



Automated and Accurate Genome Annotation with BRAKER, GALBA, & TSEBRA

2023 Workshop on Genomics, Cesky Krumlov

Who we are

Gene Prediction

BRAKER

BRAKER1: RNA-Seq

BRAKER2: Proteins

TSEBRA

BRAKER3: RNA-Seq +
Proteins

Accuracy Results

Availability

GALBA

GALBA: Proteins

Accuracy Results

Availability

Annotation Quality

Genome Browsers

Descriptive Statistics

BUSCO

OMArk

Hands on Lab Session

Katharina J. Hoff & Natalia Nenasheva

Contact: katharina.hoff@uni-greifswald.de

Short CV

2022 Habilitation (Greifswald)

2009 Ph.D. Molecular Biology (Göttingen)

2005 B.Sc. Plant Biotechnology (Hanover, stays abroad: Budapest & Alnarp)

Research

- eukaryotic genome annotation, metagenomics
- best known for: **BRAKER** & other **Gaius-Augustus** software
- 27 peer-reviewed research articles with currently 4,759 citations
- ~1 Mio € grants → **open PhD student position!**

Teaching

- currently 1(+2) PhD students, 1 MSc student, 1 BSc student
- applied bioinformatics, programming, statistics, & data science

Short CV

- 2022 - now** Institute for Mathematics and Computer Science, University of Greifswald
- 2021 - 2022** Bioinformatician at the Genotek, Moscow, Russia
- 2020 - 2022** M.Sc. in Applied Mathematics and Physics (the Moscow Institute of Physics and Technology, Department of Biological and Medical Physics)
- 2019 - 2020** Junior bioinformatician at the Boston Gene, Moscow, Russia
- 2016 - 2020** B.Sc. in Applied Mathematics and Physics (the Moscow Institute of Physics and Technology, Department of Biological and Medical Physics)

Research

- gene prediction in eukaryotic metagenomes
- affiliated with POMPU

Teaching

- Python, Bash, and other data science tools

We are Looking for a PhD Student!

Amazing Project

- develop novel methods for pathogen screening in dairy farm environments
- NGS, AI, & applied bioinformatics
- interdisciplinary team

Position

- full-time research assistant
- 3 years, starting on July 1st 2023

Application Deadline

- send cover letter, CV, copies of degree certificates to katharina.hoff@uni-greifswald.de until June 3rd 2023



Contents

1 Who we are

2 Gene Prediction

3 BRAKER

BRAKER1: RNA-Seq

BRAKER2: Proteins

TSEBRA

BRAKER3: RNA-Seq + Proteins

Accuracy Results

Availability

4 GALBA

GALBA: Proteins

Accuracy Results

Availability

5 Annotation Quality

Genome Browsers

Descriptive Statistics

BUSCO

OMArk

6 Hands on Lab Session

Who we are

Gene Prediction

BRAKER

BRAKER1: RNA-Seq

BRAKER2: Proteins

TSEBRA

BRAKER3: RNA-Seq +
Proteins

Accuracy Results

Availability

GALBA

GALBA: Proteins

Accuracy Results

Availability

Annotation Quality

Genome Browsers

Descriptive Statistics

BUSCO

OMArk

Hands on Lab Session

Eukaryotic Gene

Automated and
Accurate Genome
Annotation with
BRAKER, GALBA, &
TSEBRA

Katharina J. Hoff &
Natalia Nenasheva



Who we are

Gene Prediction

BRAKER

BRAKER1: RNA-Seq

BRAKER2: Proteins

TSEBRA

BRAKER3: RNA-Seq +
Proteins

Accuracy Results

Availability

GALBA

GALBA: Proteins

Accuracy Results

Availability

Annotation Quality

Genome Browsers

Descriptive Statistics

BUSCO

OMArk

Hands on Lab Session

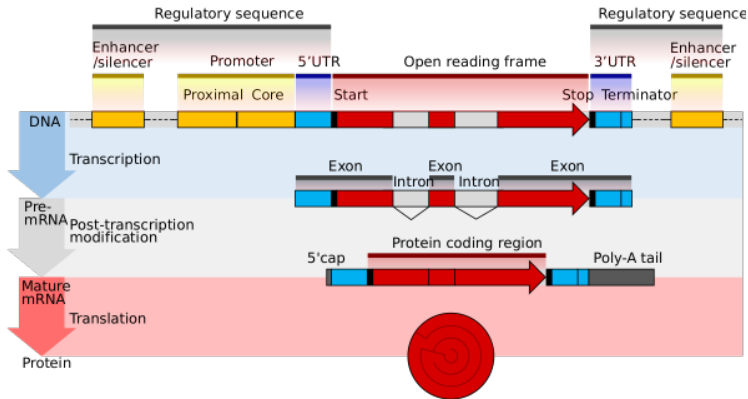


Image: Wikipedia, CC BY 4.0

Information for Genome Annotation

- genome sequence: **mathematical model**
- observed mRNA expression (parts)
- known proteins of relatives

