Sequencing Plan June 2023

Spreadsheet: <https://docs.google.com/spreadsheets/d/1fGlcBtHMxDxXAEOFDwigc_M6qqk_vagNiNNSzAIAcM4/edit?usp=sharing>

QC all gDNA samples first please and contact Spri clean up if needed before proceeding to library prep. Sequencing is novaseq S4 2X150bp end. For samples LMD129, LMD229, LMD329, LMD429, LMD529, LMD629, LMD729, LMD929, LMD1129, LMD1329 and LMD1529 (11) do 26M read pairs. For samples LMD829, LMD1029, LMD1229, and LMD1429 (4) do 80M read pairs. Do the same thing both Sam church and Namrata have been asking for from other projects in the Dunn lab please (:

For species ID - Genome skimming and mitochondrial genome for COI/16S/18S

* Early Cyanea from 2020/2021 year
* Late cyanea from 2020/2021 year
* Early Cyanea from 2023 year
* Late cyanea from 2023 year
* Aurelia polyps in culture
* Clytia polyps in culture (if available)

For Cyanea capillata genome assembly - PacBio IsoSeq & Genome skimming

* One individual multiple tissues
* This is the late and bay species

For Cyanea fulva genome assembly - Send to Mark for Hi-Fi, PacBio IsoSeq & Genome skimming

* One individual multiple tissues
* This is the early and river species

For Cyanea pop gen studies - Illumina Genome skimming

* 10 individuals per site for early season = 30
* 10 individuals per site for mid-season = 30
* 10 individuals per site for late-season = 30
* If possible get 10 individuals of both yellow and red but 10 in general as minimum
* Total per year = 90 samples
* Concatenate from last season = TBD
* Repeat next season if applicable

Total budget all sequencing

* Illumina $100 per sample x 90 samples pop gen = $9,000
* Illumina $100 per sample x 8 samples ID/genome = $800
* PacBio sequencing $ per sample x 2 samples (repeat if necessary for quality) = $

Where funding is coming from

* Casey Waterman? = $
* Mary Beth funding = $
* YIBS grant ($5,000) = $

To do right now june16-2023 = 24 illumina & 2 pacbio

Types of sequencing mentioned above

Genome skimming

NovaSeq S4 paired-end 2x150 line at YCGA is $5,297

<https://medicine.yale.edu/keck/ycga/services/illuminaprices/>

Generates 2400 Gb of data, Call it $2/Gb, Library prep is $58

<https://www.illumina.com/systems/sequencing-platforms/novaseq/specifications.html>

IsoSeq PacBio

Extract RNA and get decent values, Send plenty of volume (30ul)

Once sent they QC it and start reaction with 500ng RNA (in 7ul)

Targeted Library size (kb) = ask - weird - full length mRNA

Number of SMRT Cells = 1

By year/season/location/species my collection

* Early = March, April, May
* Early/Mid = May
* Mid = May, June, July
* Mid/Late = July
* Late = July, August, September, October

| **Year** | **Season** | **Location** | **Species** | **#** | **PCR** | **Illumina** | **To Do** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 2020 | Mid/Late | Niantic | Chrysaora | 2 | 1 |  |  |
| 2020 | Late | Groton | **Cyanea** | 1 | 1 |  | NA |
| 2020 | Late | Cedar Island Marina | Chrysaora | 4 | 1 |  |  |
| 2020 | Late | Cedar Island Marina | **Cyanea** | 6 | 1 |  | 4 |
| 2020 | Late | Culture | Chrysaora | 3 | 1 |  |  |
| 2020 | Late | Culture | **Cyanea** | 4 | 1 | 2 |  |
| 2021 | Early | Norwalk | Chrysaora | 1 |  |  |  |
| 2021 | Early | Norwalk | **Cyanea** | 2 |  |  | 2 |
| 2021 | Early | Cedar Island Marina | **Cyanea** | 1 | 1 |  | 1 |
| 2021 | Early | Lighthouse point park | **Cyanea** | 2 |  | 1 | 1 |
| 2021 | Early/Mid | Sound School | **Cyanea** | 2 |  |  | 2 |
| 2021 | Early/Mid | Cedar Island Marina | **Cyanea** | 6 | 1 |  | 4 |
| 2021 | Late | Cedar Island Marina | **Cyanea** | 1 | 1 |  | 1 |
| 2021 | Late | Horse Island | Chrysaora | 1 |  |  |  |
| 2022 | Early | Horse Island | **Cyanea** | 2 | 2 |  | 2 |
| 2023 | Early | Cedar Island Marina | **Cyanea** | 3 |  |  | 3 |
| 2023 | Early | NH Harbor | **Cyanea** | 4 |  |  | 4 |
| 2023 | Early/Mid | Cedar Island Marina | **Cyanea** | 18 |  |  | 10 |

