Jun 30, 2025

🌐 Generation and High-Throughput Sequencing Validation of DINO-SL Based cDNA Libraries for Yeast Two-Hybrid Assays

Tania Islas-Flores[1], Edgardo Galán-Vásquez[2], Marco A. Villanueva[1]

[1]Instituto de Ciencias del Mar y Limnología, Unidad Académica de Sistemas Arrecifales, Universidad Nacional Autónoma de México, UNAM, Prolongación Avenida Niños Héroes S/N, Puerto Morelos, Quintana Roo 77580, México;

[2]Departamento de Ingeniería de Sistemas Computacionales y Automatización, Instituto de Investigación en Matemáticas Aplicadas y en Sistemas, Universidad Nacional Autónoma de México, UNAM, Circuito Escolar 3000, Ciudad Universitaria, Ciudad de México 04510, México.

Tania Islas-Flores
UASA-ICML-UNAM

Protocol status: Working
We use this protocol and it's working

Created: June 19, 2025

Last Modified: June 30, 2025

Protocol Integer ID: 220608

Keywords: Two Hybrid System, DINO-SL, Symbiodinium microadriaticum, cDNA Library, high performance, dinoflagellate nuclear mrna, cdna libraries for yeast two, validation of dino, reverse transcriptase, function of the reverse transcriptase, rna transcript, better library performance for photosynthetic dinoflagellate, applicable to other dinoflagellate, symbiodinium microadriaticum casskb8, rna, other dinoflagellate, end of rna transcript, functions in each of the cdna library, de novo assembly, photosynthetic dinoflagellate, bioinformatics procedure, dino spliced leader, sequencing validation, dino, types of cdna library, cdna library, yeast, yeast two, sequencing, based cdna library, number of transcript, transcript, coding sequence

Abstract

To illustrate the application of protocol validation for cDNA libraries used in the Yeast Two-Hybrid assay, two types of cDNA libraries from *Symbiodinium microadriaticum* CassKB8 were generated. One using the post-transcriptional addition of a Dino Spliced Leader (Dino-SL) to the 5′ end of dinoflagellate nuclear mRNAs, and the other using the conventional SMART technology (Switching Mechanism at 5′ end of RNA Transcript) which is based on the template switching function of the reverse transcriptase (Islas-Flores et al. 2021). The performance of both libraries on the Yeast-Two Hybrid System was determined by scoring the yield, number and length of sequences, number of transcripts, and representation. We validated high-throughput sequencing by employing a Bioinformatics procedure that includes data control and cleaning, *de novo* assembly, identification of coding regions, and annotation of homologous sequences. Finally, we related those sequences to their biological functions at the level of motif and coding sequence in order to identify the proteins and their functions in each of the cDNA libraries. This allows to determine the optimum methodology to yield a better library performance for photosynthetic dinoflagellates such as *S. microadriaticum* and potentially applicable to other dinoflagellates.

Guidelines

1. **General Considerations**
- Molecular biology techniques including RNA extraction, cDNA synthesis, PCR, yeast transformation, and basic bioinformatics is required.
- All equipment must be calibrated and functioning properly before starting.
- Work in an RNase-free environment; treat surfaces, pipettes, and plasticware with RNase decontaminants when handling RNA.

2. Biosafety and Reagent Handling
- Use appropriate personal protective equipment (PPE) when handling hazardous chemicals like DEPC, chloroform, isopropanol, and 2-mercaptoethanol.
- Perform DEPC treatments and TRI Reagent manipulations in a chemical fume hood.
- Dispose of hazardous and biological waste according to institutional safety guidelines.

3. RNA Quality and Integrity
- Use freshly prepared and DEPC-treated solutions for RNA extraction.
- Immediately freeze *S. microadriaticum* cell pellets in liquid nitrogen and store at −70 °C until use.
- Assess RNA purity (A260/A280 ratio ~2.0) and integrity via gel electrophoresis before proceeding with cDNA synthesis.

4. Library Construction
- Use high-quality mRNA (as the purified with Oligotex kit) for cDNA synthesis.
- Include appropriate controls for SMART and DINO-SL libraries.
- Confirm cDNA amplification by LD-PCR with a clear smear on agarose gel.
- Purify cDNA to exclude fragments smaller than 200 bp.

5. Yeast Transformation
- Prepare competent yeast (Y187) cells under strict OD_{600} control to ensure transformation efficiency.
- Use high-purity SmaI linearized plasmids and ds cDNA with optimal concentration.

