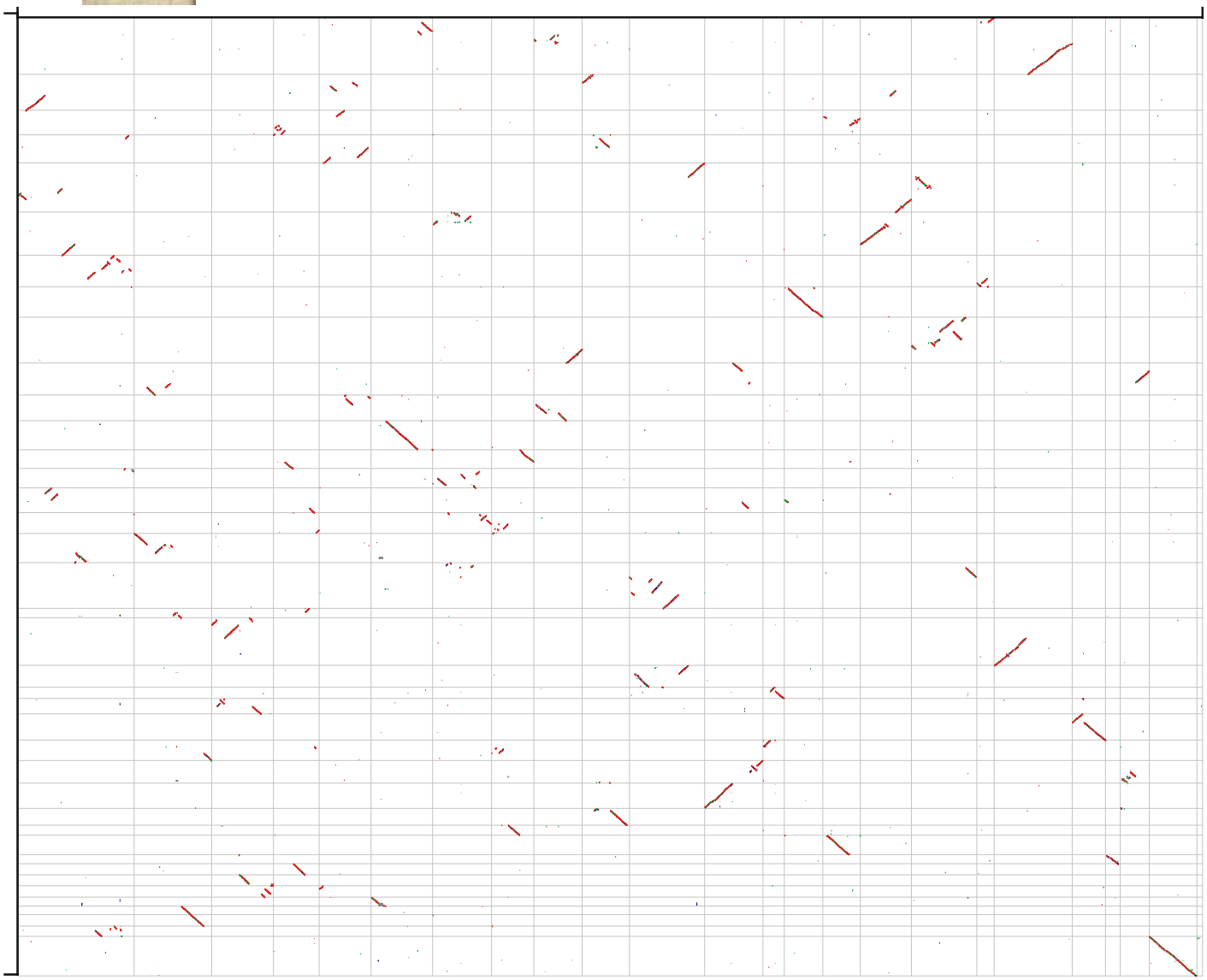


K_S Trees: predict…

- Lineage-specific paralogues
- Orthologues
- Retrotransposed pseudogenes
- Genes absent from assembly
- Conserved Synteny