Structural Genome Annotation Problem

Automated and
Accurate Genome
Annotation with
BRAKER, GALBA, &
TSEBRA

Katharina J. Hoff &
Natalia Nenasheva



Who we are

Gene Prediction

BRAKER

BRAKER1: RNA-Seq

BRAKER2: Proteins

TSEBRA

BRAKER3: RNA-Seq +
Proteins

Accuracy Results

Availability

GALBA

GALBA: Proteins

Accuracy Results

Availability

Annotation Quality

Genome Browsers

Descriptive Statistics

BUSCO

OMArk

Hands on Lab Session

Input

- genome assembly
- extrinsic evidence, e.g. from RNA-Seq, protein sequences

Output

- protein-coding genes: CDSexon-intron structures (. gff)

Task: find and predict gene structures of protein-coding genes

**Correct Gene
Structure**



**Gene Prediction
Tool**



**RNA-Seq
Alignments**



**Protein
Alignments**



Gene Finders Need Training Examples

Automated and
Accurate Genome
Annotation with
BRAKER, GALBA, &
TSEBRA

Katharina J. Hoff &
Natalia Nenasheva



Who we are

Gene Prediction

BRAKER

BRAKER1: RNA-Seq

BRAKER2: Proteins

TSEBRA

BRAKER3: RNA-Seq +
Proteins

Accuracy Results

Availability

GALBA

GALBA: Proteins

Accuracy Results

Availability

Annotation Quality

Genome Browsers

Descriptive Statistics

BUSCO

OMArk

Hands on Lab Session

Gene Finders Need Training Examples

Automated and
Accurate Genome
Annotation with
BRAKER, GALBA, &
TSEBRA

Katharina J. Hoff &
Natalia Nenasheva



Who we are

Gene Prediction

BRAKER

BRAKER1: RNA-Seq

BRAKER2: Proteins

TSEBRA

BRAKER3: RNA-Seq +
Proteins

Accuracy Results

Availability

GALBA

GALBA: Proteins

Accuracy Results

Availability

Annotation Quality

Genome Browsers

Descriptive Statistics

BUSCO

OMArk

Hands on Lab Session

The BRAKER Team

University of Greifswald & Georgia Tech University

Automated and
Accurate Genome
Annotation with
BRAKER, GALBA, &
TSEBRA

Katharina J. Hoff &
Natalia Nenasheva



Who we are

Gene Prediction

BRAKER

BRAKER1: RNA-Seq

BRAKER2: Proteins

TSEBRA

BRAKER3: RNA-Seq +
Proteins

Accuracy Results

Availability

GALBA

GALBA: Proteins

Accuracy Results

Availability

Annotation Quality

Genome Browsers

Descriptive Statistics

BUSCO

OMArk

Hands on Lab Session



Lars Gabriel



Alexandre Lomsadze, Katharina Hoff, Tomáš Brůna



Mario Stanke



Mark Borodovsky

Also: Simone Lange, Matthis Ebel, Hannah Thierfeldt, Anica Hoppe



BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS FREE

Katharina J. Hoff ✉, Simone Lange, Alexandre Lomsadze, Mark Borodovsky ✉, Mario Stanke

Bioinformatics, Volume 32, Issue 5, 1 March 2016, Pages 767–769,

<https://doi.org/10.1093/bioinformatics/btv661>

Who we are

Gene Prediction

BRAKER

BRAKER1: RNA-Seq

BRAKER2: Proteins

TSEBRA

BRAKER3: RNA-Seq +
Proteins

Accuracy Results

Availability

GALBA

GALBA: Proteins

Accuracy Results

Availability

Annotation Quality

Genome Browsers

Descriptive Statistics

BUSCO

OMArk

Hands on Lab Session

- spliced alignments of RNA-Seq are used by GeneMark-ET and AUGUSTUS
- 1,157 citations (Google Scholar)

Whole-Genome Annotation with BRAKER

**Katharina J. Hoff, Alexandre Lomsadze, Mark Borodovsky,
and Mario Stanke**

in Kollmar M. (eds) *Gene Prediction. Methods in Molecular Biology*,
vol 1962. Humana, New York, NY, 2019

GeneMark-ET uses RNA-Seq for Training

Anchors from RNA-Seq for training

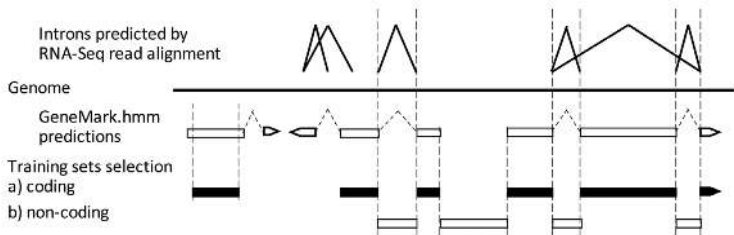


Figure 3. Selection of elements of training set in GeneMark-ET for the next iteration. The new training set of protein-coding regions is comprised from exons with at least one 'anchored splice site' as well as long exons predicted *ab initio* (> 800 nt).

