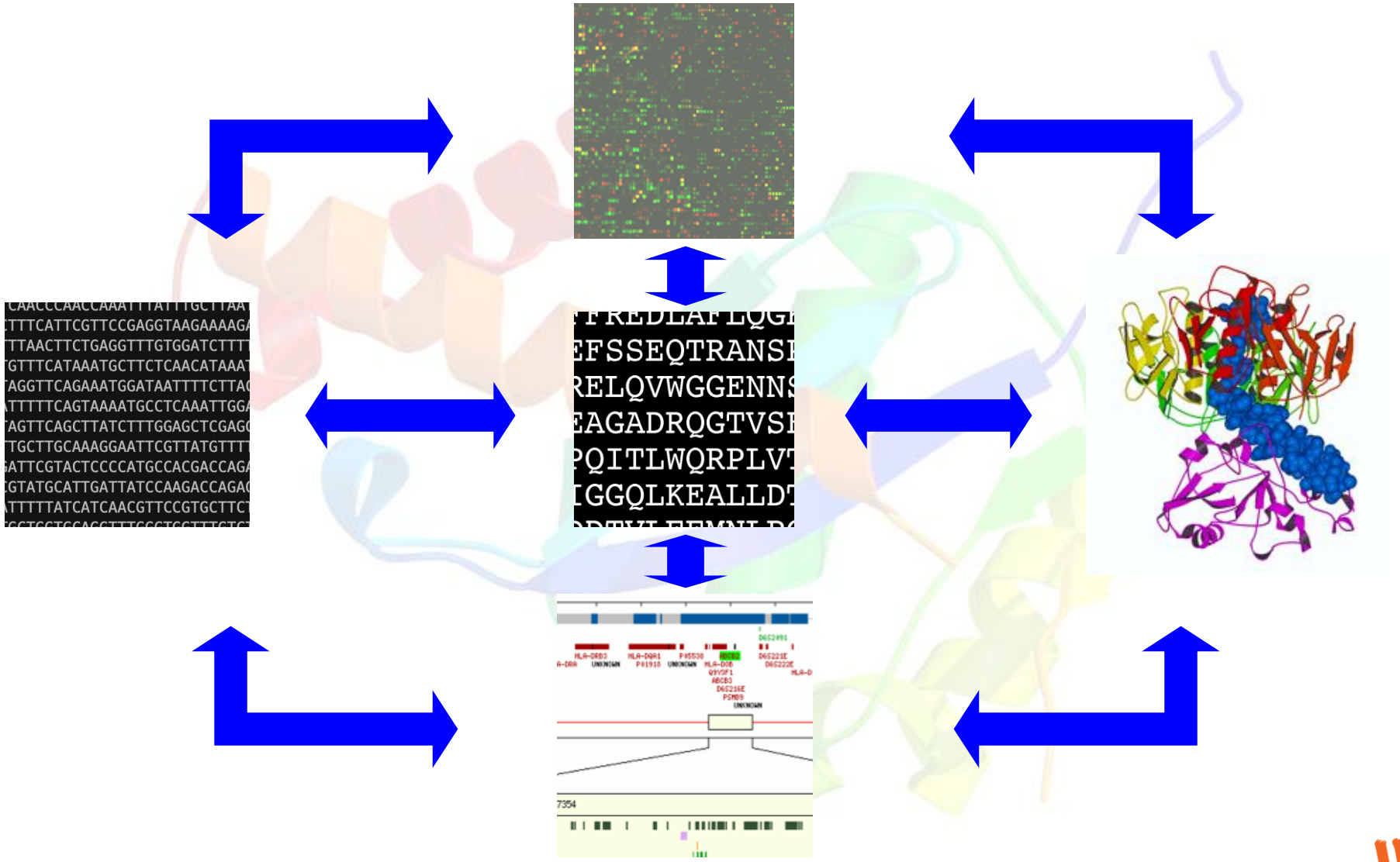
Bioinformatics Databases

Michael Landi & Andreas Gisel

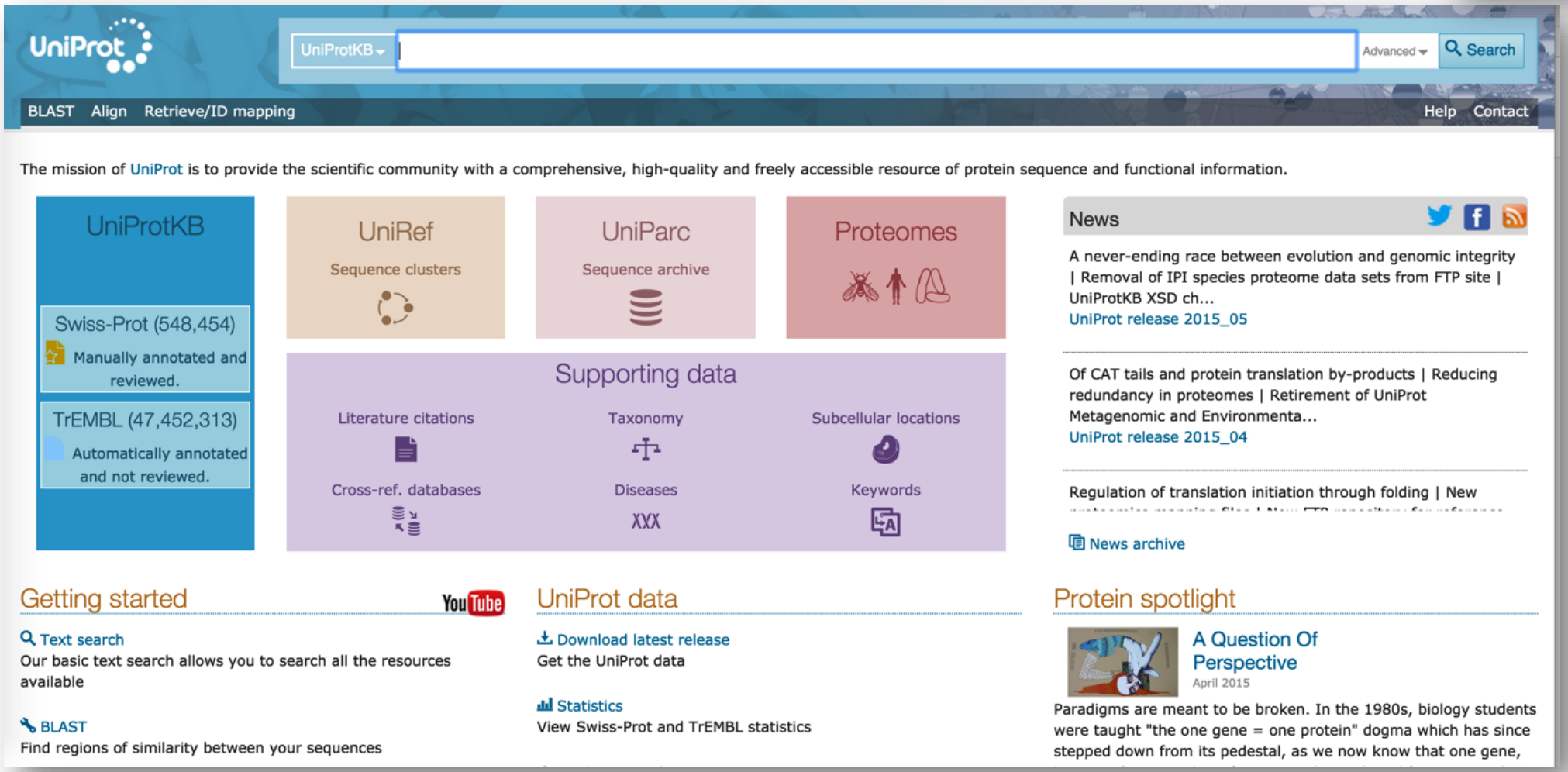
IITA- Bioinformatics
Kenya & Nigeria

African Star Apple Assembly workshop 2 – 6 June 2025

Bioinformatics Databases

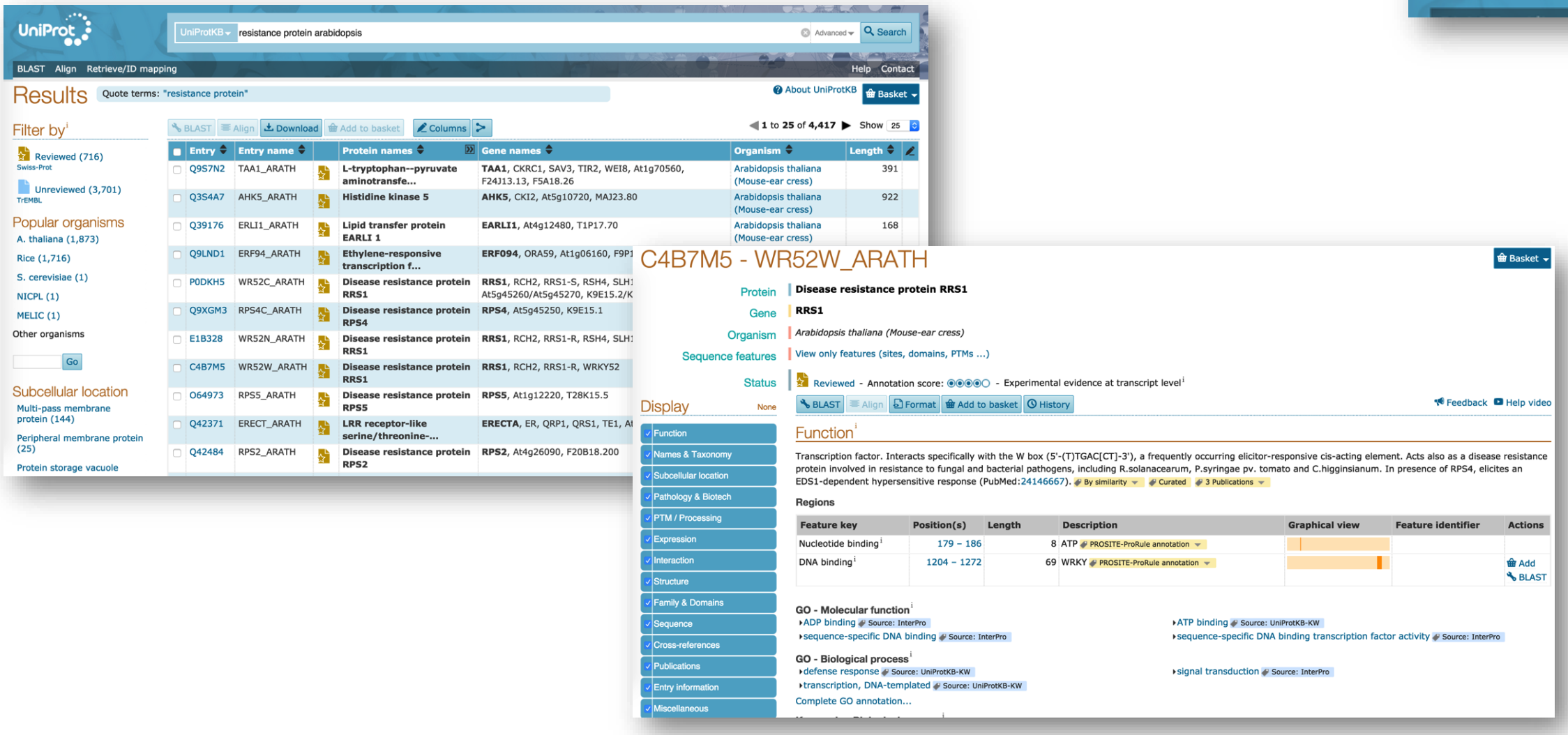


Bioinformatics Databases



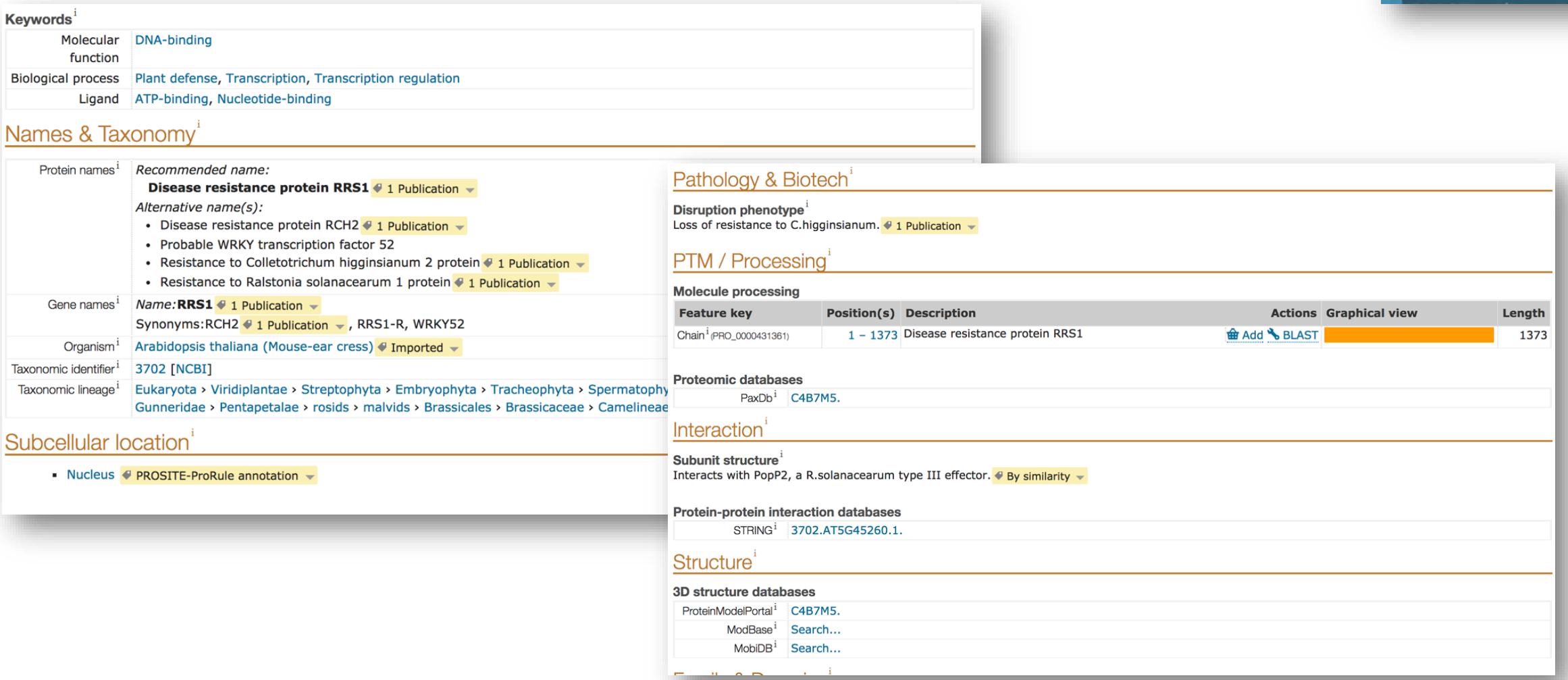
IITA Transforming African Agriculture CGIAR

Bioinformatics Databases



IITA Transforming African Agriculture CGIAR

Bioinformatics Databases



Keywords

Molecular function	DNA-binding
Biological process	Plant defense, Transcription, Transcription regulation
Ligand	ATP-binding, Nucleotide-binding

