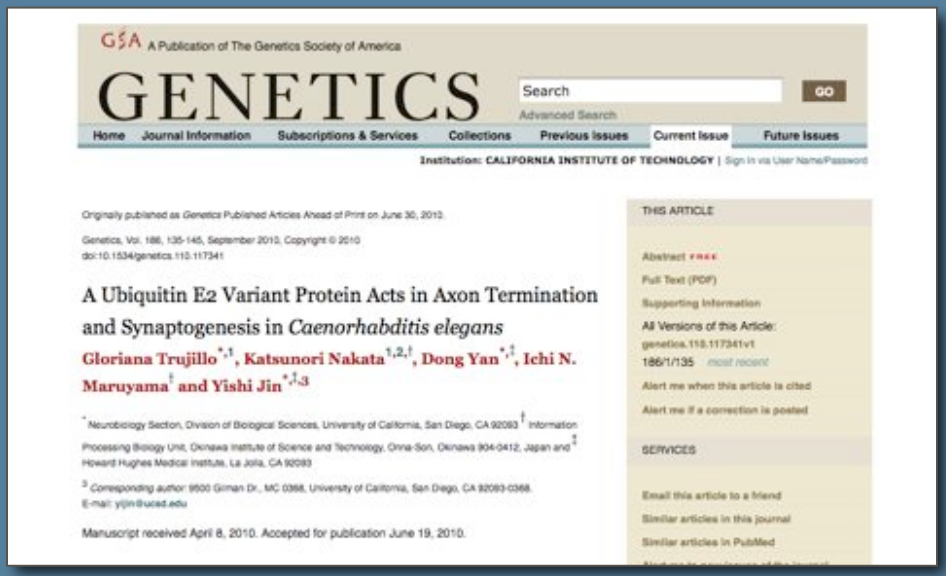
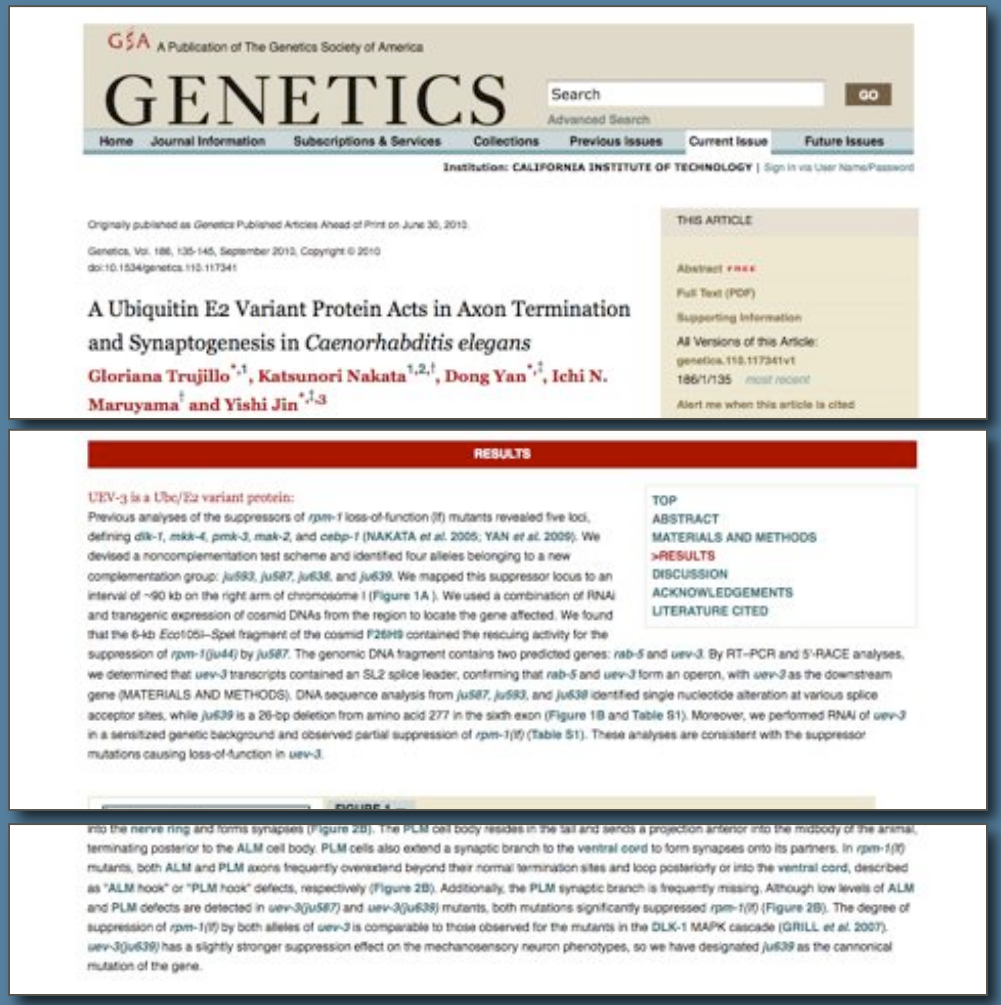
Publishing Interactive Articles: Integrating Journals and Biological Databases



Karen Yook

Genetics Society of America, Dartmouth Journal Services, Textpresso and WormBase

Goal: hyperlink entities in a published paper to their database resource page



GSA A Publication of The Genetics Society of America

GENETICS

Search GO

Advanced Search

Home Journal Information Subscriptions & Services Collections Previous Issues Current Issue Future Issues

Institution: CALIFORNIA INSTITUTE OF TECHNOLOGY | Sign In via User Name/Password

Originally published as *Genetics* Published Articles Ahead of Print on June 30, 2010.

Genetics, Vol. 186, 135–145, September 2010, Copyright © 2010

doi:10.1534/genetics.110.117341

A Ubiquitin E2 Variant Protein Acts in Axon Termination and Synaptogenesis in *Caenorhabditis elegans*

Gloriana Trujillo*,†,1, Katsunori Nakata†,2,†, Dong Yan*,†,‡, Ichi N. Maruyama‡ and Yishi Jin*,†,3

THIS ARTICLE

Abstract FREE

Full Text (PDF)

Supporting Information

All Versions of this Article:

