SequenceServer past, present and future

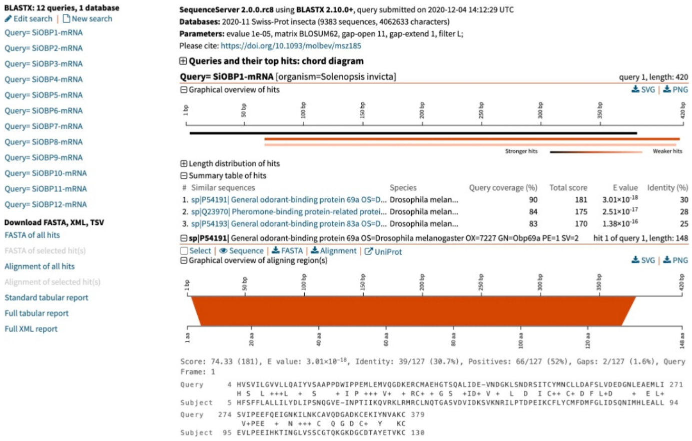
Tadas Tamošauskas & Yannick Wurm & many others
Pragmatic Genomics & Queen Mary University of London

Biologists ❤ SequenceServer BLAST

Open-source core aims to be the world's best BLAST interface

- Intuitive browser-based graphical user interface.
- "AI" interface that reduces user error.
- Visualizations designed for faster better interpretation.
- For individuals or teams.
- Easy to set up by Docker or ruby gem.

Intuitive BLAST report with plenty of export & integration flexibility



BLASTX: 12 queries, 1 database
Edit search | New search
Query= SiOBP1-mRNA
Query= SiOBP2-mRNA
Query= SiOBP3-mRNA
Query= SiOBP4-mRNA
Query= SiOBP5-mRNA
Query= SiOBP6-mRNA
Query= SiOBP7-mRNA
Query= SiOBP8-mRNA
Query= SiOBP9-mRNA
Query= SiOBP10-mRNA
Query= SiOBP11-mRNA
Query= SiOBP12-mRNA

Download FASTA, XML, TSV
FASTA of all hits
FASTA of selected hit(s)
Alignment of all hits
Alignment of selected hit(s)
Standard tabular report
Full tabular report
Full XML report

SequenceServer 2.0.0.rc8 using **BLASTX 2.10.0+**, query submitted on 2020-12-04 14:12:29 UTC
Databases: 2020-11 Swiss-Prot insecta (9383 sequences, 4062633 characters)
Parameters: evalue 1e-05, matrix BLOSUM62, gap-open 11, gap-extend 1, filter L;
Please cite: https://doi.org/10.1093/molbev/msz185

⊞ **Queries and their top hits: chord diagram**

Query= SiOBP1-mRNA [organism=Solenopsis invicta] query 1, length: 420

⊟ Graphical overview of hits

⊞ Length distribution of hits

⊟ Summary table of hits

#	Similar sequences	Species	Query coverage (%)	Total score	E value	Identity (%)
1.	sp|P54191| General odorant-binding protein 69a OS=D...	Drosophila melan...	90	181	3.01×10^{-18}	30
2.	sp|Q23970| Pheromone-binding protein-related protei...	Drosophila melan...	84	175	2.51×10^{-17}	28
3.	sp|P54193| General odorant-binding protein 83a OS=D...	Drosophila melan...	83	170	1.38×10^{-16}	25

⊟ **sp|P54191|** General odorant-binding protein 69a OS=Drosophila melanogaster OX=7227 GN=Obp69a PE=1 SV=2 hit 1 of query 1, length: 148

☐ Select | Sequence | FASTA | Alignment | UniProt

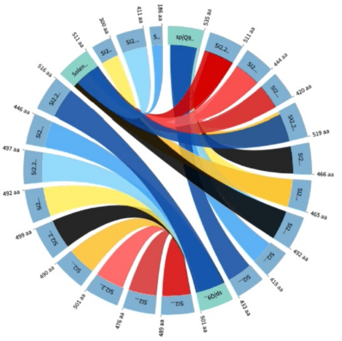
⊟ Graphical overview of aligning region(s)

```
Score: 74.33 (181), E value: 3.01×10-18, Identity: 39/127 (30.7%), Positives: 66/127 (52%), Gaps: 2/127 (1.6%), Query
Frame: 1

Query     4 HVSVILGVVLLQAIYVSAAPPDWIPPEMLEMVQGDKERCMAEHGTSQALIDE-VNDGKLSNDRSITCYMNCLLDAFSLVDEDGNLEAEMLI 271
            H S  L +++L  +  S     + I P +++ V+  + RC+ + G S  +ID+ V +  L  D  I C++ C+ D F L+D    +  E L+
Subject   5 HFSFFLALLILYDLIPSNQGVE-INPTIIKQVRKLRMRCLNQTGASVDVIDKSVKNRILPTDPEIKCFLYCMFDMFGLIDSQNIMHLEALL 94

Query   274 SVIPEEFQEIGNKILNKCAVQDGADKCEKIYNVAKC 379
             V+PEE  +  N +++ C  Q G D C+  Y   KC
Subject  95 EVLPEEIHKTINGLVSSCGTQKGKDGCDTAYETVKC 130
```

