

RESEARCH WORKSPACE MANUAL

Daphnia Browser

Detailed English User Guide

A complete photo-lyase walkthrough from gene search to experimental design

Example query	photo-lyase
Example gene	LOC116919804 · NCBI GeneID 116919804
Representative protein	XP_032781741.2 · 523 aa
Reference assembly	Daphnia magna (NIES) · GCF_020631705.1
Guide version	2026.09.03.1

Prepared for the Sen Xu Lab private research workspace

1. Purpose and scope

Daphnia Browser integrates seven Daphnia genome annotations, all-genome gene search, exact identifier conversion, browser-local gene lists, published RNA-seq heatmaps, cached protein evidence, exploratory protein phylogeny, sequence export, primer screening, and CRISPR guide discovery.

Scientific use

The browser is an evidence-integration workspace. Similarity, domain signatures, predicted amplicons, and tree proximity are hypotheses or supporting evidence—not experimental confirmation, definitive gene naming, or proof of biological mechanism.

Before you begin

- Use a modern Chromium, Edge, Firefox, or Safari browser. The Jetstream2 pilot requires an account.
- Registration requests only an ID and password plus the laboratory registration code. Do not place the shared registration code in publications or screenshots.
- Analysis jobs can take seconds to minutes. Keep the page open until the status changes from running to complete.

Workspace map

Species / assembly	Choose among seven <i>Daphnia</i> genome annotations, grouped by species.
Search	Search all seven annotations at once or only the active assembly. Accepts gene names, GeneIDs, locus tags, legacy Dmagna IDs, transcript/protein IDs, aliases, and product keywords.
Low-load Workbench	Exact ID conversion, My Gene List, multi-gene expression heatmaps, printable reports, and a precomputed domain browser.
Overview	Coordinates, exon/CDS/UTR structure, identifiers, isoforms, and source annotation.
Expression	Bulk anatomical-compartment atlas, embryo expression, DESeq2 statistics, and batch Excel export.
Protein evidence	Reviewed-homolog annotation candidates and their similarity evidence.
Exploratory tree	Choose an across-animal reference tree or a fixed top-25 DIAMOND-hit tree for any of the seven <i>Daphnia</i> genomes; neither mode is an orthology test.
Domains & cross-assembly	InterProScan protein features, same-species correspondence, and optional multi-species orthology evidence.
Sequence tools	FASTA/GenBank/protein export, NCBI BLAST handoff, primer screening, and guide design.

2. Quick start: search photo-lyase

1. Keep Search scope set to All genomes. The Dataset menu remains the assembly that will open when you select a result.
2. Type photo-lyase exactly in the search box and select Search.
3. Review the genome-grouped candidates. Select LOC116919804 in the *Daphnia magna* (NIES) group to open the detailed photo-lyase example.

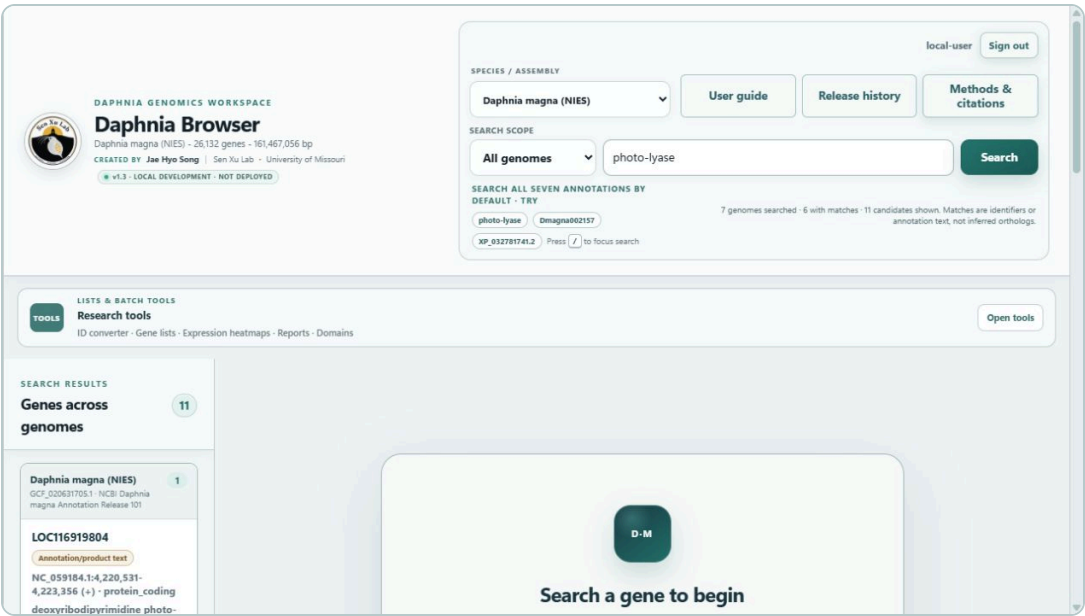


Figure 1. All-genome annotation search for photo-lyase, with candidates grouped by genome and match class.

How to read the grouped results

Exact identifier is the strongest lookup class. Name or alias text and annotation/product text are lexical annotation matches. Results across genomes are not inferred homologs or orthologs, and a genome with no text match may still contain a homolog under another annotation label.

Choose Current genome when you want the earlier assembly-specific workflow. In that mode, the first matching gene opens automatically and all subsequent sequence, primer, and CRISPR operations remain tied to the selected assembly.

What the result means

Genomic location	NC_059184.1:4,220,531–4,223,356 on the plus strand
Gene length	2,826 bp
Isoforms	3 transcripts and 3 proteins
Products	Deoxyribodipyrimidine photo-lyase isoforms X1/X2/X3
Annotation note	The RefSeq model was derived by automated computational analysis using Gnomon.

If the search returns nothing

Check the selected species/assembly and the exact spelling. Identifiers and product text differ among annotations. Try a stable identifier native to the selected assembly.

