Dsembler User Manual

Aporva Gupta

July 12th, 2021

Contents

Online Interface							
Input							
Output							
Excel File							
FASTA file							
Previous Work							
Sign up							
Login							

1. Launching Dsembler

Dsembler can be installed via GitHub.

```
# obtain dsembler package from GitHub
$ git clone https://github.com/sblabkribb/dsembler.git
```

Docker (recommended)

Users should have docker installed on their computer to run Dsembler locally. Docker can be downloaded from https://docs.docker.com/get-docker/. Ensure your BIOS settings are compatible with the docker application. Once docker is installed successfuly, run the following command in your terminal (or Powershell in Windows).

A Dsembler image will be built locally using the Dockerfile present in the Dsembler source folder. Alternatively, pull the image from docker using the following command

```
# build the docker image
$ docker build -t dsembler:latest .
OR
```

```
# pull image from dockerhub
$ docker pull sblabkribb/dsembler:latest
```

To run the web application, run the following command in your terminal or powershell.

Run docker container: Bash

```
$ docker run -d --rm -v $(pwd):/app --publish 5000:5000 --name dsembler dsembler
```

Run docker container: Powershell

```
> docker run -d --rm -v %cd%:/app --publish 5000:5000 --name dsembler dsembler
```

This will run the Dsembler application on your localhost via port 5000.

Python Dsembler can be run from the user's python environment as well. The user is responsible for installing the required packages (see requirements.txt) to run the program. Bash

```
$ export FLASK_APP=app
$ flask run
```

Powershell

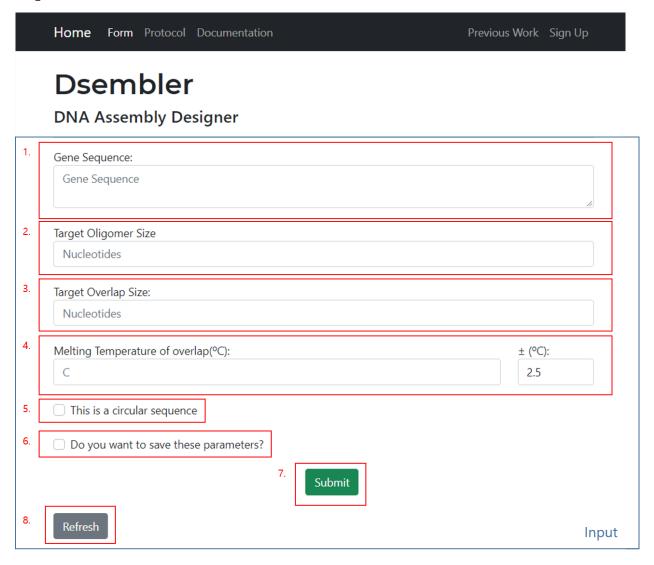
```
> $env:FLASK_APP = "app"
> flask run
```

This would publish the application on your localhost:5000 port

2. Online Interface

Dsembler has a simple, easy to use user interface.

Input



- 1. **Gene Sequence:** Users can input the gene sequence as plain text. The Application will not run if non-DNA elements are used (spaces and newlines are removed by the program automatically).
- 2. Oligomer size: Input the target oligomer size. The oligomers must be at least 50bp
- 3. Overlap size: Input the target overlap size. Overlaps must be at least 20bp and must be at least 30bp smaller than the target oligomer size.
- 4. Melting Temperature & Temperature range: The target melting temperature is

inputted in ${}^{\circ}$ C. An acceptable range is required from the user, i.e \pm n ${}^{\circ}$ C. The default and minimum range is melting temp \pm 2.5 ${}^{\circ}$ C. The recommended range is between 52 - 58 ${}^{\circ}$ C. **Note:** the melting temperature is measured using the Melting Temperature of overlaps are calculated based on the Nearest Neighbour Equation and the Sugimoto (1996) thermodynamic table.

- 5. **Sequence Orientation:** Check the box if the Sequence you are using is circular. If left unchecked, the sequence is processed as a linear sequence
- 6. Save Data: If you'd like to save your parameters, check this box. An input form for your user ID is provided if checked. Note: you should have signed up to access previous records (See Sign up).
- 7. **Submit:** Click here to initiate oligomer generation.
- 8. **Refresh:** Click this to refresh the form.

Output

Two buttons appear on the same page once the target parameters are submitted: - Download the excel sheet output

- Download the FASTA file

Excel File

The data is filled into the excel as the following columns.