6. Library Quality Control
- Estimate the number of independent clones by serial dilution plating.
- Aim for at least ≥10^6 independent transformants per library to ensure diversity.
- Harvest and store transformed cells in aliquots at −80 °C with glycerol.

7. Sequencing and Bioinformatics
- Use Illumina paired-end sequencing with sufficient depth (~10 million reads/library).
- Perform stringent quality control of sequencing reads using tools like *Fastp*.
- Use *Trinity* for de novo assembly, *TransDecoder* for ORF prediction, and *BLAST* for homology analysis.
- Identify unique transcripts and conserved domains using HMMER and Pfam.
- Annotate gene functions using KEGG pathways via GhostKOALA.
- Investigate translation potential via Kozak sequence motif analysis using DNA-pattern.

8. Documentation and Reproducibility
- Maintain a detailed lab notebook with reagent lot numbers, sample IDs, and deviations from the protocol.
- Back up sequencing and analysis files, and store raw data in public repositories where possible.
- Include all code used for data processing and analysis in the supplementary information or a GitHub repository.

9. Troubleshooting Tips

- Low RNA yield: Repeat extraction with increased bead beating or ensure cell lysis.
- Poor cDNA synthesis: Check RNA integrity and reverse transcriptase activity.
- Low transformation efficiency: Verify yeast cell OD, plasmid purity, and PEG/LiAc freshness.
- Incomplete assemblies: Optimize trimming thresholds and confirm input read quality.

Materials

Cell cultures

Symbiodinium microadriaticum CassKB8 cell cultures (clade A) are maintained in ASP-8A medium under photoperiod cycles of 12 h light/dark at 26°C, at a light intensity of 80-120 μmole photon m^{-2} sec^{-1} under aerated conditions.

RNA extraction

-2 ml screw cap polypropylene tubes containing 200 μl volume of 425–600 μm glass beads.

-TRI reagent (Sigma, St. Louis, MO, USA).

-Chloroform.

-Bead beater (BioSpec).

-3 M sodium acetate, pH 5.2.

-Isopropanol.

-80% ethanol.

-Nuclease-free water.

-1X TAE buffer (40 mM Tris, 20 mM acetate and 1 mM EDTA, pH 8).

-Spectrophotometer.

-RNase-free DNase I.

-Thermocycler.

-RNAeasy kit (Qiagen, Valencia, CA, USA)

-B-mercaptoethanol..

-OLIGOTEX kit (Qiagen, Valencia, CA, USA).

Library construction

- Matchmaker Library Construction kit (Clontech Laboratories Inc., Mountain View, CA, USA).

-10 μM DINO-SL oligo (5′-TTCCACCCAAGCAGTGGTATCAACGCAGAGTGGCCATTATGGCCCCGTAGCCATTTTGGCTCAAG-3′)

-10 μM 5′ PCR Primer (5′-TTCCACCCAAGCAGTGGTATCAACGCAGAGTGG-3′).

-10 μM 3′ PCR Primer (5′-GTATCGATGCCCACCCTCTAGAGGCCGAGGCGGCCGACA-3′).

-CHROMA SPIN+TE-400 Columns (Clontech, Mountain View, CA, USA)

Yeast transformation

-Yeastmaker Yeast Transformation System 2 (Clontech, Mountain View, CA, USA).

-0.9% NaCl.

-DMSO.

-SmaI linearized pGADT7-Rec plasmid.

-Y187 strain.

-YPDA-km broth and agar plates.

-150 mm SD/-Leu-km agar plates

Harvest and storage of transformants

-YPDA broth-25% glycerol.

-Sterile glass beads.

-1.8 ml sterile cryovials

-Hemocytometer

-SD/-Leu-km agar plates

Library sequencing

-2 cryovials.

-Zymoprep Yeast Plasmid Miniprep II kit (Zymo Research, Irvine, CA, USA).

-10 µM AD-screen-Fwd (5´- CTATTCGATGATGAAGATACCCCACCAAACCCA-3').

-10 µM AD-screen-Rv (5'- GTGAACTTGCGGGGTTTTTCAGTATCTACGATT-3').

-ADVANTAGE 2 POLYMERASE MIX (Clontech, Mountain View, CA, USA).

-Wizard Ⓡ SV Gel and PCR Clean-Up System (Promega, Madison, WI, USA).

Bioinformatic analysis

-A computer with at least the following characteristics is required: Linux Ubuntu 20.04, 8 TB hard drive, 64 GB RAM, i9-109000 cpu 2.80GH x20, NVDIA RTX3070.