Names & Taxonomy

Protein names	*Recommended name:* **Disease resistance protein RRS1** 1 Publication *Alternative name(s):* • Disease resistance protein RCH2 1 Publication • Probable WRKY transcription factor 52 • Resistance to Colletotrichum higginsianum 2 protein 1 Publication • Resistance to Ralstonia solanacearum 1 protein 1 Publication
Gene names	*Name:***RRS1** 1 Publication Synonyms:RCH2 1 Publication, RRS1-R, WRKY52
Organism	Arabidopsis thaliana (Mouse-ear cress) Imported
Taxonomic identifier	3702 [NCBI]
Taxonomic lineage	Eukaryota › Viridiplantae › Streptophyta › Embryophyta › Tracheophyta › Spermatophy Gunneridae › Pentapetalae › rosids › malvids › Brassicales › Brassicaceae › Camelineae

Subcellular location

- Nucleus PROSITE-ProRule annotation

Pathology & Biotech

Disruption phenotype
Loss of resistance to C.higginsianum. 1 Publication

PTM / Processing

Molecule processing

Feature key	Position(s)	Description	Actions	Graphical view	Length
Chain (PRO_0000431361)	1 – 1373	Disease resistance protein RRS1	Add BLAST		1373

Proteomic databases

PaxDb	C4B7M5.

Interaction

Subunit structure
Interacts with PopP2, a R.solanacearum type III effector. By similarity

Protein-protein interaction databases

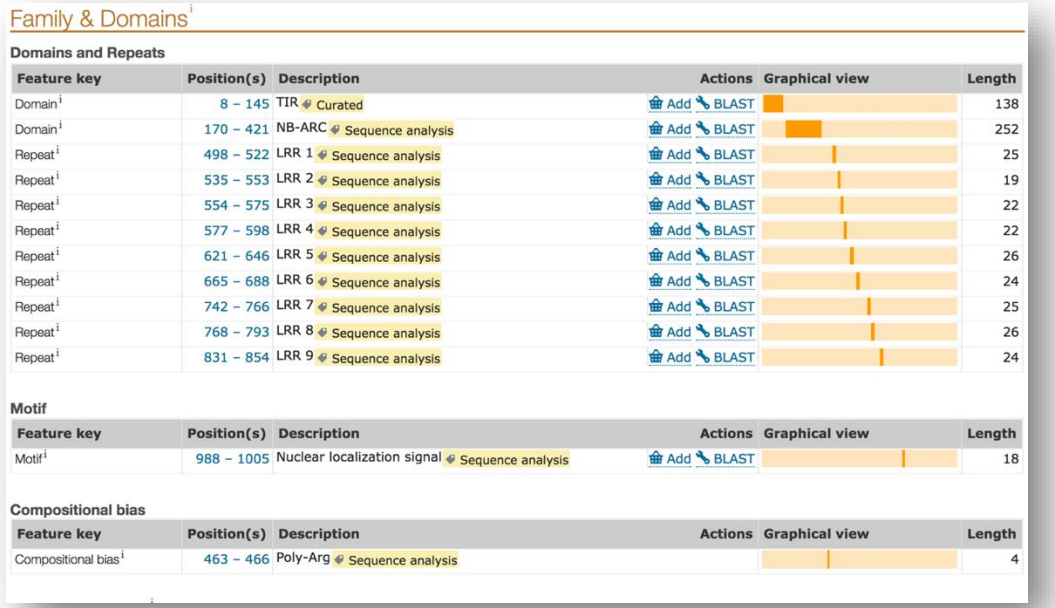
STRING	3702.AT5G45260.1.

