MSA Class Logos – Step-by-step

MSA Class Logos is a free web server allowing users to group amino acid sequences into classes and generate sequence logos for each class as well as the entire multiple sequence alignment. The web server is intended to assist scientists with little or no programming skills. Input files are a multiple sequence alignment in FASTA format (required) and, optionally, a file specifying the class each sequence belongs to in CSV format. After file upload, one reference sequence needs to be selected. This sequence will be displayed separately for comparison in the *Class Specific Logos* tab. This can be any sequence from the multiple sequence alignment and is just for convenience in case it is interesting to compare a particular sequence to the sequence logos of each class. If this is not important, any sequence can be selected since it does not affect the generation of the logos and, therefore, the results displayed. There are two tabs with different results available. The first one shows the complete multiple sequence alignment with all sequences sorted according to the classes defined by the user and a sequence logo considering all sequences. This alignment and sequence logo are explorable directly on the web page. The second tab shows sequence logos separated for each defined class, which allows the identification of positions that are conserved in only a subset of sequences but not in all. In other words, the second view (tab) allows a more detailed analysis of a multiple sequence alignment that permits the identification of class-specific conserved residues in evolutionary and/or functionally variable protein groups. Results can be reviewed visually directly on the web page or downloaded as tables in CSV format. Additionally, sections of each sequence logo can be downloaded in SVG and used to generate figures for publications or presentations.

This step-by-step description of how to use MSA Class Logos is based on the example files available on the web server.

Index:

1 File preparation 2

1.1 FASTA Alignment File (required) 2

1.2 CSV Class File (optional) 2

1.2.1 Option A: Using a CSV File (recommended) 2

1.2.2 Option B: Manual Class Management 3

2 Uploading files (Option A: using a CSV class file) 4

2.1 Selection of reference sequence and revision of class assignation 4

3 Uploading files (Option B: Manual Class Management) 6

4 Results based on the entire multiple sequence alignment (*All Sequences* tab) 7

4.1 Visualization 7

4.2 Download of amino acid frequencies per position 8

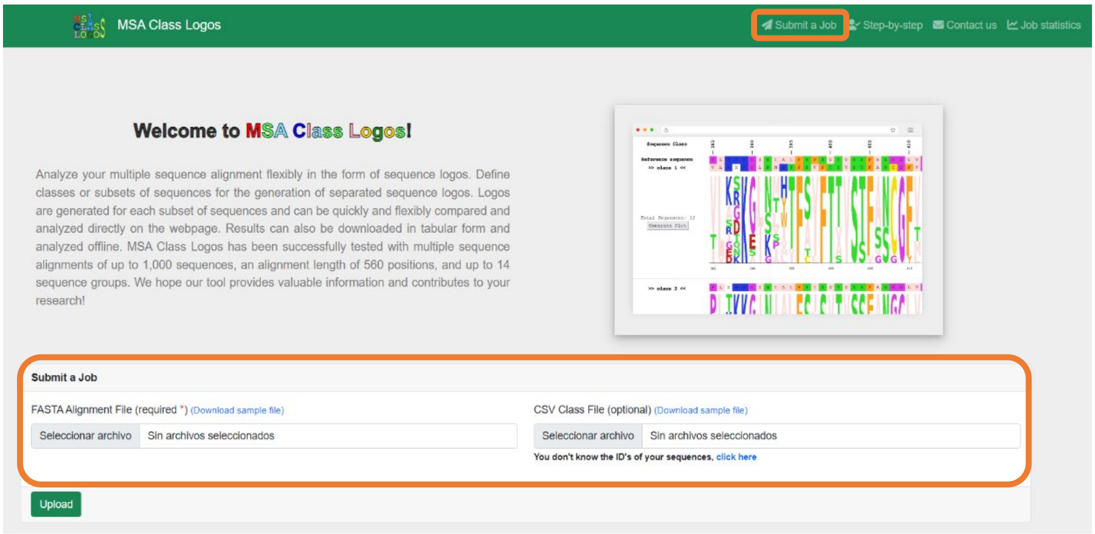
4.3 Downloading the selected sequence logo regions 9

5 Results per defined sequence class (*Class Specific Logos* tab) 10

1 File preparation

Two types of files can be used with MSA Class Logos: (1) a multiple sequence alignment file in FASTA format (required), and (2) a file specifying the class each sequence belongs to in CSV format (optional). Note: MSA Class Logos is not able to align sequences. This step needs to be performed previously using external algorithms like MAFFT, MUSCLE, or CLUSTALW, among many others.

Files can be uploaded directly on the first page after loading the MSA Class Logos webpage. Should you not see the first page, you can click on 'Submit a Job' to reach the initial page (**Figure 1**).



Welcome to MSA Class Logos!

Analyze your multiple sequence alignment flexibly in the form of sequence logos. Define classes or subsets of sequences for the generation of separated sequence logos. Logos are generated for each subset of sequences and can be quickly and flexibly compared and analyzed directly on the webpage. Results can also be downloaded in tabular form and analyzed offline. MSA Class Logos has been successfully tested with multiple sequence alignments of up to 1,000 sequences, an alignment length of 560 positions, and up to 14 sequence groups. We hope our tool provides valuable information and contributes to your research!

Submit a Job

FASTA Alignment File (required *) (Download sample file)

Seleccionar archivo | Sin archivos seleccionados

CSV Class File (optional) (Download sample file)

Seleccionar archivo | Sin archivos seleccionados

You don't know the ID's of your sequences, click here

Figure 1: Submitting a job.

1.1 FASTA Alignment File (required)

For the tutorial, click on ***Download sample file*** next to ***FASTA Alignment File (required *)***. A FASTA file will be downloaded (**Figure 2**).

