

Tools for genome annotation SAPP/GBOL/Empusa

Jasper Koehorst
jasper.koehorst@wur.nl

Laboratory of Systems and Synthetic Biology
Fairbydesign.nl

POSTER + DEMO 1 & 2



WAGENINGEN
UNIVERSITY & RESEARCH

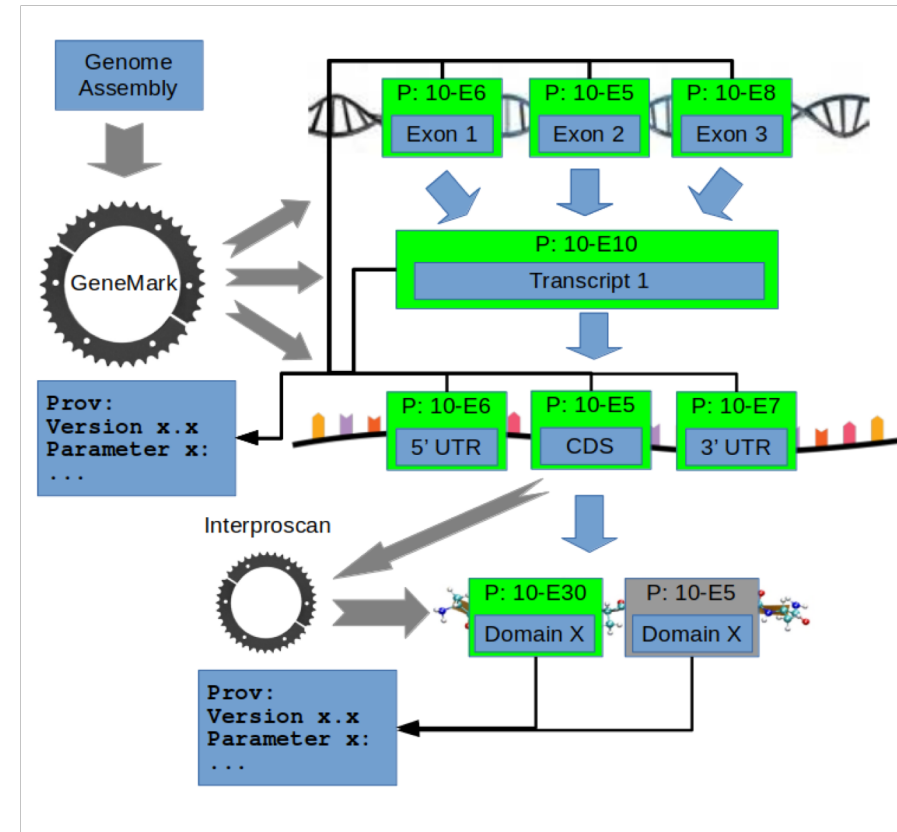


FAIR Genome Annotation

Example questions e.g:

- What (genes) distinguish species that have a desired trait?
- Which enzymes are there that can catalyze reaction X (maybe with different cofactors?)