Structure

3D structure databases

ProteinModelPortal	C4B7M5.
ModBase	Search...
MobiDB	Search...

Bioinformatics Databases



Family & Domains

Domains and Repeats

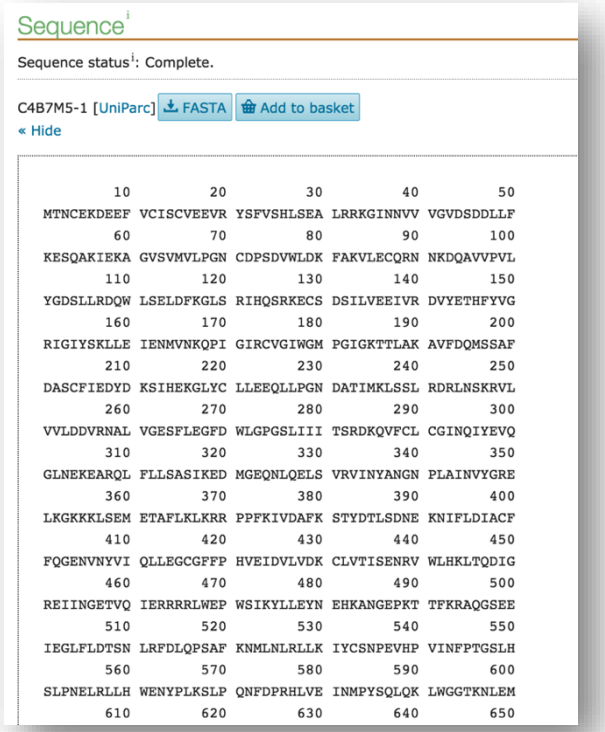
Feature key	Position(s)	Description	Actions	Graphical view	Length
Domain	8 – 145	TIR Curated	Add BLAST		138
Domain	170 – 421	NB-ARC Sequence analysis	Add BLAST		252
Repeat	498 – 522	LRR 1 Sequence analysis	Add BLAST		25
Repeat	535 – 553	LRR 2 Sequence analysis	Add BLAST		19
Repeat	554 – 575	LRR 3 Sequence analysis	Add BLAST		22
Repeat	577 – 598	LRR 4 Sequence analysis	Add BLAST		22
Repeat	621 – 646	LRR 5 Sequence analysis	Add BLAST		26
Repeat	665 – 688	LRR 6 Sequence analysis	Add BLAST		24
Repeat	742 – 766	LRR 7 Sequence analysis	Add BLAST		25
Repeat	768 – 793	LRR 8 Sequence analysis	Add BLAST		26
Repeat	831 – 854	LRR 9 Sequence analysis	Add BLAST		24

Motif

Feature key	Position(s)	Description	Actions	Graphical view	Length
Motif	988 – 1005	Nuclear localization signal Sequence analysis	Add BLAST		18

Compositional bias

Feature key	Position(s)	Description	Actions	Graphical view	Length
Compositional bias	463 – 466	Poly-Arg Sequence analysis			4



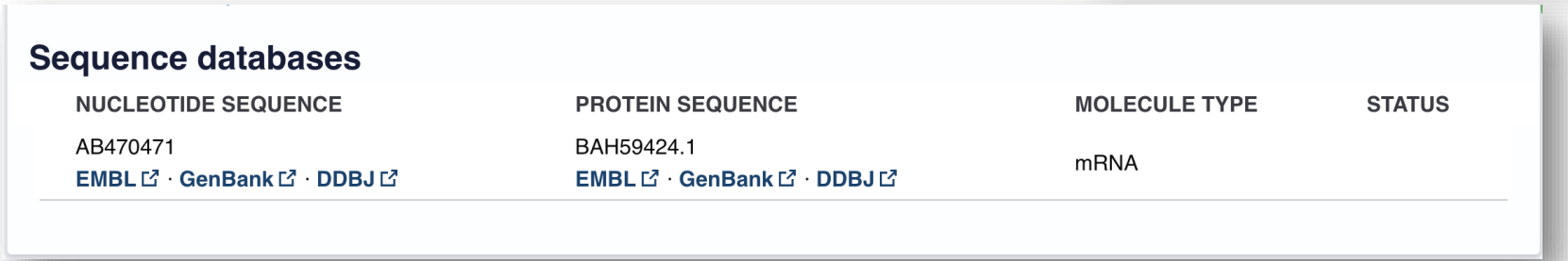
Sequence

Sequence status: Complete.

C4B7M5-1 [UniParc] FASTA Add to basket

« Hide

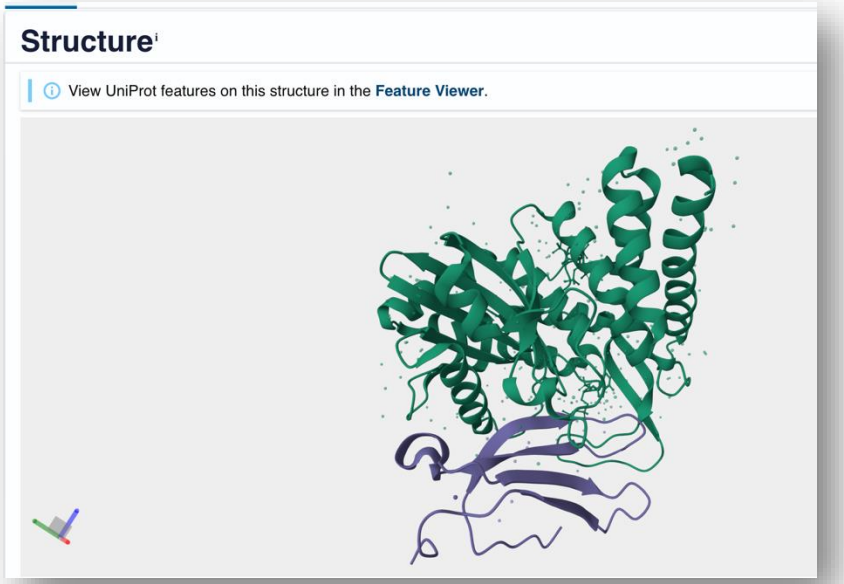
```
        10         20         30         40         50
MTNCEKDEEF VCISCVEEVR YSFVSHLSEA LRRKGINNVV VGVDSDDLLF
        60         70         80         90        100
KESQAKIEKA GVSVMVLPGN CDPSDVWLDK FAKVLECQRN NKDQAVVPVL
       110        120        130        140        150
YGDSLLRDQW LSELDFKGLS RIHQSRKECS DSILVEEIVR DVYETHFYVG
       160        170        180        190        200
RIGIYSKLLE IENMVNKQPI GIRCVGIWGM PGIGKTTLAK AVFDQMSSAF
       210        220        230        240        250
DASCFIEDYD KSIHEKGLYC LLEEQLLPGN DATIMKLSSL RDRLNSKRVL
       260        270        280        290        300
VVLDDVRNAL VGESFLEGFD WLGPGSLIII TSRDKQVFCL CGINQIYEVQ
       310        320        330        340        350
GLNEKEARQL FLLSASIKED MGEQNLQELS VRVINYANGN PLAINVYGRE
       360        370        380        390        400
LKGKKKLSEM ETAFLKLKRR PPFKIVDAFK STYDTLSDNE KNIFLDIACF
       410        420        430        440        450
FQGENVNYVI QLLEGCGFFP HVEIDVLVDK CLVTISENRV WLHKLTQDIG
       460        470        480        490        500
REIINGETVQ IERRRRLWEP WSIKYLLEYN EHKANGEPKT TFKRAQGSEE
       510        520        530        540        550
IEGLFLDTSN LRFDLQPSAF KNMLNLRLLK IYCSNPEVHP VINFPTGSLH
       560        570        580        590        600
SLPNELRLLH WENYPLKSLP QNFDPRHLVE INMPYSQLQK LWGGTKNLEM
       610        620        630        640        650
```