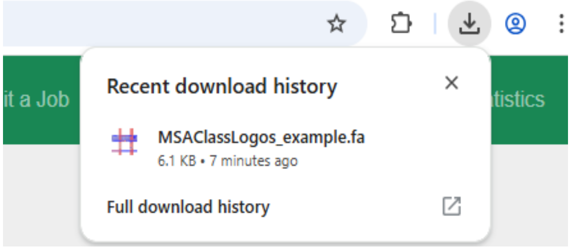


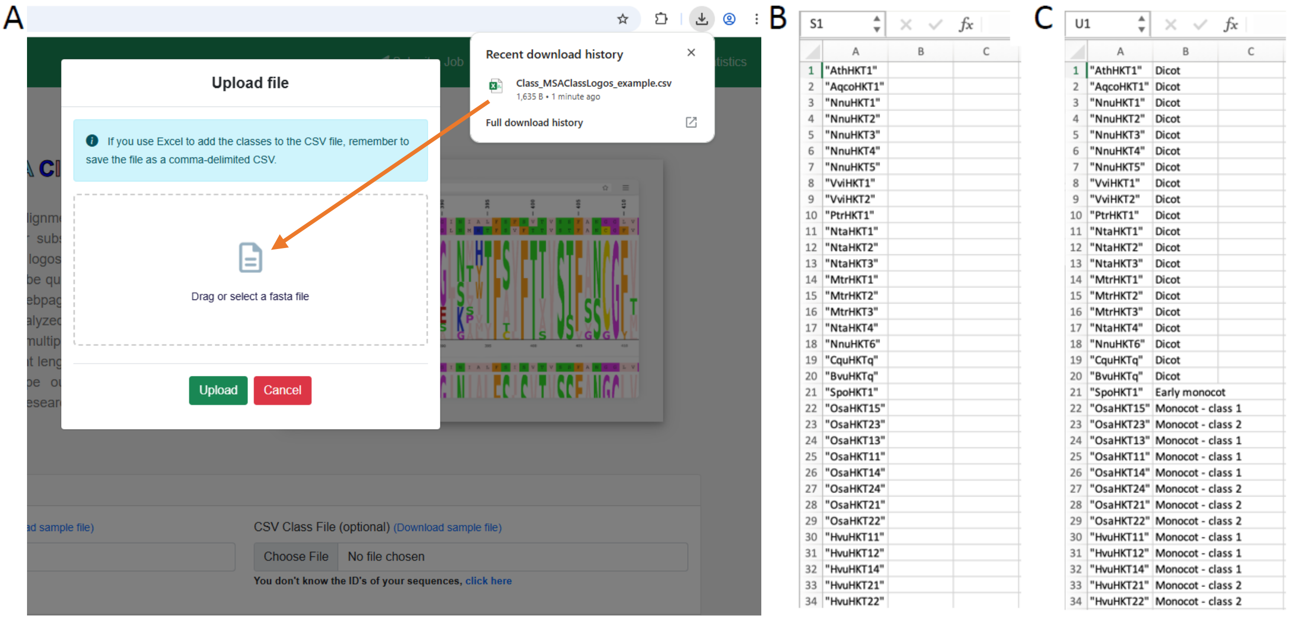
Figure 2: Download example sequence alignment.

1.2 CSV Class File (optional)

You have two options for managing sequence classes:

1.2.1 Option A: Using a CSV File (recommended)

The class file has a simple structure of two columns. Column 1 contains the sequence IDs used in the alignment file, column 2 contains the name of the class each sequence belongs to. Sequence IDs in alignment and class file must be identical. It is recommended that sequence IDs be extracted using the script implemented on the webpage. Press on '*click here*' in ***You don't know the ID's of your sequences, click here***. A window opens where the previously downloaded alignment file can be dragged and dropped into. Click ***Upload*** (**Figure 3A**).



Upload file

If you use Excel to add the classes to the CSV file, remember to save the file as a comma-delimited CSV.

Drag or select a fasta file

Recent download history

Class_MSAClassLogos_example.csv

1,635 B • 1 minute ago

Full download history

CSV Class File (optional) (Download sample file)

You don't know the ID's of your sequences, click here

S1

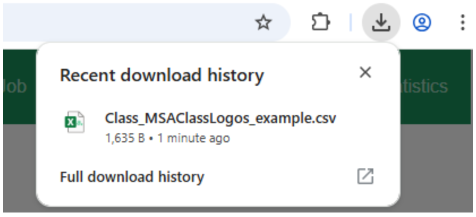
	A	B	C
1	"AthHKT1"		
2	"AqcoHKT1"		
3	"NnuHKT1"		
4	"NnuHKT2"		
5	"NnuHKT3"		
6	"NnuHKT4"		
7	"NnuHKT5"		
8	"VviHKT1"		
9	"VviHKT2"		
10	"PtrHKT1"		
11	"NtaHKT1"		
12	"NtaHKT2"		
13	"NtaHKT3"		
14	"MtrHKT1"		
15	"MtrHKT2"		
16	"MtrHKT3"		
17	"NtaHKT4"		
18	"NnuHKT6"		
19	"CquHKTq"		
20	"BvuHKTq"		
21	"SpoHKT1"		
22	"OsaHKT15"		
23	"OsaHKT23"		
24	"OsaHKT13"		
25	"OsaHKT11"		
26	"OsaHKT14"		
27	"OsaHKT24"		
28	"OsaHKT21"		
29	"OsaHKT22"		
30	"HvuHKT11"		
31	"HvuHKT12"		
32	"HvuHKT14"		
33	"HvuHKT21"		
34	"HvuHKT22"		

U1

	A	B	C
1	"AthHKT1"	Dicot	
2	"AqcoHKT1"	Dicot	
3	"NnuHKT1"	Dicot	
4	"NnuHKT2"	Dicot	
5	"NnuHKT3"	Dicot	
6	"NnuHKT4"	Dicot	
7	"NnuHKT5"	Dicot	
8	"VviHKT1"	Dicot	
9	"VviHKT2"	Dicot	
10	"PtrHKT1"	Dicot	
11	"NtaHKT1"	Dicot	
12	"NtaHKT2"	Dicot	
13	"NtaHKT3"	Dicot	
14	"MtrHKT1"	Dicot	
15	"MtrHKT2"	Dicot	
16	"MtrHKT3"	Dicot	
17	"NtaHKT4"	Dicot	
18	"NnuHKT6"	Dicot	
19	"CquHKTq"	Dicot	
20	"BvuHKTq"	Dicot	
21	"SpoHKT1"	Early monocot	
22	"OsaHKT15"	Monocot - class 1	
23	"OsaHKT23"	Monocot - class 2	
24	"OsaHKT13"	Monocot - class 1	
25	"OsaHKT11"	Monocot - class 1	
26	"OsaHKT14"	Monocot - class 1	
27	"OsaHKT24"	Monocot - class 2	
28	"OsaHKT21"	Monocot - class 2	
29	"OsaHKT22"	Monocot - class 2	
30	"HvuHKT11"	Monocot - class 1	
31	"HvuHKT12"	Monocot - class 1	
32	"HvuHKT14"	Monocot - class 1	
33	"HvuHKT21"	Monocot - class 2	
34	"HvuHKT22"	Monocot - class 2	

Figure 3: (**A**) Extracting sequence IDs using the script implemented on MSAClassLogos. (**B**) Sequence IDs in first column. (**C**) Class names associated to each sequence ID in second column.

Open the class file in, for example, Excel. The first column contains the sequence IDs extracted from the alignment file (**Figure 3**B). The second column needs to be filled by the user with specific class identifiers (**Figure 3**C). That may be a number or a word. All sequences with the same class identifier will be grouped later. Save the file as a **comma delimited CSV format**. For the tutorial, you can download the example class file by clicking on ***Download sample file*** next to ***CSV Class File (optional)*** (**Figure 4**).