genetics.110.117341v1

186/1/135 *most recent*

Alert me when this article is cited

RESULTS

TOP

ABSTRACT

MATERIALS AND METHODS

>RESULTS

DISCUSSION

ACKNOWLEDGEMENTS

LITERATURE CITED

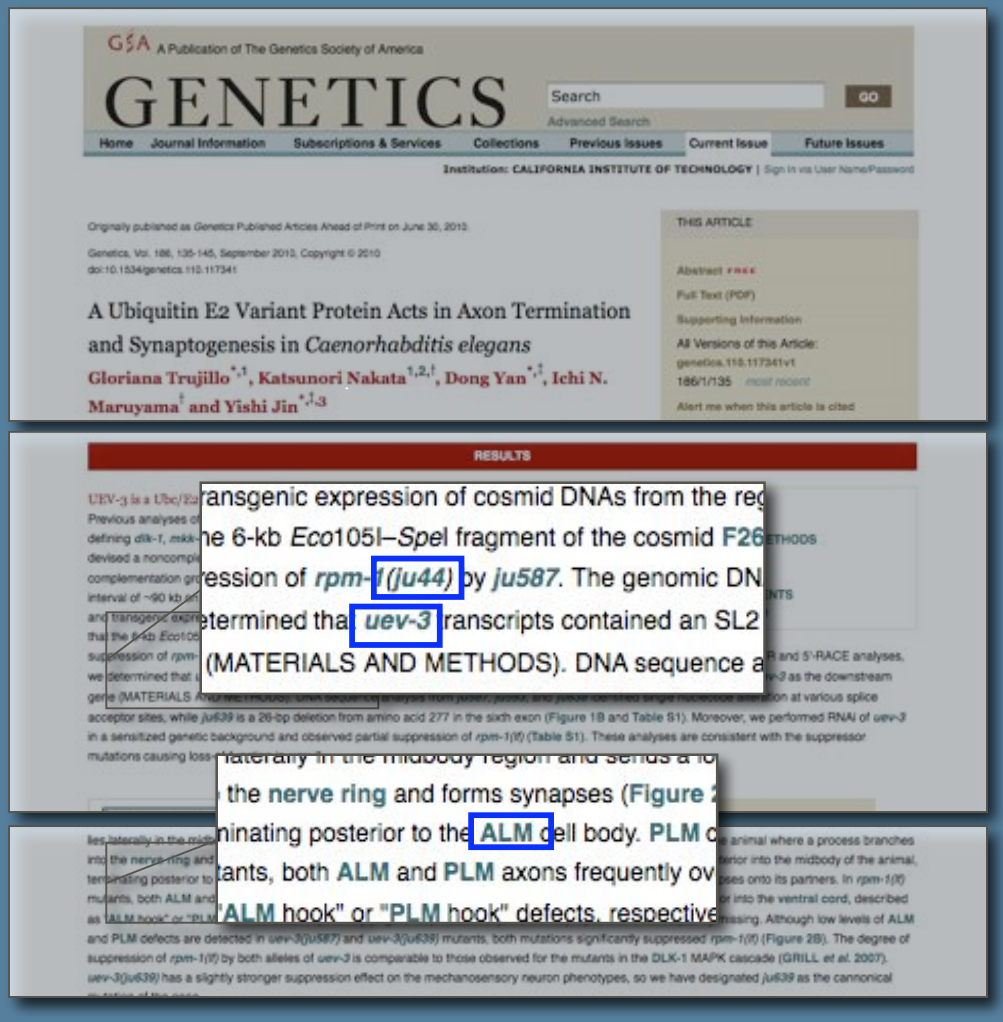
UEV-3 is a Ubc/E2 variant protein:

Previous analyses of the suppressors of *rpm-1* loss-of-function (lf) mutants revealed five loci, defining *dlk-1*, *mkk-4*, *pmk-3*, *mak-2*, and *cebp-1* (NAKATA *et al.* 2005; YAN *et al.* 2009). We devised a noncomplementation test scheme and identified four alleles belonging to a new complementation group: *ju593*, *ju587*, *ju638*, and *ju639*. We mapped this suppressor locus to an interval of ~90 kb on the right arm of chromosome I (Figure 1A). We used a combination of RNAi and transgenic expression of cosmid DNAs from the region to locate the gene affected. We found that the 6-kb *Eco*105I–*Spe*I fragment of the cosmid F26H9 contained the rescuing activity for the suppression of *rpm-1(ju44)* by *ju587*. The genomic DNA fragment contains two predicted genes: *rab-5* and *uev-3*. By RT–PCR and 5′-RACE analyses, we determined that *uev-3* transcripts contained an SL2 splice leader, confirming that *rab-5* and *uev-3* form an operon, with *uev-3* as the downstream gene (MATERIALS AND METHODS). DNA sequence analysis from *ju587*, *ju593*, and *ju638* identified single nucleotide alteration at various splice acceptor sites, while *ju639* is a 26-bp deletion from amino acid 277 in the sixth exon (Figure 1B and Table S1). Moreover, we performed RNAi of *uev-3* in a sensitized genetic background and observed partial suppression of *rpm-1(lf)* (Table S1). These analyses are consistent with the suppressor mutations causing loss-of-function in *uev-3*.

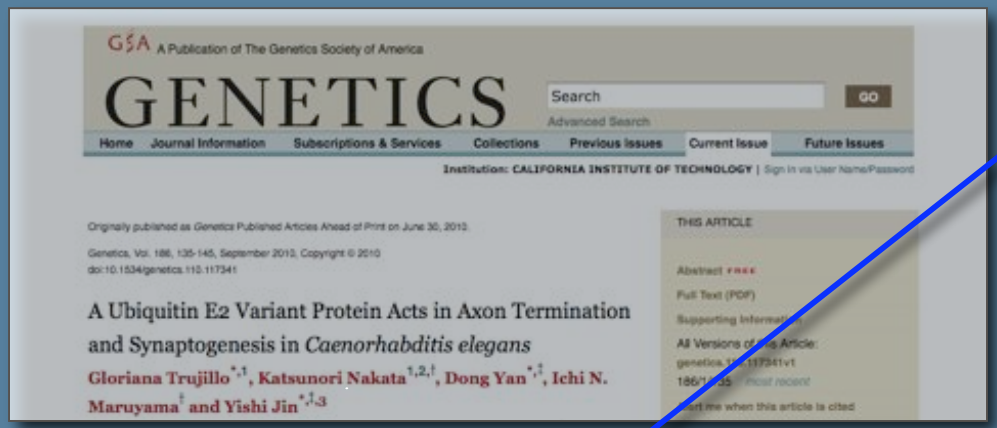
into the nerve ring and forms synapses (Figure 2B). The PLM cell body resides in the tail and sends a projection anterior into the midbody of the animal, terminating posterior to the ALM cell body. PLM cells also extend a synaptic branch to the ventral cord to form synapses onto its partners. In *rpm-1(lf)* mutants, both ALM and PLM axons frequently overextend beyond their normal termination sites and loop posteriorly or into the ventral cord, described as "ALM hook" or "PLM hook" defects, respectively (Figure 2B). Additionally, the PLM synaptic branch is frequently missing. Although low levels of ALM and PLM defects are detected in *uev-3(ju587)* and *uev-3(ju638)* mutants, both mutations significantly suppressed *rpm-1(lf)* (Figure 2B). The degree of suppression of *rpm-1(lf)* by both alleles of *uev-3* is comparable to those observed for the mutants in the DLK-1 MAPK cascade (GRILL *et al.* 2007). *uev-3(ju639)* has a slightly stronger suppression effect on the mechanosensory neuron phenotypes, so we have designated *ju639* as the cannonical mutation of the gene.