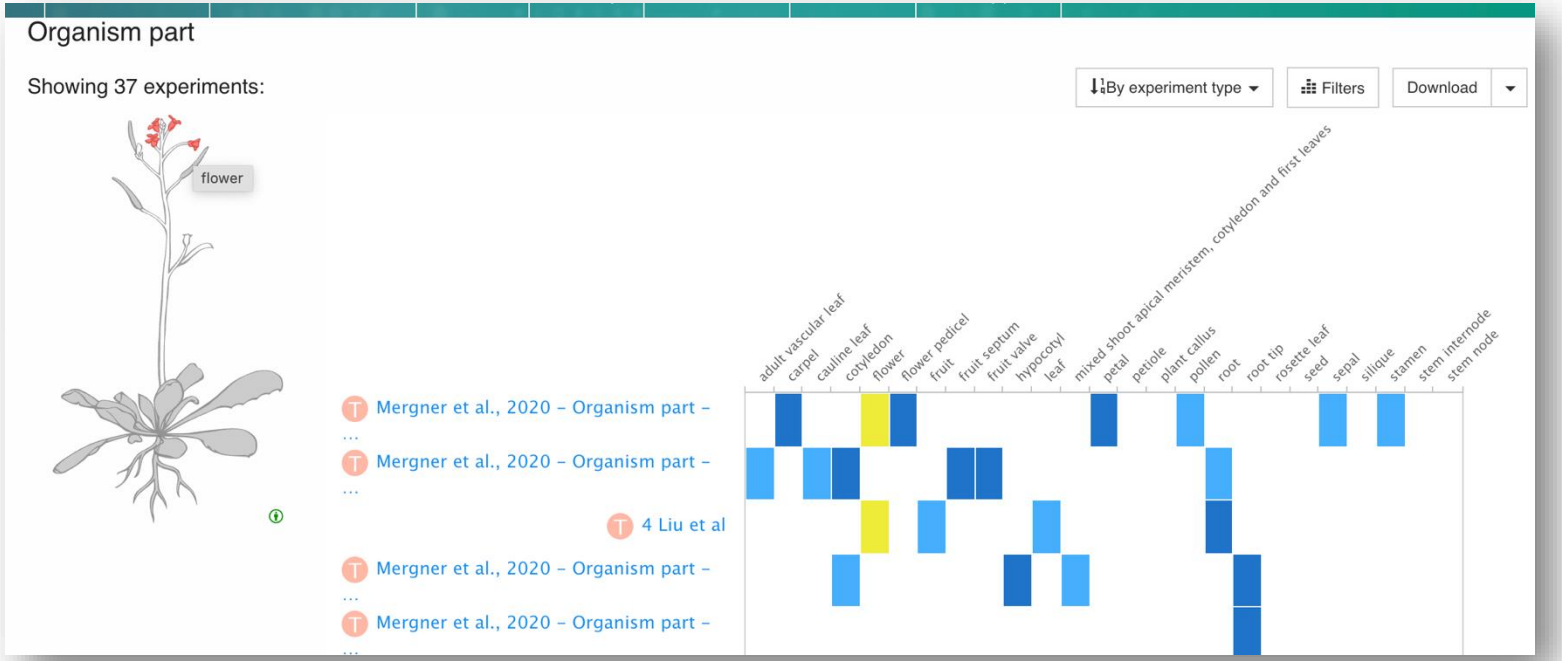


Sequence databases

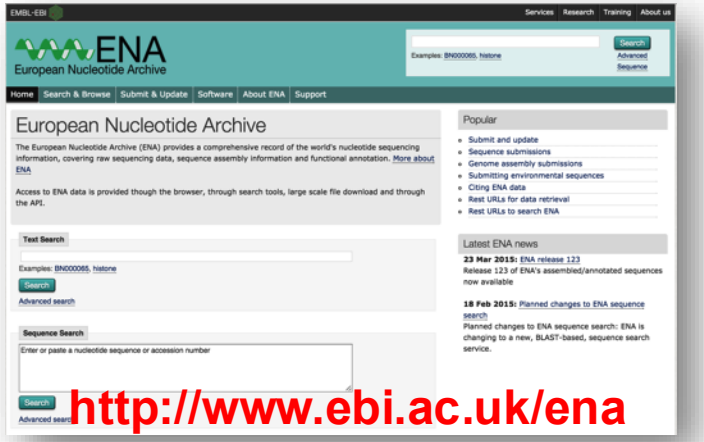
NUCLEOTIDE SEQUENCE	PROTEIN SEQUENCE	MOLECULE TYPE	STATUS
AB470471 EMBL · GenBank · DDBJ	BAH59424.1 EMBL · GenBank · DDBJ	mRNA	