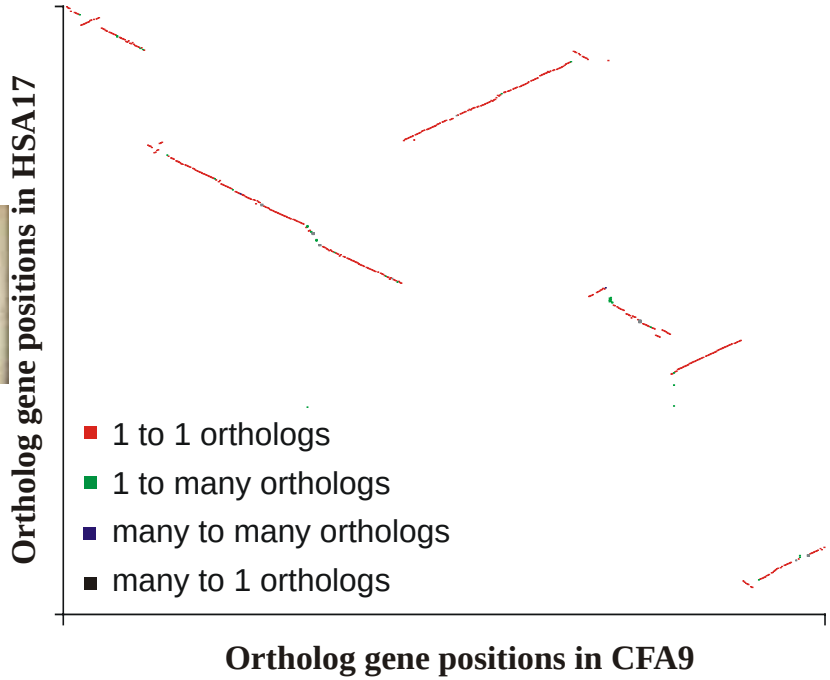
Leo Goodstadt

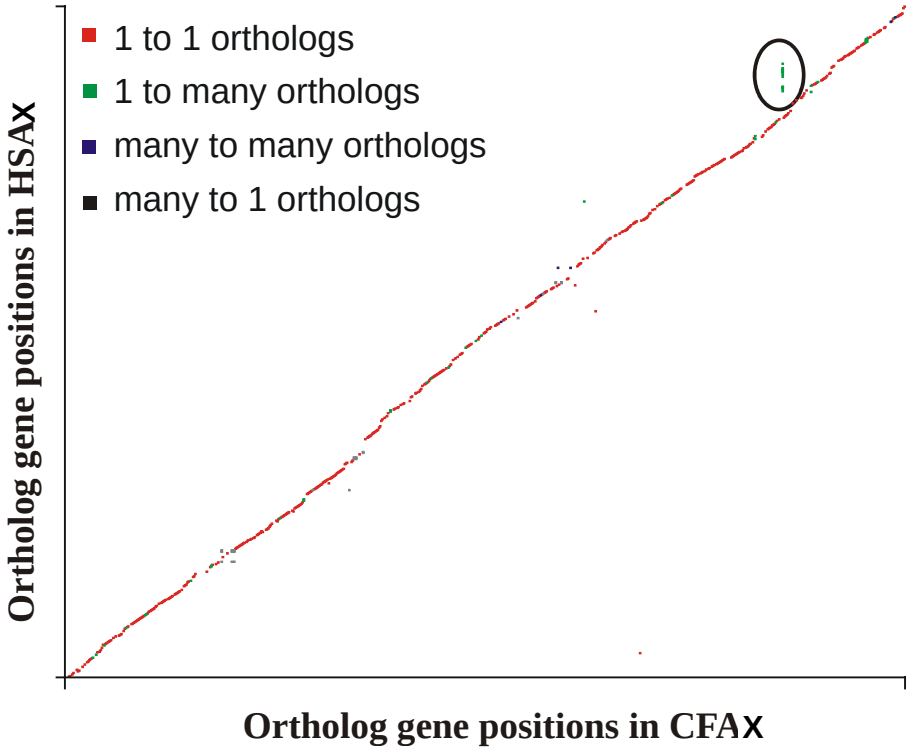
Ortholog gene positions in the human genome