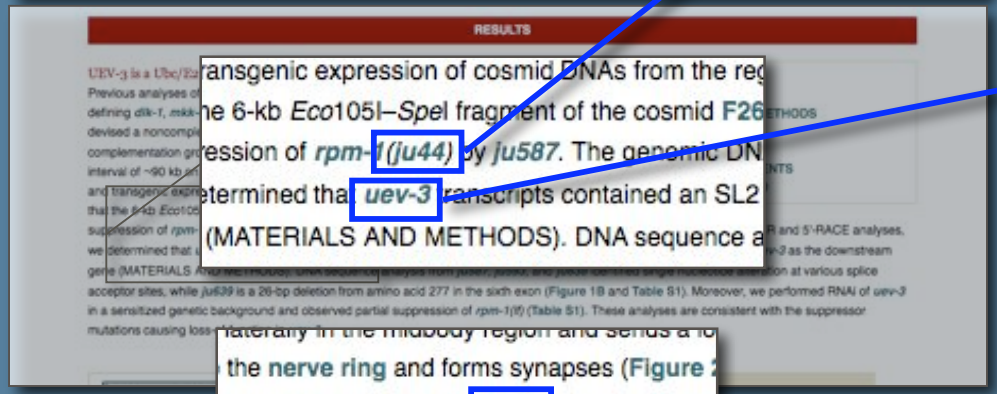
Goal: hyperlink entities in a published paper to their database resource page

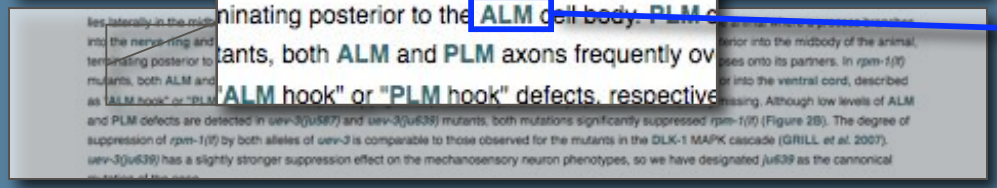
Goal: hyperlink entities in a published paper to their database resource page

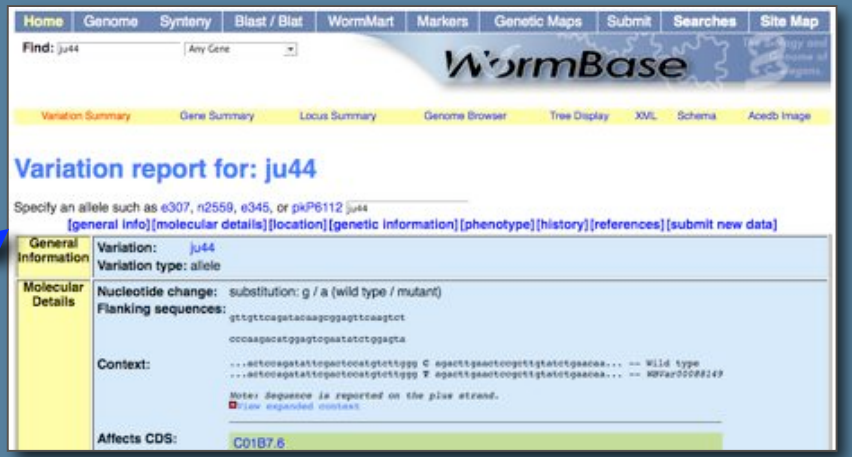


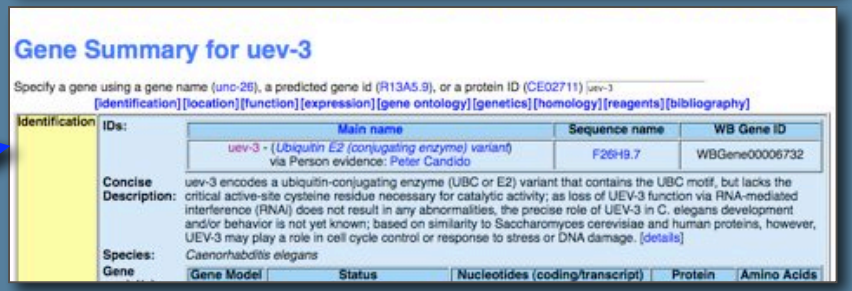
Goal: hyperlink entities in a published paper to their database resource page