Bioinformatics Databases

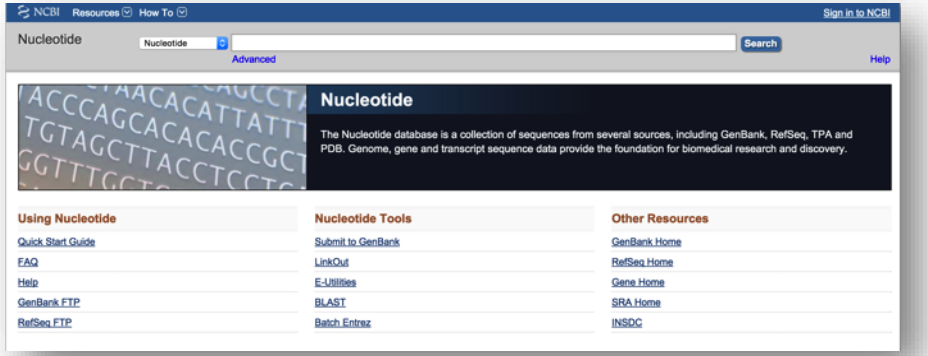




Bioinformatics Databases



http://www.ebi.ac.uk/ena



http://www.ncbi.nlm.nih.gov/nucleotide/

http://www.ddbj.nig.ac.jp/

Bioinformatics Databases

OrthoDB

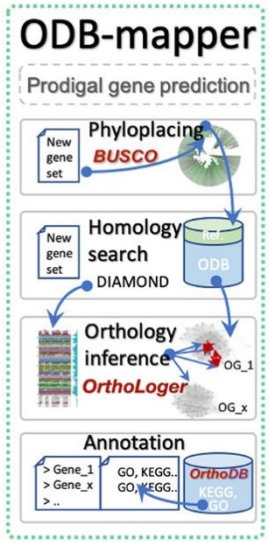
the hierarchical catalogue of orthologs

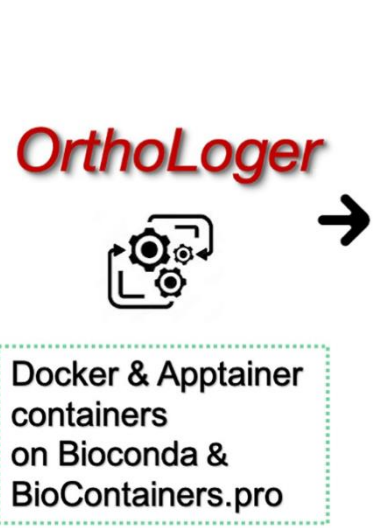
IITA Transforming African Agriculture CGIAR