Consistently annotated genomes that can be mined



WAGENINGEN
UNIVERSITY & RESEARCH

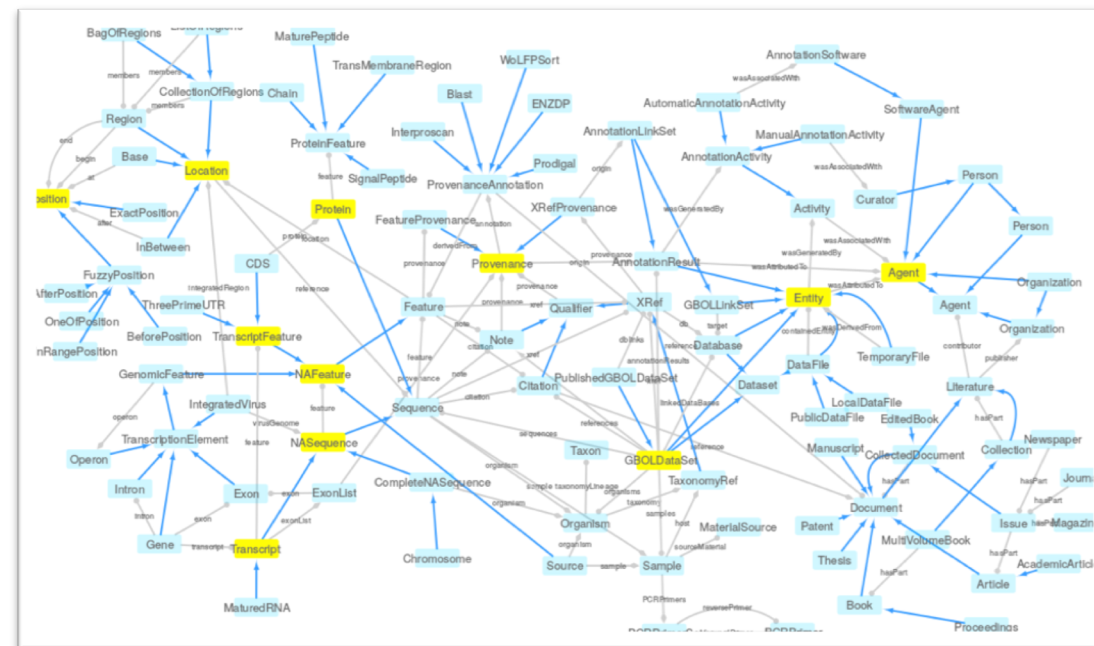
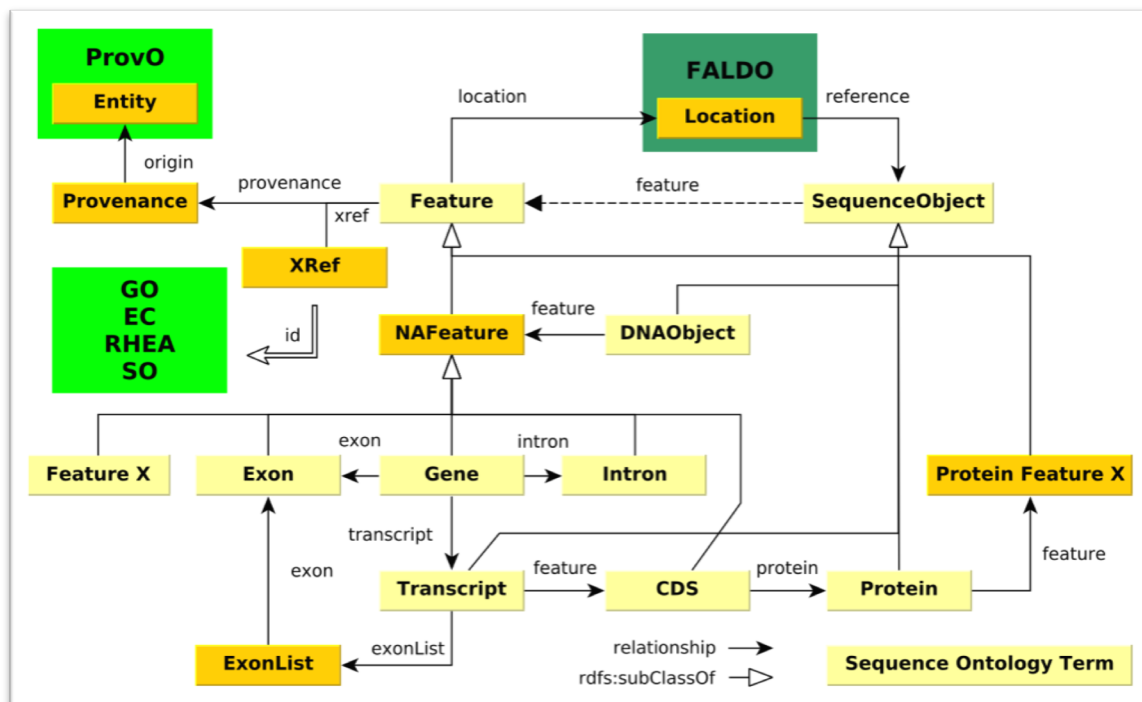


Requirements for genome mining

- A semantic annotation platform that incorporates common tools and stores the results in “proper” format. **SAPP**
- A definition of the “proper format”: definitions of biological terms and their relationships: **GBOL ontology**
- Interface to use the ontology: **GBOL stack**
- Tools to develop all of these: **Empusa**: code generator

GBOL: Genome Biology Ontology Language

Sub domain	Classes	Properties
Genomic locations	16	17
Genes		
transcripts and features	114	133
Document structure	27	107
Dataset-wise provenance	22	54
Element-wise provenance	5	9
BIBO	59	90