Recent download history

Class_MSAClassLogos_example.csv

1,635 B • 1 minute ago

Full download history

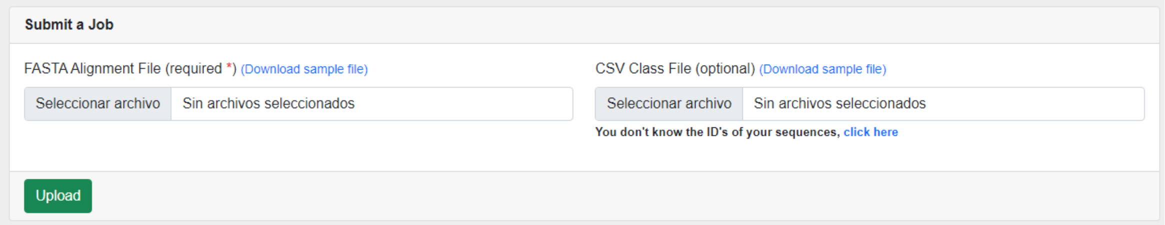
Figure 4: Download example class file.

1.2.2 Option B: Manual Class Management

This can be done directly on the web page via an interactive interface for class management and requires no file. See below for details (**3 Uploading files (Option B: Manual Class Management)**).

2 Uploading files (Option A: using a CSV class file)

On the first page of the webpage or after clicking on ***Submit a Job*** you can upload the example alignment and class file (**Figure 5**).



Submit a Job

FASTA Alignment File (required *) (Download sample file)

Seleccionar archivo | Sin archivos seleccionados

CSV Class File (optional) (Download sample file)

Seleccionar archivo | Sin archivos seleccionados

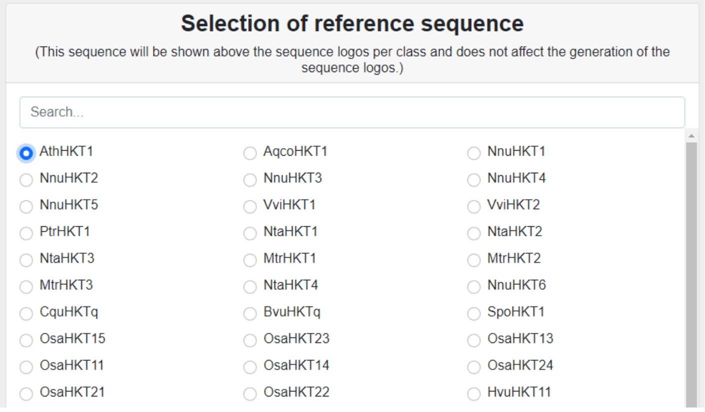
You don't know the ID's of your sequences, click here

Upload

Figure 5: Uploading sequence alignment and class file.

2.1 Selection of reference sequence and revision of class assignation

After uploading the files, one reference sequence needs to be selected (**Figure 6**). This can be any sequence from the multiple sequence alignment and is just for convenience in case it is interesting to compare a certain sequence to the sequence logos of each class. For comparison, the selected sequence will be displayed separately from the sequence logos. If this is not important, any sequence can be selected since it does not affect the generation of the logos and, therefore, the results displayed. Note the search function, which is helpful in case of large alignments. Click ***Select and Next***.



Selection of reference sequence

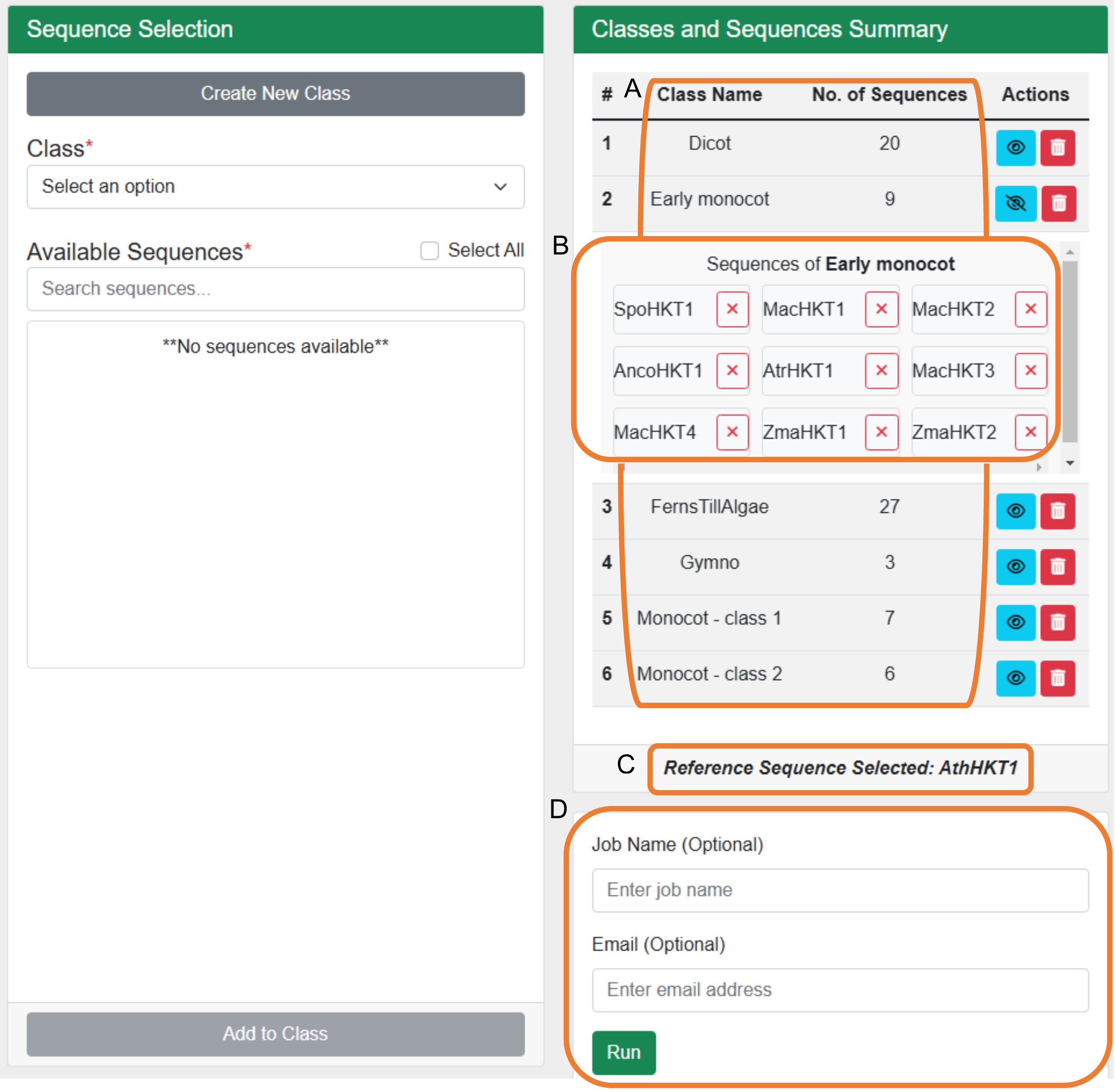
(This sequence will be shown above the sequence logos per class and does not affect the generation of the sequence logos.)