Troubleshooting

Safety warnings

⛔ **Note 1.** DEPC treatments must be done in a fume extraction hood for a minimum of 8 h with continuous agitation. In addition, personal protective equipment must be used when handling DEPC. To inactivate the DEPC, it must be sterilized twice for 15 min.

Note 2. To avoid heating the sample during mixing in the bead beater, pre-chill the bead beater and keep at 4°C, break the cells for 1 min, return to ice 1 min, and repeat until complete three times.

Note 3. The 5 min incubation at room temperature allows the dissociation of protein complexes from nucleic acids. This is the only step that is carried out at room temperature. With this exception, the entire procedure should be carried out at 4°C.

Note 4. If after the first chloroform extraction the interface looks white, repeat the chloroform extraction until it almost disappears. It is important to be aware that with each extraction RNA is lost.

Note 5. To obtain higher yield of RNA, pass the filtrate through the column again.

Note 6. Repeat the elution with 30 μl more of water to recover as much RNA as possible.

Note 7. To improve the binding of the mRNA to the Oligotex resin, incubate an additional 30 min at room temperature.

Note 8. To obtain higher yield of mRNA, repeat the elution with the same volume. To concentrate the mRNA, precipitate with ethanol and 3 M sodium acetate, pH 5.2 (3:0.1 vol).

Note 9. Do the necessary LD-PCR reactions to obtain the minimum concentration of 3 μg of ds cDNA.

Note 10. Apply the LD-PCR volume of reaction at the center of the surface of the gel matrix to prevent sticking to the column walls.

Note 11. In order to prevent cell precipitation, it is important to grow the yeasts in a large area and a small volume of culture medium to allow for proper agitation and oxygenation.

Note 12. If the libraries density is $<2\times10^7$ cells/ml, centrifuge to reduce the solution volume.

Note 13. If the supernatant appears cloudy, collect the supernatant in another tube and centrifuge for an additional 5 min.

Note 14. By using low volume columns, the eluted volume can be reduced to 10 μl and obtain a higher concentration of plasmid.

Note 15. To maximize plasmid yield, run the filtrate through the column again.

Note 16. To recover more plasmid, perform a second elution.

Before start

To render solutions and labware nuclease-free, treat all solutions and materials with DEPC water (0.1% DEPC in ultrapure water; **see Note 1**), inactivate the DEPC by sterilization of all materials and solutions. The Tris-HCl solution must be prepared with sterile inactivated DEPC water. **RNA extraction procedures must be carried out at 4°C.**

RNA Extraction

1   Resuspend the cells in TRI-reagent and transfer them to a vial with glass beads. Break the cells in the bead beater for 3 min at 42 of velocity (***see*** **Note 2**). Centrifuge 10 min, 20,000 x g at 4°C.

2   Transfer the organic red phase to a nuclease-free tube, add 600 µl chloroform, mix and hold at room temperature for 5 min (***see*** **Note 3**). Transfer to ice 10 min and centrifuge 10 min, 20,000 x g at 4°C (***see*** **Note 4**).

3   Transfer the organic upper phase to a nuclease-free tube, add 2 vol isopropanol and 1/10 vol 3 M sodium acetate, pH 5.2. Mix, incubate 10 min on ice and centrifuge 15 min at 20,000 x g at 4°C.

4   Wash the pellet 3 times with 80% ethanol by centrifugation for 5 min and resuspension between each wash. Let the pellet dry and resuspend in 100 µl nuclease-free water.

5   Quantify, check integrity on an agarose gel, and store at -80°C.

6   **DNase treatment.** Thaw the RNA on ice and add 10 µL of 10X DNase I reaction buffer and 1 µl (2 U) of DNase I per 5 to 8 µg of total RNA. Adjust to a final volume of 100 µl with nuclease-free water.

7   Incubate 10 min at 37°C and inactivate the DNase I with 1 µL of 0.5 M EDTA (5 mM final concentration) for 10 min at 75°C.

8   **RNA CleanUp.** To 100 µl of RNA add 350 µl of RLT buffer with 1% of 2-mercaptoetanol, mix and add
250 µl ethanol.

9   Transfer to a mini spin column, incubate 5 min on ice and centrifuge 15 s at 20,000 x g, 4°C (***see*** **Note 5**).

10 Wash the column with 500 µl of RPE buffer, centrifuge 15 s at 20,000 x g. Repeat the wash and centrifuge the empty column for 2 min.

