GBrowse.org
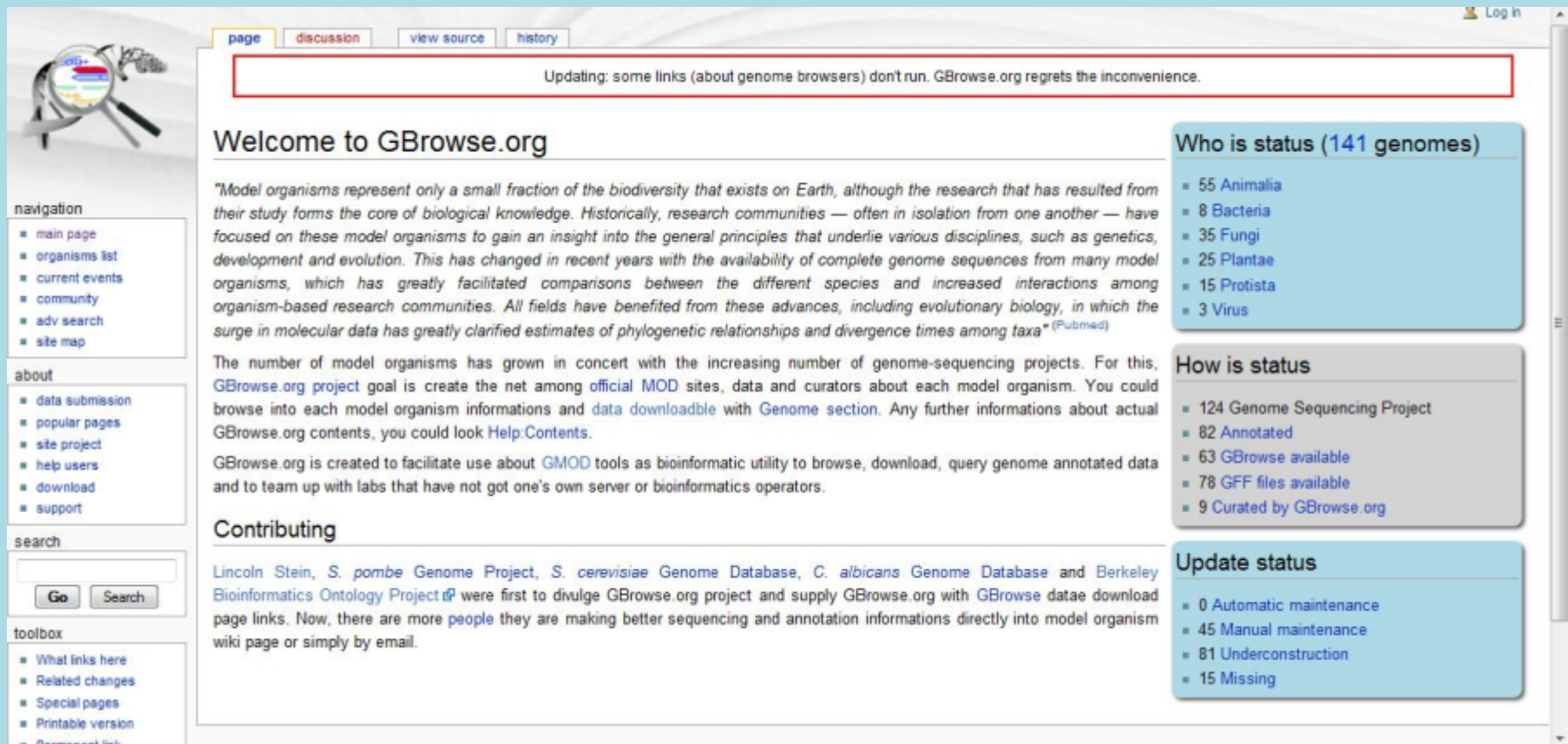
Alessandra Bilardi

CRIBI Biotech Center

Padua University

alessandra@bilardi.net

Home page

Gbrowse.org goals

1 genome

1 gbrowse available

Gbrowse.org goals

1 genome

$\updownarrow$

1 gbrowse available

Users type
- add tracks into gbrowse
- download gff and conf files

Gbrowse.org goals

1 genome

↕

1 gbrowse available

↕ ↕ ↕ ↕ ↕
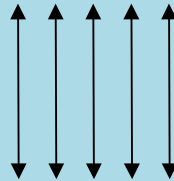
1 ftp site available

Automatic adding and update

GBrowse.org automatic scripts

GenBank entries → conversion by → GFF3 files → create → CONF files & intra GBrowse