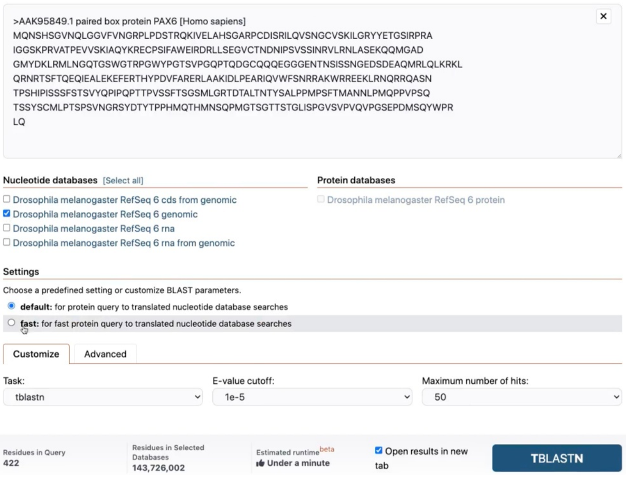
Circos-style chord diagrams of all queries & hits



Over past two years, we've dramatically improved:

- Handling of large results in browser
- Added automatic FASTQ to FASTA conversion
- Added an API with BLASTing from python or R.
- Modernised the look.
- Eliminated many bugs.
- Security.
- Tech stack (ruby 2-> 3; modern react; fewer & newer JS dependencies; reducing jQuery; migrating to Tailwind...).

"Automagic" protein vs nucleotide, FASTA vs FASTQ detection & algorithm selection



```
>AAK95849.1 paired box protein PAX6 [Homo sapiens]
MQNSHSGVNQLGGVFVNGRPLPDSTRQKIVELAHSGARPCDISRILQVSNGCVSKILGRYYETGSIRPRA
IGGSKPRVATPEVVSKIAQYKRECPSIFAWEIRDRLLSEGVCTNDNIPSVSSINRVLRNLASEKQQMGAD
GMYDKLRMLNGQTGSWGTRPGWYPGTSVPGQPTQDGCQQQEGGGENTNSISSNGEDSDEAQMRLQLKRKL
QRNRTSFTQEQIEALEKEFERTHYPDVFARERLAAKIDLPEARIQVWFSNRRAKWRREEKLRNQRRQASN
TPSHIPISSSFSTSVYQPIPQPTTPVSSFTSGSMLGRTDTALTNTYSALPPMPSFTMANNLPMQPPVPSQ
TSSYSCMLPTSPSVNGRSYDTYTPPHMQTHMNSQPMGTSGTTSTGLISPGVSVPVQVPGSEPDMSQYWPR
LQ
```

Nucleotide databases [Select all]
☐ Drosophila melanogaster RefSeq 6 cds from genomic
☑ Drosophila melanogaster RefSeq 6 genomic
☐ Drosophila melanogaster RefSeq 6 rna
☐ Drosophila melanogaster RefSeq 6 rna from genomic

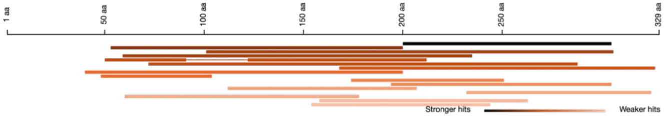
Protein databases
☐ Drosophila melanogaster RefSeq 6 protein

Settings
Choose a predefined setting or customize BLAST parameters.
● **default:** for protein query to translated nucleotide database searches
○ **fast:** for fast protein query to translated nucleotide database searches

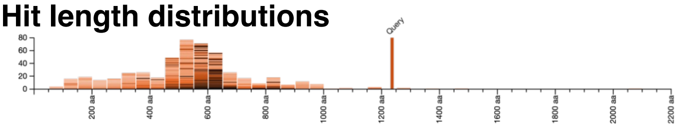
Task: tblastn
E-value cutoff: 1e-5
Maximum number of hits: 50

Residues in Query 422
Residues in Selected Databases 143,726,002
Estimated runtime beta: Under a minute
☑ Open results in new tab

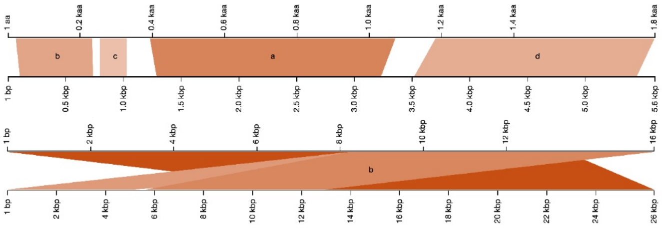
Alignment region & strength overview



Hit length distributions



Pairwise "kablammo" alignments



History & future

- Started in 2010 when Yannick needed to share the first ant genome sequences.
- Deployed in hundreds of institutions worldwide.
- >250 citations; thousands of daily users.
- SequenceServer Cloud service since 2022:
  - Aim: accelerate biological discovery.
  - Hosted secure servers, with smart compute scaling.
  - Many functionalities beyond BLAST (DNA Visualizer, SRA-BLAST, Graphical configuration, CDD functional annotation, History/audit trails, Sharing, Annotation...).
- Income enables maintenance, updates, and improvements of open-source core.