Cluster Num	ber Oligomer Numbe	r Sequence (5' to 3')	Overlap length	Oligomer length	Melting Temperature of overlap	Overlap Score Cause of errors Repeat Sequence
Cluster 1	oligomer 1	ATGGCTTTGGCAAATGAAATCATTGGGTCATCTTTTCTTCACAAAACATCCTCAGTTTCTTCTCAGTTTCAGGGA	25	75	54.4	7 1 G
Cluster 1	oligomer 2	CTTTTCGCAACTGGCACCACCCAAATAGGCCTGAAGAACTGTTTTCCCTGAAACTGAGAAGAAACTGAG	19	69	53.3	0.17 L
Cluster 1	oligomer 3	TGGTGCCAGTTGCGAAAAGGCAAGTTGTAGCACAACTGAGGAAGGCAGTCAACTCCCCTGTCGCGGC	15	67	53.9	1 1 G
Cluster 1	oligomer 4	CCTTGTACTTCACAGGGTTTTTCATCAAGGGGGACTACCTTTACAAGATCCTCAGAAATAGCCGCGACAGGGGAG	26	74	53.9	1 G
Cluster 1	oligomer 5	TGATGAAAAACCTGTGAAGTACAAGGTTCGGGCTGTGGTAACCATCCGTAATAAAAACAAAGAAGATATAAAAAGAGAC	27	78	47.0	5 6.44 L
Cluster 1	oligomer 6	CAACATTTTGACCAATCCTATCTGTCAAAGCATCCAAATGTTTCACTATAGTCTCTTTTATATCTTCTTTTTTTT	28	3 77	53.5	3 0
Cluster 1	oligomer 7	TTGACAGATAGGATTGGTCAAAATGTTGTGTTACAACTGATATCTACAGAAATAGATCCTAAGACCAACGCACC	22	74	53.6	9 0
Cluster 2	oligomer 1	CTTGAGATTAGTTTTTTTGACCAATCTTTAAGTACAGCTTCGTTGGATTTCTTTGGTGCGTTGGTCTTAGGATCTA	29	77	51.6	7 1.83 L
Cluster 2	oligomer 2	AAGATTGGTCAAAAAAAACTAATCTCAAGGCTGAGCGTGTGAACTATATAGCTGAGTTCCTTTTAACTTCTGACTTTG	28	3 78	55.2	1 0
Cluster 2	oligomer 3	AGAAAAATTCTTGTTGGTGTTTGTTATTTGTGATTGCGCCGGGTTCACCAAAGTCAGAAGTTAAAAGGAACTCAGC	29	79	53.8	1 0
Cluster 2	oligomer 4	TACAAACAACACCAACAAGAATTTTTCTTGGAAACTATTACCATTGAACAATTTGCAAATGATCCAATTCACTTTCC	26	78	53.6	3 0
Cluster 2	oligomer 5	AATTCTCTTAGCTGGATGATCCTTTCTGGACTGGACCCATGAATTGCAAGGAAAGTGAATTGGATCATTTGCAAA	27	7 75	52.8	9 0.61 L
Cluster 2	oligomer 6	AGAAAGGATCATCCAGCTAAGAGAATTTTCTTTAGTAACAAGCCTTATTTGCCAGGAGAAACACCCGC	18	68	55.3	1 G
Cluster 2	oligomer 7	CCTTTCCCATCTCCTCTAATGTCCTTCAACTCTATCTCCCGCAATTTCTTAATGCCCGCGGGTGTTTCTCCTGGC	25	75	54.7	7 1 G
Cluster 2	oligomer 8	AGGACATTAGAGGAGATGGGAAAGGTGAGAGAAAACTCTCAGATCGTGTGTATGACTTTTGACGTTTACAACGATCTC	27	77	54.0	0
Cluster 2	oligomer 9	TCCTCCCAATCTAGGTCTAGCATATTCTATTCCCTTGTCTGGGTTTCCGAGATCGTTGTAAACGTCAAAGTCATA	25	5 75	53.6	4 0
Cluster 2	oligomer 10	ATATGCTAGACCTAGATTGGGAGGAGAAAATTCCGTATCCTCGTAGGTGTAGGACTGGTCGCGCTC	18	68	54.3	1 G
Cluster 2	oligomer 11	CATTGGCAGTGGCTTCTCAACTCTTGACTCTGCAGTCATGTCAGTGTCACTTGGAGCGCGACCAGTCCTAC	21	71	53.2	3 0.22 L
Cluster 3	oligomer 1	GTTGAGAAGCCACTGCCAATGTATGTTCCTAGAGATGAGCAATTTGAAGAATCCAAACAACAACTTTCTCACTCG	25	76	54.5	3 0
Cluster 3	oligomer 2	AATTGATGCTTTTAAACTAGGAATAAGATTGTGCAAGACTGCTTTCAGCCTCCCGAGTGAGAAAGTTGTTTGT	29	78	50.0	4 3.46 L
Cluster 3	oligomer 3	ATCTTATTCCTAGTTTAAAAGCATCAATTCTGAGTAATAAGCATGATTTCCATGGCTTTTCTGATATTGACAGTCTTTA	29	79	54.3	7 0
Cluster 3	oligomer 4	AGTAACTCATCTTGGAGGCCTAGTTTAAGTAGCACCCCTTCAGAATAAAGACTGTCAATATCAGAAAAGCCATG	24	1 74	54.8	4 0
Cluster 3	oligomer 5	ACTAGGCCTCCAAGATGAGTTACTAAAAAAAACTTCCATTACCACGGGTTGTCTCAGAATCATCACAAGGACTTCTAAG	28	3 78	54.5	5 0
luster 3	oligomer 6	AGCCAAGCAAATTTATCCTTGCTTAAAATCTTGGGAGTATTGTATCTTAGAAGTCCTTGTGATGATTCTGAGA	2	3 73	52.3	7 1.13 L
Cluster 3	oligomer 7	AGCAAGGATAAATTTGCTTGGCTTAGGGATGATGAGTTTGCGCGACAGGCTATAGCAGGTGTAAATCCAGTGAATATTG	29	79	53.5	9 0
Cluster 3	oligomer 8	TCCGGATCTAAATTCGACACCGGAGGGAAAACTTTTAGACGTTCAATATTCACTGGATTTACACCTGCTATA	22	72	54.4	9 0
Cluster 3	oligomer 9	CGGTGTCGAATTTAGATCCGGATGTCTACGGTCCTTTAGAATCGAGTCTTAAGGAAGAACATATTTTGGGACAGATAAA	27	7 79	52.0	3 1.47 L
Cluster 3	oligomer 10	ACAATAAAGAGCTTGTTCTCATCGAGTGCCTGTTGAACTGTCATCCCATTTATCTGTCCCAAAATATGTTCTTCC	27	7 75	54.5	4 0