- employs unsupervised training
- includes in training introns and exons anchored by mapped RNA-Seq reads
- does not require RNA-Seq reads assembly
- does not use RNA-Seq information in the *prediction* step



AUGUSTUS uses RNA-Seq for Prediction

Who we are

Gene Prediction

BRAKER

BRAKER1: RNA-Seq

BRAKER2: Proteins

TSEBRA

BRAKER3: RNA-Seq +
Proteins

Accuracy Results

Availability

GALBA

GALBA: Proteins

Accuracy Results

Availability

Annotation Quality

Genome Browsers

Descriptive Statistics

BUSCO

OMArk

Hands on Lab Session

- requires “prior data” for training
- uses intron information from RNA-seq for *prediction*
- no RNA-Seq assembly required

BRAKER1 Gene F1 Accuracy

Automated and
Accurate Genome
Annotation with
BRAKER, GALBA, &
TSEBRA

Katharina J. Hoff &
Natalia Nenasheva



Who we are

Gene Prediction

BRAKER

BRAKER1: RNA-Seq

BRAKER2: Proteins

TSEBRA

BRAKER3: RNA-Seq +
Proteins

Accuracy Results

Availability

GALBA

GALBA: Proteins

Accuracy Results

Availability

Annotation Quality

Genome Browsers

Descriptive Statistics

BUSCO

OMArk

Hands on Lab Session



BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database

Tomáš Brůna^{1,†}, Katharina J. Hoff^{2,3,†}, Alexandre Lomsadze⁴, Mario Stanke^{2,3,1} and Mark Borodovsky^{4,5,*†}

Who we are

Gene Prediction

BRAKER

BRAKER1: RNA-Seq

BRAKER2: Proteins

TSEBRA

BRAKER3: RNA-Seq +
Proteins

Accuracy Results

Availability

GALBA

GALBA: Proteins

Accuracy Results

Availability

Annotation Quality

Genome Browsers

Descriptive Statistics

BUSCO

OMArk

Hands on Lab Session

- spliced alignments of a large number of proteins
- 428 citations (Google Scholar)

Evidence Usage by GeneMark-EP+ & AUGUSTUS During Prediction

BRAKER2 Gene F1 Accuracy

Automated and
Accurate Genome
Annotation with
BRAKER, GALBA, &
TSEBRA

Katharina J. Hoff &
Natalia Nenasheva



Who we are

Gene Prediction

BRAKER

BRAKER1: RNA-Seq

BRAKER2: Proteins

TSEBRA

BRAKER3: RNA-Seq +
Proteins

Accuracy Results

Availability

GALBA

GALBA: Proteins

Accuracy Results

Availability

Annotation Quality

Genome Browsers

Descriptive Statistics

BUSCO

OMArk

Hands on Lab Session

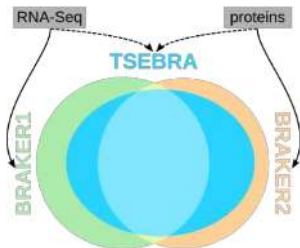
SOFTWARE

Open Access



TSEBRA: transcript selector for BRAKER

Lars Gabriel^{1,2}, Katharina J. Hoff^{1,2}, Tomáš Brůna³, Mark Borodovsky^{4,5} and Mario Stanke^{1,2*} 



- run both BRAKER1 and BRAKER2
- combine & increase accuracy
- 34 citations (Google Scholar)

