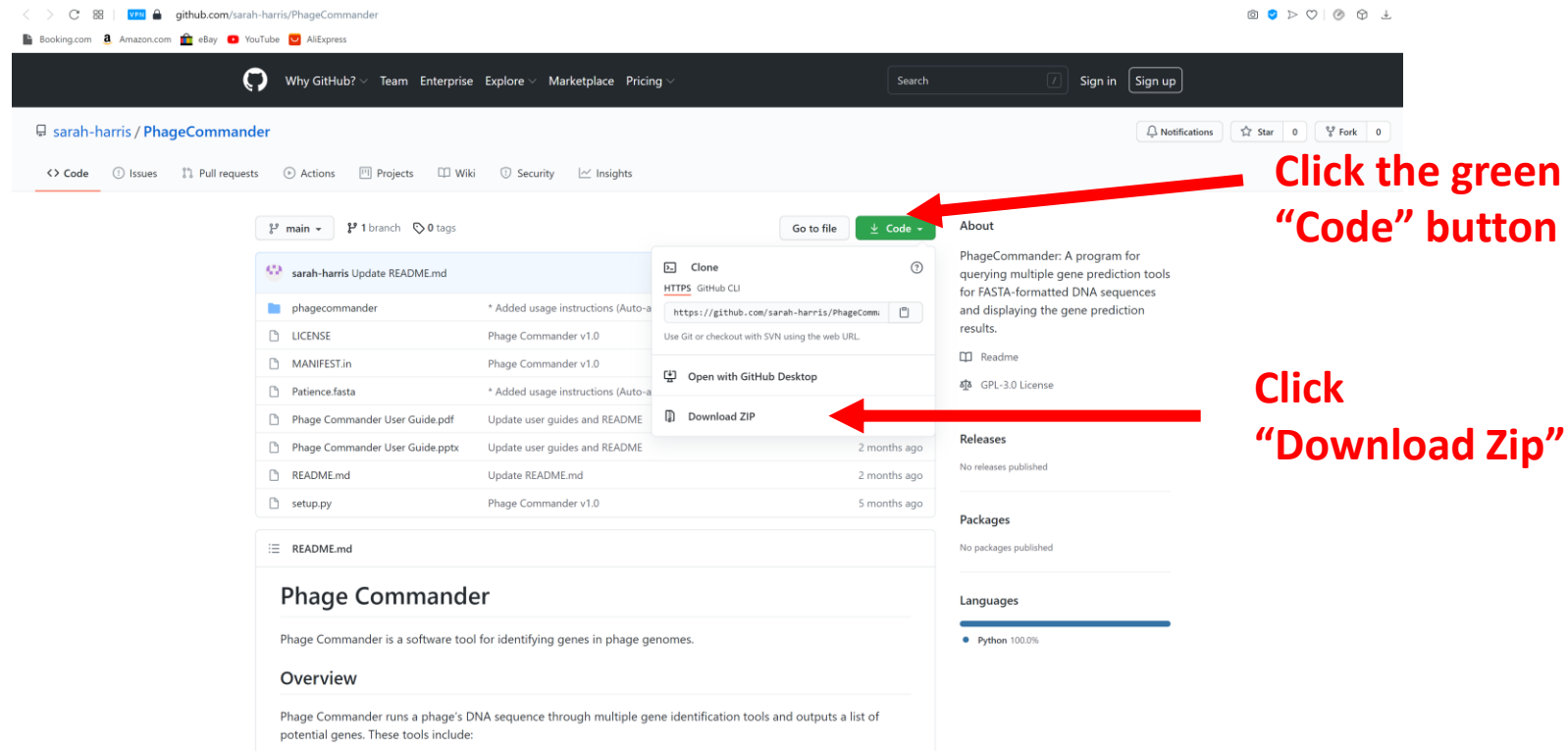


Introduction to Phage Commander

- **Phage Commander** is an application for identifying genes in phage genomes using multiple programs
- Phage Commander runs a phage's DNA sequence through gene identification tools and outputs a list of potential genes. These tools include:
 - RAST, Prodigal, Metagene, Glimmer, Genemark, Genemark.hmm, Genemark S, Genemark S2, Genemark Heuristic, and Aragorn (for tRNA genes)
- Phage Commander's output can be exported in Excel format (.xlsx) or NCBI GenBank format (.gb)
- A paper describing Phage Commander in detail is here:
<https://www.biorxiv.org/content/10.1101/2020.11.11.378802v1>