Figure 1: Excel Output File Example

Cluster	Oligomer	Oligomer	Overlap	Overlap	Overlap	Score	Repeats
Num-	Num-	Sequence (5'	Length	Melting	Score	Faults	Se-
ber	ber	to 3')		Temperature			quences

Score of each oligomer's overlap:

- **H**: Overlap melting temperature is higher than the specified target temperature (score = overlap_temp temp_range_high)
- L: Overlap melting temperature is higher than the specified target temperature (score = temp_range_low overlap_temp)
- T: The overlap of each oligomer is not the smallest/largest within the possible oligomer range, and there is 'T' at the 3' end (score = 1)
- **G**: The last 5 bp of 3' end have 3 consecutive 'G', 'C', or a combination of both. (score = 1)
- R: There are repeats (10bp) within oligomer overlaps within each cluster (score = 10) Overlap score is 0 if there are no possible areas of errors.

FASTA file

- Oligomer_1.1 Overlap Length: 25, Oligomer Length: 75, Overlap Melting Temperature: 54.47 ATGGCTTTGGCAAATGAAATCATTGGGTCATCTTTTCTTCACAAAACATCCTCAGTTTCT TCTCAGTTTCAGGGA
- 2. >Oligomer_1.2 Overlap Length: 19 Oligomer Length: 69, Overlap Melting Temperature: 53.33 CTTTTCGCAACTGGCACCAAATAGGCCTGAAGAACTGTTTTCCCTGAAACTGAGAA GAAACTGAG
- >Oligomer_1.3 Overlap Length: 15, Oligomer Length: 67 Overlap Melting Temperature: 53.91 TGGTGCCAGTTGCGAAAAGGCAAGTTGTAGCACAACTGAGGAAGGCAGTCAACTCCCCTG TCGCGGC
- 4. >Oligomer_1.4 Overlap Length: 26, Oligomer Length: 74, Overlap Melting Temperature: 53.99 CCTTGTACTTCACAGGGTTTTTCATCAAGGGGGACTACCTTTACAAGATCCTCAGAAATAG CCGCGACAGGGGAG
 - >Oligomer_1.5 Overlap Length: 27, Oligomer Length: 78, Overlap Melting Temperature: 47.06 TGATGAAAAACCTGTGAAGTACAAGGTTCGGGCTGTGGTAACCATCCGTAATAAAAACAA
- AGAAGATATAAAAGAGAC

Figure 2: FASTA file output (Same example as Excel file; viewed on Notepad)