11 Transfer the column to a nuclease-free tube, add 40 µl of nuclease free water, incubate on ice 2 min and centrifuge 1 min at 20,000 x g (***see*** **Note 6**). Quantify, check integrity on an agarose gel, and store at -80°C.

mRNA Purification

12   Bring the RNA to a 500 µl vol with nuclease-free water. Heat the OBB buffer to 70°C and add 500 µl. Add 30 µl of Oligotex resin heated to 37°C, mix and incubate 8 min at 70°C (**see Note 7**).

13   Centrifuge at 20,000 x g for 3 min at 4°C. Discard the supernatant and resuspend the Oligotex resin with 400 µl of OW2 buffer. Transfer the mix to a spin column into a collector tube and centrifuge 1 min at 20,000 x g and 4°C. Transfer the column to another collector tube, add 400 µl of OW2 buffer and centrifuge 1 min at 20,000 x g, 4°C. Discard the flowthrough and centrifuge again the column for 1 extra min.

14   Transfer the column to a nuclease-free tube, add 120 µl of 70°C OEB buffer to the resin. Incubate the column at 70°C for 2 min. Centrifuge 1 min at 20,000 x g (**see Note 8**).

15   Quantify and store at -80°C.

cDNA Synthesis

16   Use a 0.2 ml PCR tube to mix 0.3 µg of mRNA with 1 µl of 10 µM CDSIII/6 oligo, nuclease-free water until 4 µl final volume, mix and heat to 72°C, 8 min and quickly transfer to 4°C, 2 min.

17   In another tube add 2.0 µl 5X first-strand buffer, 1.0 µl 100 mM DTT, 1.0 µl 10 mM dNTP Mix, 1.0 µl SMART MMLV Reverse Transcriptase.

18   Add the mixture to the mRNA-oligo and incubate 10 min at 25°C, then after 10 min at 42°C, add 1 µl of SMART III oligo (only for the SMART Library), and hold the reaction 60 min at 42°C. To end the reaction heat at 75°C, 10 min.

LD-PCR

19   Prepare 2 reactions of 100 µl for each library. Use 4 µl of cDNA mixed with 140 µl nuclease-free water, 20 µl 10X Advantage buffer, 4 µl 50X dNTP mix, 4 µl 10 µM DINO-SL oligo (for the DINO-SL library) or 4 µl 10 µM 5′ LD-PCR oligo (for the SMART library), 4 µl 10 µM 3′ LD-PCR oligo, 20 µl 10X melting solution, 4 µl 50X advantage polymerase mix.

20   Add 100 µl of the mix in two 200 µl tubes for PCR.

21   Set the thermocycler with the following program: 1. 95°C, 30 s; 2. 95°C, 10 s; 3. 68°C, 6 min + 5 s each cycle; repeat 20 times from 2 to 3; 4. 68°C, 5 min; 5. 4°C, 5 min.

22   Analyze 7 µl of each LD-PCR reaction in a 1.2% agarose gel and quantify until you get at least 3 µg of ds cDNA (**see Note 9**).

ds cDNA Fractionation

23   Use one CHROMA SPIN TE-400 column for each LD-PCR reaction (93 µl of ds cDNA). Mix by flipping the column to resuspend the gel in the equilibration buffer. Then, remove the column cap and break from the bottom.

24   Place the column in a 2 ml collector tube and centrifuge for 5 min at 700 x g to eliminate the buffer.

25   Use another collection tube to place the column and apply the LD-PCR reaction on the gel matrix (**see Note 10**). Centrifuge for 5 min at 700 x g. Combine the filtrates and add 3 vol of ethanol and 1/10 vol of 3 M sodium acetate, pH 5.2. Incubate 1 h at -20°C.

26   Centrifuge at 20,000 x g for 20 min at room temperature, discard the supernatant and air dry the pellet for 10 min. Resuspend the ds cDNA in 20 µl nuclease-free water.

27   Quantify and analyze on an agarose gel.

Preparing Y187 Competent Cells

28   Streak the Y187 strain from a frozen stock in a YPDA-km agar plate, incubate upside down for 3 d at 30°C.

29   Inoculate three separate tubes with 3 ml of YPDA-km with three 2-3 mm Y187 colonies (in 10 ml culture tubes). Incubate with agitation overnight (ON) at 30°C and 250 rpm.

30   Add 10 µl of the fastest growing ON culture to 100 ml of YPDA-km, divide the inoculum into two 0.5 L flasks (50 mL in each flask) and incubate at 30°C with 250 rpm agitation (**see Note 11**). After 18 h check the OD600 until it reaches 0.15-0.3 (18-21 h).