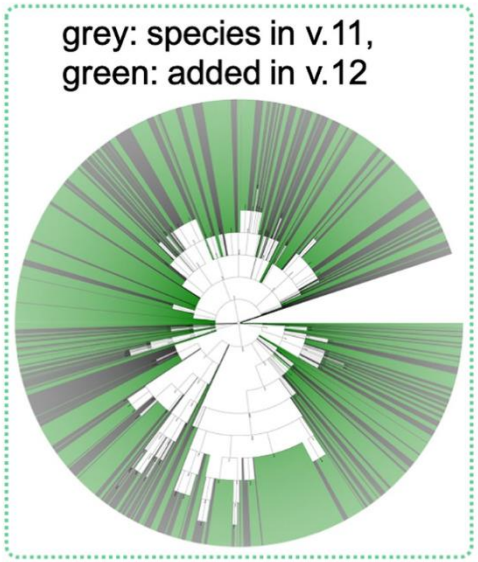
Bioinformatics Databases

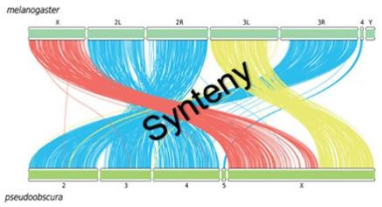
OrthoDB

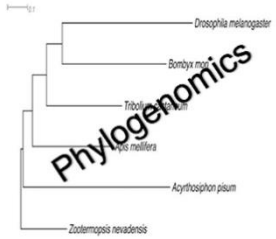
OrthoDB graphical abstract

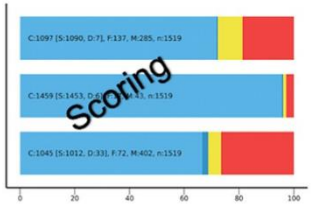








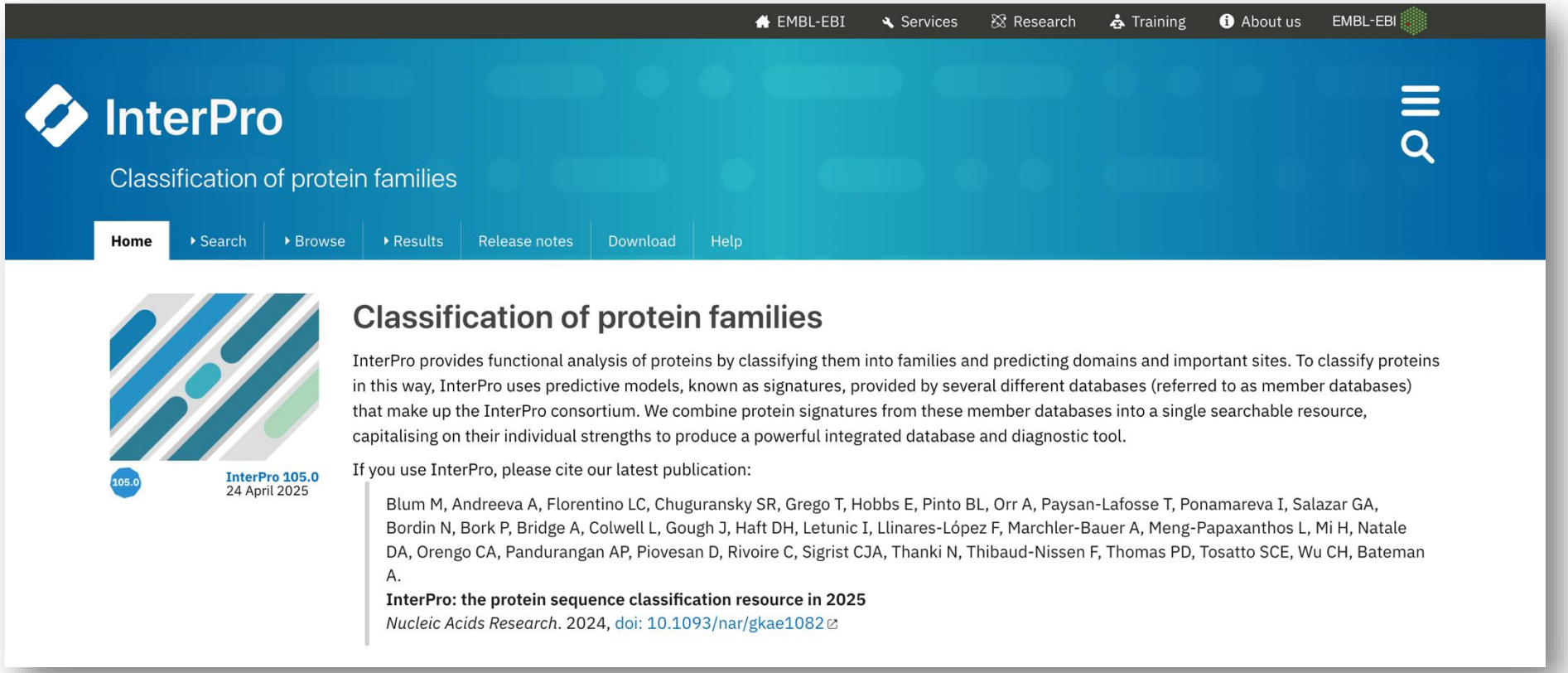




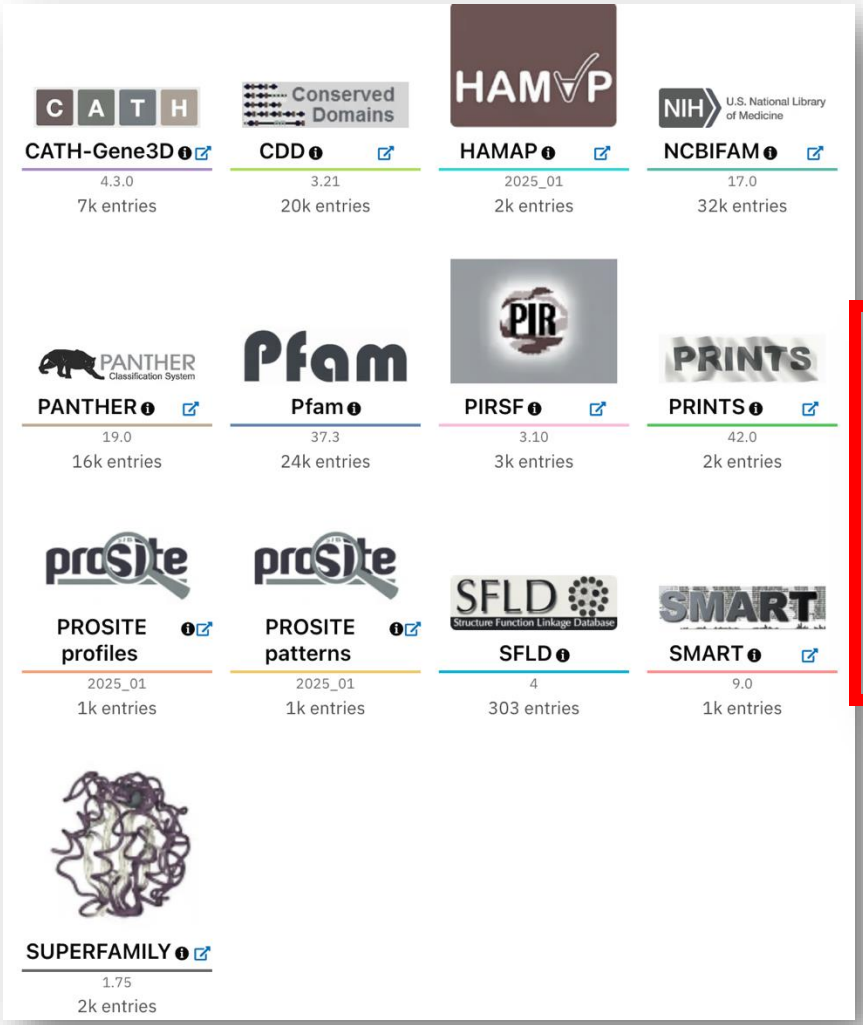
	v10	v12 update
Eukaryota	67	109
Bacteria	83	332
Archaea	16	29

IITA Transforming African Agriculture CGIAR

Bioinformatics Databases



Bioinformatics Databases



InterPro

InterProScan

InterProScan is the software package that allows sequences to be scanned against InterPro's member database signatures.

Users who have novel nucleotide or protein sequences that they wish to functionally characterise can use InterProScan to run the scanning algorithms against the InterPro database in an integrated way.

Download the latest version (5.74-105.0)