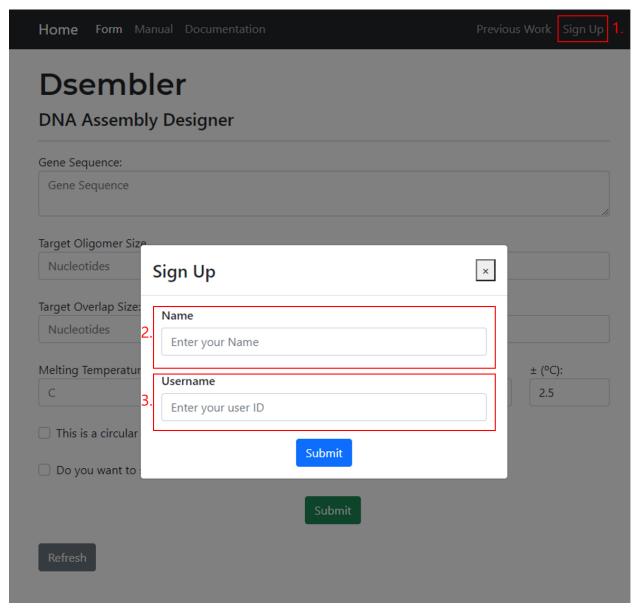
The oligomer FASTA file can be used for easy visualization of all oligomers. It can allow for amendments on other DNA visualising software, such as SnapGene. The first sequence is the user's inputted sequence with the parameters as the sequence description. Similar to the Excel file (Fig.2), the Cluster and Oligomer number (1), the overlap length (2), The Oligomer length (3), the Overlap Melting Temperature(4), and the Oligomer sequence in a 5' to 3' orientation (5) is given for every oligomer.

Previous Work

A simple SQLite based storage is used to provide users to store and access any previous work.

Sign up

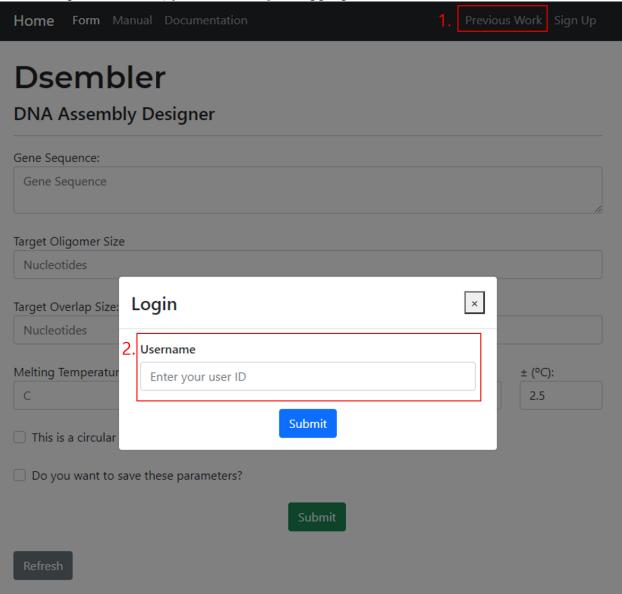
First, you must register into the system by signing up.



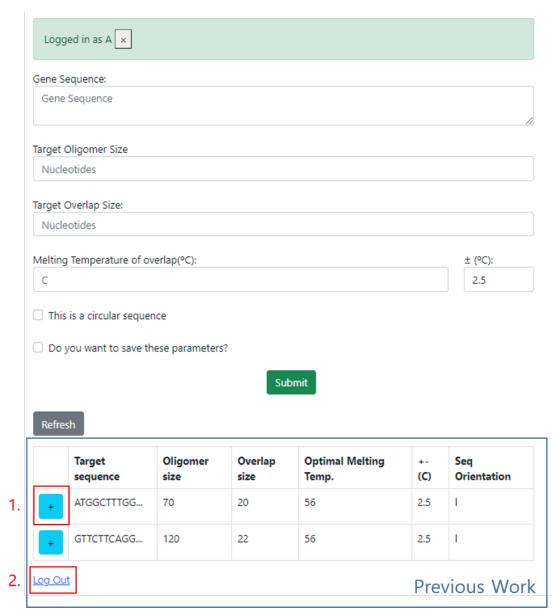
As the webpage does not cater to having password based logins, only the user's name and preferred username is registered. Only unique ID's are registered per container. Success and Failure messages are given appropriately.

Login

To access previous work, you must use your appropriate ID.



Once Logged in, you can find your previous queries, and can access the oligomers for the same.



The oligomers for a specific query can be accessed by clicking on the blue '+' button, which will generate the oligomers in a similar manner as shown in Output.

3. Contact

To report bugs and provide suggestions for the Dsembler User Interface and functionality, please contact the maintainer, Aporva Gupta (aporva03@kribb.re.kr)