Search...

● AthHKT1 ○ AqcoHKT1 ○ NnuHKT1
○ NnuHKT2 ○ NnuHKT3 ○ NnuHKT4
○ NnuHKT5 ○ VviHKT1 ○ VviHKT2
○ PtrHKT1 ○ NtaHKT1 ○ NtaHKT2
○ NtaHKT3 ○ MtrHKT1 ○ MtrHKT2
○ MtrHKT3 ○ NtaHKT4 ○ NnuHKT6
○ CquHKTq ○ BvuHKTq ○ SpoHKT1
○ OsaHKT15 ○ OsaHKT23 ○ OsaHKT13
○ OsaHKT11 ○ OsaHKT14 ○ OsaHKT24
○ OsaHKT21 ○ OsaHKT22 ○ HvuHKT11

Figure 6: Selection of reference sequence.

On the following page, a summary of all classes (A), sequences (B), and the selected reference sequence (C) is given (**Figure 7**). Since a CSV file was uploaded, the assigned classes will be displayed. If any error is detected you can correct a specific sequence by clicking on the X next to it or delete a whole class by clicking on the bin next to it. The removed sequences can then be assigned manually to the corresponding class(es). Optionally, a job title and an email address can be provided (**Figure 7D**). In this case, you will be notified once the job is done. The email contains a link that leads to the results, which will be stored for one month on the server. Generally, the calculation of the sequence logos takes a few minutes, depending on the number of sequences in the alignment and the length of the alignment. MSA Class Logos has been successfully tested with multiple sequence alignments of up to 1,000 sequences with an alignment length of 560 positions and up to 14 sequence groups. If everything is well selected press ***Run***.



Class*

Select an option

Available Sequences*

□ Select All

Search sequences...

No sequences available

#	Class Name	No. of Sequences	Actions
1	Dicot	20	👁 🗑
2	Early monocot	9	👁 🗑
3	FernsTillAlgae	27	👁 🗑
4	Gymno	3	👁 🗑
5	Monocot - class 1	7	👁 🗑
6	Monocot - class 2	6	👁 🗑

Sequences of **Early monocot**

SpoHKT1 ×	MacHKT1 ×	MacHKT2 ×
AncoHKT1 ×	AtrHKT1 ×	MacHKT3 ×
MacHKT4 ×	ZmaHKT1 ×	ZmaHKT2 ×

Reference Sequence Selected: AthHKT1

Job Name (Optional)

Enter job name

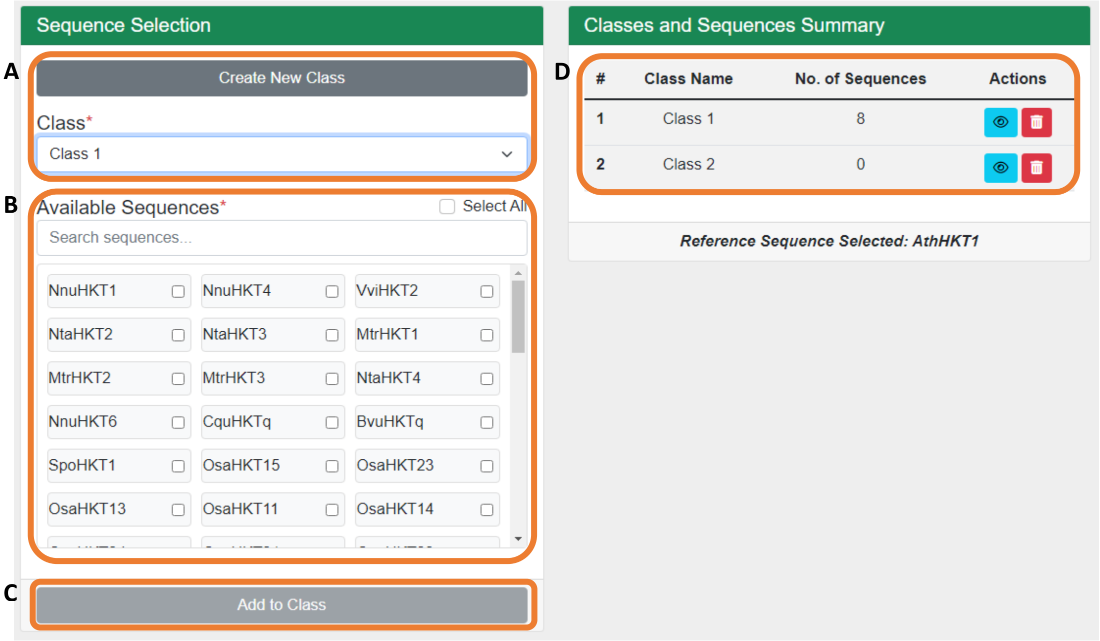
Email (Optional)

Enter email address

Figure 7: Summary and revision of sequences and their assigned classes.

3 Uploading files (Option B: Manual Class Management)

If you choose not to use a CSV file, MSA Class Logos provides an interactive interface for class management directly on the web page. Classes can be created and managed manually, and sequences can be added or removed as needed. Selecting only the ***FASTA alignment file*** and pressing ***Upload*** opens a warning asking whether you are sure to continue without using the CSV class file. Hit ***Yes, continue***. After confirmation, you have to select a reference sequence as described in Option A and then you are led to the class management panel where classes can be created and sequences assigned to them (**Figure 8**). Create new classes by clicking the ***Create New Class*** button and assigning a name to each class, e.g., Class 1 (**Figure 8A**). Select sequences from your alignment by activating the checkboxes (**Figure 8B**) and press ***Add to class*** to assign them to classes (**Figure 8C**). The interface allows real-time modification of class compositions. Sequences can be removed from classes and be re-assigned to others as needed (**Figure 8D**). This flexible approach is particularly useful when exploring different classification schemes. All class assignments can be modified until you proceed to the next step of the analysis. Once all sequences are assigned to classes, the option to add a ***Job Name*** and provide an ***Email*** address appear (**Figure 7**). If everything is well selected, press **Run**.