Available genome assemblies

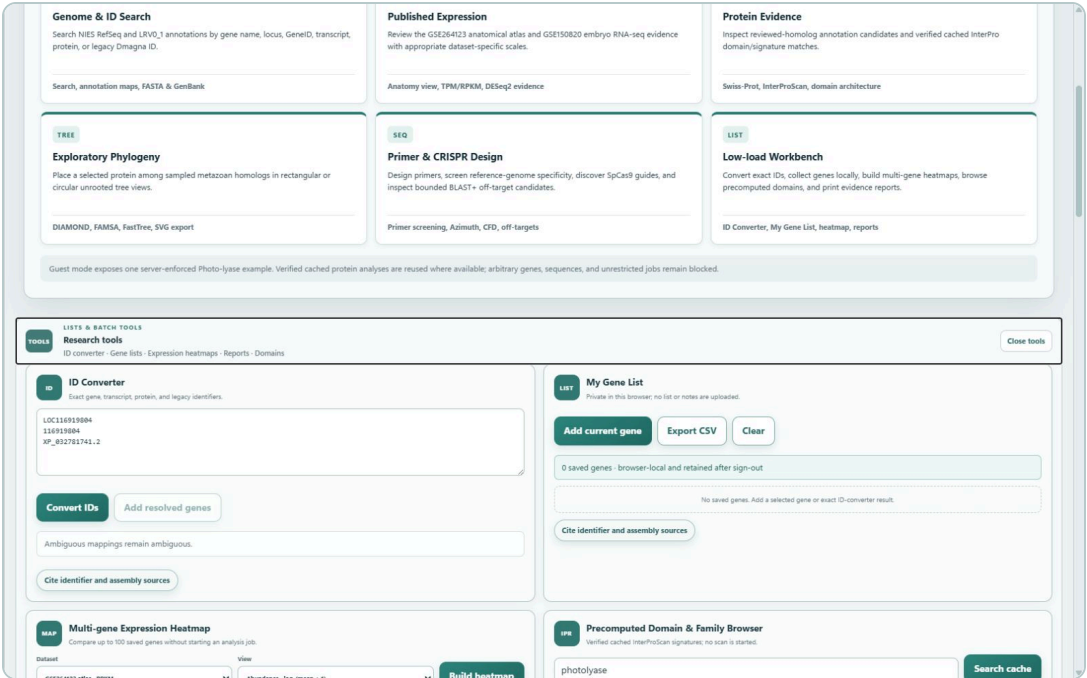
The local Version 1.4 candidate contains seven independently indexed genome annotations. Genome search, sequence export, local primer specificity, CRISPR screening, and within-genome phylogeny use the assembly explicitly selected for each operation.

Daphnia magna (NIES)	GCF_020631705.1 · NCBI Daphnia magna Annotation Release 101
Daphnia magna (LRVO_1)	GCA_030254905.1 · GenBank plus legacy Dmagna-ID crosswalk
Daphnia magna (Xinb3; APZ42 IDs)	GCA_001632505.1 · daphmag2.4; APZ42 is the locus-tag prefix
Daphnia carinata (CSIRO-1)	GCF_022539665.2 · GCF_022539665.2-RS_2023_09
Daphnia sinensis (WSL)	GCA_013167095.2 · submitter annotation
Daphnia pulicaria (SC F1-1A)	GCF_021234035.1 · NCBI Annotation Release 100
Daphnia pulex (KAP4)	GCF_021134715.1 · NCBI Annotation Release 100

Expression-data boundary

GSE264123 and GSE150820 are D. magna studies. The browser does not project their expression values onto D. carinata, D. sinensis, D. pulicaria, or D. pulex genes. A missing expression panel is not evidence of zero expression.

Low-load Research Workbench



Low-load Workbench with exact ID conversion and browser-local gene-list tools.

Expand the workbench below the search header for batch-oriented tools. These functions use exact indexes, bundled expression databases, verified cached signatures, or browser-local storage; opening them does not start a heavy analysis job.

- ID Converter accepts up to 500 exact stable identifiers. One-to-many matches remain ambiguous; product keywords and partial strings are never promoted to an ID mapping.
- My Gene List is stored only in the current browser and remains after sign-out. Export CSV or clear it on a shared computer.
- Expression Heatmap keeps GSE264123 RPKM and GSE150820 TPM in separate views. Gray NA means no mapped record, not zero expression; only an explicit zero in the published matrix is treated as zero. Row z-scores show within-gene pattern only.
- The Domain and Family Browser covers verified precomputed InterProScan proteins only. Details retain protein length, sequence-MD5 verification, member database, signature/IPR accession, coordinates, and score when available. Missing means not precomputed, not absence of a domain.
- Gene evidence reports are printable snapshots of evidence already loaded in the current browser session. Reports include assembly, annotation, selected protein, method/dataset citation keys, and a session-status caveat; creating one starts no DIAMOND, InterProScan, phylogeny, AlphaFold, OrthoFinder, or external API request.

3. Read the Gene Overview

The Gene Annotation Map summarizes exon, CDS, UTR, and intron coordinates in gene direction. Use it to verify which isoform and genomic interval you are exporting or targeting.

Recommended checks

- Confirm chromosome/scaffold, strand, and coordinate interval before designing primers or guides.
- Compare transcript isoforms. LOC116919804 has proteins XP_032781741.2 (523 aa), XP_032781742.2 (452 aa), and XP_045027373.1 (481 aa).
- Open NCBI Gene when you need the external RefSeq record, current annotation history, or cross-links.
- Choose a comparison assembly and use Find similar gene candidates. The browser aligns every annotated protein isoform to the selected target proteome with DIAMOND, collapses protein hits to genes, and shows up to five ranked candidates with identity, bilateral coverage, E value, bit score, and reciprocal-best-hit evidence.

Similarity is not orthology

The top candidate can be a paralog, partial or split gene model, fusion/fission product, or domain-only match. Reciprocal best hit strengthens sequence correspondence but does not prove duplication-aware orthology, conserved function, synteny, or gene-name equivalence. Review the ranked candidates before opening one.

Coordinate rule

Published-reference coordinates remain the primary coordinate system. Never attach an unchanged GFF directly to an indel-adjusted consensus FASTA without coordinate liftover.

Identifier guide

LOC116919804	RefSeq-style gene symbol/identifier used by this gene model.
116919804	NCBI GeneID.
XP_032781741.2	Representative RefSeq protein accession, isoform X1.
Dmagna...	Legacy LRVO_1 annotation identifier; availability depends on the old-annotation crosswalk.
Dapma7bEVm...	A maker/EvidenceModeler-style internal annotation ID from another annotation set.