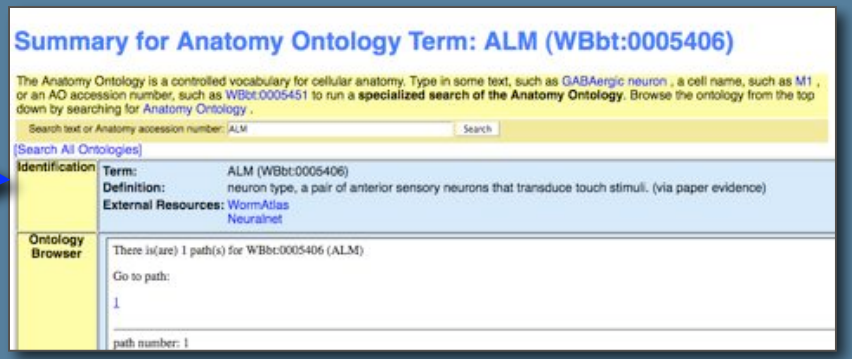




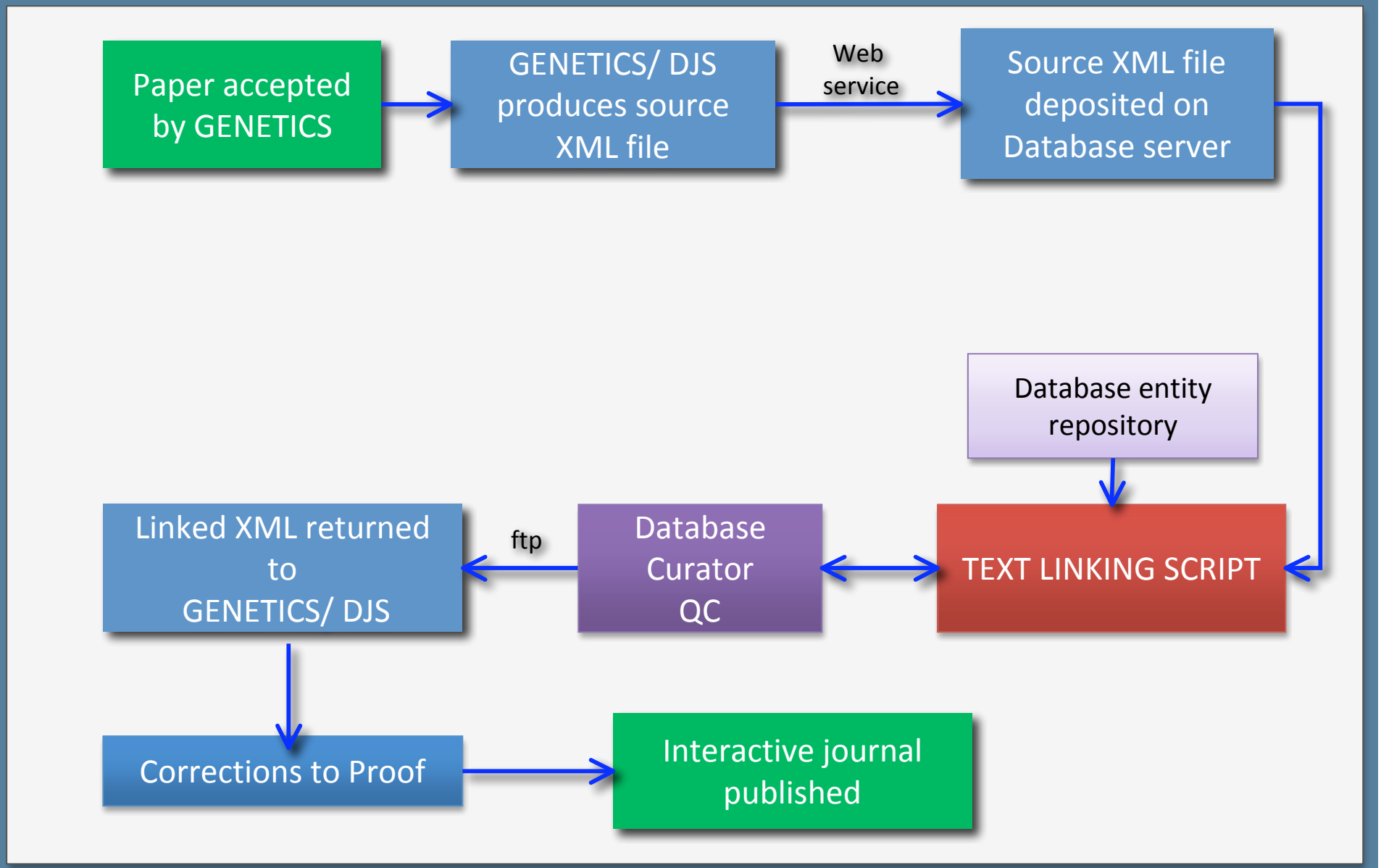








GSA-Database markup pipeline



What is needed for this to work

- Textpresso -> Arun
- Source file from the journal (XML format)
- List of entities that exist in the database
- Stable database URL constructor
- Database curator for manual quality control (QC)
- Method for ensuring URL stability

What is needed:

- List of entities that exist in the database, these are mined directly from the database (WormBase)

Entity Class	Examples	# of WormBase entities
Clone	A1D10, C02H12, GAP1_10	204,605
Gene	mec-15, F25B3.5, y110a7a.10	537,027
Variation (alleles)	ad820sd, cb19860	274,649
Transgene	adEx1290, cmIs6, jeIn2	6,546
Rearrangement	hDp23, stDf47, sT1	709
Protein (to gene pages)	MEC-15, ABF-1	
Strain	BC954, IE35413	28,826
Anatomy	AWC, body wall muscle	6,563
Phenotype	Muv, Lon, Dpy	122
	Total	1,059,047

What is needed:

- Stable database URL constructor:

For WormBase-

http://www.wormbase.org/db/get?name=NAME;class=CLASS

NAME is the entity string (lin-11, e189, gonad)

CLASS is the class the entity belongs to (gene, variation, anatomy)

For Saccharomyces Genome Database (SGD)-

http://www.yeastgenome.org/cgi-bin/locus.fpl?dbid=SGDID

SGDID is the SGD identification for the entity

What is needed:

- Database curator for manual quality control:

  Checks all links made by the linking script

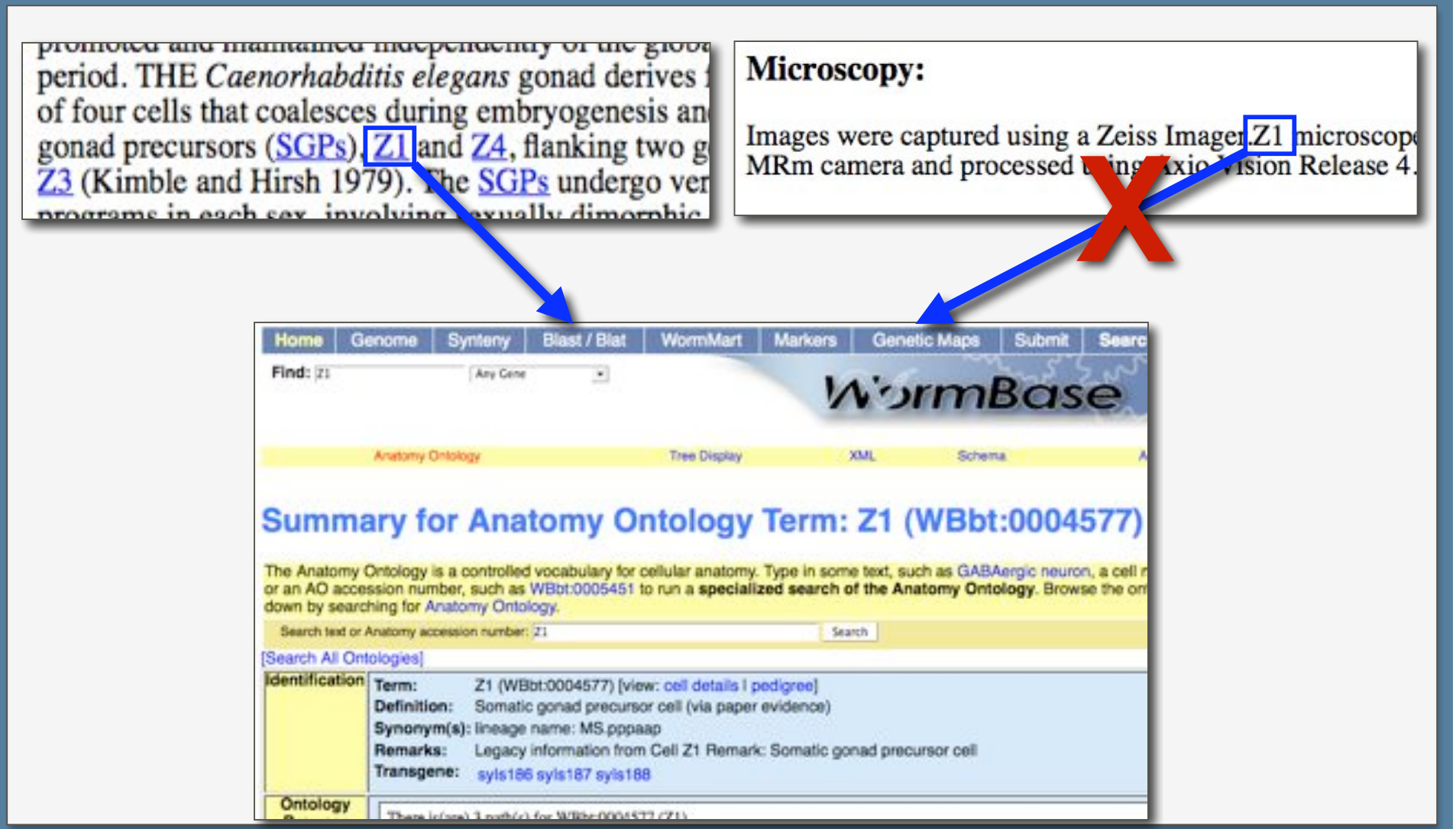
  Checks for ambiguities

  - adds them to an exclusion list
  - provides the right entity NAME/ID

  Checks for entities that did not get linked

  - author writing shortcuts
  - science jargon
  - new entities not in the database yet