GFF3 files → validation → all GFF3 correct

all GFF3 correct → no → errors log

all GFF3 correct → yes → GFF3 checked

EFetch Entrez Utility
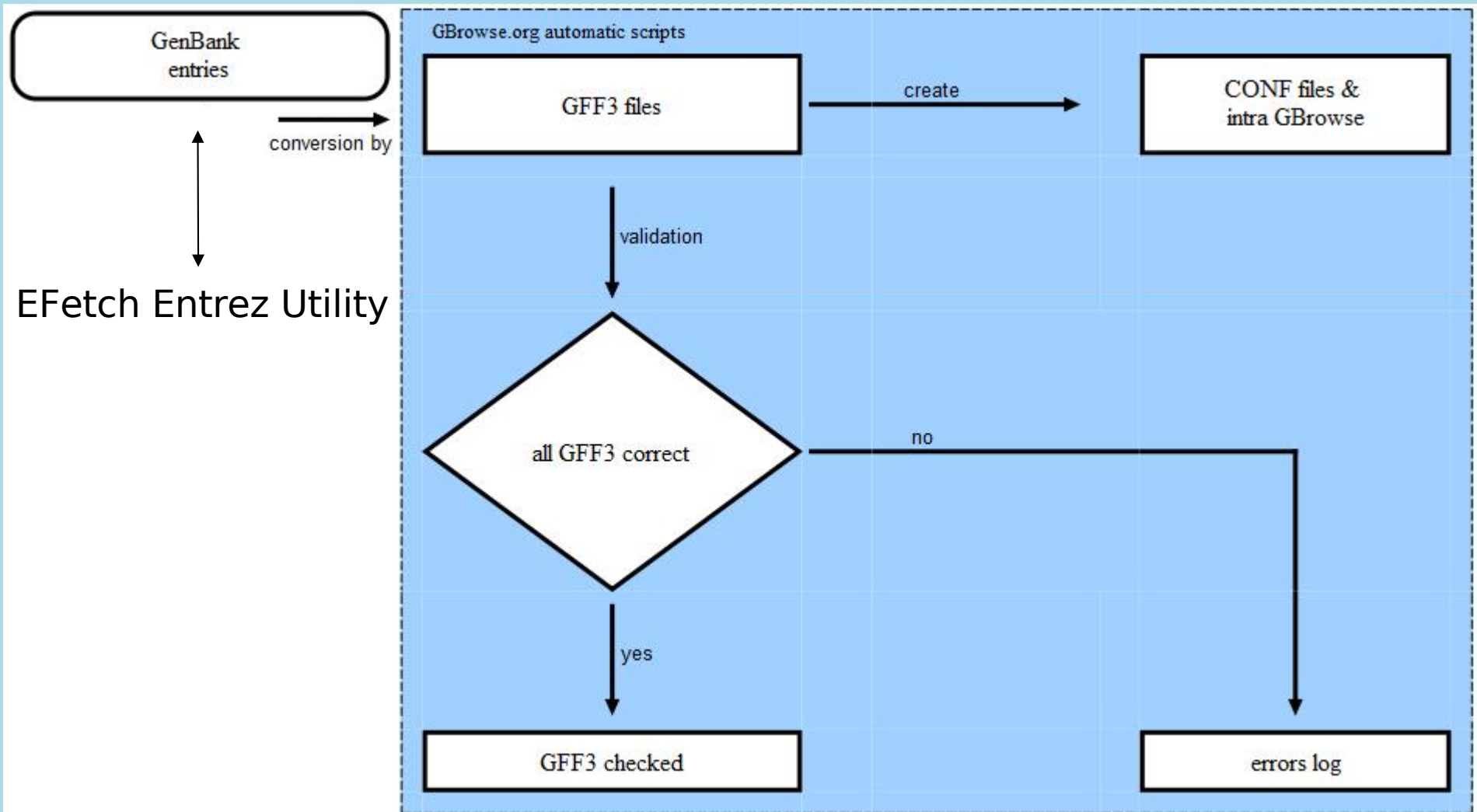
Manual step
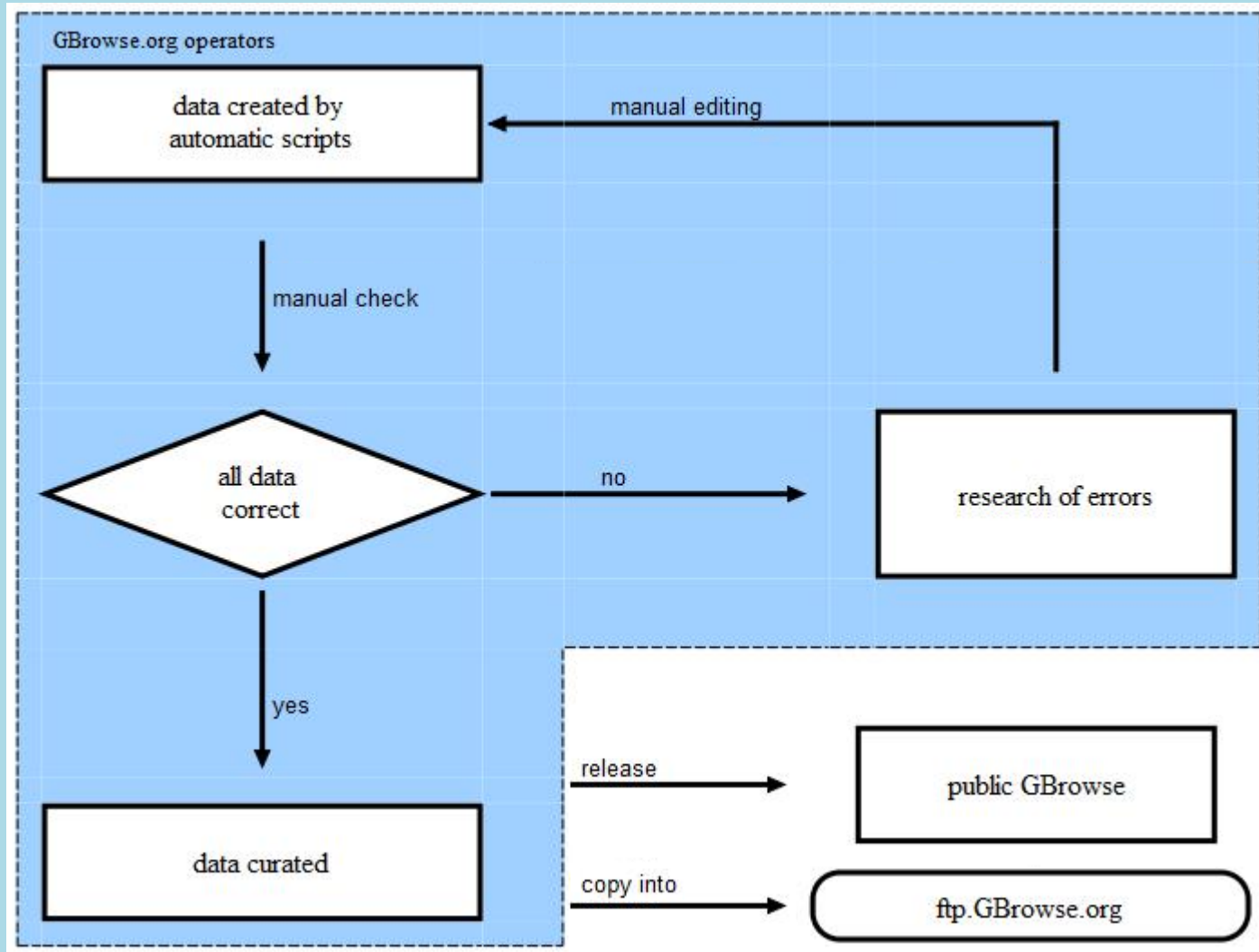
Genome page