31   Centrifuge the cultures 5 min at 800 x g at room temperature, discard the supernatant and resuspend the cells (from the fastest growing culture) with 200 ml of 2X YPDA-km and transfer to a 1 L flask. Incubate at 30°C and 250 rpm, until the OD_{600} reaches 0.4-0.5 (it takes 4-5 h).

32   Centrifuge the culture 5 min at 800 x g, discard the supernatant and resuspend the cells with 120 ml sterile deionized water. Centrifuge 5 min at 800 x g and resuspend the cells with 6 ml of fresh 1.1X TE/LiAC. Transfer to 1.6 ml tubes and centrifuge at 20,000 x g, 15 s

at room temperature. Discard the supernatant and resuspend with 2.4 ml of 1.1X TE/LiAc, transfer to ice.

Y187 Transformation

33 In a 1.6 ml tube add 160 µl of 10 mg/ml denatured herring testes carrier DNA and boil at 95-100°C for 5 min, quickly transfer to ice for 5 min and boil again 5 min, transfer to ice.

34 Prechill 4 sterile 15 ml tubes on ice (2 tubes for each library). Add to each tube 6.5 µl of 0.5 µg/µl of pGADT7-Rec vector (*sma* I linearized) and 60 µl (20 µg) of ds cDNA for the SMART library, or 60 µl (18 µg) of the DINO-SL ds cDNA, mix gently by tapping. Meanwhile, boil the herring testes carrier DNA again (quickly cool on ice and spin). Add 40 µl denatured herring testes carrier DNA to the mix of plasmid-ds cDNAs. Mix gently by tapping.

35 Add to each tube 600 µl of Y187 competent cells, mix gently, add 2.5 ml of PEG/LiAC and mix gently. Incubate at 30°C for 45 min (gently vortexing every 15 min).

36 Add 170 µl of DMSO, mix gently. Incubate in a water bath for 30 min at 42°C (gently vortexing every 10 min).

37 Centrifuge the tubes at 800 x g for 5 min at room temperature. Discard the supernatant and carefully resuspend the cells with 3 ml 2X YPDA. Incubate at 30°C for 90 min at 250 rpm.

38 Centrifuge at 800 x g, 5 min at room temperature. Discard the supernatant and resuspend with 0.9% NaCl at a final volume of 5 ml.

39 Combine the two transformations from each library in one tube. Make 1:10 and 1:100 dilutions of each library transformation and streak 100 µl in 100 mm agar SD/-Leu-km to calculate the number of independent clones and the transformation efficiency. Spread with glass beads the remainder of the transformations on 150 mm agar SD/-Leu-km plates (200 µl of transformed cells/plate on 50 plates).

40 Incubate the plates upside down at 30°C until colonies appear (3-6 d).

Harvesting Transformants

41 Transfer the plates to 4°C for 3 h. Add 5 ml of YPDA-25% glycerol-km per plate and 5-8 glass beads. Shake gently the plates to detach the colonies and recover the cells with a micropipette. Repeat with all the plates. The yield should be ~ 250 ml for each library.

42 Determine the cell density of the libraries by counting cells in a hemocytometer with a 1:10 dilution (5.9×10^8 and 6×10^8 cells/ml for the SMART and DINO-SL libraries respectively) (**see Note 12**).

43 Take 10 µl of each library and make a dilution "A" by adding 990 µl of YPDA-km. Take 10 µl of dilution A and add 990 µl of YPDA-km to obtain a dilution "B". From dilution A spread 100 µl, 50 µl and 10 µl + 50 µl of YPDA-km on duplicate agar SD/-Leu-km plates. For dilution B spread only 100 and 50 µl also in duplicate plates. Incubate the plates upside down for 3-6 d and count the colonies to calculate the cfu/ml.

44 Aliquot the libraries in 1 ml aliquots and store at -80°C.

Libraries Plasmid Extraction

45 Thaw two aliquots of each library in a water bath at 25°C. Centrifuge at 16,000 x g for 2 min at room temperature. Discard the supernatant and resuspend with 1 ml resuspension solution, centrifuge at 16,000 x g for 2 min and discard the supernatant.

46 Resuspend with 200 µl of digestion buffer, add 3 µl of zymolyase and incubate at 37°C for 5 h.

47 Add 200 µl of lysis buffer and mix quickly by inverting the tube. Incubate 5 min at room temperature and combine with 400 µl of neutralizing buffer, mix by inverting the tube.