NnuHKT1 □	NnuHKT4 □	VviHKT2 □
NtaHKT2 □	NtaHKT3 □	MtrHKT1 □
MtrHKT2 □	MtrHKT3 □	NtaHKT4 □
NnuHKT6 □	CquHKTq □	BvuHKTq □
SpoHKT1 □	OsaHKT15 □	OsaHKT23 □
OsaHKT13 □	OsaHKT11 □	OsaHKT14 □

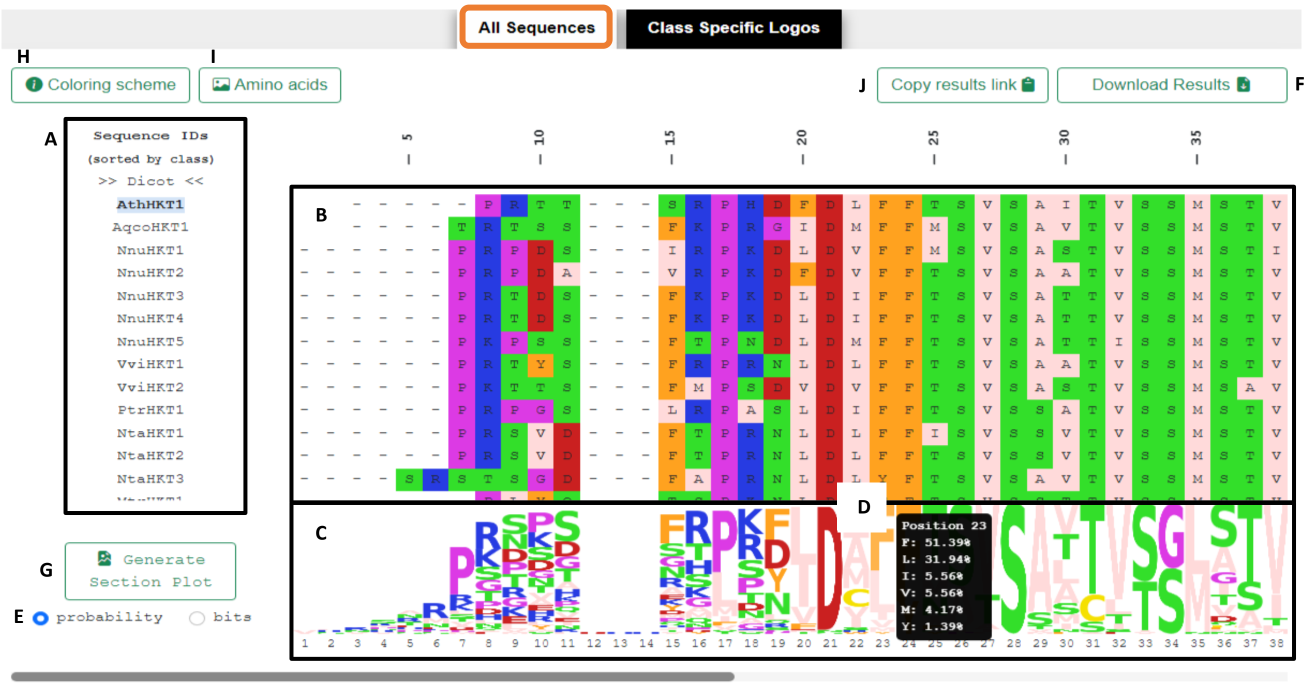
#	Class Name	No. of Sequences	Actions
1	Class 1	8	👁 🗑
2	Class 2	0	👁 🗑

Reference Sequence Selected: AthHKT1

Figure 8: Class management interface. (**A**) New class creation panel. (**B**) List of sequence from the alignment with checkboxes for selection. (**C**) System for assigning selected sequences to created classes. (**D**) Summary table showing classes with their sequence counts and buttons to view class contents and manage classes.

4 Results based on the entire multiple sequence alignment (*All Sequences* tab)

4.1 Visualization



```
    - - - - - P R T T - - - S R P H D F D L F F T S V S A I T V S S M S T V
    - - - - T R T S S - - - F K P R G I D M F F M S V S A V T V S S M S T V
- - - - - - P R P D S - - - I R P K D L D V F F M S V S A S T V S S M S T I
- - - - - - P R P D A - - - V R P K D F D V F F T S V S A A T V S S M S T V
- - - - - - P R T D S - - - F K P K D L D I F F T S V S A T T V S S M S T V
- - - - - - P R T D S - - - F K P K D L D I F F T S V S A T T V S S M S T V
- - - - - - P K P S S - - - F T P N D L D M F F T S V S A T T I S S M S T V
- - - - - - P R T Y S - - - F R P R N L D L F F T S V S A A T V S S M S T V
- - - - - - P K T T S - - - F M P S D V D V F F T S V S A S T V S S M S A V
- - - - - - P R P G S - - - L R P A S L D I F F T S V S S A T V S S M S T V
- - - - - - P R S V D - - - F T P R N L D L F F I S V S S V T V S S M S T V
- - - - - - P R S V D - - - F T P R N L D L F F T S V S S V T V S S M S T V
- - - - S R S T S G D - - - F A P R N L D L Y F T S V S A V T V S S M S T V
```

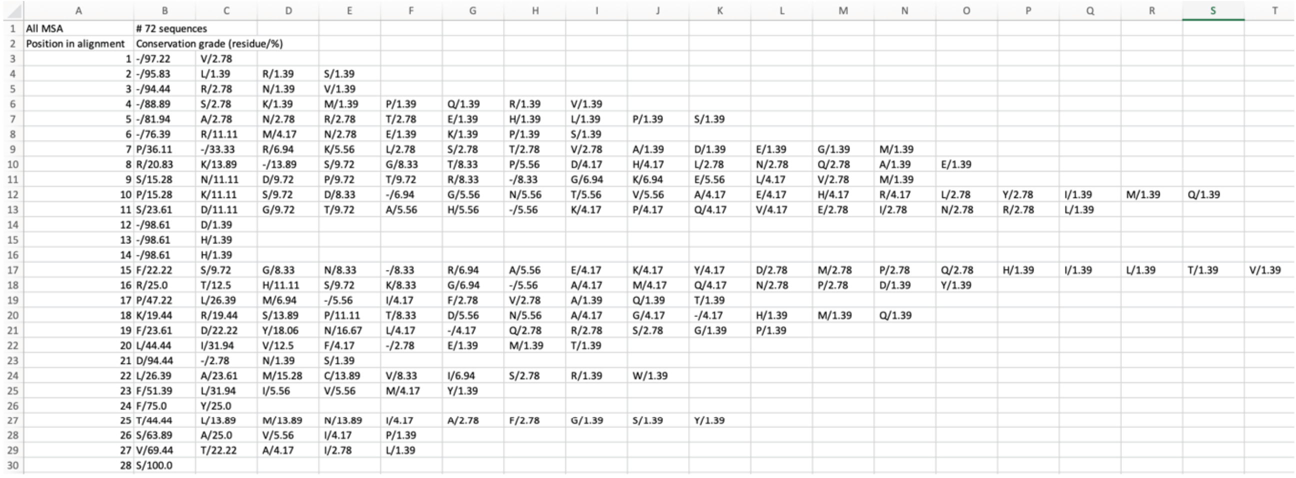
Figure 9: Results for MSA of all sequences. Sequences are sorted according to the assigned classes (**A**). MSA is coloured according to the Zappo colouring scheme (**B, H**). Sequence logos are represented below the MSA using the same colouring scheme (**C**). When moving the mouse over a position of the sequence logos, amino acid frequencies for this position are displayed (**D**). Logos can be analyzed in two views, probability and bits (**E**). MSA and sequence logos can be explored by using the scroll bars on the bottom and right. Results can be downloaded for offline analysis in tabular form (**F**). Graphical representation of sequence logos can be exported in SVG format (**G**). A chart of amino acid structures is available for quick structure references (**I**). Button to keep the link to the results obtained, in case the user did not provide an e-mail address previously (**J**).