Arabidopsis thaliana

Image	Browser	Download	Site dedicated	Page dedicated
	■ *Arabidopsis thaliana* TAIR8 ☑ ■ ATIDB ☑ ■ Gramene ☑ ■ MIPS PLantsDB SynBrowse ☑	■ *Arabidopsis thaliana* TAIR8 ☐ ■ Gramene ☐	■ The Arabidopsis Information Resource ☑ ■ ATIDB ☑	■ wikipedia ■ ChromDB ☑

Sequencing

Genomic sequencing of *A. thaliana* ecotype Columbia was done by a multinational effort collectively termed "The Arabidopsis Genome Initiative" (AGI). The Strategy employed was based on the use of BAC, YAC, cosmid, or P1 clones. Participating members were:

- TIGR ☑ - sequencing and annotation of Chromosome 1, plus parts of 2 and 3, whole genome analysis[1][2][3].
- Kazusa DNA Research Institute ☑ - sequencing and annotation of chromosome 3 and 5[4]. In addition, They also sequenced the complete chloroplast genome of Arabidopsis[5].
- ESSA - consisting of 17 european laboratories. They sequenced parts of chromosome 4 and 5[6]. MIPS ☑, as the informatics coordinator within the ESSA project, was responsible for data handling and analysis.
- SSP consortium ☑ - consisting of Stanford University, Salk Institute, and University of California, Berkeley. Sequencing and annotation of Chromosome 1[1].
- CSH/WU/ABI consortium - consisting of Cold Spring Harbor Laboratories ☑, Washington University School of Medicine ☑, and ABI-Perkin Elmer ☑. They sequenced Chromosomes 4 and 5 and annotated chromosome 4[6].
- GENOSCOPE ☑ - Chromosome 3 and BAC Ends[3].

Sequencing was completed in 2000[7]. The genome has five chromosomes and a total size of approximately 135-megabases. The current golden path length is 119,186,497 bp (TAIR8 release).

Genome page

Annotation

The Arabidopsis genome was initially annotated by the Arabidopsis Genome Initiative (AGI) and later re-annotated by TIGR in collaboration with MIPs and The Arabidopsis Information Resource (TAIR). TAIR[8] has now assumed primary responsibility for maintaining the Arabidopsis genome annotation in North America following (TIGR's) final genome release in 2004. Details of the latest genome release can be found at TAIRs Genome annotation portal ⮥. The TAIR8 release ⮥ contains 27,235 protein coding genes, 4759 pseudogenes or transposable elements and 1288 ncRNAs (33,282 genes in all, 38,963 gene models).

The curation process includes annotation of genes with Gene Ontology (GO; function, process and cellular component) and Plant Ontology (structure and developmental stage) terms with appropriate evidence codes and references. In addition, gene symbols, alleles, phenotypes and germplasm information are captured from the literature and a free text gene description summarizing a gene's important features is composed by curators. The TAIR gene structure annotation pipeline incorporates both manual and automated methods to update gene features such as exons and UTRs and add new genes based on new transcript evidence. Computational functional annotation pipelines assign GO terms to genes based on the presence of protein domains or signal sequences and generate a short phrase describing a gene's function