Installing Phage Commander (Windows and Mac)

- Download Phage Commander from GitHub repository:
<https://github.com/sarah-harris/PhageCommander>



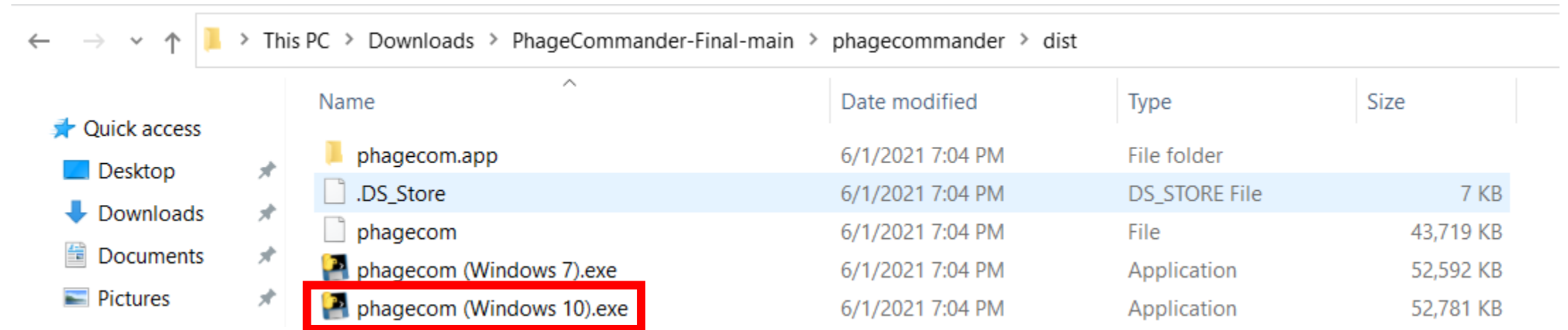
The screenshot shows the GitHub repository page for `sarah-harris/PhageCommander`. The repository is in the `main` branch and contains several files, including `phagecommander`, `LICENSE`, `MANIFEST.in`, `Patience.fasta`, `Phage Commander User Guide.pdf`, `Phage Commander User Guide.pptx`, `README.md`, and `setup.py`. The `README.md` file is selected, showing the project overview. Two red arrows point to the `Code` button and the `Download ZIP` option in the dropdown menu.

Click the green "Code" button

Click "Download Zip"

Installing Phage Commander (Windows and Mac)

- Go to your downloads folder and unzip Phage Commander with your favorite unzip program (I like 7-zip <https://www.7-zip.org/>)
- After unzipping, move the “PhageCommander-Final-main” folder to your desired location, open it, click on the “phagecommander” subfolder, and then the “dist” subfolder, and then the .exe file to run