Manual QC resolves ambiguities that an automated script cannot resolve



Manual QC is needed to decipher authors' writing shortcuts

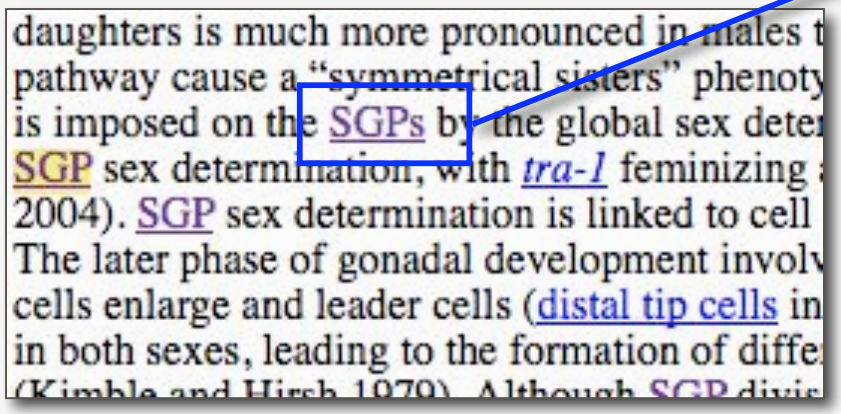
Analysis of *zig* gene knockout alleles: Of the eight *zig* genes only one, *zig-4*, was previously analyzed through gene deletion analysis (Aurello *et al.* 2002). Through generating deletion alleles by either transposon mobilization (*zig-1*) or deletion library screening, conducted by the *C. elegans* knockout consortia (*zig-2*, *-3*, *-5*, *-6*, *-7*, and *-8*), we have obtained knockout alleles for each of the so far uncharacterized, seven remaining *zig* genes in *C. elegans*. Each allele is a predicted molecular null allele (Figure 2C) and none affects viability or produces any

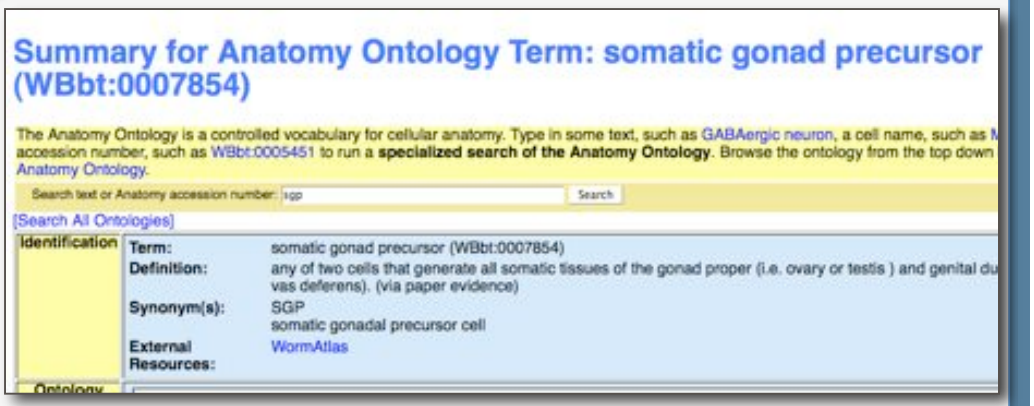
All genes are hyperlinked to their respective web page

(*zig-2, zig-3, zig-5, zig-6, zig-7,*

Manual QC maps informal terminology (jargon) to formal science terms

SGPs = somatic gonad precursor cells


daughters is much more pronounced in males t
pathway cause a "symmetrical sisters" phenoty
is imposed on the SGPs by the global sex deter
SGP sex determination, with *tra-1* feminizing
2004). SGP sex determination is linked to cell
The later phase of gonadal development involv
cells enlarge and leader cells (distal tip cells in
in both sexes, leading to the formation of diffe


Summary for Anatomy Ontology Term: somatic gonad precursor (WBbt:0007854)

The Anatomy Ontology is a controlled vocabulary for cellular anatomy. Type in some text, such as GABAergic neuron, a cell name, such as M accession number, such as WBbt:0005451 to run a **specialized search of the Anatomy Ontology**. Browse the ontology from the top down Anatomy Ontology.

Search text or Anatomy accession number: sgp Search

[Search All Ontologies]

Identification		
	Term:	somatic gonad precursor (WBbt:0007854)
	Definition:	any of two cells that generate all somatic tissues of the gonad proper (i.e. ovary or testis) and genital du vas deferens). (via paper evidence)
	Synonym(s):	SGP somatic gonadal precursor cell
	External Resources:	WormAtlas
Ontology		

By adding acronyms and science-field jargon as synonyms to the canonical science term in the database, we capture the current language of our community while making the science more understandable to the wider science community.

What is needed:

- Method for ensuring URL stability :

  A list of entity classes, names, URLs, and webpage status is automatically generated after the final manual QC stage has been done.

  This list provides an easy way to test the links for viability in the future.

WB Paper ID : WBPaper00036368

Genetics DOI: 10.1534/genetics.110.115501

Entity class	Entity name	link	live/silent
Anatomy	anterior gonad arms	http://www.wormbase.org/db/get?name=anterior gonad arm;class=Anatomy_name	live
Anatomy	germ cell	http://www.wormbase.org/db/get?name=germ cell;class=Anatomy_name	live
Anatomy	germ cells	http://www.wormbase.org/db/get?name=germ cell;class=Anatomy_name	live
Anatomy	germ line	http://www.wormbase.org/db/get?name=germ line;class=Anatomy_name	live
Anatomy	germ lines	http://www.wormbase.org/db/get?name=germ line;class=Anatomy_name	live
Anatomy	spermatocyte	http://www.wormbase.org/db/get?name=spermatocyte;class=Anatomy_name	live
Clone	zip-1	http://www.wormbase.org/db/get?name=zip-1;class=Clone	live
Gene/Protein	-2	http://www.wormbase.org/db/get?name=SYP-2;class=Gene	live
Gene/Protein	-3	http://www.wormbase.org/db/get?name=SYP-3;class=Gene	live
Gene/Protein	-4	http://www.wormbase.org/db/get?name=SYP-4;class=Gene	silent
Gene/Protein	F26D2.2	http://www.wormbase.org/db/get?name=F26D2.2;class=Gene	live
Gene/Protein	HIM-3	http://www.wormbase.org/db/get?name=HIM-3;class=Gene	live
Gene/Protein	HIM-8	http://www.wormbase.org/db/get?name=HIM-8;class=Gene	live
Gene/Protein	RAD-51	http://www.wormbase.org/db/get?name=RAD-51;class=Gene	live
Gene/Protein	SYP-1	http://www.wormbase.org/db/get?name=SYP-1;class=Gene	live