References

1. ↑ [1.0] [1.1] Theologis A et al. (2000) Sequence and analysis of chromosome 1 of the plant Arabidopsis thaliana. *Nature* **408**: 816-20 PubMed
2. ↑ Lin X et al. (1999) Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana. *Nature* **402**: 761-8 PubMed
3. ↑ [3.0] [3.1] Salanoubat M et al. (2000) Sequence and analysis of chromosome 3 of the plant Arabidopsis thaliana. *Nature* **408**: 820-2 PubMed
4. ↑ Tabata S et al. (2000) Sequence and analysis of chromosome 5 of the plant Arabidopsis thaliana. *Nature* **408**: 823-6 PubMed
5. ↑ Sato S et al. (1999) Complete structure of the chloroplast genome of Arabidopsis thaliana. *DNA Res* **6**: 283-90 PubMed
6. ↑ [6.0] [6.1] Mayer K et al. (1999) Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana. *Nature* **402**: 769-77 PubMed
7. ↑ Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature* **408**: 796-815 PubMed
8. ↑ Swarbreck D et al. (2008) The Arabidopsis Information Resource (TAIR): gene structure and function annotation. *Nucleic Acids Res* **36**: D1009-14 PubMed

Categories: GBrowse | Ensembl | SynBrowse | GFF3 | Genome Database | Sequenced | Assembled | Annotated | Plantae | Manual maintenance | Genome

Categories

```
[−] SiteMap
  [+] About
  [+] Community
  [+] Genome
  [+] Help
  [+] Support
  [+] Template
  [+] Tool
  [+] Workflow
```

Genome **Has a::feature**

[−] **SiteMap**
  [+] **About**
  [+] **Community**
  [−] Genome
   [+] **Annotation**
   [+] **Assembling**
   [+] **Browser**
   [+] **Download**
   [+] **Link**
   [+] **Sequencing**
   [+] **Status**
   [+] **Taxonomy**

Concepts **Is a::status**

Concept pages	Form pages	Property pages
▪ Concept:Annotation		
▪ Concept:Assembling		
▪ Concept:Browser Is a::browser	▪ Form:Organism	▪ Property:Has a
▪ Concept:File Is a::file	▪ Form:User	▪ Property:Is a
▪ Concept:Sequencing	▪ Form:User2	
▪ Concept:Site Is a::site		
▪ Concept:Status		
▪ Concept:Taxonomy		

Each category **Is a::**

[−] **Annotation**
  [+] **Annotated**
  [+] **Not Annotated**

[−] **Assembling**
  [+] **Assembled**
  [+] **In Assembly**
  [+] **Not Assembled**

[−] **Browser**
  [+] **Ensembl**
  [+] **GBK2GFF3**
  [+] **GBrowse**
  [+] **Genome Viewer**
  [+] **Java Genome Viewer**
  [+] **SynBrowse**
  [+] **UCSC**

[−] **Download**
  [+] **AGP**
  [+] **CONF**
  [+] **GFF1**
  [+] **GFF2**
  [+] **GFF3**
  [+] **GTF**
  [+] **MYSQL**
  [+] **TXT**

[−] **Link**
  [×] **Page dedicated**
  [+] **Site dedicated**

[−] **Sequencing**
  [+] **In progress**
  [+] **Not Sequenced**
  [+] **Sequenced**

[−] **Status**
  [×] **Automatic maintenance**
  [+] **Manual maintenance**
  [+] **Missing**
  [+] **Underconstruction**

[−] **Taxonomy**
  [+] **Animalia**
  [×] **Archaea**
  [+] **Bacteria**
  [+] **Fungi**
  [+] **Plantae**
  [+] **Protista**
  [+] **Virus**

Form

(gbrowse.org)

Multi-category search

(gbrowse.org)

Automatic create and update

EFetch Entrez Utility

$\updownarrow$

Create genome page with genome sequencing status

Manual step

(gbrowse.org)

Future

- complete automations

Future

- complete automations

- test and edit links

Future

- complete automations

- test and edit links

- edit sequencing and annotation methods

Future

- complete automations

- test and edit links

- edit sequencing and annotation methods

- generate gbrowses and pages about all genomes with sequencing completed

Future

- complete automations

- test and edit links

- edit sequencing and annotation methods

- generate gbrowses and pages about all genomes with sequencing completed

- divide gbrowses and genome pages in different sites (?!?)

doubt and suggest..

..ehm, do you ask me by email?