48 Centrifuge at 16,000 x g for 3 min (**see Note 13**) and transfer the supernatant to a low volume column (see **Note 14**). Incubate for 5 min at room temperature and centrifuge 30 s at 16,000 x g (**see Note 15**). Discard the flowthrough and wash the column 3 times with 500 µl of DNA wash buffer, centrifuge 1 min at 16,000 x g. Spin-dry the column 3 additional min, transfer the column to a sterile tube and add 10 µl of elution buffer at the center of the resin in the column, incubate at 60°C for 2 min, and centrifuge at 16,000 x g for 3 min (**see Note 16**).

49 Store the plasmids at -20°C.

PCR of Plasmids from the Libraries

50 In a 0.5 ml tube for each library combine 20 µl of 10X Advantage 2 PCR buffer, 4 µl of 10 mM dNTP mix, 4 µl of AD-Screen Fwd primer, 4 µl of AD-Screen Rv primer, 4 µl of 50X Advantage 2 polymerase mix, 9 µl of the library plasmid, and 155 µl of nuclease-free water. Mix and transfer 100 µl in two 0.2 ml PCR tubes.

51 Program the thermocycler as follows: 1. 95°C, 3 min; 2. 95°C, 30 s; 3. 56°C, 30 s; 4. 68°C, 6 min; repeat from 2-4 for 30 times; 5. 68°C, 10 min; 7. 4°C, 5 min.

52 Analyze 5 μl from each PCR reaction in a 0.5% agarose gel.

53 Combine the two PCR reactions, add 200 μl of membrane binding solution, mix and transfer to a low volume column, incubate for 5 min at room temperature and centrifuge at 16,000 x g for 1 min.

54 Wash the column 3 times with 500 μl of washing solution with centrifugations at 16,000 x g for 1 min. Spin-dry the column for 3 additional min.

55 Add 30 μl of nuclease-free water to the center of the resin in the column, incubate at 65°C for 1 min and centrifuge at 16,000 x g for 1 min.

56 Analyze on a 0.5% agarose gel and quantify.

Sequencing of Libraries

57 20 μg in a 100 μl volume of PCR products (of each library) are sent for sequencing.

58 Sequencing is carried out by Illumina (2 × 75 cycles, 10 million reads).

Bioinformatic Analysis

59 **Read quality control and cleaning.** Fastp was used to perform quality control and cleaning on each of the Fastq files from the libraries. The latest version of Fastp v0.32.2 can be downloaded and installed using Anaconda (Chen et al. 2018). Anaconda can be installed from https://docs.anaconda.com/free/anaconda/install/index.html, with the following lines:

This is a comment, and they explain the function of the code
More information about fastp can be found on the github
https://github.com/OpenGene/fastp
Install fastp in Conda terminal
conda install -c bioconda fastp

```
# If you have single end data (Example Smart library)
fastp -i SMART_S62_R1_001.good.fastq -o out.Smart.R1.fq -h SmartFastp.html
# If you have paired end data
fastp -i SMART_S62_R1_001.good.fastq -I SMART_S62_R2_001.good.fastq -o
out.Smart.R1.fq -O out.Smart.R2.fq -h SmartFastp.html
```

60 ***De novo* assembly.** Trinity is used for the *de novo* assembly using sequences free of adapters, overrepresented sequences, low quality sequences, and sequences of low complexity (Haas et al. 2013).

```
# In Linux command line
# Download Trinity
# https://github.com/trinityrnaseq/trinityrnaseq/releases

Trinity --seqType fq --left out.Smart.R1.fq --right out.Smart.R2.fq --CPU 20 --
max_memory 32G --min_contig_length 150

#The above command generates a file named Trinity.fasta
#Rename file
mv Trinity.fasta SmartTrinity.fasta
```

61 **Identification of potential coding regions.** To find potential coding regions from the assembled transcripts, TransDecoder is then employed.

```
# In the Linux command line
# More information about TransDecoder can be found on the github
# https://github.com/TransDecoder/TransDecoder/wiki

#Identifies all long ORFs
TransDecoder.LongOrfs -t SmartTrinity.fasta

#Identifies the ORFs that are most likely to be coding
TransDecoder.Predict -t SmartTrinity.fasta
```