Who we are

Gene Prediction

BRAKER

BRAKER1: RNA-Seq

BRAKER2: Proteins

TSEBRA

BRAKER3: RNA-Seq +
Proteins

Accuracy Results

Availability

GALBA

GALBA: Proteins

Accuracy Results

Availability

Annotation Quality

Genome Browsers

Descriptive Statistics

BUSCO

OMArk

Hands on Lab Session

BRAKER1 + BRAKER2 → TSEBRA

Automated and
Accurate Genome
Annotation with
BRAKER, GALBA, &
TSEBRA

Katharina J. Hoff &
Natalia Nenasheva



Who we are

Gene Prediction

BRAKER

BRAKER1: RNA-Seq

BRAKER2: Proteins

TSEBRA

BRAKER3: RNA-Seq +
Proteins

Accuracy Results

Availability

GALBA

GALBA: Proteins

Accuracy Results

Availability

Annotation Quality

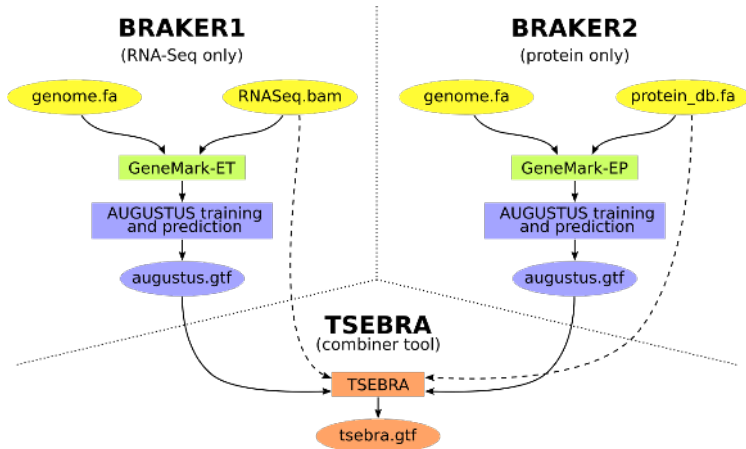
Genome Browsers

Descriptive Statistics

BUSCO

OMArk

Hands on Lab Session



TSEBRA: BRAKER1 + BRAKER2 Gene F1 Accuracy

Automated and
Accurate Genome
Annotation with
BRAKER, GALBA, &
TSEBRA

Katharina J. Hoff &
Natalia Nenasheva



Who we are

Gene Prediction

BRAKER

BRAKER1: RNA-Seq

BRAKER2: Proteins

TSEBRA

BRAKER3: RNA-Seq +
Proteins

Accuracy Results

Availability

GALBA

GALBA: Proteins

Accuracy Results

Availability

Annotation Quality

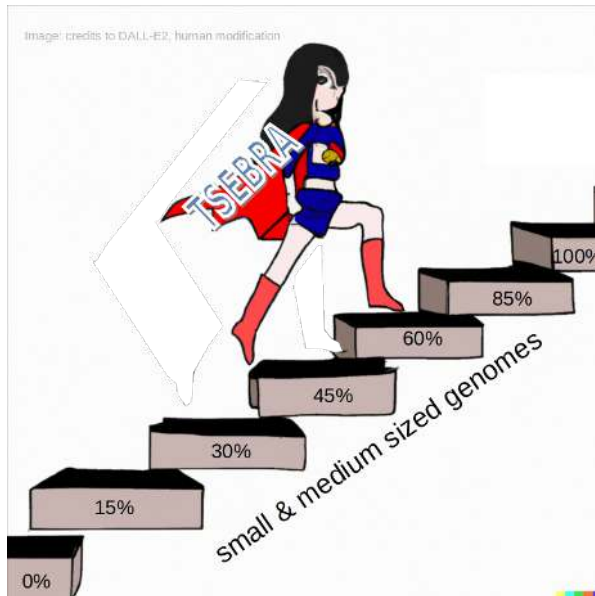
Genome Browsers

Descriptive Statistics

BUSCO

OMArk

Hands on Lab Session



BRAKER3 Gene F1 Accuracy - Climbing the Top

Automated and
Accurate Genome
Annotation with
BRAKER, GALBA, &
TSEBRA

Katharina J. Hoff &
Natalia Nenasheva



Who we are

Gene Prediction

BRAKER

BRAKER1: RNA-Seq

BRAKER2: Proteins

TSEBRA

BRAKER3: RNA-Seq +
Proteins

Accuracy Results

Availability

GALBA

GALBA: Proteins

Accuracy Results

Availability

Annotation Quality

Genome Browsers

Descriptive Statistics

BUSCO

OMArk

Hands on Lab Session

BRAKER3: Using RNA-Seq and Protein Evidence with GeneMark-ETP, AUGUSTUS and TSEBRA

Automated and
Accurate Genome
Annotation with
BRAKER, GALBA, &
TSEBRA

Katharina J. Hoff &
Natalia Nenasheva



Who we are

Gene Prediction

BRAKER

BRAKER1: RNA-Seq

BRAKER2: Proteins

TSEBRA

BRAKER3: RNA-Seq +
Proteins

Accuracy Results

Availability

GALBA

GALBA: Proteins

Accuracy Results

Availability

Annotation Quality

Genome Browsers

Descriptive Statistics

BUSCO

OMArk

Hands on Lab Session

- spliced aligned and **assembled** RNA-Seq
- large protein database
- combines GeneMark-ETP and AUGUSTUS gene sets with TSEBRA

Manuscript by Lars Gabriel et. al in preparation

BRAKER3: Using RNA-Seq and Protein Evidence with GeneMark-ETP, AUGUSTUS and TSEBRA