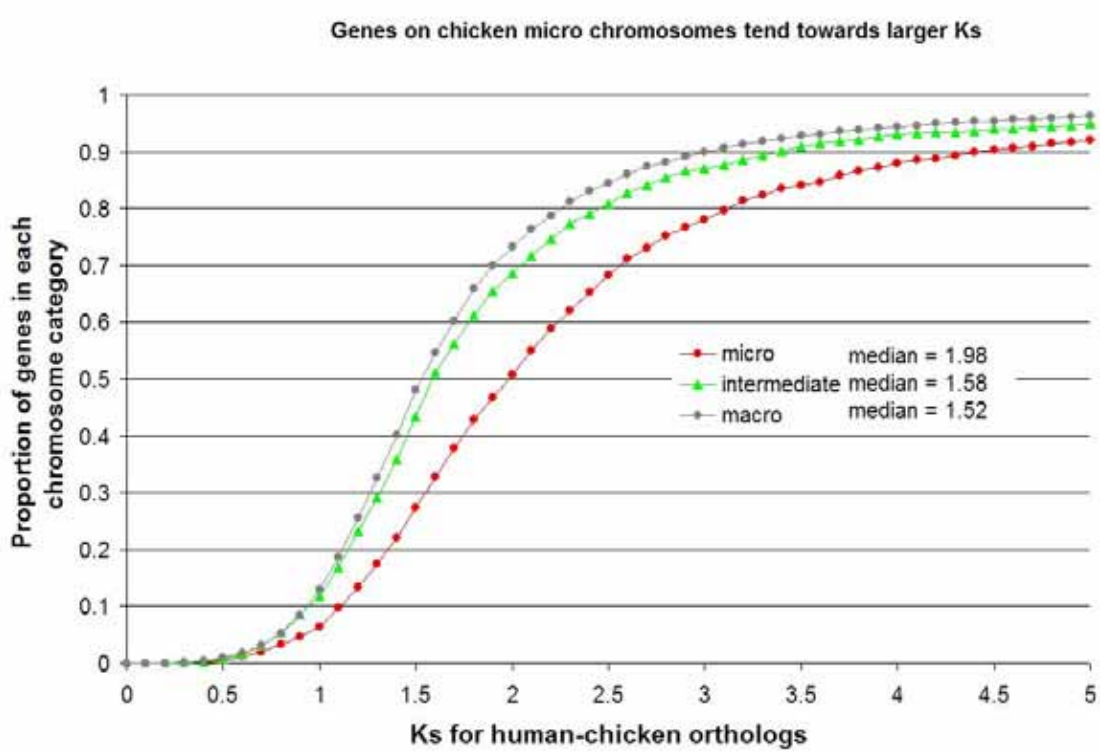
13,326: 73.2%

Leo Goodstadt

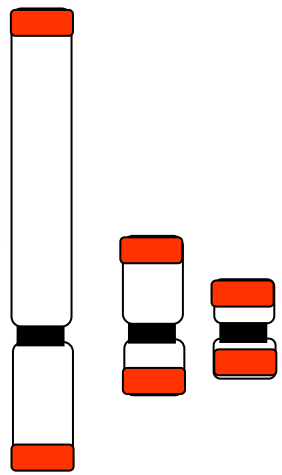




Variation in K_S



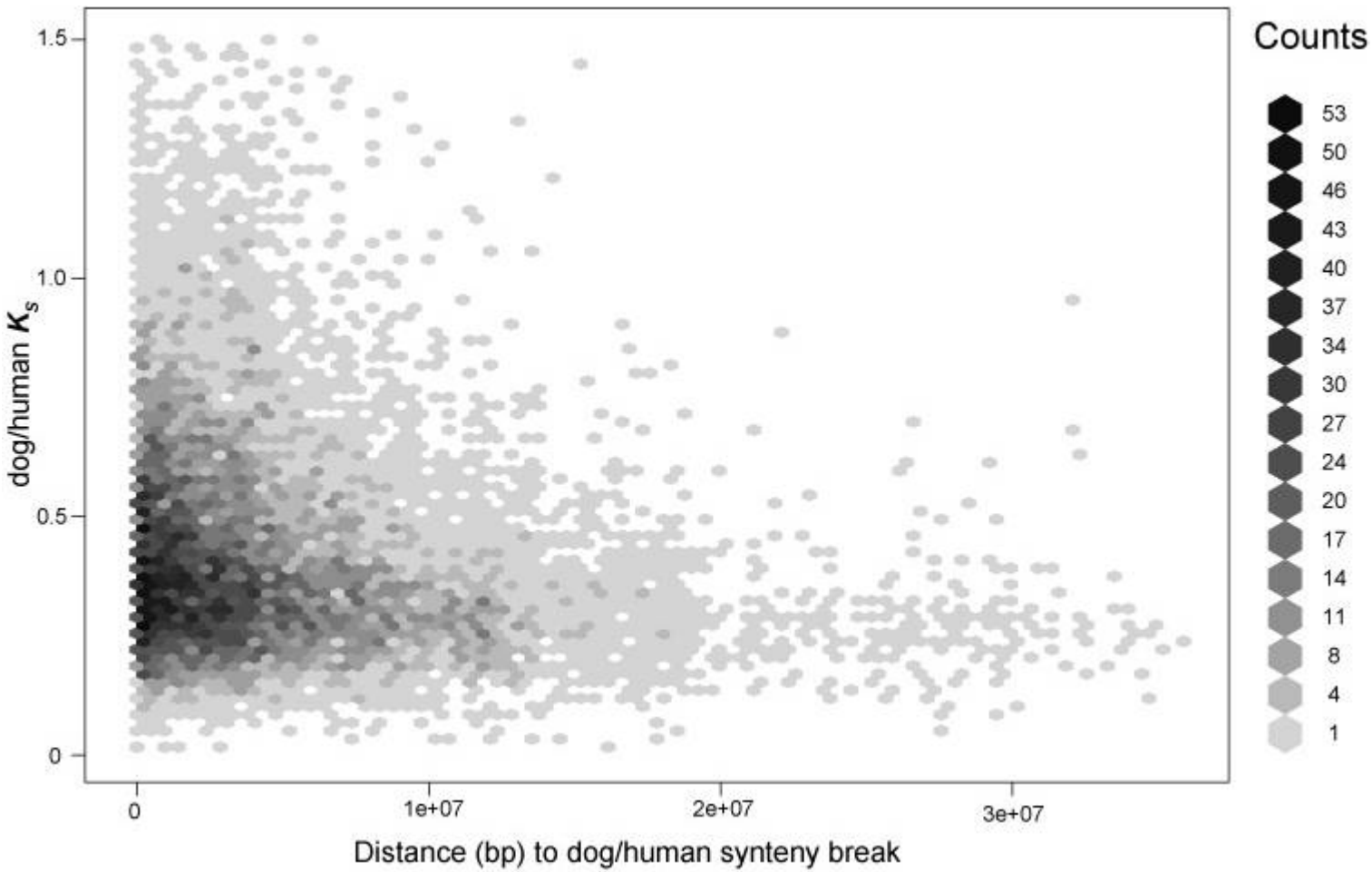
10Mb subtelomeric genes indistinguishable from microchromosomal genes, with respect to K_s



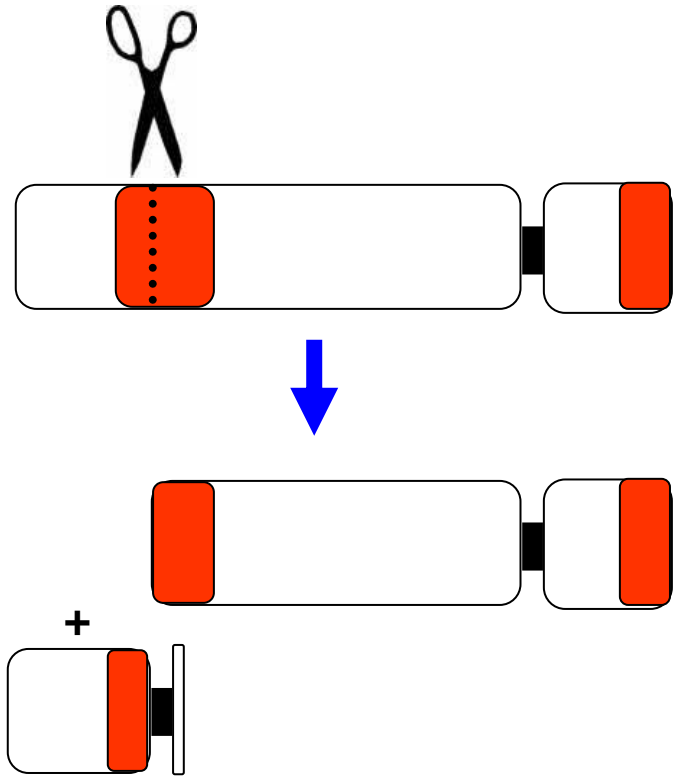
K_S variation

Species	**Subtelomeric / Subtelomeric**	**Subtelomeric / Interstitial**	**Interstitial / Subtelomeric**	**Interstitial / Interstitial**
Chicken / Human	3.15	1.95	2.40	1.53
Dog / Human	0.69	0.35	0.67	0.31
Dog / Chicken	2.15	2.08	1.98	1.64