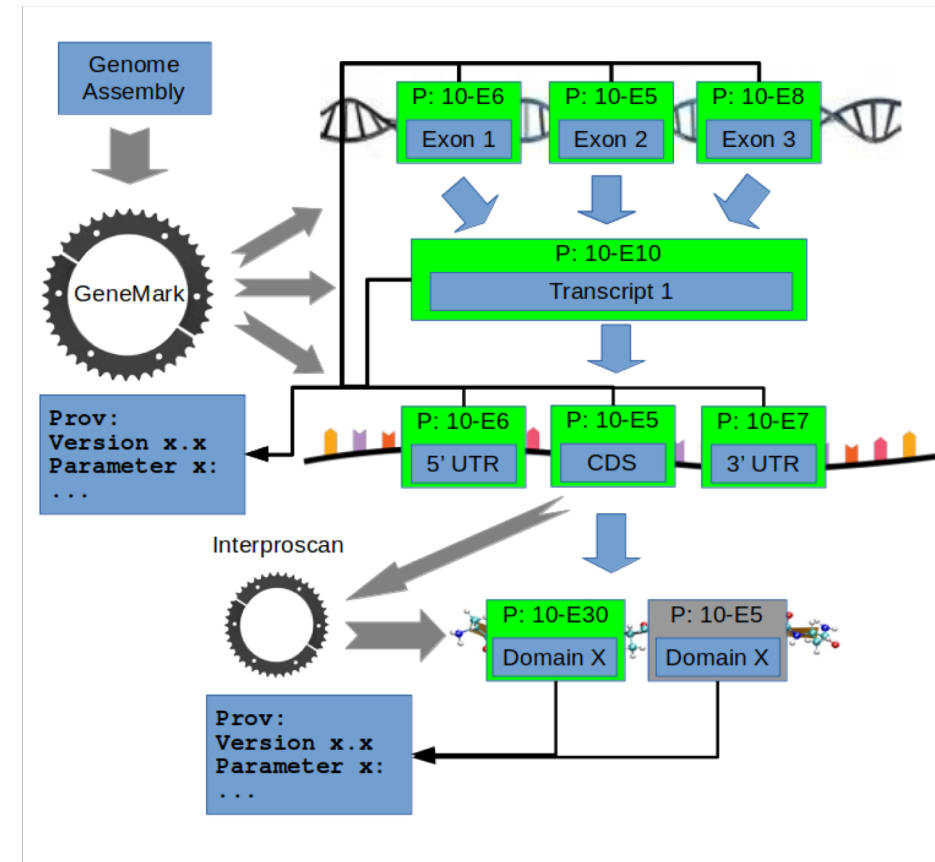


Embedded with existing ontologies.

Van Dam et al. Journal of biomedical semantics 2015

SAPP: Annotation information storage

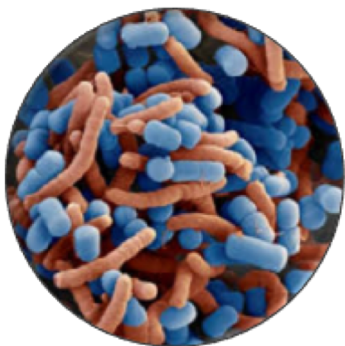
- Wrapper to commonly used annotation tools (prokaryotes and eukaryotes) that generates FAIR data
- Examples:
 - Uniform annotation of over 80 000 bacterial species.
 - Uniform annotation of eukaryotes e.g. (human, fish, insects, plants, ...)



Modular design of SAPP

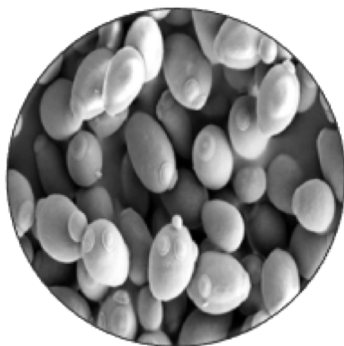
Conversion types

- EMBL / GenBank
- FASTA
- GFF
- QTL
- VCF
- ...



Genetic elements

- Gene prediction
- tRNA/rRNA
- Crispr
- ...



Functional annotation

- BLAST
- Enzyme predictions
- Domain annotation
- Signal peptides
- Transmembrane
- Localization
- ...