Automated and
Accurate Genome
Annotation with
BRAKER, GALBA, &
TSEBRA

Katharina J. Hoff &
Natalia Nenasheva



Who we are

Gene Prediction

BRAKER

BRAKER1: RNA-Seq

BRAKER2: Proteins

TSEBRA

**BRAKER3: RNA-Seq +
Proteins**

Accuracy Results

Availability

GALBA

GALBA: Proteins

Accuracy Results

Availability

Annotation Quality

Genome Browsers

Descriptive Statistics

BUSCO

OMArk

Hands on Lab Session

BRAKER3 Experiments

Automated and
Accurate Genome
Annotation with
BRAKER, GALBA, &
TSEBRA

Katharina J. Hoff &
Natalia Nenasheva



Who we are

Gene Prediction

BRAKER

BRAKER1: RNA-Seq

BRAKER2: Proteins

TSEBRA

BRAKER3: RNA-Seq +
Proteins

Accuracy Results

Availability

GALBA

GALBA: Proteins

Accuracy Results

Availability

Annotation Quality

Genome Browsers

Descriptive Statistics

BUSCO

OMArk

Hands on Lab Session

Experiments

Accuracy assessment using genome-wide predictions in 6 species:

Species	Genome Size (Mb)	# Genes in Annotation
<i>Arabidopsis thaliana</i> (thale cress)	119	27,444
<i>Caenorhabditis elegans</i> (nematode)	100	20,172
<i>Drosophila melanogaster</i> (fruit fly)	137	13,928
<i>Gallus gallus</i> (chicken)	1,040	17,279
<i>Mus musculus</i> (mouse)	2,650	22,378
<i>Solanum lycopersicum</i> (tomato)	772	33,562

Accuracy metrics

Specificity [Sp]: Percentage of correctly found genes/transcripts/exons in the **predicted gene set**.

Sensitivity [Sn]: Percentage of correctly found genes/transcripts/exons in the **reference annotation**.

Harmonic Mean [F1]:
$$\frac{2 \cdot \text{Sensitivity} \cdot \text{Specificity}}{\text{Sensitivity} + \text{Specificity}}$$

INPUT

GeneMark-ETP

AUGUSTUS

TSEBRA

OUTPUT

BRAKER3 Accuracy in Small Genomes

Automated and
Accurate Genome
Annotation with
BRAKER, GALBA, &
TSEBRA

Katharina J. Hoff &
Natalia Nenasheva



Who we are

Gene Prediction

BRAKER

BRAKER1: RNA-Seq

BRAKER2: Proteins

TSEBRA

BRAKER3: RNA-Seq +
Proteins

Accuracy Results

Availability

GALBA

GALBA: Proteins

Accuracy Results

Availability

Annotation Quality

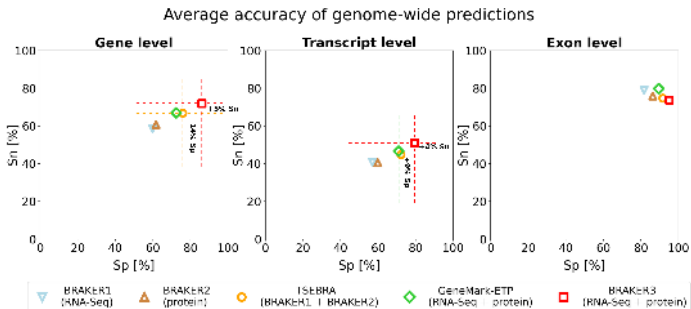
Genome Browsers

Descriptive Statistics

BUSCO

OMArk

Hands on Lab Session



Species: *D. melanogaster*
A. thaliana
C. elegans

Extrinsic evidence:

- paired RNA-Seq short reads
- OrthoDB v.10 clade partitions
(order excluded)

INPUT

GeneMark-ETP

AUGUSTUS

TSEBRA

OUTPUT

BRAKER3 Accuracy in Larger Genomes

Automated and
Accurate Genome
Annotation with
BRAKER, GALBA, &
TSEBRA

Katharina J. Hoff &
Natalia Nenasheva



Who we are

Gene Prediction

BRAKER

BRAKER1: RNA-Seq

BRAKER2: Proteins

TSEBRA

BRAKER3: RNA-Seq +
Proteins

Accuracy Results

Availability

GALBA

GALBA: Proteins

Accuracy Results

Availability

Annotation Quality

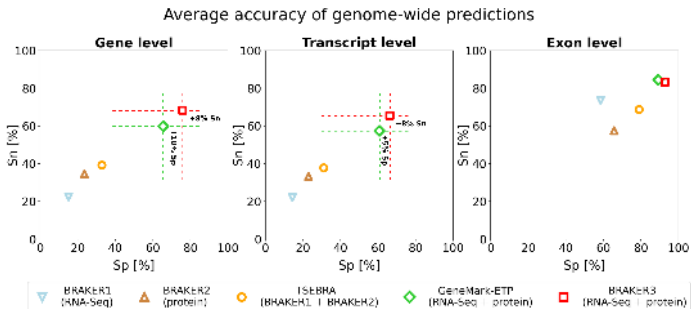
Genome Browsers

Descriptive Statistics

BUSCO

OMark

Hands on Lab Session



Species: *M. musculus*
G. gallus
S. lycopersicum

Extrinsic evidence:

- paired RNA-Seq short reads
- OrthoDB v.10 clade partitions
(order excluded)

INPUT

GeneMark-ETP

AUGUSTUS

TSEBRA

OUTPUT



Who we are

Gene Prediction

BRAKER

BRAKER1: RNA-Seq

BRAKER2: Proteins

TSEBRA

BRAKER3: RNA-Seq +
Proteins

Accuracy Results

Availability

GALBA

GALBA: Proteins

Accuracy Results

Availability

Annotation Quality

Genome Browsers

Descriptive Statistics

BUSCO

OMArk

Hands on Lab Session

Usage & Runtime

Command line

```
braker.pl --genome=genome.fa --prot_seq=protein_db.fa \  
--rnaseq_sets_ids=RNA_ID1, RNA_ID2 \  
--rnaseq_sets_dirs=/path/to/RNASeq/
```

Runtime

- average for *A. thaliana*, *C. elegans*, *D. melanogaster*, *G. gallus*, *M. musculus*, *S. lycopersicum*.
- with 48 threads:

	BRAKER1	BRAKER2	GM-ETP	BRAKER3
Runtime (h)	06:26	09:01	06:03	17:55



Who we are

Gene Prediction

BRAKER

BRAKER1: RNA-Seq

BRAKER2: Proteins

TSEBRA

BRAKER3: RNA-Seq +
Proteins

Accuracy Results

Availability

GALBA

GALBA: Proteins

Accuracy Results

Availability

Annotation Quality

Genome Browsers

Descriptive Statistics

BUSCO

OMArk

Hands on Lab Session

Availability

GitHub

<https://github.com/Gaius-Augustus/BRAKER>

Docker/Singularity

```
singularity build braker3.sif \  
docker://teambraker/braker3:latest
```

```
singularity exec braker3.sif braker.pl [OPTIONS]
```

Licenses

- BRAKER: Artistic License
- GeneMark-ETP: License for GeneMark family software
http://topaz.gatech.edu/genemark/license_download.cgi

The GALBA Team

DOE Joint Genome Institute & Harvard Medical School
& University of Greifswald

Automated and
Accurate Genome
Annotation with
BRAKER, GALBA, &
TSEBRA

Katharina J. Hoff &
Natalia Nenasheva



Who we are

Gene Prediction

BRAKER

BRAKER1: RNA-Seq

BRAKER2: Proteins

TSEBRA

BRAKER3: RNA-Seq +
Proteins

Accuracy Results

Availability

GALBA

GALBA: Proteins

Accuracy Results

Availability

Annotation Quality

Genome Browsers

Descriptive Statistics

BUSCO

OMArk

Hands on Lab Session



Tomáš Brůna



Heng Li



Lars Gabriel



Natalia Nenasheva



Matthis Ebel



Mario Stanke



Katharina Hoff

Also: Daniel Honsel, & Steffen Herboldt



Who we are

Gene Prediction

BRAKER

BRAKER1: RNA-Seq

BRAKER2: Proteins

TSEBRA

BRAKER3: RNA-Seq +
Proteins

Accuracy Results

Availability

GALBA

GALBA: Proteins

Accuracy Results

Availability

Annotation Quality

Genome Browsers

Descriptive Statistics

BUSCO

OMArk

Hands on Lab Session

Miniprot

Bioinformatics, 39(1), 2023, btad014

<https://doi.org/10.1093/bioinformatics/btad014>

Advance Access Publication Date: 17 January 2023

Original Paper

OXFORD

Genome analysis

Protein-to-genome alignment with miniprot

Heng Li  ^{1,2}

“Miniprot is a fast protein-to-genome aligner comparable to existing tools in accuracy. Its primary use case is to assist gene annotation.”

GALBA

Do we need another pipeline?

- no RNA-Seq, large genome
- proteins of few or 1 reference species
- fully open source