The first result tab shows the entire multiple sequence alignment with all sequences sorted according to the classes defined by the user (**Figure 9A**). Amino acids are colored according to the Zappo scheme (for more details, click on ***Coloring Scheme*** on the upper left (H)) (B). Below the sequence alignment is the sequence logo considering all sequences (C). When moving the mouse over an amino acid position in the sequence logo, the frequency in which amino acids occur in this position is shown (D). Gaps are indicated as "–" and are considered in the calculation of frequencies. The view of the sequence logo can be switched between the representation of amino acid frequencies (probability) and bits, which consider the entropy of each position (E). On the top left, clicking on ***Amino acids*** shows the 20 structures of proteinogenic amino acids for quick reference (I). Results can be explored directly on the webpage by scrolling through the sequence alignment and sequence logo. For users who haven't registered their email, a ***Copy results link*** button is available in the top right corner, allowing to save and share results by copying the unique URL to the clipboard (J).

4.2 Download of amino acid frequencies per position

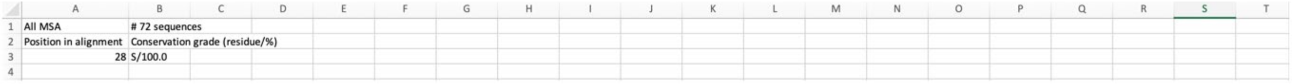
Results can be visually inspected on the webpage but can also be downloaded as tables in CSV format from the ***Download Results*** button (**Figure 9F**). The file ***Amino acid frequencies complete alignment*** contains all amino acid frequencies for each position of the alignment and, therefore, contains all information that is visualized on the webpage in tabular form (**Figure 10A**). All other files containing a specified percentage in their name, contain only a subset of positions applying the indicated filter. For example, the file ***Positions 100% conserved*** reports only those positions of the alignment that contain amino acids with a frequency of 100% (**Figure 10B**). In other words, 100% conserved positions. The file ***Positions ≥90% conserved*** reports all positions where one amino acid appears with a frequency of 90% or higher (**Figure 10C**). And so on. The CSV files can be downloaded and further processed or filtered in, for example, Excel.

A Amino acid frequencies complete alignment (CompleteAlignment_ConservationAll.csv)



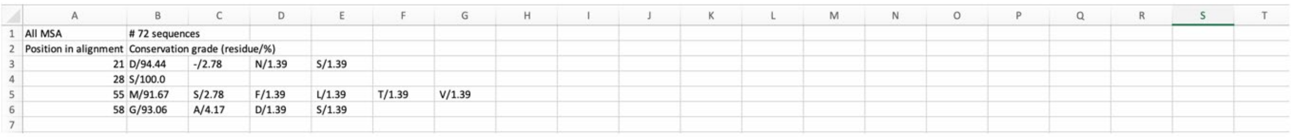
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	All MSA	# 72 sequences																		
2	Position in alignment	Conservation grade (residue/%)																		
3	1	-/97.22	V/2.78																	
4	2	-/95.83	L/1.39	R/1.39	S/1.39															
5	3	-/94.44	R/2.78	N/1.39	V/1.39															
6	4	-/88.89	S/2.78	K/1.39	M/1.39	P/1.39	Q/1.39	R/1.39	V/1.39											
7	5	-/81.94	A/2.78	N/2.78	R/2.78	T/2.78	E/1.39	H/1.39	L/1.39	P/1.39	S/1.39									
8	6	-/76.39	R/11.11	M/4.17	N/2.78	E/1.39	K/1.39	P/1.39	S/1.39											
9	7	P/36.11	-/33.33	R/6.94	K/5.56	L/2.78	S/2.78	T/2.78	V/2.78	A/1.39	D/1.39	E/1.39	G/1.39	M/1.39						
10	8	R/20.83	K/13.89	-/13.89	S/9.72	G/8.33	T/8.33	P/5.56	D/4.17	H/4.17	L/2.78	N/2.78	Q/2.78	A/1.39	E/1.39					
11	9	S/15.28	N/11.11	D/9.72	P/9.72	T/9.72	R/8.33	-/8.33	G/6.94	K/6.94	E/5.56	L/4.17	V/2.78	M/1.39						
12	10	P/15.28	K/11.11	S/9.72	D/8.33	-/6.94	G/5.56	N/5.56	T/5.56	V/5.56	A/4.17	E/4.17	H/4.17	R/4.17	L/2.78	Y/2.78	I/1.39	M/1.39	Q/1.39	
13	11	S/23.61	D/11.11	G/9.72	T/9.72	A/5.56	H/5.56	-/5.56	K/4.17	P/4.17	Q/4.17	V/4.17	E/2.78	I/2.78	N/2.78	R/2.78	L/1.39			
14	12	-/98.61	D/1.39																	
15	13	-/98.61	H/1.39																	
16	14	-/98.61	H/1.39																	
17	15	F/22.22	S/9.72	G/8.33	N/8.33	-/8.33	R/6.94	A/5.56	E/4.17	K/4.17	Y/4.17	D/2.78	M/2.78	P/2.78	Q/2.78	H/1.39	I/1.39	L/1.39	T/1.39	V/1.39
18	16	R/25.0	T/12.5	H/11.11	S/9.72	K/8.33	G/6.94	-/5.56	A/4.17	M/4.17	Q/4.17	N/2.78	P/2.78	D/1.39	Y/1.39					
19	17	P/47.22	L/26.39	M/6.94	-/5.56	I/4.17	F/2.78	V/2.78	A/1.39	Q/1.39	T/1.39									
20	18	K/19.44	R/19.44	S/13.89	P/11.11	T/8.33	D/5.56	N/5.56	A/4.17	G/4.17	-/4.17	H/1.39	M/1.39	Q/1.39						
21	19	F/23.61	D/22.22	Y/18.06	N/16.67	L/4.17	-/4.17	Q/2.78	R/2.78	S/2.78	G/1.39	P/1.39								
22	20	L/44.44	I/31.94	V/12.5	F/4.17	-/2.78	E/1.39	M/1.39	T/1.39											
23	21	D/94.44	-/2.78	N/1.39	S/1.39															
24	22	L/26.39	A/23.61	M/15.28	C/13.89	V/8.33	I/6.94	S/2.78	R/1.39	W/1.39										
25	23	F/51.39	L/31.94	I/5.56	V/5.56	M/4.17	Y/1.39													
26	24	F/75.0	Y/25.0																	
27	25	T/44.44	L/13.89	M/13.89	N/13.89	I/4.17	A/2.78	F/2.78	G/1.39	S/1.39	Y/1.39									
28	26	S/63.89	A/25.0	V/5.56	I/4.17	P/1.39														
29	27	V/69.44	T/22.22	A/4.17	I/2.78	L/1.39														
30	28	S/100.0																		