4. Inspect published RNA-seq evidence

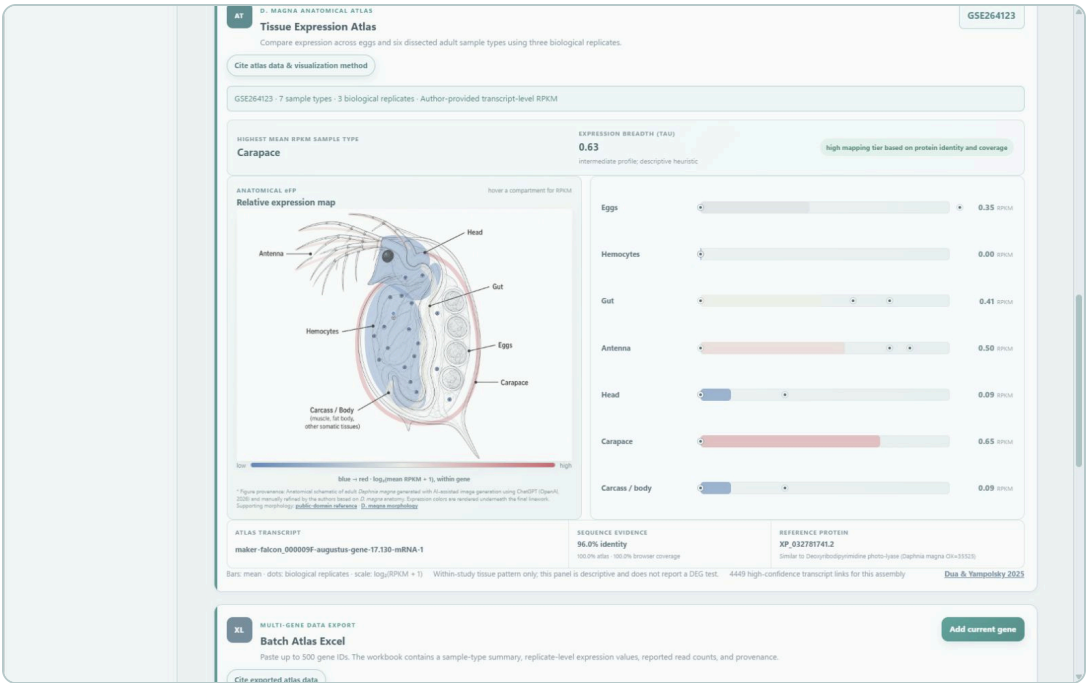


Figure 2. Tissue Expression Atlas for the photo-lyase example.

Anatomical atlas (GSE264123)

For LOC116919804, the mapped atlas transcript is maker-falcon_000009F-augustus-gene-17.130-mRNA-1. Protein-sequence mapping is 96.0% identity with 100% atlas and browser coverage. The displayed highest mean compartment is carapace (0.65 RPKM), and Tau is 0.63.

Interpret with the read depth

The compartment means are very low for this example and individual replicates include zeros. The eFP color is a within-gene visualization of bulk dissected samples; it is not in situ localization, single-cell expression, compartment purity evidence, or a DEG test.

Embryo dataset (GSE150820)

The four descriptive group means are WT female 13.8 TPM, Line A male 12.8 TPM, Line B male 13.2 TPM, and WT male 13.9 TPM. In the default WT male versus WT female contrast, the shrunk log2 fold change is -0.008 and BH-adjusted P value is 0.9926; this example does not support differential expression in that contrast.

- Bars are group means; dots are biological replicates.
- TPM is descriptive. Use the DESeq2 panel and multiple-testing-adjusted P value for inferential reporting.
- Line A/B combine genotype and sexual phenotype; those effects cannot be separated by this design.

Zero versus missing

A displayed 0 TPM/RPKM is a value explicitly present in the published quantification matrix. A gene or sample absent from that matrix is shown as gray NA and must not be interpreted as no expression.

Batch Atlas Excel

Paste up to 500 identifiers and select Download Excel. The workbook contains sample-type means, replicate-level RPKM/read counts, mapping evidence, and provenance. Use one identifier per line when preparing a reproducible gene list.

5. Evaluate reviewed-protein similarity

Tree

Exploratory tree

LAB

Domains & cross-assembly

SEQ

Sequence tools

QA

FAST REVIEWED EVIDENCE

Quick Protein Annotation

Search the selected protein against UniProtKB/Swiss-Prot (reviewed) to identify annotation and gene-symbol candidates.

Cite DIAMOND & UniProtKB/Swiss-Prot

UniProtKB/Swiss-Prot (reviewed) search complete - cached result - 6 top hits

TOP REVIEWED: HOMOLOG ANNOTATION CANDIDATE

GENE SYMBOL CANDIDATE

high sequence-similarity list

Deoxyribodipyrimidine photo-lyase

PHR

A top sequence-similarity hit is an annotation candidate, not a functional or gene-symbol assignment.

Deoxyribodipyrimidine photo-lyase

Photorhabdus luminescens: PHR

Q28811

85.6% identity 90.7% query coverage 88.5% subject coverage 3.40e-228 E-value

Deoxyribodipyrimidine photo-lyase

Campylobacter jejuni: phr

P34205

82.4% identity 86.4% query coverage 81.7% subject coverage 5.85e-209 E-value

Deoxyribodipyrimidine photo-lyase

Anaerostipes thalassiae: PHR1

Q25809

48.7% identity 54.3% query coverage 51.3% subject coverage 7.85e-158 E-value

Deoxyribodipyrimidine photo-lyase

Oryza sativa subsp. japonica: PHR

Q8F5A2

50.2% identity 82.8% query coverage 87.3% subject coverage 1.09e-169 E-value

Deoxyribodipyrimidine photo-lyase

Methanorhabdus thermotrophicus (strain ATCC 29096 / DSM 1053 / JCM 10044 / NBRC 100330 / DSM 82 / Marburg): phr

P32783

37.3% identity 82.4% query coverage 84.4% subject coverage 5.15e-87 E-value

Deoxyribodipyrimidine photo-lyase

Methanorhabdus marburgensis (strain ATCC BAA-327 / DSM 2133 / JCM 14631 / NBRC 100331 / DSM 82 / Marburg): phr

P38818

36.5% identity 82.4% query coverage 84.4% subject coverage 5.83e-85 E-value

Figure 3. Quick Protein Annotation for XP_032781741.2.