K_S peaks towards synteny breaks

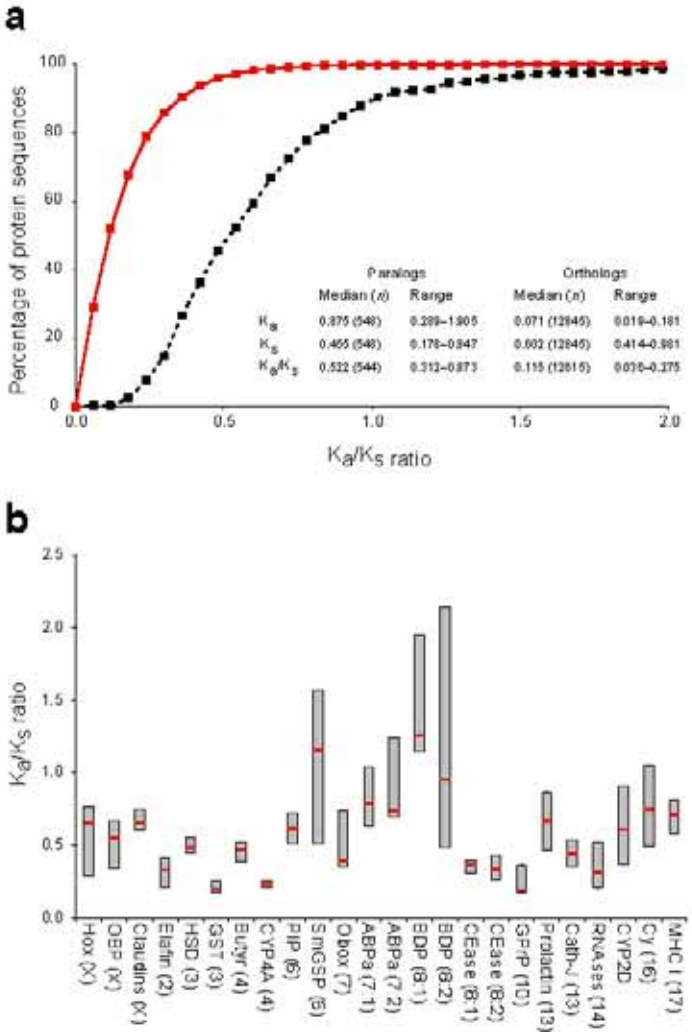


Fragile breakage model

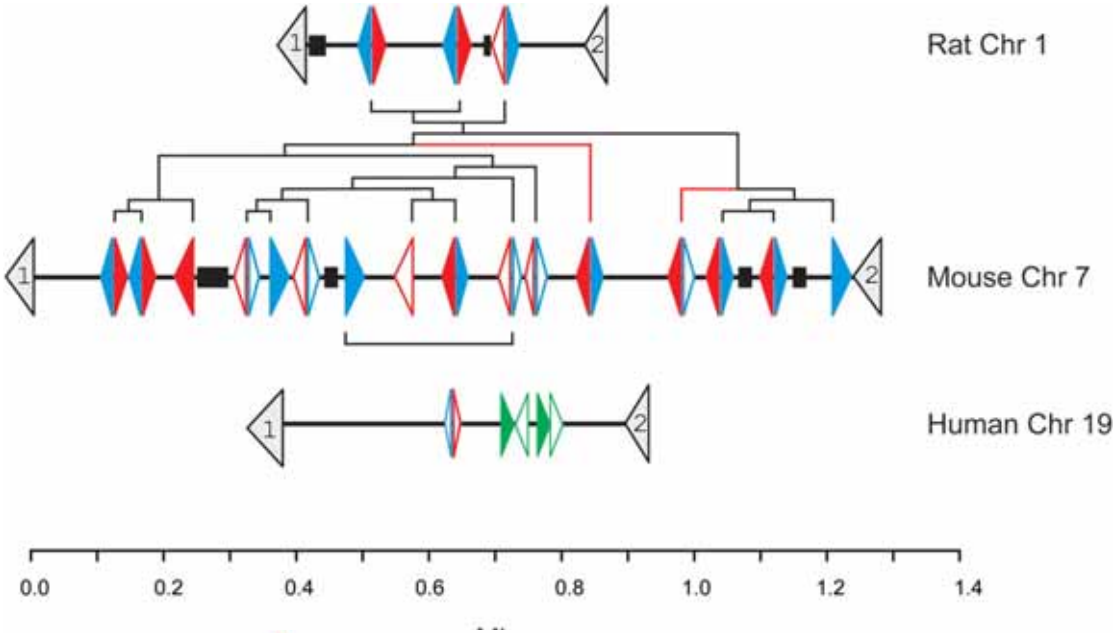


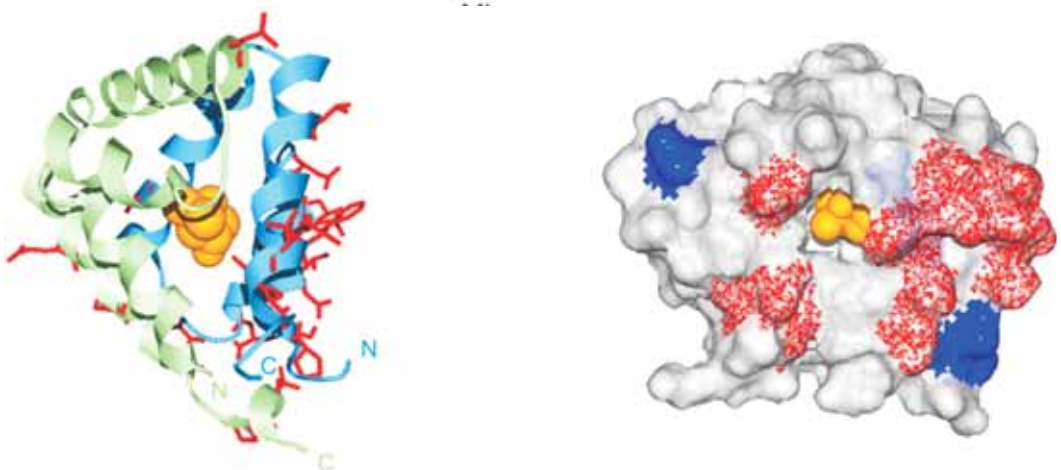
It still remains unclear *WHY* the human and cattle karyotypes are relatively stable, whereas those of dogs, mice and rats are not.