Who we are

Gene Prediction

BRAKER

BRAKER1: RNA-Seq

BRAKER2: Proteins

TSEBRA

BRAKER3: RNA-Seq +
Proteins

Accuracy Results

Availability

GALBA

GALBA: Proteins

Accuracy Results

Availability

Annotation Quality

Genome Browsers

Descriptive Statistics

BUSCO

OMArk

Hands on Lab Session

Donor proteins from

dsim *D. simulans*

combo *D. ananassae*,
D. pseudoobscura,
D. willistoni,
D. virilis,
D. grimshawi

Accuracy of GALBA with Different Protein Donors

Automated and
Accurate Genome
Annotation with
BRAKER, GALBA, &
TSEBRA

Katharina J. Hoff &
Natalia Nenasheva



Who we are

Gene Prediction

BRAKER

BRAKER1: RNA-Seq

BRAKER2: Proteins

TSEBRA

BRAKER3: RNA-Seq +
Proteins

Accuracy Results

Availability

GALBA

GALBA: Proteins

Accuracy Results

Availability

Annotation Quality

Genome Browsers

Descriptive Statistics

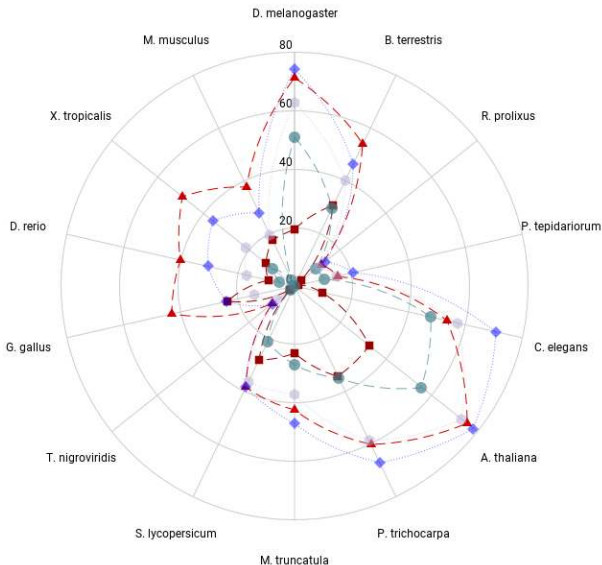
BUSCO

OMArk

Hands on Lab Session

Gene F1 of GALBA in 14 Species

■ miniprot ▲ GALBA ● GeneMark-ES ● GeneMark-EP ◆ BRAKER2



Automated and
Accurate Genome
Annotation with
BRAKER, GALBA, &
TSEBRA

Katharina J. Hoff &
Natalia Nenasheva



Who we are

Gene Prediction

BRAKER

BRAKER1: RNA-Seq

BRAKER2: Proteins

TSEBRA

BRAKER3: RNA-Seq +
Proteins

Accuracy Results

Availability

GALBA

GALBA: Proteins

Accuracy Results

Availability

Annotation Quality

Genome Browsers

Descriptive Statistics

BUSCO

OMArk

Hands on Lab Session

GALBA: Gene F1 Accuracy

Automated and
Accurate Genome
Annotation with
BRAKER, GALBA, &
TSEBRA

Katharina J. Hoff &
Natalia Nenasheva



Who we are

Gene Prediction

BRAKER

BRAKER1: RNA-Seq

BRAKER2: Proteins

TSEBRA

BRAKER3: RNA-Seq +
Proteins

Accuracy Results

Availability

GALBA

GALBA: Proteins

Accuracy Results

Availability

Annotation Quality

Genome Browsers

Descriptive Statistics

BUSCO

OMArk

Hands on Lab Session



Who we are

Gene Prediction

BRAKER

BRAKER1: RNA-Seq

BRAKER2: Proteins

TSEBRA

BRAKER3: RNA-Seq +
Proteins

Accuracy Results

Availability

GALBA

GALBA: Proteins

Accuracy Results

Availability

Annotation Quality

Genome Browsers

Descriptive Statistics

BUSCO

OMArk

Hands on Lab Session

Availability

GitHub

<https://github.com/Gaius-Augustus/GALBA>

Docker/Singularity

```
singularity build galba.sif \  
                        docker://katharinahoff/galba:latest
```

```
singularity exec galba.sif galba.pl [OPTIONS]
```

Licenses

- GALBA: Artistic License
- miniprot: MIT License

Did We Do a Good Job?

Automated and
Accurate Genome
Annotation with
BRAKER, GALBA, &
TSEBRA

Katharina J. Hoff &
Natalia Nenasheva



Who we are

Gene Prediction

BRAKER

BRAKER1: RNA-Seq

BRAKER2: Proteins

TSEBRA

BRAKER3: RNA-Seq +
Proteins

Accuracy Results

Availability

GALBA

GALBA: Proteins

Accuracy Results

Availability

Annotation Quality

Genome Browsers

Descriptive Statistics

BUSCO

OMArk

Hands on Lab Session



- UCSC Genome Browser, MakeHub
- JBrowse
- ...



Describe Your Annotation

Automated and
Accurate Genome
Annotation with
BRAKER, GALBA, &
TSEBRA

Katharina J. Hoff &
Natalia Nenasheva



Who we are

Gene Prediction

BRAKER

BRAKER1: RNA-Seq

BRAKER2: Proteins

TSEBRA

BRAKER3: RNA-Seq +
Proteins

Accuracy Results

Availability

GALBA

GALBA: Proteins

Accuracy Results

Availability

Annotation Quality

Genome Browsers

Descriptive Statistics

BUSCO

OMArk

Hands on Lab Session

- number of genes
- number of transcripts
- ratio of mono-exonic to multi-exonic genes
- median number of exons per transcript
- maximal number of exons per transcript
- median transcript length
- ...

If possible, compare to annotated close relatives.
Consider effect of individual annotation pipelines.