B Positions 100% conserved (CompleteAlignment_Conservation100.csv)



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	All MSA	# 72 sequences																		
2	Position in alignment	Conservation grade (residue/%)																		
3	28	S/100.0																		
4																				

C Positions ≥90% conserved (CompleteAlignment_Conservation90.csv)

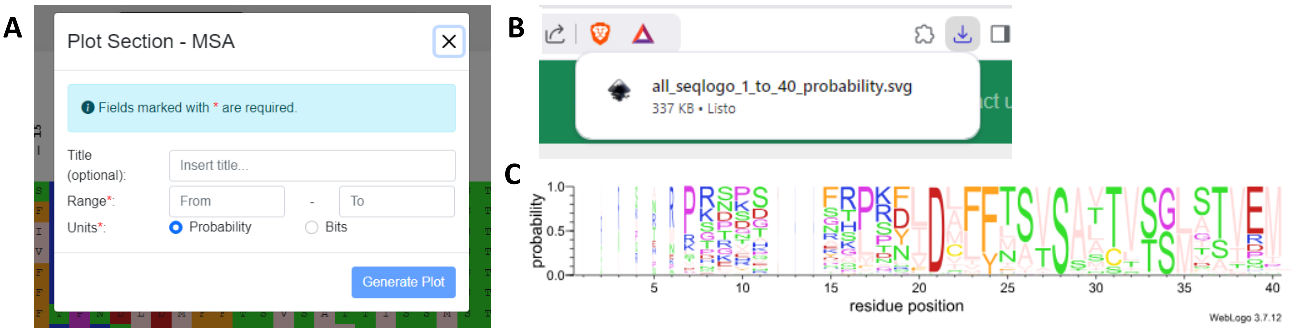


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	All MSA	# 72 sequences																		
2	Position in alignment	Conservation grade (residue/%)																		
3	21	D/94.44	-/2.78	N/1.39	S/1.39															
4	28	S/100.0																		
5	55	M/91.67	S/2.78	F/1.39	L/1.39	T/1.39	V/1.39													
6	58	G/93.06	A/4.17	D/1.39	S/1.39															
7																				

Figure 10: Structure of downloadable results of amino acid frequencies per MSA position. (**A**) File of frequencies of all MSA position. Column A states the position of the MSA. From column B onwards, amino acids and their frequency of appearance in this particular position are indicated (amino acid / frequency in %). The more cells contain amino acid frequency information per MSA position, the higher the variability of this position or the less conserved is this position. In case of a 100% conserved position only one cell will contain information. Compare positions 15 (highly variable) and 28 (highly conserved). For simplicity only the first 28 positions of the sequence alignment are shown. (**B**) This file contains only the positions that are 100% conserved in the MSA. In this example, this is only one position. (**C**) This file contains only the positions with amino acid frequencies of 90% or higher in the MSA. In this example, this are only four positions.

4.3 Downloading the selected sequence logo regions

Once an interesting region has been identified, the sequence logo of these positions can be downloaded as an image in SVG format by clicking on the ***Generate Section Plot*** button to the left of the sequence logo (**Figure 9G**). A window opens in which the start and end position of the logos can be specified, as well as the logos view in probability or bits (**Figure 11**). Pressing ***Generate Plot*** generates the SVG file. This takes a moment, depending on the length of the sequence range. Please wait while the file is generated. Sequence logo plots are generated by WebLogo3.7.12. The scalable vector graphics (SVG) file can be further processed in programs suitable for generating high-resolution figures for scientific publications and presentations.



Plot Section - MSA

Fields marked with * are required.

Title (optional): Insert title...

Range*: From - To

Units*: ● Probability ○ Bits

Generate Plot

all_seqlogo_1_to_40_probability.svg
337 KB • Listo

Figure 11: Generation and export of sequence logo plot. (**A**) The sequence logo range that shall be exported can be specified, for example position 1 to 40. (**B**) Within a short time (depending on the selected length of the sequence logo) an SVG file is generated. (**C**) The SVG files are generated using WebLogo and can be incorporated into publication figures or scientific presentation, for example.

5 Results per defined sequence class (*Class Specific Logos* tab)

The visualization of the results is slightly different in this tab, while the downloadable files and the generation of the sequence logo plots follow the same scheme described in the previous section. The initially selected reference sequence is the top sequence displayed (**Figure 12A**). Below the reference sequence are represented a sequence logo (**Figure 12B**) and a consensus sequence for each sequence class specified initially in the class file (**Figure 12C**). The name of the class and the number of sequences selected for this class are indicated on the left (**Figure 12D**). The consensus sequence shows amino acids with the highest frequency for each position in case this frequency is higher than 50%. If the amino acid with the highest frequency has a value below 50% the position is marked with an X and considered as nonconserved. The sequence logos have the same characteristics as in the results page described above, except that in this section you can export a given position range of all classes at once or separated per class. The advantage of the visualization per class is that sequences grouped according to sequence similarity, functional similarity or other aspects can be analyzed in more detail and contrasted and compared against other groups, or the complete alignment. This can be done quick and easy directly on the webpage or by analyzing the downloadable frequency tables. For example, positions that are conserved in only some classes but not others can be easily identified in the ***Class Specific Logos*** tab although they appear as seemingly not highly conserved in the sequence logos based on the entire alignment (***All Sequences*** tab).