Select View cached result first. If a sequence-matched cache exists, it is displayed immediately. If it does not, a separate Run one DIAMOND search button appears. The cached example supports deoxyribodipyrimidine photolyase as the top reviewed-homolog candidate, with PHR as a candidate gene symbol.

Top hit	Q28811 · Potorous tridactylus · PHR
Identity	61.6%
Query coverage	90.1%
Subject coverage	88.5%
E value	3.49×10^{-220}
Evidence tier	High sequence-similarity tier

Do not overstate the result

A reviewed homolog supports an annotation candidate. It does not by itself assign the Daphnia gene symbol, demonstrate enzymatic activity, or establish orthology. Inspect coverage, domain architecture, multiple hits, and phylogenetic context.

Run only when needed

A gene or protein selection never starts DIAMOND automatically. Start one queued search only when no verified cached result is available or the sequence/database changed. The result is cached by protein sequence for later use.

6. Build and read the protein phylogeny

1. Open Exploratory tree and choose Across animal reference proteomes or Within one selected Daphnia genome.
2. Choose any of the seven Daphnia species/assemblies as the target proteome. When it matches the query dataset, the result is a within-genome tree; otherwise, the source query is placed among the selected target proteome's top hits. The sample size is fixed at the 25 highest-bit-score gene-level hits.
3. Wait for DIAMOND ultra-sensitive top-hit ranking, one-best-isoform-per-target-gene collapsing, FAMSA alignment, removal of columns with more than 50% gaps, and FastTree WAG+Gamma inference. The query gene is excluded only when source and target are the same dataset. Version 2026.09.03.1 applies no biological similarity threshold to candidate inclusion and requires at least 50 retained alignment columns for a tree.
4. Switch between Rectangular and Circular layouts. The topology is the same; only the display changes.

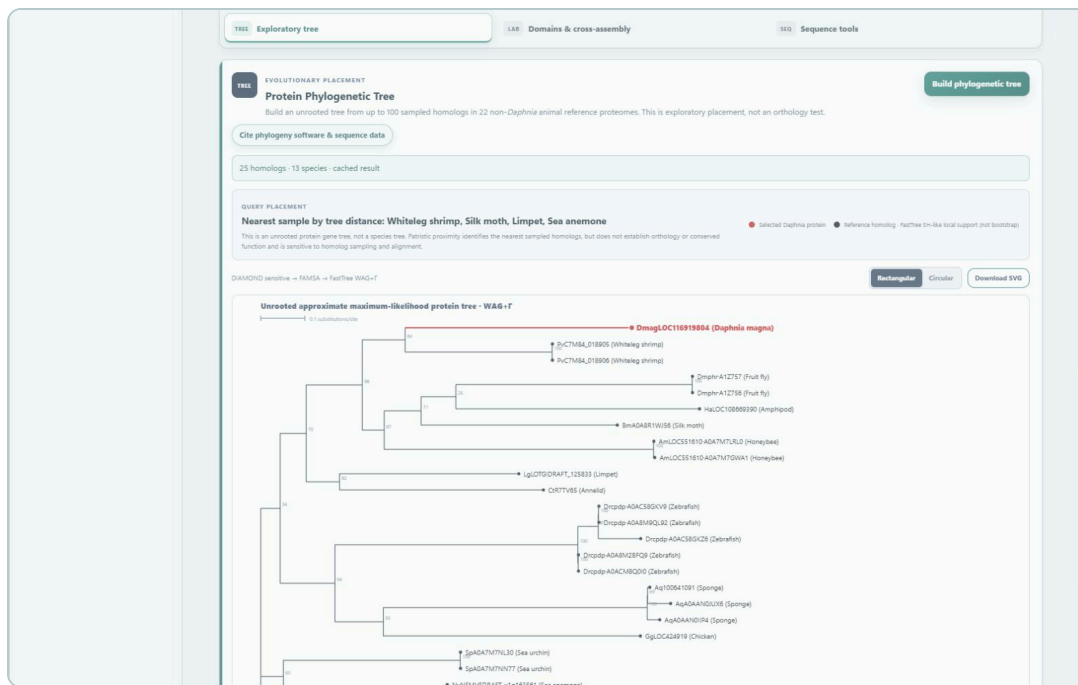


Figure 4. Rectangular-tree controls and query-placement summary.

The current photo-lyase run contains 25 homologs from 13 species. The query is red; reference homologs are black. The nearest sampled homologs by patristic distance include whiteleg shrimp, silk moth, limpet, and sea anemone.

Cached-example provenance

The displayed 25-homolog example was generated by an earlier DIAMOND sensitive-mode cache. The deployed code now uses very-sensitive mode for a new or refreshed analysis, so the candidate set and topology may change after refresh. Report the method stored with the exported result, not only the current software default.

Target-proteome top-hit mode

This mode ranks DIAMOND local alignments by bit score, excludes the query gene only when the source and target dataset are identical, collapses target isoforms to the highest-ranked protein per annotated gene, and retains the first 25 genes. Cross-dataset runs are labeled target-proteome query placement and report the actual source/target dataset and species counts. No E-value, identity, coverage, or aligned-length rule is used as a biological inclusion threshold. DIAMOND uses ultra-sensitive mode, default low-complexity masking, and a permissive E-value reporting ceiling of 1e9 so weak local matches can be ranked.

- Tree tips show gene ID, aligned-region percent identity, and query/target coverage in the rectangular view; the circular view keeps a shorter identity label. Hovering a tip also shows protein ID, product, aligned amino acids, E-value, and bit score.
- Teal means a well-supported local match, amber means partial/domain-limited, and gray means weak/unresolved. These are descriptive evidence flags and do not exclude candidates.

- The candidate table remains available when fewer than three target gene models are returned or fewer than 50 alignment columns remain after gap filtering; a tree is not inferred from inadequate alignment evidence. Gap-rich alignment-column removal is tree QC, not hit filtering.