62 **Identification of homologous sequences.** The candidate coding regions were used for BLAST analysis (Altschul et al. 1990) against the genome draft of *S. microadriaticum* from *Stylophora pistillata* (Aranda et al. 2016). At present, the genome from *S. microadriaticum* CassKB8 has also become available (González-Pech et al. 2021) and could be used as reference.

```
#Download reference genome of S. microadriaticum
# http://smic.reefgenomics.org/download/
```

```
#Download Blast from NCBI
# https://ftp.ncbi.nlm.nih.gov/blast/executables/LATEST/
#Instruction about install
# https://www.ncbi.nlm.nih.gov/books/NBK569861/
# Smic.genome.annotation.pep.longest.fa

# In Linux command line
# Make a database from S. microadriaticum
makeblastdb -in Smic.genome.annotation.pep.longest.fa -out Smicroadriaticum -
parse_seqids -dbtype prot

# Make a blast using blastp
blastp -db Smicroadriaticum -query SmartTrinitytransdecoderpep.fa -evalue 10e-5 -
qcov_hsp_perc >90 -outfmt 6 -out SmartTrinitypep.tab

# We need run the same code for the Dino library to obtain the file
# DinoTrinitypep.tab, that contain the homologues between
# DinoTrinitytransdecoderpep.fa and reference genome.
```

63   **Identification of unique genes in each library.** After the Blast, each library can contain more than one homolog with the reference genome, to identify the coding sequences that correspond to a unique protein in the reference genome, the unique proteins in each library will be filtered using python.

```
#Using Python 3.9
import os
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from matplotlib_venn import venn2

# Read homologous proteins from Blasp to the Smart library.
# We will identify the unique proteins in the Smart library with respect to the proteins of
the reference organism.

smartHomo =pd.read_csv('SmartTrinitypep.tab', sep='\t')
df=smartHomo[['Protein']]
SMARTunq=df.drop_duplicates()

SMARTunq["subject"] = 'aa'
SMARTunq["identity"] = 0
for row_tuple in SMARTunq.itertuples():
    temp=0
```

```python
    for row_tuple1 in smartHomo.itertuples():
        if row_tuple.Protein == row_tuple1.Protein:
            if row_tuple1.identity > temp:
                subject=row_tuple1.subject
                temp=row_tuple1.identity
    SMARTunq.at[row_tuple.Index, 'subject'] = subject
    SMARTunq.at[row_tuple.Index, 'identity'] = temp
    subject=0

ddf=SMARTunq[['subject']]
SMARTunq2=ddf.drop_duplicates()
SMARTunq2["Protein"] = 'aa'
SMARTunq2["identity"] = 0
for row_tuple in SMARTunq2.itertuples():
    temp=0
    for row_tuple1 in SMARTunq.itertuples():
        if row_tuple.subject == row_tuple1.subject:
            if row_tuple1.identity > temp:
                Protein=row_tuple1.Protein
                temp=row_tuple1.identity
    SMARTunq2.at[row_tuple.Index, 'Protein'] = Protein
    SMARTunq2.at[row_tuple.Index, 'identity'] = temp
    Protein=0

SMARTunq2.to_csv("SMARTUniq.txt", sep='\t', index=False, header=True)
```

The same is carried out for the Dino-SL library.

```python
# Continuation of the previous python code
# Read homologous proteins from Blasp to the Dino library.
# We will identify the unique proteins in the Dino library with respect to the proteins of
the reference organism.
dinoHomo=pd.read_csv('DinoTrinitypep.tab', sep='\t')
df=dinoHomo[['Protein']]
DINOunq=df.drop_duplicates()

DINOunq["subject"] = 'aa'
DINOunq["identity"] = 0
for row_tuple in DINOunq.itertuples():
    temp=0
    for row_tuple1 in dinoHomo.itertuples():
        if row_tuple.Protein == row_tuple1.Protein:
```

```python
                if row_tuple1.identity > temp:
                    subject=row_tuple1.subject
                    temp=row_tuple1.identity

    DINOunq.at[row_tuple.Index, 'subject'] = subject
    DINOunq.at[row_tuple.Index, 'identity'] = temp
    subject=0

ddf=DINOunq[['subject']]
DINOunq2=ddf.drop_duplicates()

DINOunq2["Protein"] = 'aa'
DINOunq2["identity"] = 0
for row_tuple in DINOunq2.itertuples():
    temp=0
for row_tuple1 in DINOunq.itertuples():
        if row_tuple.subject == row_tuple1.subject:
            if row_tuple1.identity > temp:
                Protein=row_tuple1.Protein
                temp=row_tuple1.identity
    DINOunq2.at[row_tuple.Index, 'Protein'] = Protein
    DINOunq2.at[row_tuple.Index, 'identity'] = temp
    Protein=0

DINOunq2.to_csv("DINOUniq.txt", sep='\t', index=False, header=True)
```