Rapid Evolution in Gene Clusters



Gene Duplication leading to Genome remodelling





Cattle: 7 α genes, and 11 β genes (3 of which are pseudogenes)

Directed Finishing?

Androgen-binding proteins: Emes et al. (2004) *Genome Res.* 14(8):1516-29

To Do List?

Chris Ponting
Leo Goodstadt
Caleb Webber
Andreas Heger

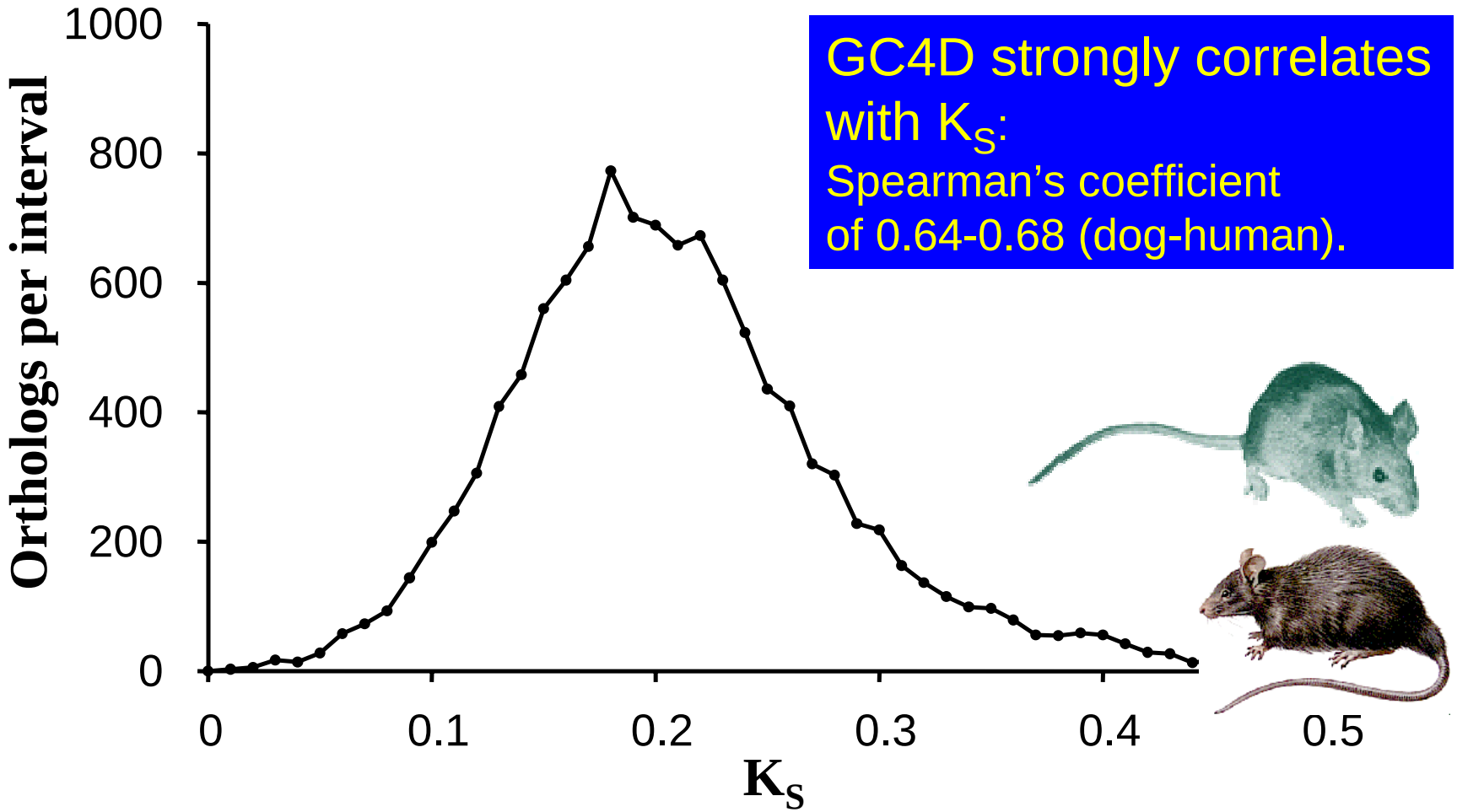
- Genes unique to cattle
- Complex orthology & paralogy relationships.
- QC on genes and assembly (pseudogenes and absent genes)
- Gene Reprediction
- Evolution of the bovine karyotype
- Low & high evolutionary rates: K_A/K_S & K_S.
- Intronic rates
- Adaptive evolution within genes

Issues

- Gene Build: 6X coverage or more?
- Timing: ASAP
- Lineage-specific gene & lineage-specific biology: interactions with community.