Tree interpretation

This unrooted exploratory tree contains up to 25 highest-bit-score DIAMOND target-proteome hits after collapsing target isoforms by gene; the query gene is excluded only for an identical source/target dataset, and no biological similarity threshold is used for inclusion. Some leaves may be weak, domain-only, low-complexity, partial, fusion/fission, or unrelated whole-protein matches and should not be assumed to belong to one protein family. Percent identity applies only to the aligned region, and DIAMOND rank reflects local-alignment score rather than evolutionary distance. Tree topology, branch length, and SH-like local support do not establish paralogy, orthology, conserved function, duplication order or time, or justify gene-name transfer.

7. Use the circular tree and export SVG

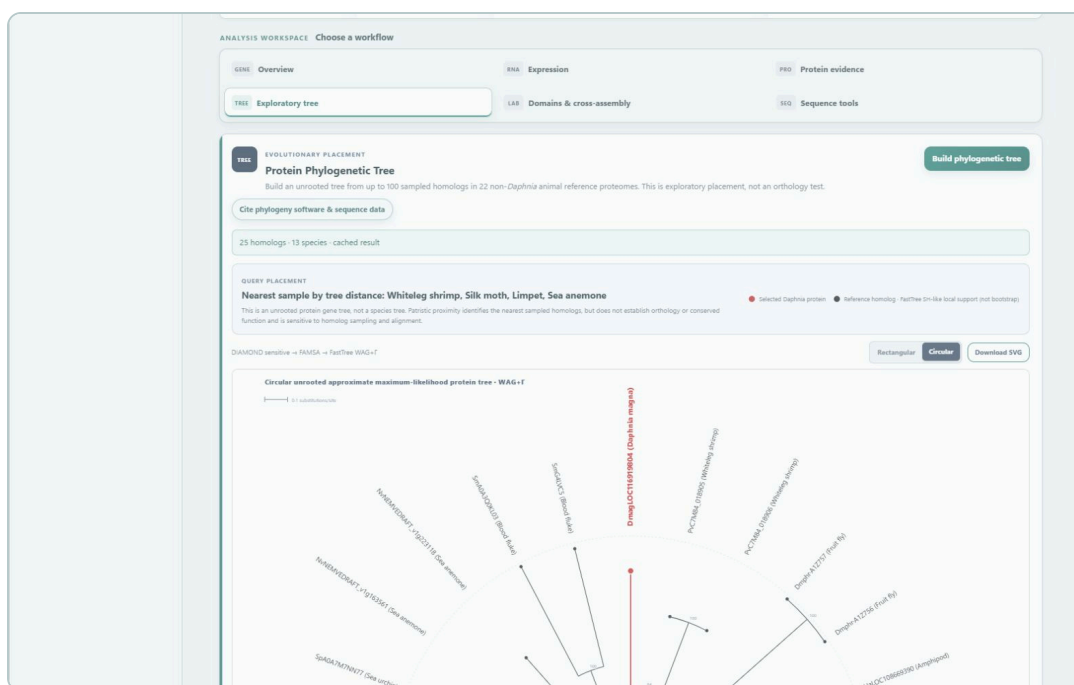


Figure 5. Circular layout; the selected *Daphnia* protein is highlighted in red.

Circular layout is useful for surveying many tips; rectangular layout is usually easier for reading labels and branch order. Select Download SVG for a publication-ready vector that retains the analysis provenance in SVG metadata.

A defensible reporting sentence

Example

“XP_032781741.2 grouped within the sampled animal photolyase homolog set in an unrooted WAG+Gamma FastTree analysis; this exploratory placement was not treated as a duplication-aware orthology assignment.”

What to report

- Query accession and sequence version.
- Reference-proteome/database release and candidate-selection thresholds.
- DIAMOND, FAMSA, and FastTree versions and analysis parameters.
- Analysis scope, source and target assembly, gene-collapsed DIAMOND-hit count, sampled count, visible similarity metrics, alignment columns retained, Newick tree, and whether a root/outgroup was used.

8. Inspect InterProScan protein features

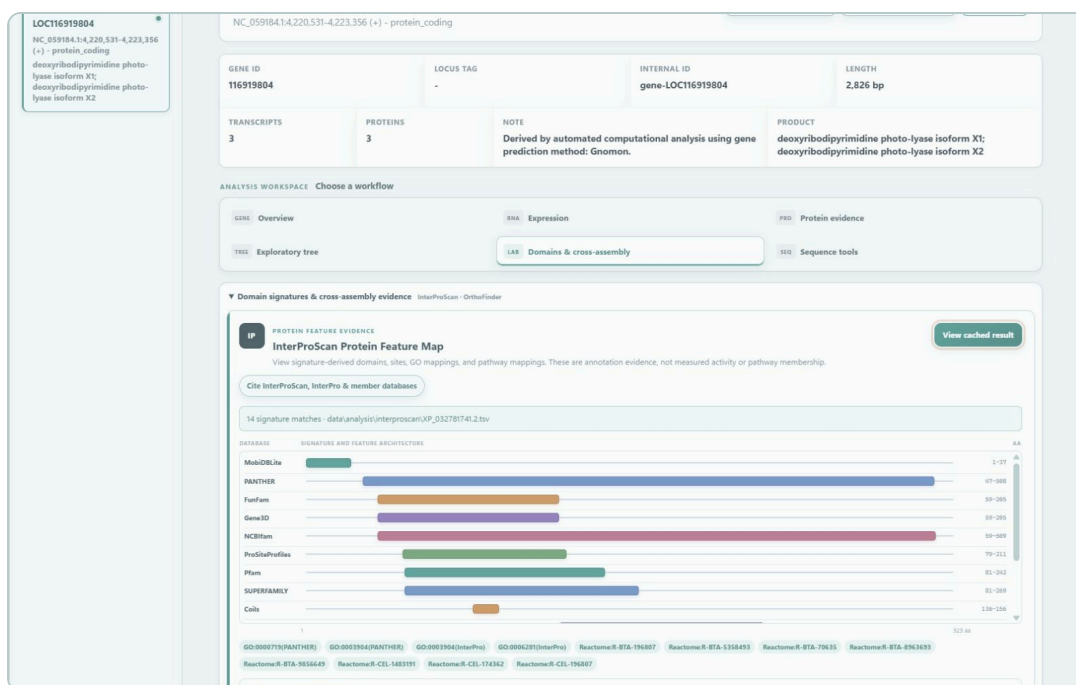


Figure 6. Domains & cross-assembly workspace showing InterProScan and cross-assembly evidence panels.