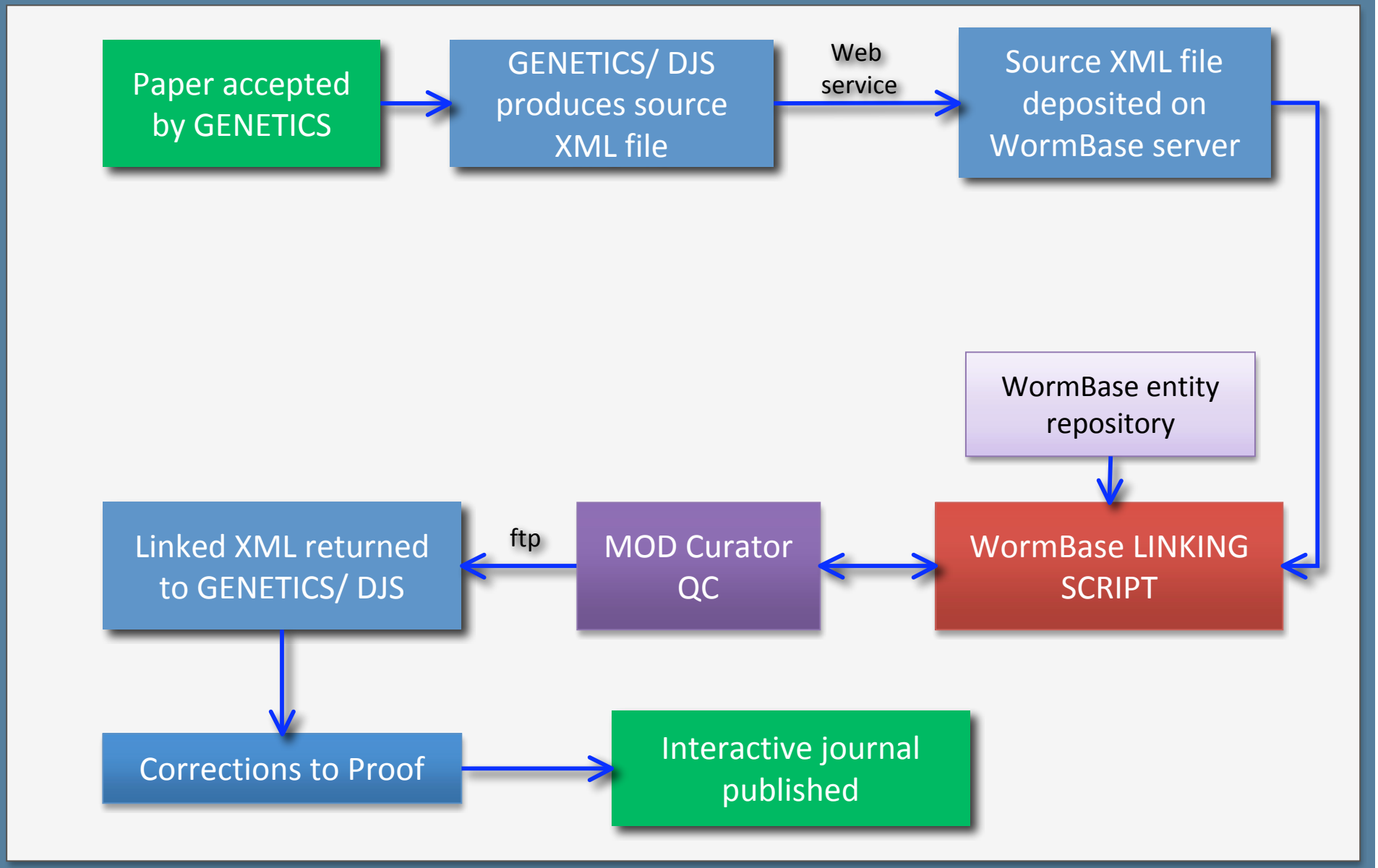
WormBase creates silent links to entities not yet in the database

GENETICS helps WormBase identify new entities by sending the author an entity declaration form

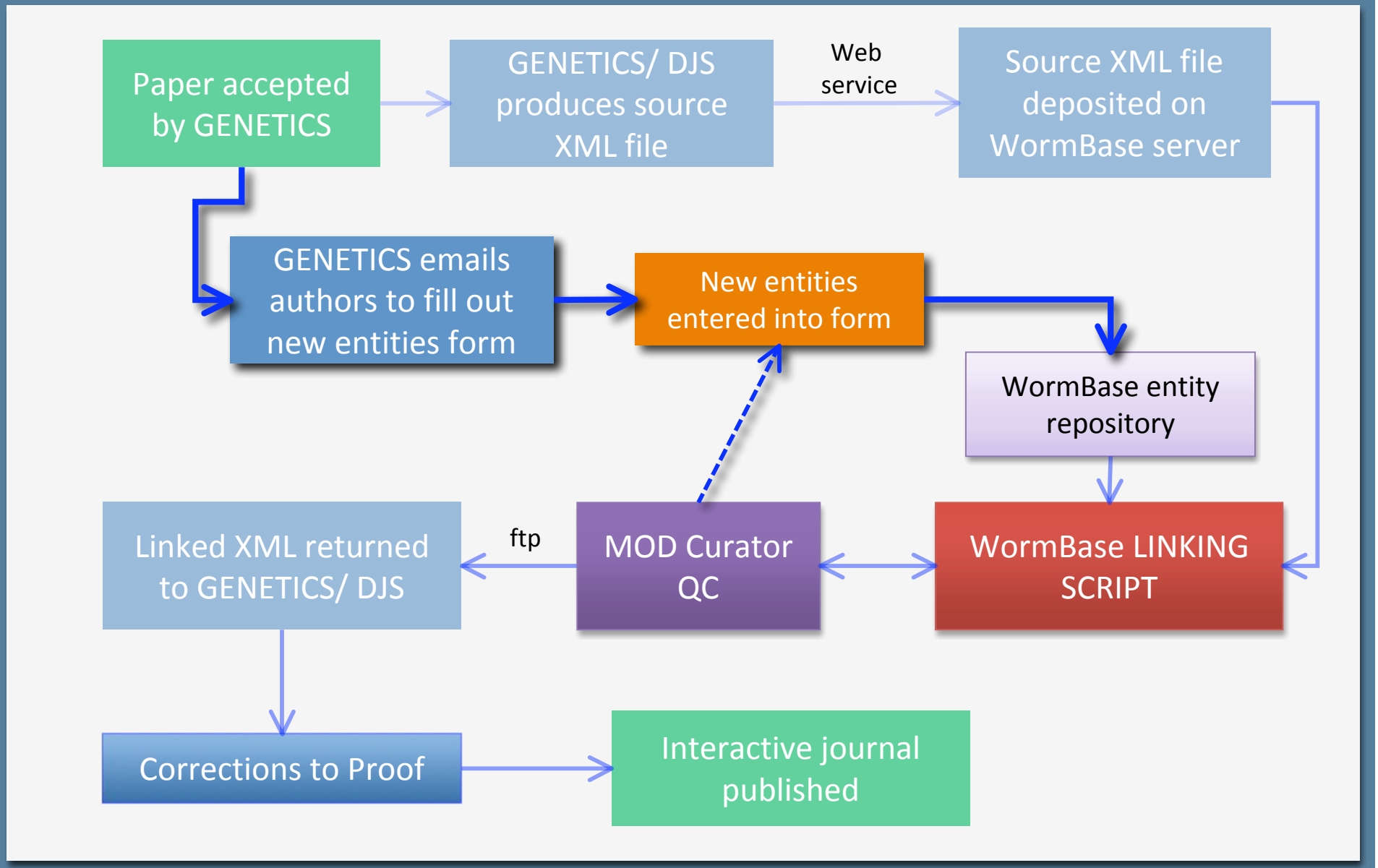
Manual QC identifies new entities not declared by the author