Casey Slack notes - June 21, 2023

* Chrysaora
  + Published Chrysaora quinquecirrha genome not chromosome scale but reasonable completeness (80% eukaryotic BUSCO) and contiguity (scaffold N50 length of 733.65 Kb)
  + <https://www.frontiersin.org/articles/10.3389/fgene.2020.00535/full>
  + So unless it turns out to be distantly related, I can use this as a reference for this species
  + No need for IsoSeq/HiFi/HiC
  + Genome size is 330.67 Mb and 20x would be 7Gb = $52 sequencing cost
  + Each chrysaora is $100 sample prep and $52 sequencing = $150
  + Start with 2-3 Chrysaora quinquecirrha specimens for Illumina and make sure it maps well and then we evaluate next steps based on that
  + If too different we need to consider doing IsoSeq or a genome to build a reference or skipping this species
* Cyanea
  + Paper (<https://link.springer.com/article/10.1007/s12562-016-1050-4>) suggests Cyanea capillata genome is about 400 Mb so 20x coverage is 8 Gb = $64
  + Illumina genome skimming will cost $170 per sample

Other

* Technical stuff
  + No cyanea capillata genome
  + Contiguity - how big pieces
    - N50
    - Length - size of piece of middle
  + Completely - how big
  + Illumina 150 base chunks
  + Physalia 90% readings map
  + Then few more sample
  + 3 one year 3 another year
  + variation/mapping
  + Start 1 from each year
  + $150
  + 20x coverage
  + $300 up front and max all 6
  + How map going
  + Work through workflow with good genome
* Cyanea
  + C.f.fulva is a maybe
  + Really diff or similar
  + Capillata
    - Preliminary is small
    - Japan paper bigger
    - No reference genome
    - Need to map
    - 2 in progress in darwin tree - HiC
    - Maybe 300 Megs
    - 354 Mg
    - But from HiC
    - Do the mitochondrial genome comparisons
    - SRA accession
    - Check page is updated often
    - Two kinds pacbio
    - Variants only protein coding genes
    - Only exons
    - Ignore anything else
    - No genome
    - Transcriptome
    - mRNA
    - Pacbio isoseq
    - Messenger rna and full length
    - Need gene models anyway
    - $500 sample prep $1500 for sequencing
    - Only need to do once
    - Gene models
    - Map illumina reads to those genes
    - Can we use one reference for both
    - Do pacbio for one of them
    - Skimming for both fulva and capillata $150
    - # samples 2-3 samples
    - See if it works for both
    - If not do second transcriptome
    - No fraction of map reads
    - Perfect mapping would be 20% reads
    - After mapping - how deep mapping in - how deep do they stack - 20 fold
    - Eye ball it
    - Mapping technical thing
    - Variant calling - ID sites different between map and reference
    - Collect thousands variants and collect pop work - get variants
    - Have 1 late already
    - Work with field collected ones
    - 2-3 in years
    - Culture not really
    - Chrysaora
      * 2 - 1 per year
      * $300 for skimming
      * Illumina 20x
    - Cyanea cap
      * One from lab cross
      * 2 here
      * Pick 1 as existing
      * $300
      * Illumina 20x
    - Cyanea fulva
      * 1 illummina 80x for genome size
      * 200 sequencing for 100 sample prep
      * 1 illumina at 20x for $150
    - $1050 for illumina only - first technical
      * 600 up front
    - 2-4 weeks for turn around time minimum 11 days and max 17-18 days
    - 2-3 weeks
    - C cap
      * Iso seq
      * $2000
      * Building transcriptome
      * RNA extraction - multiple tissues pulling together
    - C fulva
      * Iso seq
      * $2000
      * Building transcriptome
      * RNA extraction - multiple tissues pulling together
    - Total 2600
    - No
      * No map well fulva to cap iso seq
      * Then get iso seq for fulva
      * Cant map - so much pop structure
    - Think about questions
      * Yes or no positive data - is a story
  + Phases
    - One - Technical
      * What do we need to get it to work
      * Minimum 2600
    - Two - yes
      * One from each year to three from each year
      * Done how many samples etc
      * 20 samples
    - Biological
      * Enough from each population
      * 6 samples from each = 30
      * Order of 20 samples of illumina
      * 3000 dollars
    - Scale
      * Locations
      * 40 more samples
      * Another 6000
    - Doesnt work as reference
      * Empirical question
      * Other genomes not available
      * For both
      * In order to move forward 2 or 3
      * Own genome
      * HiFi - HMW- DNA $2000
      * HiC - send raw tissue ~$1500
      * Do not need illumina for that
      * Genome $4k
      * Even get to two we could drop 8K
      * Start iso seq for cap to figure it out
  + MB - cover portion
  + Pop gen change YIBS
  + Local specimens in peabody - IZ
  + Publish accession numbers
  + Proceed first - good sense and where to be
  + Decide other stuff
  + Scypho diversity
  + agency /funder
  + Big expenses for this $2600
  + First
    - 2 chrysaora
    - 7 Gb
    - Ask for 8Gb
    - 8000/300 = 26
    - millions of reads
    - 26 million read pairs each
    - Wild collect 1 per year
  + First - 2
    - C.fulva 1
    - 300
    - 80x
    - 24 Gb
    - 24000/idk
    - 80 million reads
    - One sample 80 million
  + First 2 -
    - One sample isoseq for capillata
* Stopping points
* Forks in the road

Cyanea capillata

* Tree of Life que <https://tolqc.cog.sanger.ac.uk/darwin/jellyfish/Cyanea_capillata/>
  + Specimen used -

Continue later

<https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Info&id=27804>