Select View cached result first. A full InterProScan starts only after the separate Run one InterProScan analysis confirmation. Gene selection and report generation never start a scan.

The cached XP_032781741.2 result contains 14 signature matches. Major evidence includes PANTHER DNA Photolyase Class-2 (aa 47–508), NCBIfam DNA photolyase class 2 (aa 59–509), the N-terminal photolyase region, and a cryptochrome/DNA-photolyase FAD-binding-domain-like superfamily.

How to interpret the track

- Colored blocks show signature coordinates on the 523-aa protein; overlapping blocks are independent database evidence, not duplicated protein regions.
- GO terms and Reactome links are signature/curated-entry-derived mappings. They are not measured activity or direct pathway-membership evidence in Daphnia.
- Use the Cite button to cite InterProScan, InterPro, and the member databases that produced signatures you interpret.

Cache integrity

InterPro results are tied to protein sequence MD5, length, and InterPro release. Re-run only when the sequence or database release changes.

Cross-Genome Protein Similarity

This panel searches the selected annotated *Daphnia* proteome with DIAMOND instead of requiring an exact genomic seed match. Target proteins are collapsed to genes and the five highest-scoring candidates are shown. A strong label requires a reciprocal best hit and broad coverage; a partial label indicates a high-identity reciprocal match covering nearly all of one protein but only part of the other, consistent with a partial, truncated, or split model. The browser does not use exact genomic seeds for this cross-genome ranking.

Name-transfer boundary

These are homolog candidates, not confirmed orthologs. Do not transfer a gene name from the top hit without duplication-aware phylogeny, domain architecture, model completeness, and preferably synteny or other independent evidence.

9. Export sequence and launch NCBI Protein BLAST

The screenshot displays the Daphnia Browser interface with two main panels visible. The top panel, 'Sequence Export', allows users to set flanking bases and download DNA, GenBank, or protein FASTA. It includes a dropdown for 'Transcript isoform' (set to 'Gene locus / all isoforms (2,826 bp)'), input fields for 'Upstream bp' and 'Downstream bp' (both set to 1000), a checkbox for 'Return sequence in gene direction' (checked), and buttons for 'Download FASTA', 'Download GenBank', and 'Find in Other Dataset'. The bottom panel, 'PCR Primer Design & Specificity', is for designing a PCR pair around a target. It features a 'Target region sequence' text area containing a DNA sequence, a 'Design around target' button, and a 'Test my primer pair' button. Below the sequence, there are input fields for 'Product min (bp)' (250), 'Product max (bp)' (800), 'Primer min (nt)' (19), 'Primer max (nt)' (24), and 'Pairs to return' (5), along with a 'Find primer pairs' button. The interface also shows navigation tabs at the top: 'Exploratory tree', 'Domains & cross-assembly', and 'Sequence tools'.

Figure 7. Sequence export, isoform selection, NCBI Protein BLAST, and primer entry point.

DNA and protein downloads

- Select a transcript isoform or Gene locus / all isoforms.
- Set upstream/downstream flanks. Keep Return sequence in gene direction selected when you want a strand-normalized sequence.

- Download FASTA for sequence workflows or GenBank when feature annotations are required.
- Select one protein isoform or All isoforms, then Download Protein FASTA.

NCBI Protein BLAST

1. Select the protein isoform, for example XP_032781741.2.
2. Select NCBI Protein BLAST and confirm the external submission.
3. A new NCBI BLAST tab opens with the FASTA query. Choose the desired NCBI database/organism filters and run the search there.

Data boundary

NCBI Protein BLAST sends the selected protein sequence to the external NCBI service. Local DIAMOND annotation and phylogeny instead use server-side databases. If the NCBI tab does not open, allow pop-ups for the Daphnia Browser domain.

10. Design or test PCR primers

Heuristic in-silico screen: candidate sites are discovered with up to 6 total mismatches, no gaps, exact terminal 3 bases, and at most 1 mismatch in the terminal 5 bases, inward-facing F/R, F/F, R/R, and R/F products are screened across 50-10,000 bp; the selected range controls which products are returned. Experimental validation remains required.

Complete - NCBI BLAST+ 2.17.0 - blastn-short, word size 4

SEQUENCE	TARGET MAPPING	RETURNED
NIES RefSeq (GCF_020631705.1)	1 genomic locus/loci	5 ranked pairs

Target: NC_059184.1:4221784-4222103 (+) - exact match

#1 - NC_059184.1:4221696-4222256 - 561 bp ONE PERFECT SEQUENCE-MATCHED PRODUCT

Forward AATATGTCCTGTTTGGGTGGCC 24 nt - Tm 59.1°C - GC 54.2% - poor

Reverse GCTGGGACAGCACTGACCTATTTC 24 nt - Tm 59.1°C - GC 54.2% - poor

Target occupies amplicon bases 89-408. Tm Δ 0°C - extra predicted products 0 - pair score 2.694

► Predicted products (1)

► Intended amplicon sequence

#2 - NC_059184.1:4221696-4222264 - 569 bp ONE PERFECT SEQUENCE-MATCHED PRODUCT

Forward AATATGTCCTGTTTGGGTGGCC 24 nt - Tm 59.1°C - GC 54.2% - poor

Reverse AATGCAAGCTGGGACAGCAACTG 24 nt - Tm 59.1°C - GC 54.2% - poor

Target occupies amplicon bases 89-408. Tm Δ 0°C - extra predicted products 0 - pair score 2.884

► Predicted products (1)

► Intended amplicon sequence

#3 - NC_059184.1:4221696-4222154 - 459 bp ONE PERFECT SEQUENCE-MATCHED PRODUCT

Forward AATATGTCCTGTTTGGGTGGCC 24 nt - Tm 59.1°C - GC 54.2% - poor

Reverse TTACTCAAGCCGCCAGCTTTGGG 24 nt - Tm 59.1°C - GC 54.2% - poor

Target occupies amplicon bases 89-408. Tm Δ 0°C - extra predicted products 0 - pair score 2.895

► Predicted products (1)

Figure 8. Local whole-genome PCR primer design and specificity screen.