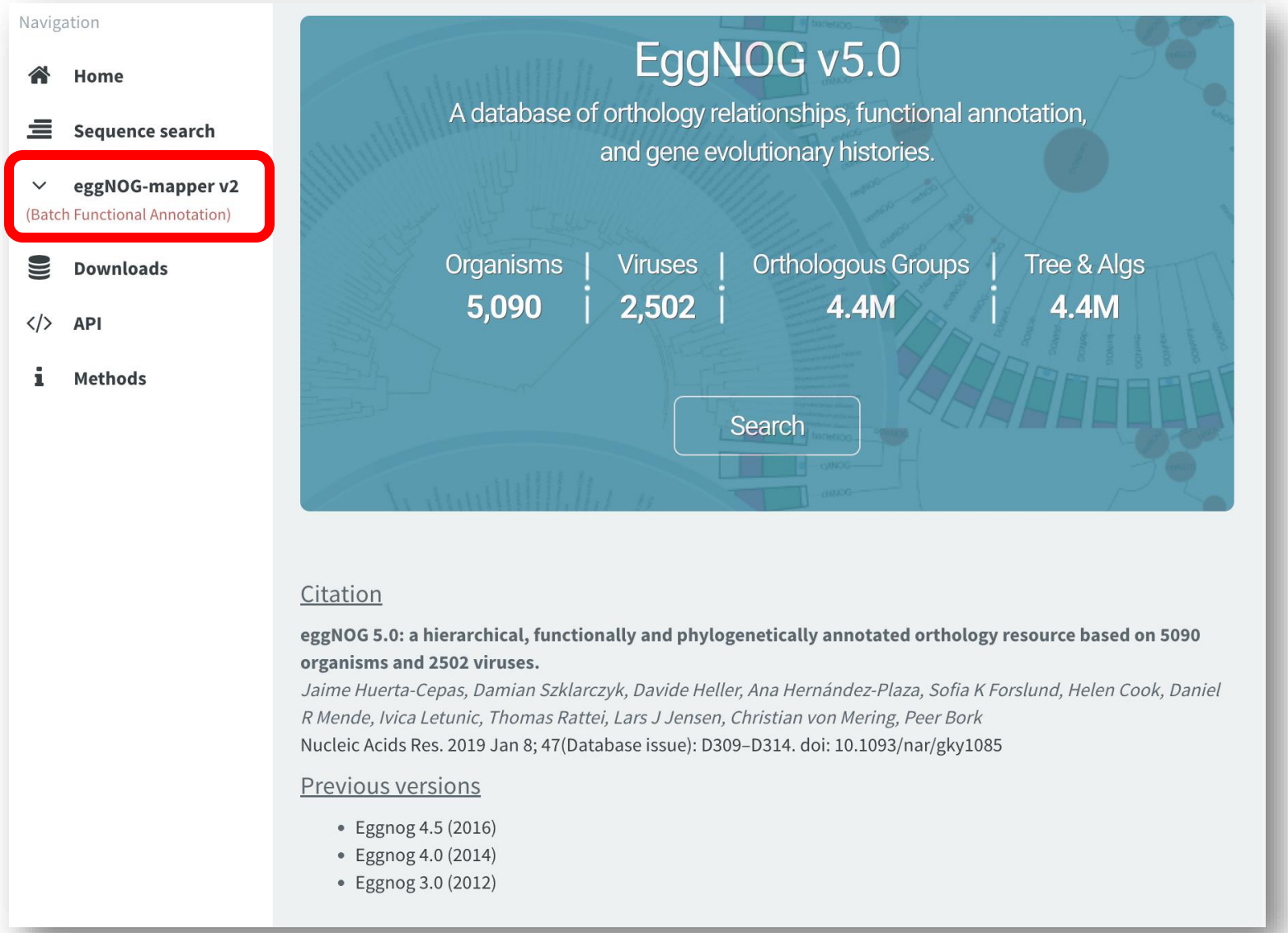
Linux (64-bit) - System requirements

- There are no versions planned for Windows or Apple operating systems. This is due to constraints in the various third-party binaries that InterProScan runs.
- Older versions of InterProScan are not supported anymore. We highly recommend you to update to the latest version.

IITA Transforming African Agriculture CGIAR

Bioinformatics Databases

EggNOG 5.0.0

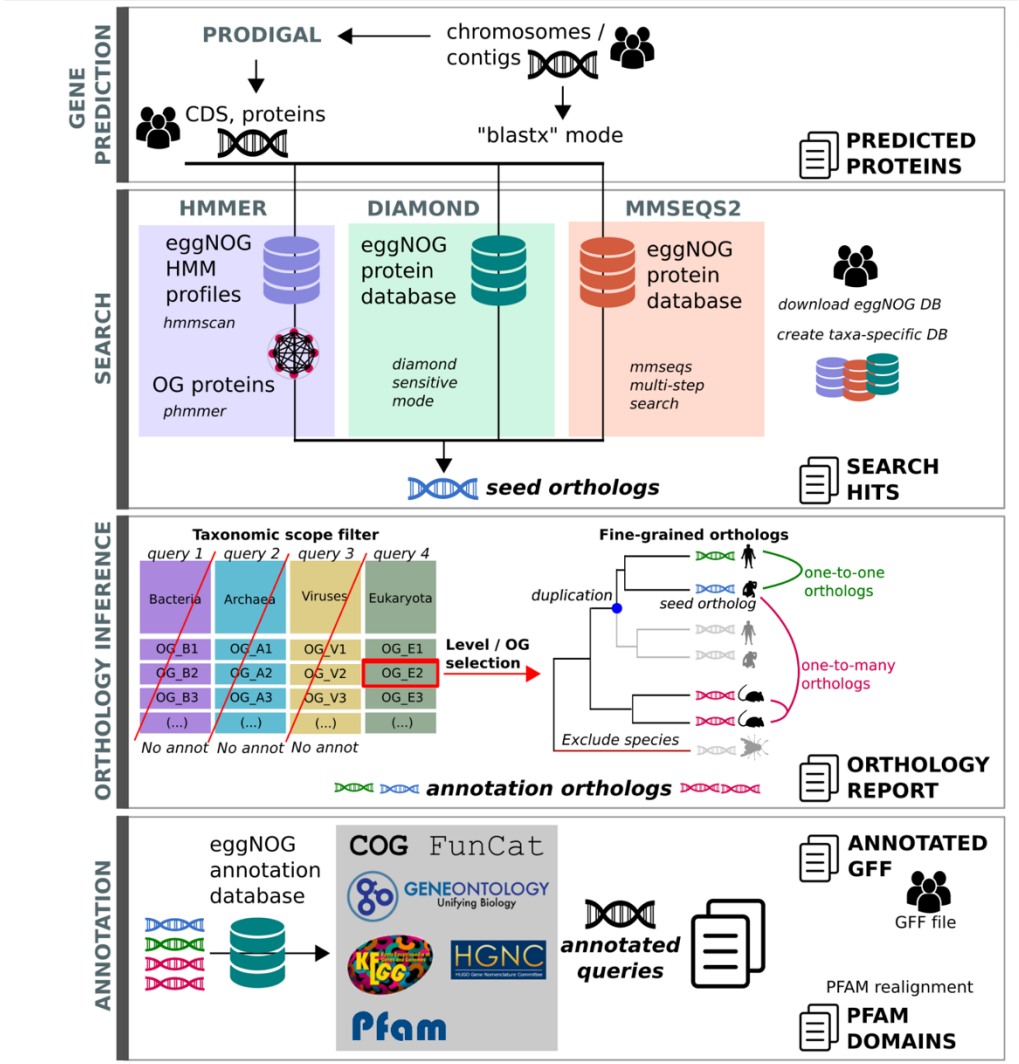


IITA Transforming African Agriculture CGIAR

Bioinformatics Databases

EggNOG 5.0.0

eggNOG-mapper v2 is a tool for functional annotation of large sets of sequences based on fast orthology assignments using precomputed **eggNOG v5.0** clusters and phylogenies.



IITA Transforming African Agriculture CGIAR

Bioinformatics Databases

Bioinformatics Databases

Bioinformatics Databases