Silent links are made using the URL constructors, which will become live once the entities are added to the database

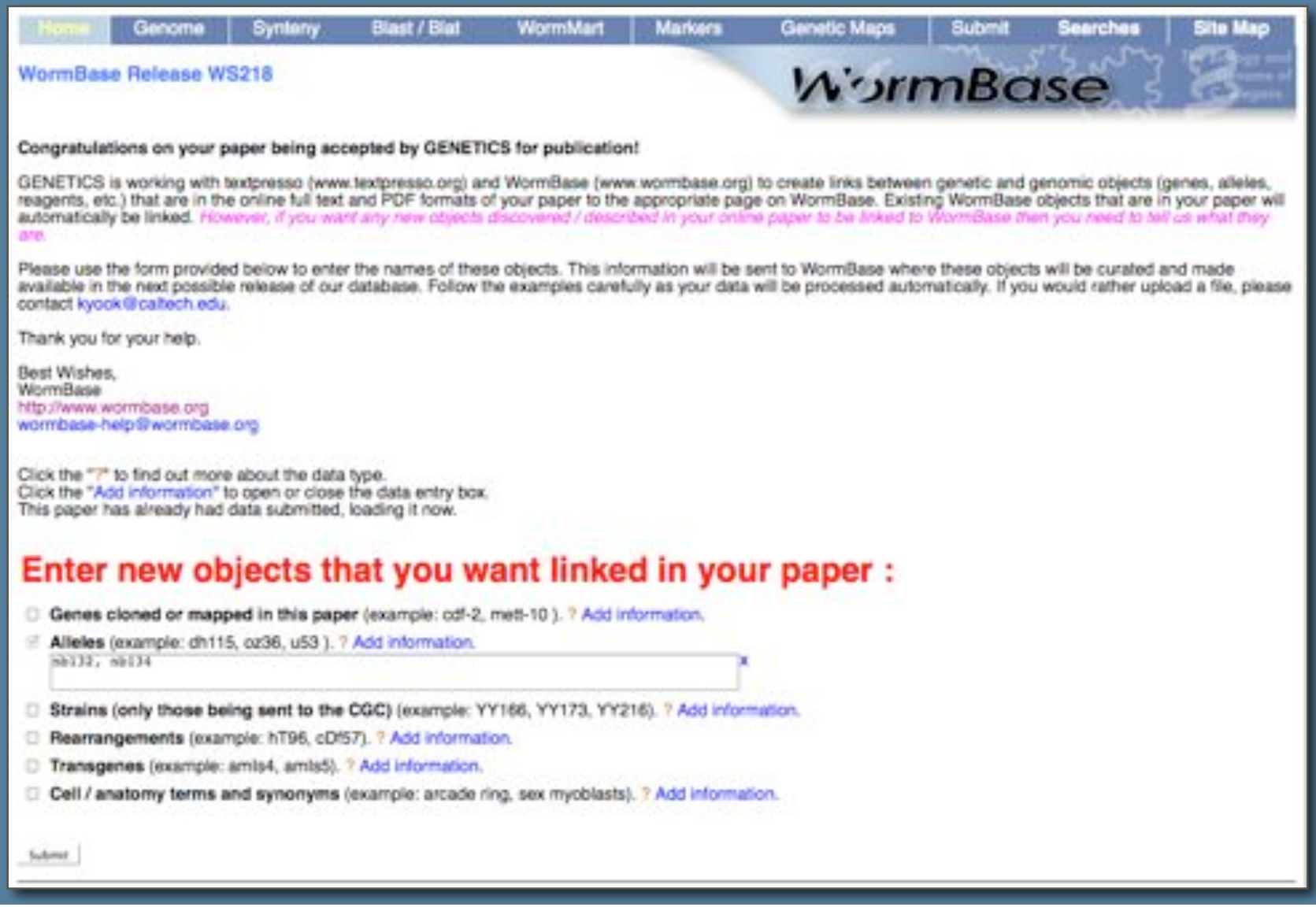
GSA-WormBase markup pipeline



GSA-WB markup pipeline includes author participation for new entities



Genetics sends authors an entity declaration form for WormBase



Home | Genome | Synteny | Blast / Blat | WormMart | Markers | Genetic Maps | Submit | Searches | Site Map

WormBase Release WS218

WormBase

Congratulations on your paper being accepted by GENETICS for publication!

GENETICS is working with textpresso (www.textpresso.org) and WormBase (www.wormbase.org) to create links between genetic and genomic objects (genes, alleles, reagents, etc.) that are in the online full text and PDF formats of your paper to the appropriate page on WormBase. Existing WormBase objects that are in your paper will automatically be linked. *However, if you want any new objects discovered / described in your online paper to be linked to WormBase then you need to tell us what they are.*

Please use the form provided below to enter the names of these objects. This information will be sent to WormBase where these objects will be curated and made available in the next possible release of our database. Follow the examples carefully as your data will be processed automatically. If you would rather upload a file, please contact kyook@caltech.edu.

Thank you for your help.

Best Wishes,
WormBase
http://www.wormbase.org
wormbase-help@wormbase.org

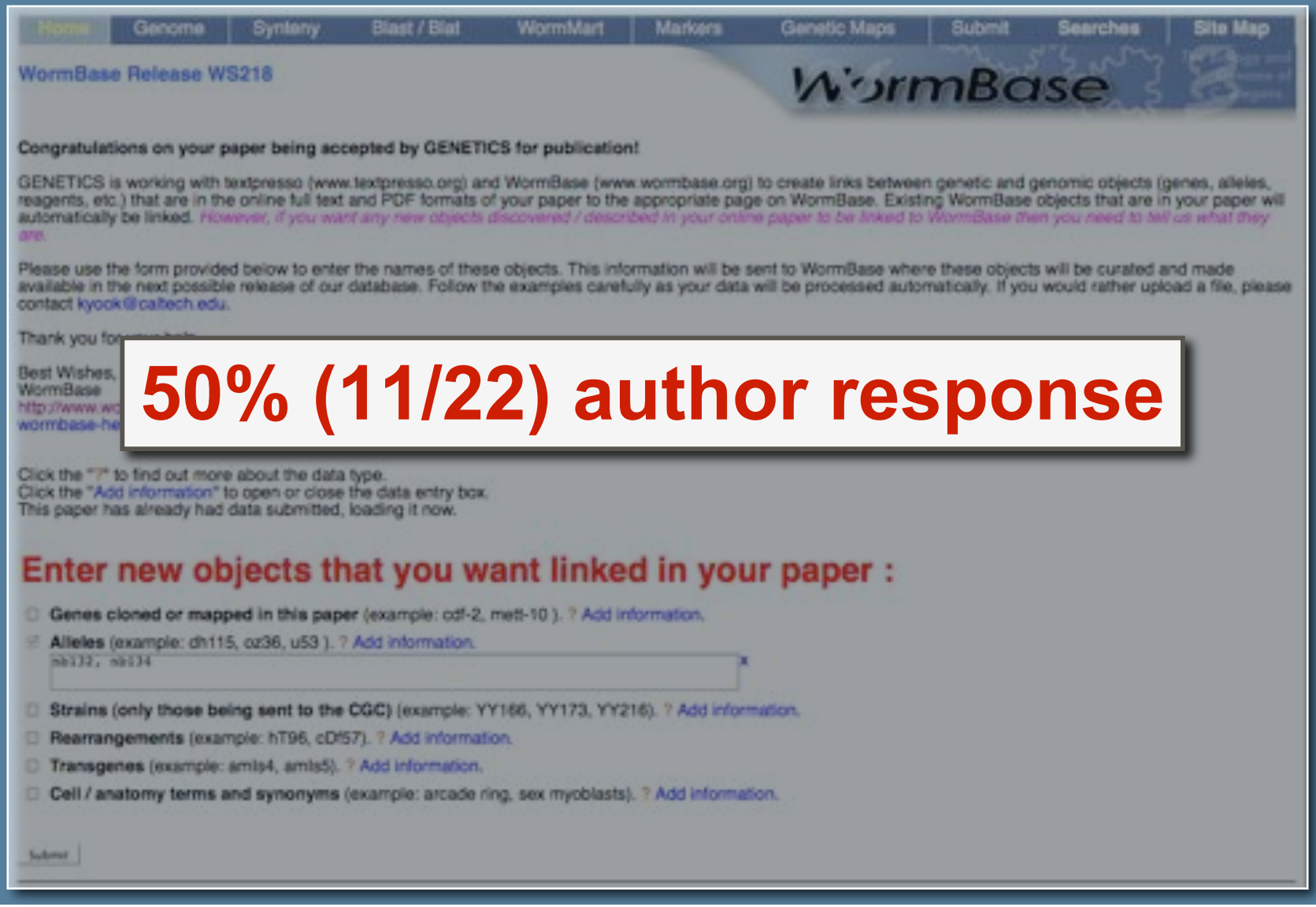
Click the "?" to find out more about the data type.
Click the "Add information" to open or close the data entry box.
This paper has already had data submitted, loading it now.