Finally, we can identify the similar proteins between the two libraries by means of a Venn diagram.

```python
# Continuation of the previous python code
# Venn Diagram
SMARTlist = SMARTunq2['subject'].tolist()
DiNOlist = DINOunq2['subject'].tolist()
venn2([set(SMARTlist), set(DiNOlist)],set_labels = ('SMART', 'DINO'),set_colors=
('purple', 'skyblue'), alpha = 0.7)
plt.title("Similar genes in libraries")
```

```bash
#Code for filtering coding regions from each library using python's list of unique proteins
# In Linux command line
# The column of unique sequence identifiers is cut, and a temporary file is created
cut -f 2 ./Blast/SMARTUniq.txt | cut -f1 -d"." > Smartuniqtemp.txt
```

> \# We search and retrieve the sequences from the original SmartTrinity.fasta file
> grep -A1 -f Smartuniqtemp.txt SmartTrinity.fasta > SmartUniq.fasta

64   **Conserved domains.** The presence of conserved domains in the assembled transcripts were identified and annotated using HMMER (Potter et al. 2018). We use the unique proteins in each library to run a HMMER search against the Pfam database and identify conserved domains.

```
# HMMER installation introduction
# http://hmmer.org/documentation.html
#For Linux: apt install hmmer
# hmmscan requires a domain database to use as a reference, we can get the latest
# Version of Pfam from the following link.
# http://ftp.ebi.ac.uk/pub/databases/Pfam/current_release/

hmmscan --cpu 10 --domtblout SmartPFAM.out / Pfam-A.hmm /UniqueSmart.fasta
```

65   **KEGG pathway.** To identify the metabolic pathways associated with each library use GhostKOALA (Kanehisa et al. 2016), https://www.kegg.jp/ghostkoala/, which allows assigning K numbers to the sequence data, this is a web service, where the filtered sequences will be uploaded for each library, and the database of genus_prokaryotes + family_eukaryotes will be used.

66   **Kozak sequence search.** Finally, the DNA-pattern program (Santana-Garcia et al. 2022) will be used to identify the Kozak sequence (RCCATGGCN), which allows to search all occurrences of a pattern within DNA sequence. The searches are done only in the direct strands, matching positions can be calculated either relative to the sequence start and 1 substitution allowed.

Protocol references

Aranda M, Li Y, Liew YJ, Baumgarten S, Simakov O et al. Genomes of coral dinoflagellate symbionts highlight evolutionary adaptations conducive to a symbiotic lifestyle. Sci Rep 2016;**6**:39734. DOI: 10.1038/srep39734.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ et al. "Basic local alignment search tool". J Mol Biol 1990;**215**:403-10. DOI: 10.1016/S0022-2836(05)80360-2.

Chen S, Zhou Y, Chen Y, Gu J et al. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 2018;**34**(17):i884-90. DOI: 10.1093/bioinformatics/bty560.

Fields S, Song OK. A novel genetic system to detect protein–protein interactions. Nature 1989;**340**:245-46. DOI: 10.1038/340245a0.

González-Pech RA, Stephens TG, Chen Y, Mohamed AR, Cheng Y et al. Comparison of 15 dinoflagellate genomes reveals extensive sequence and structural divergence in family Symbiodiniaceae and genus *Symbiodinium*. BMC Biol 2021;**19**(1):73. DOI: 10.1186/s12915-021-00994-6.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD et al. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature Protocols 2013;**8**(8):1494-512. DOI: 10.1038/nprot.2013.084.

Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol 2016;**428**(4):726-31. DOI: 10.1016/j.jmb.2015.11.006.

Islas-Flores T, Galán-Vásquez E, Villanueva MA. Screening a spliced leader-based *Symbiodinium microadriaticum* cDNA library using the Yeast-Two Hybrid System reveals a hemerythrin-like protein as a putative SmicRACK1 ligand. Microorganisms 2021;**9**:791. DOI: 10.3390/ microorganisms9040791.

Potter SC, Luciani A, Eddy SR, Park Y, Lopez R et al. HMMER web server: 2018 update. Nucleic Acids Res 2018;**46**(W1):W200-4. DOI: 10.1093/nar/gky448.

Santana-Garcia W, Castro-Mondragon JA, Padilla-Gálvez M, Nguyen NTT, Elizondo-Salas A et al. RSAT 2022: regulatory sequence analysis tools. Nucleic Acids Res 2022;**50**(W1):W670-6. DOI: 10.1093/nar/gkac312.

Van Criekinge W, Beyaert R. Yeast Two-Hybrid: State of the art. Biol Proced Online 1999;**2**:1-38. DOI: 10.1251/bpo16.