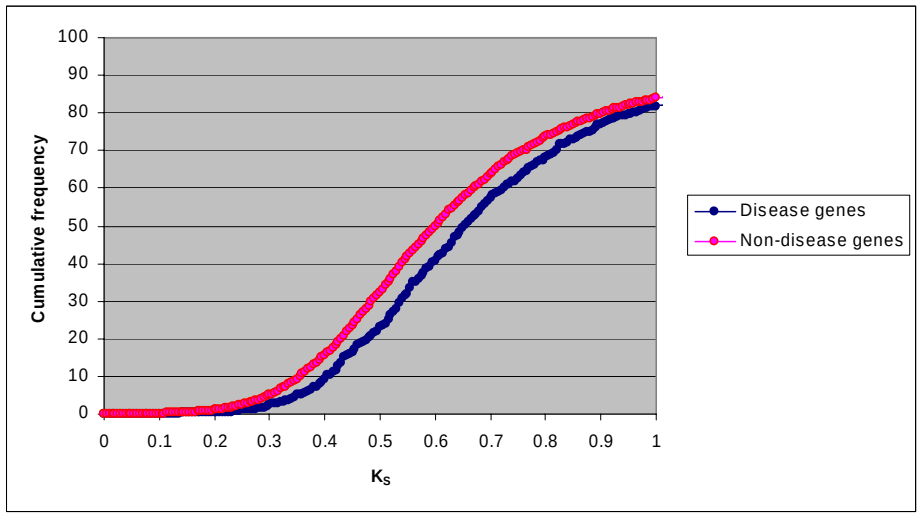
MRC Functional Genetics Unit

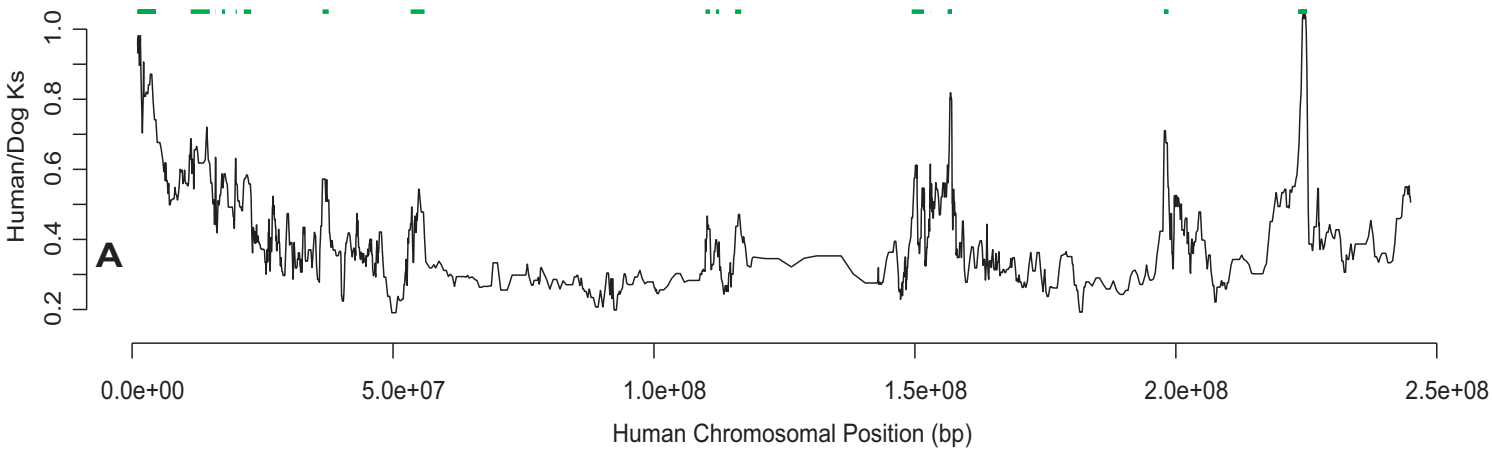
Variations in K_S values



GC4D strongly correlates with K_S:
Spearman's coefficient of 0.64-0.68 (dog-human).

K_S variation





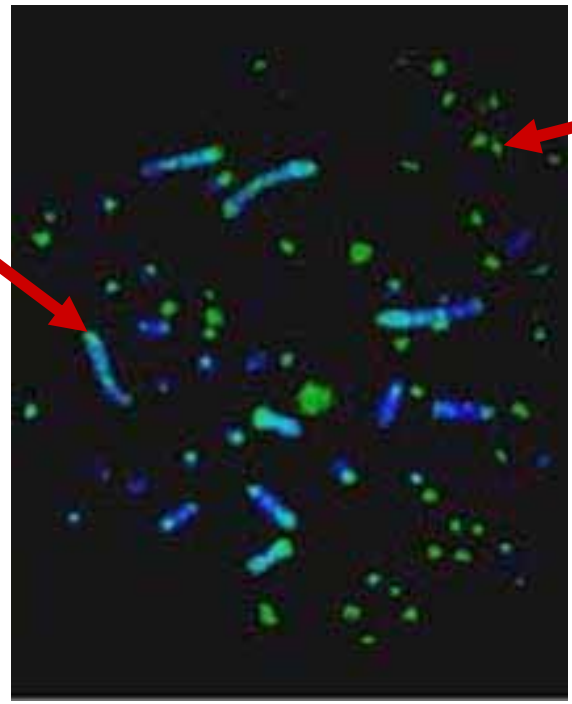
Chicken:

Large and Small Chromosomes

Large (macro-) chromosomes:

more DNA but lower gene density;

lower mutation rate; lower G+C



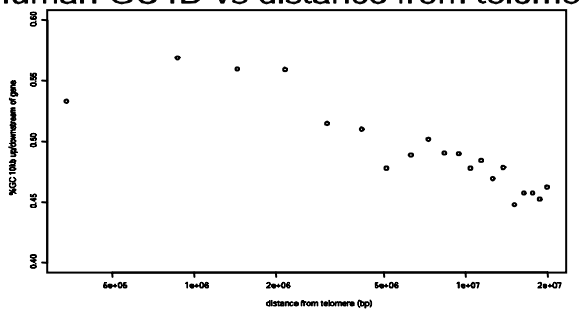
Small (micro-) chromosomes:

25% of the DNA but half the genes;

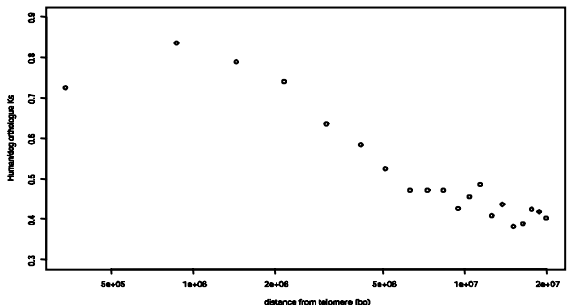
higher mutation rate; higher G+C

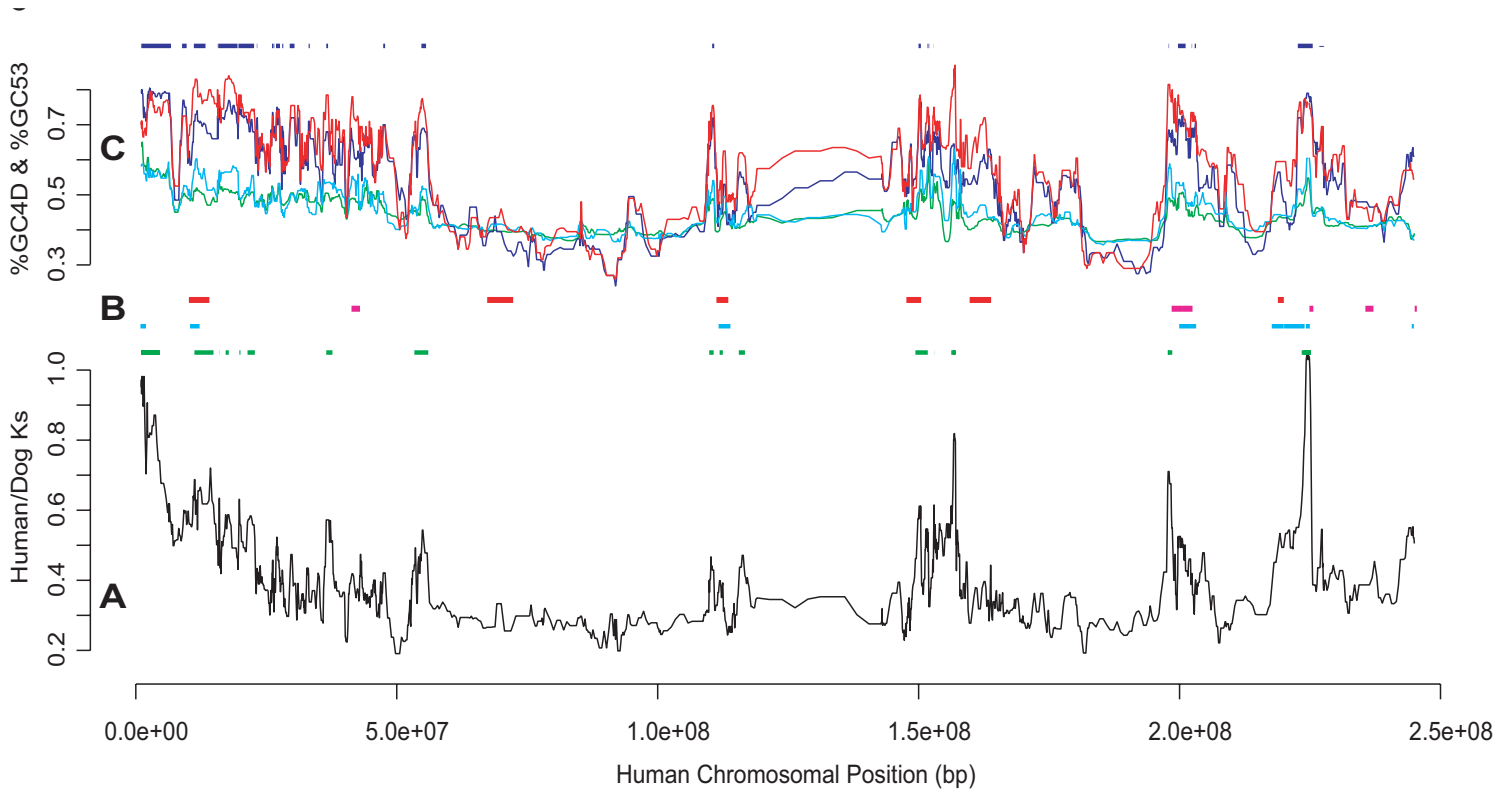
but dog telomeres are recent (derived)…?

Human GC4D vs distance from telomere



Hs-Cf Ks vs distance from human telomere




%GC4D & %GC53
0.3
0.5
0.7
C
B
Human/Dog Ks
0.2
0.4
0.6
0.8
1.0
A
0.0e+00
5.0e+07
1.0e+08
1.5e+08
2.0e+08
2.5e+08
Human Chromosomal Position (bp)

Breakpoints are *not* random: they occur preferentially in high G+C / high K_S regions

- K_S hot spots that occur in ancestral interstitial regions have preferentially relocated to dog subtelomeres (*p*-value $< 10^{-4}$).

