

Figure 1 | **The three layers of genome annotation: where, what and how?**

How *C. elegans* was annotated

- ◆ Genome size: 100 Mb
- ◆ Cosmid-by-cosmid finished sequencing strategy
- ◆ Central annotation database - ACeDB/WormBase
- ◆ Experimental distributed annotation system (DAS)

Annotation Type

Performed By

Nucleotide-level

Sequencing centers

Protein-level

Proteome, Inc., SwissProt

Process-level

Community,
WormBase curators

DasView on WormBase

DasView for C.elegans
 DasView showing 0 bp from CHROMOSOME_I/1,14972282

Features from: Wormbase - elegans

- intron
- oligo
- Homol_data
- RNAi
- CDS
- tandem
- assembly_tag
- Genomic_canonical
- GenePair_STS
- transcript
- Queryprosite
- cDNA_for_RNAi
- inverted
- Pseudogene
- exon
- Clone_right_end
- Clone
- Link
- Clone_left_end

Click Button to Save Settings

Wormbase - elegans

10000 20000

ZK673

SA:yk257c9
 yk257c9

SA:yk386h5
 yk386h5

SA:yk366f1
 yk366f10

ZK673.2 ZK673.3 ZK673.4 ZK673.5 ZK673.6 ZK673.7 ZK673.8 ZK673.10 ZK673

Wormbase - worm2

ZK673

F37H8.2

ZK673

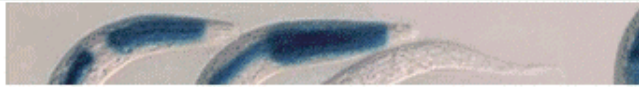
DAS Client

Wormbase - annotations	<input type="radio"/> all <input checked="" type="radio"/> none <input type="radio"/> custom
Wormbase - elegans	<input type="radio"/> all <input type="radio"/> none <input checked="" type="radio"/> custom
Wolfe - orfdb	<input type="radio"/> all <input checked="" type="radio"/> none <input type="radio"/> custom
Wolfe - wbriggdb	<input type="radio"/> all <input checked="" type="radio"/> none <input type="radio"/> custom

WormBase

[Basic Searches](#)[Expr. Pattern Searches](#)[Gene Hunter](#)[Blast Searches](#)[Advanced Searches](#)[Atlas](#)

WormBase



the priarynx to the posterior of the worms. This may correspond to the body wall muscle cells, apart from those body wall muscle cells in the head. The second component begins in late larval stages as the gonad develops. b-galactosidase staining is first diffuse, in four patches a-p along the gonad. It then becomes more discrete as the subcellular foci form although apparently still in four bands, which we think correspond to the uterine muscle cells. As the adult hermaphrodites age, these foci seem to become progressively more specifically associated with lateral positions and we hypothesise that these foci are related to the attachments of the uterine muscle cells to the hypodermal seam cells. Expression in males has not yet been examined.

- **Remark:** *Other strains injected with this plasmid are UL268 and UL269.*

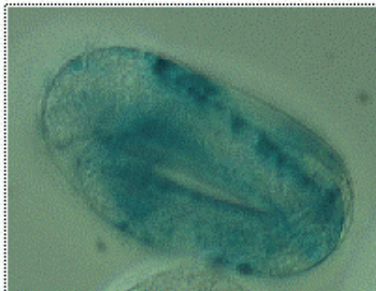
[More Details](#)

Diagrams and Images



f41c3_2_du.jpeg:

(Click image to enlarge)



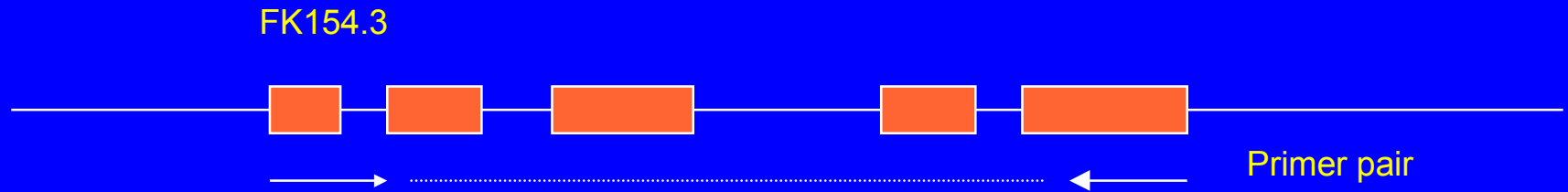
f41c3_2_emb.jpeg:

(Click image to enlarge)

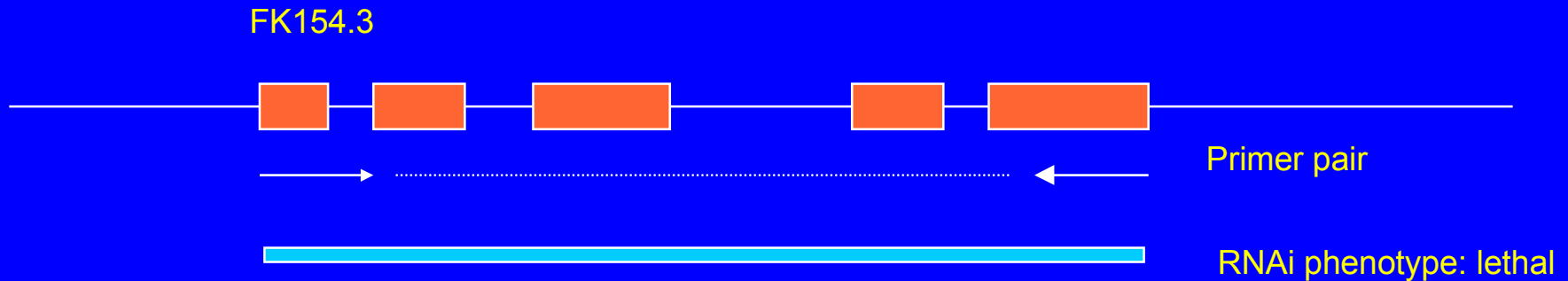
Problems with Worm Annotation

- ◆ Genome still changing even after being “finished”
- ◆ Ill-defined sequence releases
- ◆ No standard community annotation format
- ◆ Unstable identifiers
- ◆ Poor identifier tracking