Design around a target

1. Paste the region that must be fully contained inside the PCR product. A FASTA header is allowed.
2. Set product and primer-length ranges, then choose how many pairs to return.
3. Review each primer's binding sites and every predicted inward-facing product in the selected reference assembly.

Test my primer pair

Paste forward and reverse primers. The server searches conservative near matches and evaluates F/R, F/F, R/R, and R/F pairings over 50–10,000 bp. The displayed amplification likelihood describes reference-sequence matching, not a laboratory probability.

Specificity limits

The screen cannot account for every lab-stock polymorphism, non-reference allele, structural variant, assembly gap/error, polymerase chemistry, or annealing condition. A single strong reference-genome match is favorable but still requires wet-lab validation.

11. CRISPR guide discovery

Select CRISPR Design after choosing a gene, then review overlap with the relevant transcript isoforms. The browser identifies SpCas9 NGG candidates, evaluates on-target activity with the Azimuth 2.0 Rule Set 2 V3 no-position model, and evaluates reference-genome off-target candidates with the official CFD scoring tables described by Doench et al. (2016).

Official on-target score

- Rule Set 2 uses the exact 30-nt context expected by Azimuth: 4 upstream bases, the 20-nt spacer, the 3-nt PAM, and 3 downstream bases.
- The raw model output is multiplied by 100 for display. It is not a probability and, as documented by Azimuth, can rarely fall just outside 0-100. The browser preserves this behavior rather than clipping the official calculation.
- All 947 usable rows from the Azimuth repository no-position fixture agreed within 0.001; maximum absolute difference was below 6×10^{-10} .

Validated CFD implementation and off-target screen

- All shortlisted 20-nt guides are submitted in one local BLAST+ blastn-short search against the selected reference assembly.
- Candidate sites must be full-length, ungapped alignments with no more than four mismatches. The adjacent genomic PAM is checked before scoring.
- Each retained site receives the official CFD product of mismatch-identity/position weights and PAM weight (0-1). The table reports counts at $\text{CFD} \geq 0.20$ and 0.05-0.20, maximum off-target CFD, locus, strand, PAM, and mismatch count.
- The CFD implementation reproduced all seven Cas9 CFD reference cases from the crisprScore test suite after three-decimal rounding.

Interpretation boundary

Rule Set 2 and CFD are the official algorithms, but their scores are not Daphnia-specific experimental probabilities. The bounded reference-genome search does not model DNA/RNA bulges, structural variants, non-reference or lab-stock alleles, chromatin, or PCR/editing chemistry. Absence of a reported candidate is not proof of biological specificity.

Server-efficient without changing the math

The server scores a bounded shortlist, sends all guides through one BLAST process, uses at most two BLAST threads in production, shares the analysis queue, and caches identical searches. These controls avoid duplicate work; they do not approximate or simplify Rule Set 2 or CFD calculations.

Browser rank

The overall browser rank combines official Rule Set 2, official per-site CFD evidence, target annotation, and primer availability. This composite rank is a Daphnia Browser prioritization rule, not an official Azimuth or CFD output.

Automatic RNP-test primers

Primer pairs shown beside a guide are screened for exact matches in the selected reference assembly. They do not include the conservative near-match screen. Copy the pair into the separate Test my primer pair workflow before ordering primers.

Required checks

- Verify the PAM, strand, cut-site location, intended exon or functional region, and the full lab-stock target sequence.
- Inspect strong alternative sites and confirm important off-target loci in the actual experimental stock.
- Use independent review and experimental validation before editing.

Practical selection strategy

1. Prefer guides in an exon shared by biologically relevant isoforms.
2. Avoid common polymorphisms and candidates with strong alternative sites near coding genes.
3. Design independent guides and a genotyping PCR strategy before beginning the experiment.

12. Methods, citations, and reproducible export

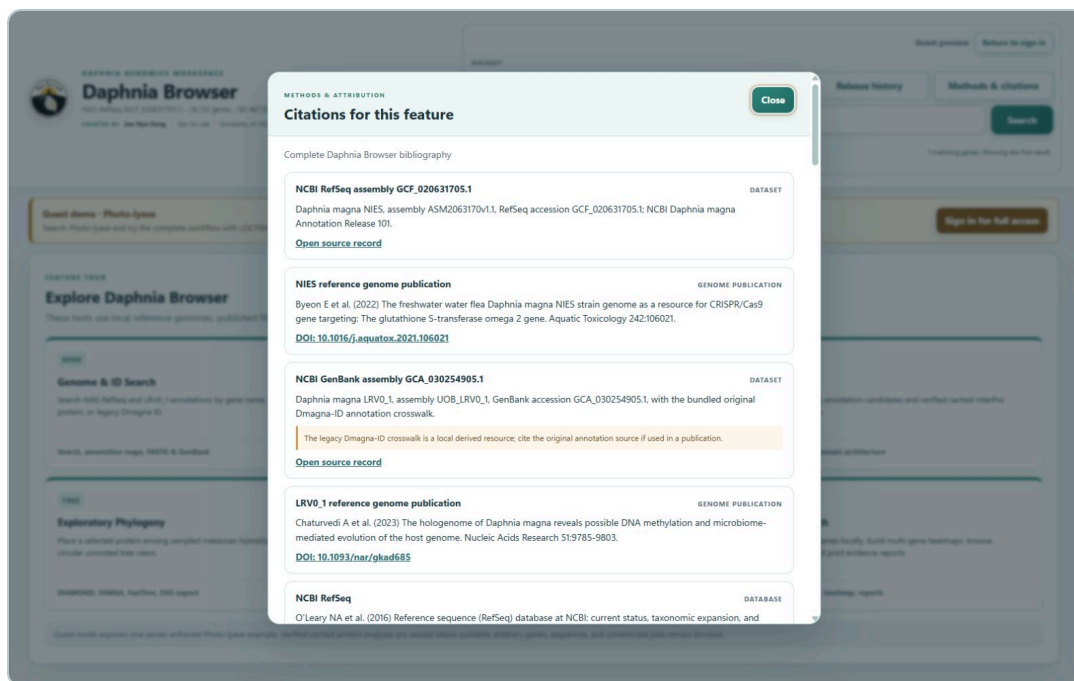


Figure 9. Central Methods & citations dialog in the deployed browser.