BUSCO: Sensitivity in Clade-Specific Conserved Genes

Automated and
Accurate Genome
Annotation with
BRAKER, GALBA, &
TSEBRA

Katharina J. Hoff &
Natalia Nenasheva



Who we are

Gene Prediction

BRAKER

BRAKER1: RNA-Seq

BRAKER2: Proteins

TSEBRA

BRAKER3: RNA-Seq +
Proteins

Accuracy Results

Availability

GALBA

GALBA: Proteins

Accuracy Results

Availability

Annotation Quality

Genome Browsers

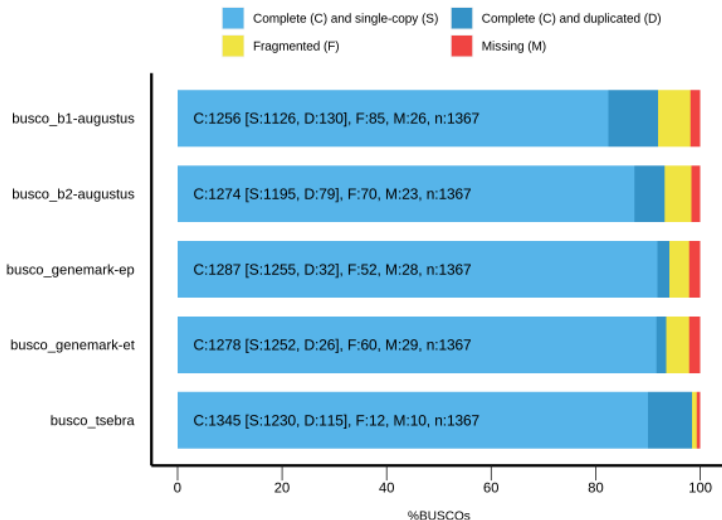
Descriptive Statistics

BUSCO

OMark

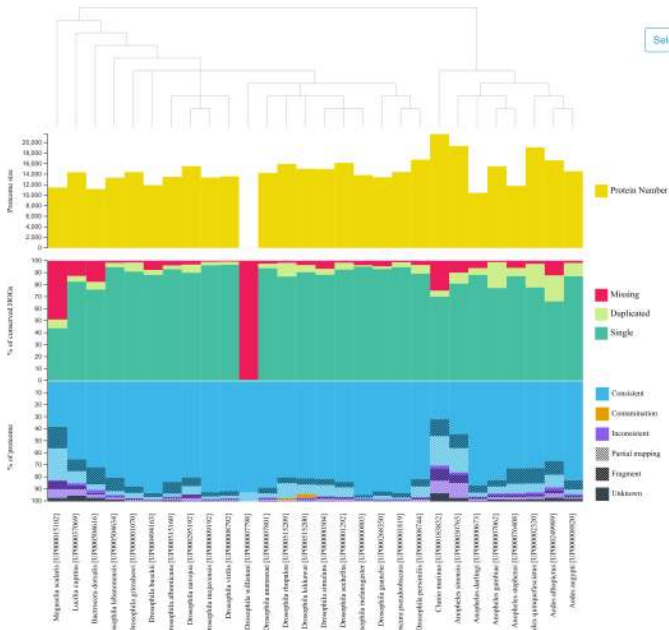
Hands on Lab Session

BUSCO Assessment Results



OMark: Sensitivity, Contaminations, & More

Select Taxon ▾





Who we are

Gene Prediction

BRAKER

BRAKER1: RNA-Seq

BRAKER2: Proteins

TSEBRA

BRAKER3: RNA-Seq +
Proteins

Accuracy Results

Availability

GALBA

GALBA: Proteins

Accuracy Results

Availability

Annotation Quality

Genome Browsers

Descriptive Statistics

BUSCO

OMArk

Hands on Lab Session

Hands on Lab Session

Organization of Session

- 1 start JupyterLab with Singularity
- 2 familiarize yourself with JupyterNotebooks (modify & run cells)
- 3 we walk you through GenomeAnnotation.ipynb
- 4 Annotate_Babesia_duncani.ipynb is for advanced learners, organized as group work

Getting Started

Go to

https://github.com/KatharinaHoff/GenomeAnnotation_Workshop2023 and follow instructions





Who we are

Gene Prediction

BRAKER

BRAKER1: RNA-Seq

BRAKER2: Proteins

TSEBRA

BRAKER3: RNA-Seq +
Proteins

Accuracy Results

Availability

GALBA

GALBA: Proteins

Accuracy Results

Availability

Annotation Quality

Genome Browsers

Descriptive Statistics

BUSCO

OMArk

Hands on Lab Session

- Research on BRAKER is supported by US National Institutes of Health grant GM128145 to M.B. and M.S.
- Research on GALBA is supported by German Research Foundation grant 277249973 to K.J.H., and by Project Data Competency granted to K.J.H. and M.S. by the government of Mecklenburg-Vorpommern, and by US National Institute of Health grant R01HG010040 to H.L., and by German Research Foundation grant 391397397 to S.H. and M.S.