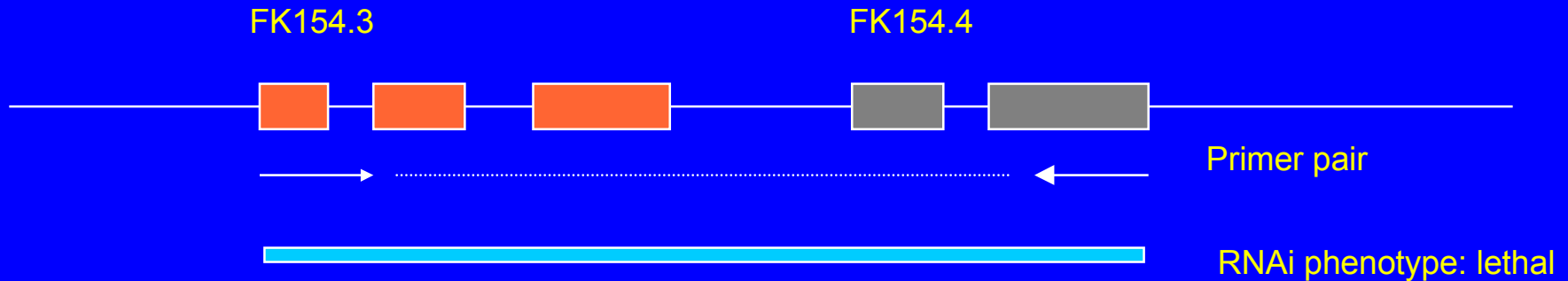
Functional Annotation on an Unstable Genome



Functional Annotation on an Unstable Genome



Functional Annotation on an Unstable Genome



How Human was Annotated

- ◆ Genome size: 3,000 Mb
- ◆ BAC-by-BAC “draft” strategy
- ◆ “Golden path” developed at UCSC
- ◆ Many groups participated in initial annotation

Annotation Type

Performed By

Nucleotide-level

Ensembl, SNP consortium
Case Western Reserve, UCSC...

Protein-level

Ensembl, SwissProt, NCBI...

Process-level

Community

Making Human Annotation Work

- ◆ Well-defined golden path builds (~quarterly)
- ◆ A common coordinate system
- ◆ Common annotation exchange format (GAF)
- ◆ A standard baseline annotation (Ensembl)
- ◆ A common database for contributing & displaying community annotation information (UCSC genome browser)

Ensembl Sequence Viewer

The screenshot shows the Ensembl Genome Server interface within a Netscape browser window. The browser title is "Netscape: Ensembl Genome Server". The address bar shows the URL: `http://www.ensembl.org/perl/contigview?chr=19&vc_star`. The main content area is titled "Detailed View" and displays a genomic track for chromosome 19, spanning from 20.54 Mb to 20.63 Mb. The track includes various annotations such as UniGene, Genscan, DNA(contigs), Transcript, mRNA, SpTrEMBL, CpG island, SNP, Mouse trace, Repeats, Tile Path, and %GC. The track is zoomed in to a 100.00 Kb scale. Navigation controls include "Zoom" buttons for 2 Mb, 1 Mb, and Window, and a "Refresh" button. The bottom of the interface shows the current chromosome (19) and nucleotide coordinates (20542291 to 20642291).

Netscape: Ensembl Genome Server

File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Shop Stop

Bookmarks Location: `http://www.ensembl.org/perl/contigview?chr=19&vc_star` What's Related

Detailed View

◀ 2 Mb ◀ 1 Mb Window Zoom Window 1 Mb 2 Mb ▶▶

Length 100.00 Kb

20.54 Mb 20.55 Mb 20.56 Mb 20.57 Mb 20.58 Mb 20.59 Mb 20.60 Mb 20.61 Mb 20.62 Mb 20.63 Mb

UniGene

Genscan

DNA(contigs) AC005759 AC008397 AC008397 AC008397 AC008397 AC008397

Transcript

Genscan

UniGene

mRNA

SpTrEMBL

CpG island

SNP

Mouse trace

Repeats

20.54 Mb 20.55 Mb 20.56 Mb 20.57 Mb 20.58 Mb 20.59 Mb 20.60 Mb 20.61 Mb 20.62 Mb 20.63 Mb

Tile Path LLNLR-229D3 CTC-251H24

%GC

View as text Dump Traces jump to UCSC Customise this view

Chr: 19 Nucleotides: 20542291 to 20642291 Turn menus On Off Refresh

Problems with Human Annotation

- ◆ Golden path misassemblies are common
- ◆ Contamination with non-human sequence
- ◆ Golden path version skew
- ◆ Some groups annotating to other coordinate systems
 - GenBank “NT contigs”
 - NCBI “yellow brick road”

Worm & Human: Lessons

- ◆ Annotation is a community process
- ◆ Annotation is a continuing process
- ◆ Essential resources:
 - Common reference sequence (assembled)
 - Well-defined release numbers
 - Baseline nucleotide-level annotation
 - Stable identifiers and identifier tracking
 - Support for 3d party annotation