WAGENINGEN
UNIVERSITY & RESEARCH

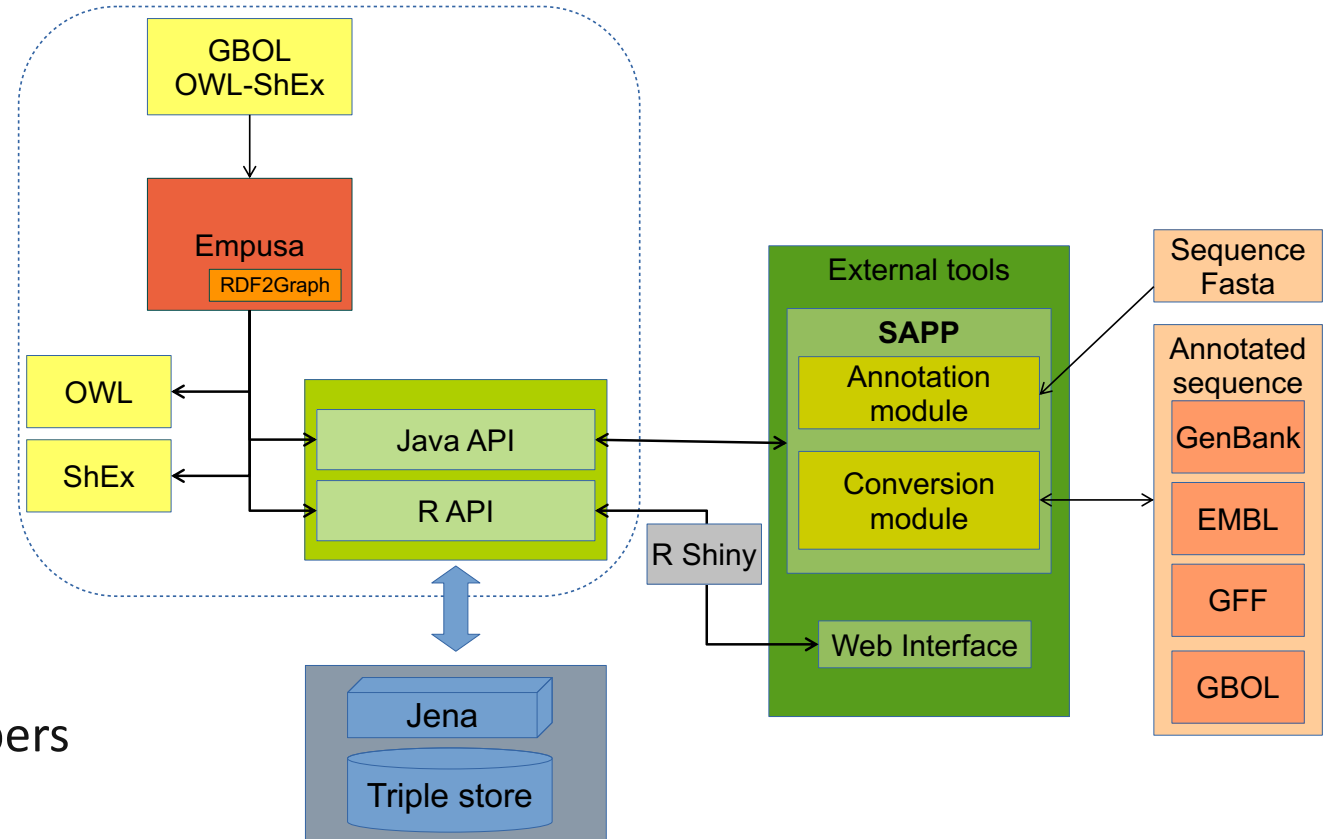


Tool development for FAIR genome annotation

- **SAPP**: an annotation platform

Developer:

- **GBOL stack:**
 - GBOL ontology (backbone)
 - Java/R Api
 - Owl/ShEx
 - Interface gate keeper
- Code generator: **Empusa** useful for developers



Availability

- **SAPP** Koehorst et al Bioinformatics 2017
<https://sapp.gitlab.io>
- **Empusa**: <https://gitlab.com/Empusa>
- **GBOL**: Documentation & namespace:
<http://gbol.life/0.1/>

SAPP: functional genome annotation and analysis through a semantic framework using FAIR principles

Jasper J Koehorst , Jesse C J van Dam, Edoardo Saccenti, Vitor A P Martins dos Santos, Maria Suarez-Diez, Peter J Schaap 

Bioinformatics, Volume 34, Issue 8, 15 April 2018, Pages 1401–1403,
<https://doi.org/10.1093/bioinformatics/btx767>

Published: 23 November 2017 **Article history** ▼



bioRxiv

THE PREPRINT SERVER FOR BIOLOGY

HOME

Search

New Results

Interoperable genome annotation with GBOL, an extendable infrastructure for functional data mining

 Jesse C.J. van Dam,  Jasper Jan J. Koehorst,  Jon Olav Vik,  Peter J. Schaap,  Maria Suarez-Diez

doi: <https://doi.org/10.1101/184747>



WAGENINGEN
UNIVERSITY & RESEARCH