Every major analysis card includes a Cite button. The central dialog separates datasets, software, databases, methods, and asset provenance. Download the BibTeX file for manuscripts or the machine-readable bibliography for automated provenance workflows.

Citation rules

- Cite the biological dataset/reference assembly and the software/database used to generate the result.
- Report stable accessions, release/version, thresholds, and analysis date—not only the paper.
- For InterProScan, also cite member databases whose signatures are interpreted.
- A method citation documents how evidence was generated; it does not independently validate the conclusion.

Atlas anatomy schematic

Anatomical schematic of adult Daphnia magna generated with AI-assisted image generation using ChatGPT (OpenAI, 2026) and manually refined by the authors based on D. magna anatomy. ChatGPT was used as an image-generation aid; the authors retain responsibility for the final figure.

13. Troubleshooting

Symptom	Recommended action
photo-lyase returns no result	Keep the hyphen and use All genomes. For an exact NIES lookup, try LOC116919804 / 116919804 / XP_032781741.2; no text hit in one annotation is not evidence that a homolog is absent.
Analysis button appears inactive	Wait for the current job; the private server limits concurrent heavy analyses to protect memory and CPU.
Session expired / 401	Return to the sign-in page and log in again; unsaved browser state may need to be recreated.
NCBI Protein BLAST does nothing	Allow pop-ups for the site, confirm the submission dialog, and verify that the browser permits forms to blast.ncbi.nlm.nih.gov.
InterProScan is slow	Use View cached result first. A full local scan starts only after the separate confirmation and can take minutes.
Primer target not found	Confirm species/assembly, orientation, sequence quality, and whether the pasted sequence belongs to the selected reference.
Downloaded file is blocked	Allow downloads for the domain and retry from the corresponding Download button.
Tree labels are crowded	Switch layouts or download SVG and inspect it in a vector editor; do not change branch topology for presentation.

Research-quality checklist

- Record assembly/accession, annotation release, isoform, and exact sequence version.
- Save raw values and provenance, not screenshots alone.
- Separate biological replicates from technical replicates.
- Treat low-count RNA-seq patterns and sequence-derived functional predictions conservatively.
- Confirm key hypotheses with independent data and wet-lab experiments.

14. Selected references

- Byeon E et al. (2022). The freshwater water flea *Daphnia magna* NIES strain genome as a resource for CRISPR/Cas9 gene targeting: The glutathione S-transferase omega 2 gene. *Aquatic Toxicology* 242:106021. <https://doi.org/10.1016/j.aquatox.2021.106021>

- Chaturvedi A et al. (2023). The hologenome of *Daphnia magna* reveals possible DNA methylation and microbiome-mediated evolution of the host genome. *Nucleic Acids Research* 51:9785–9803. <https://doi.org/10.1093/nar/gkad685>
- NCBI GenBank assembly GCA_001632505.1. *Daphnia magna* strain Xinb3, assembly daphmag2.4; Don Gilbert *Daphnia magna* Annotation 1. APZ42 is the locus-tag prefix. https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_001632505.1/
- NCBI RefSeq assembly GCF_022539665.2. *Daphnia carinata* strain CSIRO-1, assembly CSIRO_AGI_Dcar_HiC_V3; annotation GCF_022539665.2-RS_2023_09. https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_022539665.2/
- NCBI GenBank assembly GCA_013167095.2. *Daphnia sinensis* strain WSL, assembly Dsi; submitter annotation. https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_013167095.2/
- NCBI RefSeq assembly GCF_021234035.1. *Daphnia pulicaria* isolate SC F1-1A, assembly SC_F0-13Bv2; Annotation Release 100. https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_021234035.1/
- NCBI RefSeq assembly GCF_021134715.1. *Daphnia pulex* isolate KAP4, assembly ASM2113471v1; Annotation Release 100. https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_021134715.1/
- Dua A, Yampolsky LY (2025). Transcriptional atlas of *Daphnia magna*. *Comparative Biochemistry and Physiology Part D* 55:101504. <https://doi.org/10.1016/j.cbd.2025.101504>
- Nong QD et al. (2020). Two Doublesex1 mutants revealed a tunable gene network underlying intersexuality in *Daphnia magna*. *PLOS ONE* 15:e0238256. <https://doi.org/10.1371/journal.pone.0238256>
- Buchfink B, Reuter K, Drost HG (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods* 18:366–368. <https://doi.org/10.1038/s41592-021-01101-x>
- Deorowicz S, Debudaj-Grabysz A, Gudys A (2016). FAMSA: Fast and accurate multiple sequence alignment of huge protein families. *Scientific Reports* 6:33964. <https://doi.org/10.1038/srep33964>
- Price MN, Dehal PS, Arkin AP (2010). FastTree 2—Approximately Maximum-Likelihood Trees for Large Alignments. *PLOS ONE* 5:e9490. <https://doi.org/10.1371/journal.pone.0009490>
- The UniProt Consortium (2025). UniProt: the Universal Protein Knowledgebase in 2025. *Nucleic Acids Research* 53:D609–D617. <https://doi.org/10.1093/nar/gkae1010>
- Jones P et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30:1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>
- Blum M et al. (2025). InterPro: the protein sequence classification resource in 2025. *Nucleic Acids Research* 53:D444–D456. <https://doi.org/10.1093/nar/gkae1082>
- Emms DM, Kelly S (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology* 20:238. <https://doi.org/10.1186/s13059-019-1832-y>
- Camacho C et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. <https://doi.org/10.1186/1471-2105-10-421>
- Doench JG et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology* 34:184–191. <https://doi.org/10.1038/nbt.3437>
- Microsoft Research (2016). Azimuth 2.0: on-target activity prediction using the Rule Set 2 model. <https://github.com/MicrosoftResearch/Azimuth>

- Cock PJA et al. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25:1422–1423. <https://doi.org/10.1093/bioinformatics/btp163>
- Fortin JP et al. crisprScore: on-target and off-target scoring algorithms for CRISPR guide RNAs. *Bioconductor*. <https://doi.org/10.18129/B9.bioc.crisprScore>