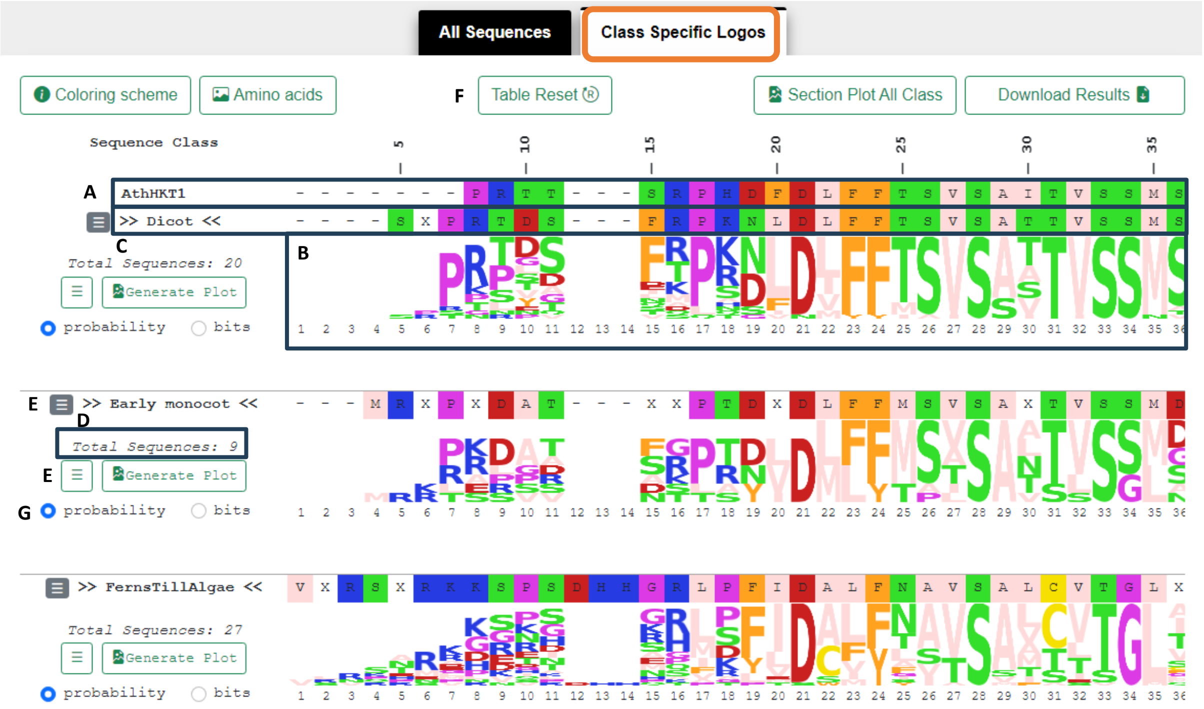


Figure 12: Results of sequence logo representation per class. Sequence logos for each class are represented based on the class assignments done during the first step (**B**). The initially selected reference sequence is shown on top (**A**). For each sequence class logo, a consensus sequence is displayed where highly variable positions (amino acid frequencies lower than 50%) are marked with an X (**C**). The colouring scheme, download results and sequence plot generation are equal to the *All Sequences* tab. The number of sequences used for the generation of sequence logos is stated on the left of each logo (**D**). The small square symbol with three horizontal lines in it enables hiding sequence logos and consensus sequences (**E**). The Table Reset button on the top reveals all hidden information (**F**). Also here, sequence logos can be viewed based on probabilities or bits (G).

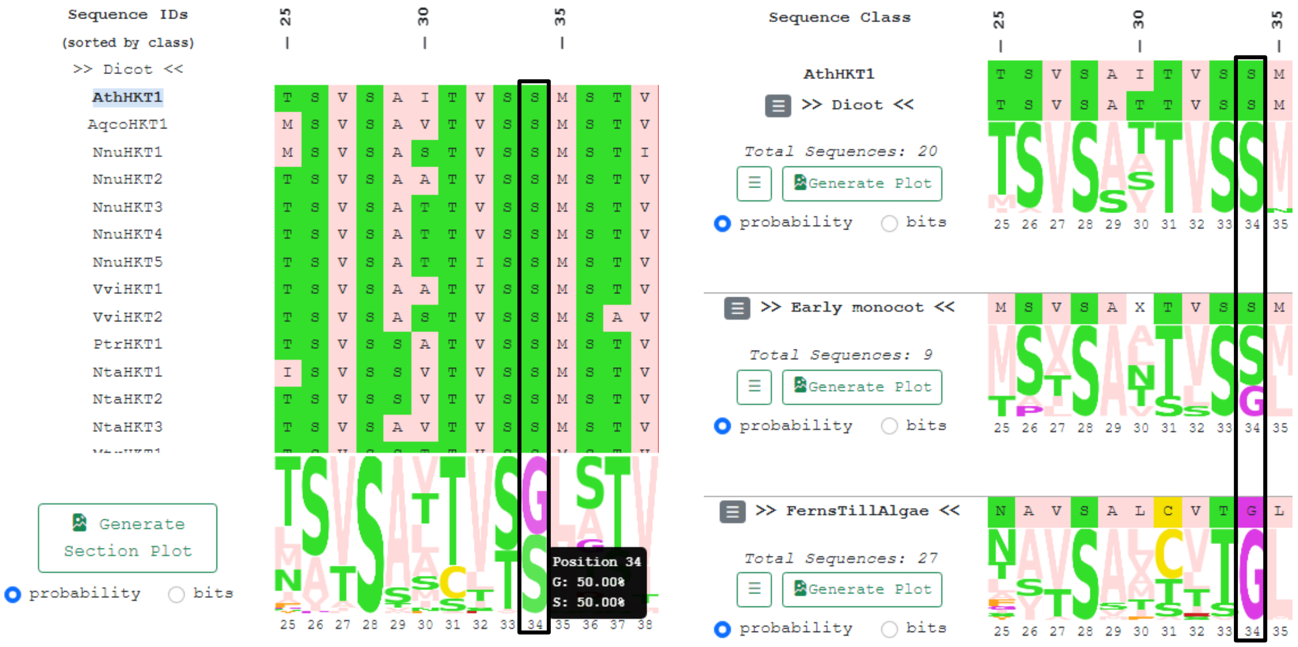


Figure 13: Comparison of All Sequence and Class Specific Logos tab on the example of position 34. (**A**) The All Sequence tab illustrates that 50% of the sequences contain serine residues, and the other 50% glycine residues. (**B**) The Class Specific Logos tab illustrates that some sequence classes conserve the serine or the glycine, while others accept both residues.

For example, position 34 in the example alignment shows that 50% of the sequences contain a serine, while the other 50% contain a glycine residue at this position as observed in the ***All Sequences*** tab (**Figure 13A**). When analyzing the ***Class Specific Logos*** tab it is evident that some sequence classes have a clear preference for one of the two residues (**Figure 13B**). The classes Dicot and Monocot - class 1 express only serine residues, and FernsTillAlgae only glycine. Monocot - class 2 have a preference for glycine, while the class Gymno and Early Monocot accept both residues in their sequences. To simplify the comparison of specific groups it is possible to hide consensus sequences and logos by clicking on the square symbol with three horizontal lines in it that appears on the left of each sequence logo and consensus sequence (**Figure 12E**). The ***Table Reset*** button on the top restores all hidden information (**Figure 12F**). Here, for clarity, we visualized only the classes Dicot, Early Monocot, and FernsTillAlgae.

Additionally, within the ***Download Results*** menu, a CSV class file containing the information of which sequences are assigned to which classes can be downloaded (***Class file csv generated***). This file can be modified or used as is for future use in MSA Class Logos when the same sequences as in the current run are used.