- K_S is significantly elevated towards dog-human synteny breakpoints (*p*-value $< 2.2 \times 10^{-16}$)

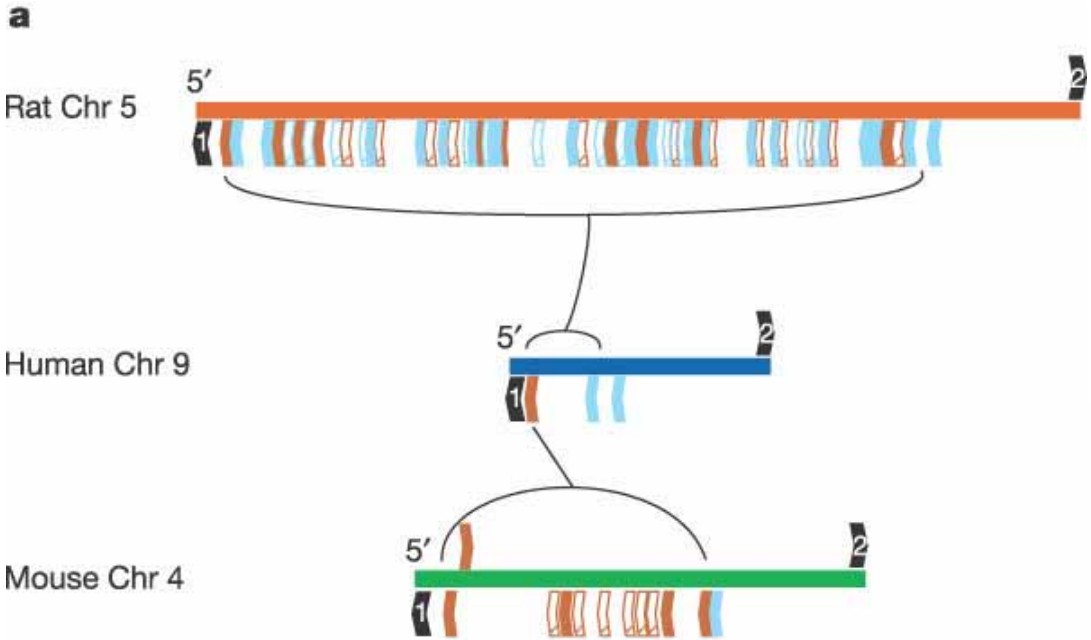
Pathogenic Breakage

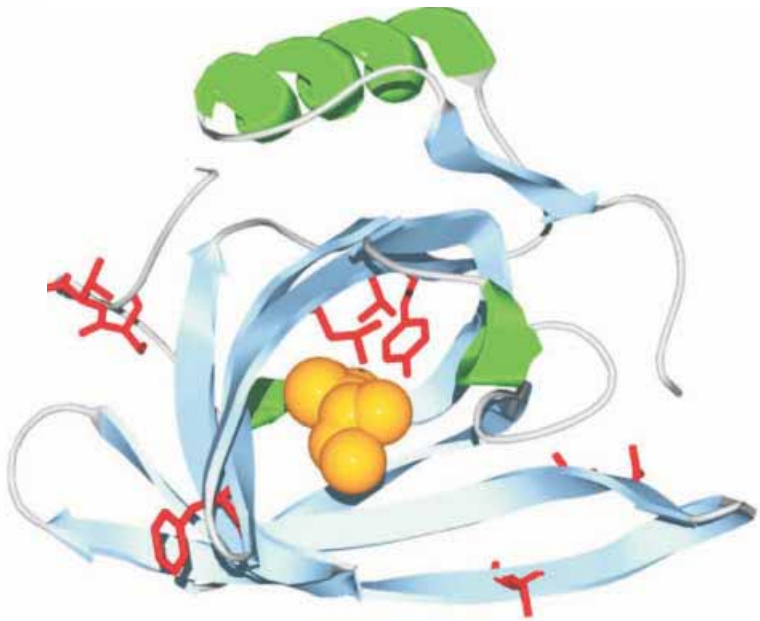
Occurs most frequently in:

- subtelomeric regions
- high G+C regions.

(Yu et al., 1978; Stoll, 1980; Aula and von Koskull, 1976; Nakagome and Chiyo, 1976; Abeysinghe et al., 2003)

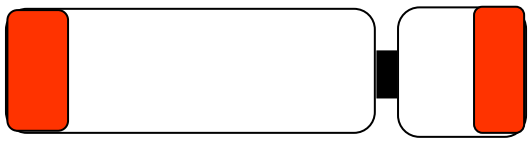
Gene Duplication leading to Genome remodelling



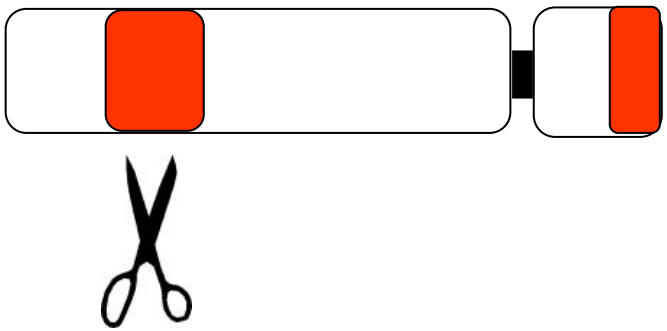


Mouse major urinary proteins & rat α2u-globulins

Two models



FAST **increase in G+C and in neutral rates in the last 60MY, since the dog karyotype shattered (DERIVED)**



stable G+C and neutral rates in the last 60MY, since the dog karyotype shattered (ANCESTRAL)

Correlation coefficients (Spearman's *rho*) between 1:1 orthologues' GC4D fractions.

	Human	Dog	Mouse	Chicken
Human	-			
Dog	0.945	-		
Mouse	0.836	0.818	-	
Chicken	0.539	0.595	0.539	-
Rat	0.825	0.810	0.937	0.524

Discussion

- Status and plans for gene prediction for bovine at Ensembl Sanger/EBI
- NCBI plans for the bovine sequence
- Using comparative genomics on protein families

- what the international bovine community wants and needs
- potential scientific partners