Enter new objects that you want linked in your paper :

- **Genes cloned or mapped in this paper** (example: odf-2, meti-10). ? Add information.
- **Alleles** (example: dh115, oz36, u53). ? Add information.
  nb132, nb134
- **Strains (only those being sent to the CGC)** (example: YY166, YY173, YY216). ? Add information.
- **Rearrangements** (example: hT96, cDf57). ? Add information.
- **Transgenes** (example: amIs4, amIs5). ? Add information.
- **Cell / anatomy terms and synonyms** (example: arcade ring, sex myoblasts). ? Add information.

Submit

Genetics sends authors an entity declaration form for WormBase



Where we stand at the moment

Caenorhabditis elegans papers marked up- **42**

35 published starting in August 2009

7 waiting to be published

5 more papers accepted, still to mark up

10 entity classes marked up

All have needed manual QC

(ambiguities, new objects, copy edit errors)

Saccharomyces cerevisiae papers marked up- **23**

2 entity classes marked up

Drosophila melanogaster papers will be ready to go live soon

Current challenges
Linking to multiple MODs in a single paper

C. elegans eIF1 carrying an equivalent yeast _sui1_ mutation allows translation initiation at a similar subset of non-AUG codons:

In contrast to what we observed with the *eif-1* mutants described above, non-AUG initiation by yeast eIF1 *sui1* mutants occurs at UUG but not with other non-AUG codons (Huang *et al*. 1997). To address whether eIF1 carrying these *sui1* mutations are able to confer a similar phenotype in our reporter system, we constructed mutations in the *C. elegans* eIF1 gene (Figure 2A) corresponding to all known *sui1* alleles (*sui1-1*, *sui1-4*, and *sui1-17)* isolated in the *His4* suppressor screens (Yoon and Donahue 1992) and *mof2-1*, which was first isolated as a mutant with an altered frameshift efficiency but later found to have a reduction in start codon recognition fidelity as well (Cui *et al*. 1998). Since *sui1* mutations are dominant or codominant in yeast (Yoon and Donahue 1992), it is possible that these mutations also behave dominantly in *C. elegans*, allowing the detection of their defects in our transgenic assay. These mutants were initially assayed with non-AUG GFP reporters carrying GUG or UUG codons. The *C. elegans* eIF1 transgene carrying the *sui1-4* mutation promoted GFP expression from both GUG and UUG reporters unlike transgenes carrying the wild-type eIF1 and the mutant eIF1 genes containing the *sui1-1*, *sui1-17*, or *mof2-1* mutations (data not shown). These results indicate that only the *sui1-4* mutant was able to confer non-AUG translation in the *C. elegans* reporter system, which is in contrast to that observed in yeast where all these mutations allowed translation to start at the UUG codon. We further assayed the ability of the *sui1-4* mutant to allow GFP expression from reporters containing other non-AUG start codons. GFP expression was observed from additional

Current challenges
Linking fly genes to FlyBase

Many fly gene names are difficult to distinguish from standard everyday words, e.g., *we*, *for*, *a*

Information-rich XML-tagged files help identify these genes

"Therefore, we tested members of major pathways that control DC such as <I>**Jun-related antigen**</I> (<I>**Jra**</I>), <I>**puckered**</I> (<I>**puc**</I>), and <I>**Src homology 2 ankyrin repeat tyrosine kinase**<I> (<I>**shark**</I>) in the DJNK signaling pathway and <I>**thickveins**<I> (<I>**tkv**</I>), <I>**schnurri**</I> (<I>**shn**</I>), and <I>**zipper**</I> (<I>**zip**</I>) in the TGF-β signaling pathway. Furthermore, we tested additional genes involved in DC such as <I>**u-shaped**</I> (<I>**ush**</I>), <I>**Epidermal growth factor receptor**</I> (<I>**Egfr**</I>), <I>**Protein kinase related to protein kinase N**</I> (<I>**Pkn**</I>), <I>**scab**</I> (<I>**scb**</I>), and <I>**Rho1**</I>…."

Benefits of published text markup

For the community:

- Less time to get to the correct database page
- Transparency of science jargon
- Term ambiguities are dealt with before publication

For the journal:

- Becomes a live portal to science information

For the author:

- Extra pair of eyes for proofing their document
- Data enter the database faster

For the database:

- Increase depth and accuracy of data in the database
- Get information before publication
- Resolve conflicts in term use before publication

Future challenges

- Extending this pipeline to other MODs / Databases
- Extending this pipeline to other journals
  - standardized XML format can help
- Increasing the number of Databases and entity classes linked in a multi-organism paper
- Fixing broken links
- Re-marking up publications

Acknowledgements

Arun Rangarajan
Karen Yook
Juancarlos Chan
Hans-Michael Muller
Paul W. Sternberg (HHMI)

Tracey DePellegrin-Connelly
Ruth Isaacson
Tim Schedl (Wash.U. School of Medicine)

Stephen Haenel
Lolly Otis
Sharon Faelten

Saccharomyces Genome Database
Marek S. Skrzypek, J. Michael Cherry (Stanford University School of Medicine)

William M. Gelbart (Harvard University)
Thom Kaufman, Kathy Matthews (Indiana University)
Raymund Stefancsik, Steven Marygold (Cambridge University, UK)