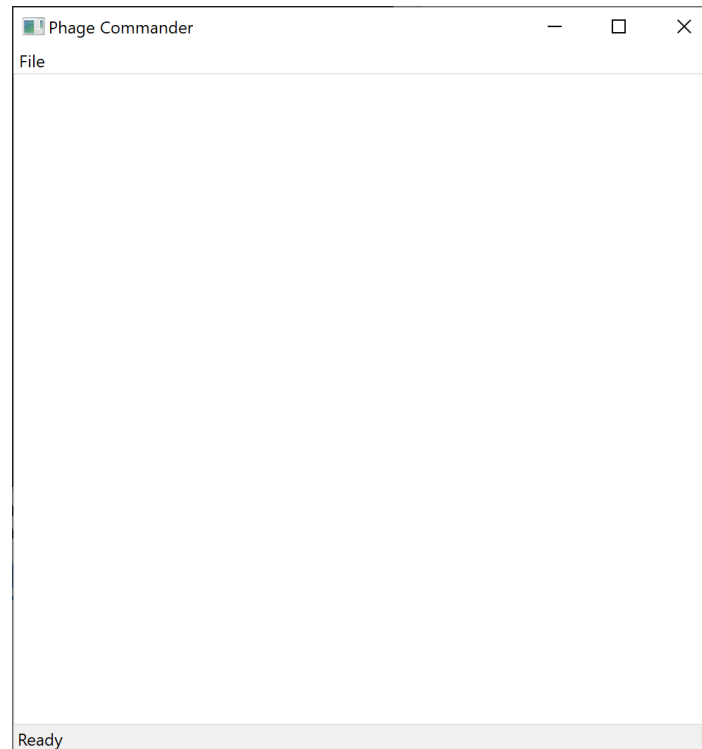


Installing Phage Commander (Linux)

- Open a shell. Navigate to folder where you have downloaded phagecommander. Type `py phagecom.py`. If you need to install any of the supporting packages, first type: `'pip install .'` (in the PhageCommander-master directory – the directory that contains `setup.py`)

Running Phage Commander

- Once you have clicked on the appropriate .exe executable, a blank Phage Commander window will appear



How to use Phage Commander

- In the Phage Commander window, click on the **File** menu, and select **New**
- A Select Gene Identification Tools window will open (see next slide). Select which tools you would like to use
 - To use **RAST**, you will need to create a login and password at <https://rast.nmpdr.org> (free and easy)
 - If using **GeneMark.hmm** you will need to select your phage's bacterial host from the drop down menu under **Species**
- Press **Open** and select your phage's fasta file - an example fasta file is included in the GitHub repository: Patience.fasta
- Press **Run Phage Commander** to run Phage Commander. Phage Commander will now run the phage genome through the selected tools – this may take several minutes
- When Phage Commander completes, press **OK**

Select Gene Identification Tools

Check the boxes of the tools you want to use

If using RAST, you will need to create a login and password on the RAST website (rast.nmpdr.org/) and enter it when prompted by Phage Commander

Select Gene Identification Programs

<input checked="" type="checkbox"/> RAST	<input checked="" type="checkbox"/> Prodigal	<input checked="" type="checkbox"/> Metagene
<input checked="" type="checkbox"/> Glimmer	<input checked="" type="checkbox"/> GeneMark	<input checked="" type="checkbox"/> GeneMark w/ Heuristics
<input checked="" type="checkbox"/> GeneMark S	<input checked="" type="checkbox"/> GeneMark S2	<input checked="" type="checkbox"/> Host-trained GeneMark

Host species (for host-trained GeneMark): Acaryochloris_marina_MBIC11017

☒ Aragorn (for tRNA)

Input genome (.fasta file): Select...

Run Phage Commander Cancel

Select host from drop down menu if using Host-trained GeneMark

Select genome (fasta file)

Run Phage Commander

In the Event of an Error

- If Phage Commander runs successfully, you will see a window saying **“Done! Query Successful”**
- Press **OK** and you will see the Phage Commander output in the form of a spreadsheet (expand the window)
- If phage commander encounters an error, the error message will display which program is causing the error (e.g. Glimmer, Aragorn, GMS2, etc.)
- To work around this, re-run Phage Commander without including the program causing the error

“Total Calls”
is the number of
programs identifying
a gene

“All” shows which genes are identified by
all programs

Genes identified
by Prodigal

Genes identified
by Glimmer

Genes identified
by MetaGene

Each row
is a gene

Phage Commander - untitled																	
	TOTAL CALLS	ALL	ONE	PRODIGAL	PRODIGAL	PRODIGAL	PRODIGAL		GLIMMER	GLIMMER	GLIMMER	GLIMMER		METAGENE	METAGENE	METAGENE	METAGENE
21	8	X		+	20008	20442	435		+	20008	20442	435		+	20008	20442	435
22	8	X		+	20442	22835	2394		+	20442	22835	2394		+	20442	22835	2394
23	8	X		+	22850	23170	321		+	22850	23170	321		+	22850	23170	321
24	8	X		+	23173	23355	183		+	23173	23355	183		+	23173	23355	183
25	8	X		+	23355	23630	276		+	23355	23630	276		+	23355	23630	276
26	7			+	23630	23824	195		+	23630	23824	195					
27	8	X		+	23885	25201	1317		+	23885	25201	1317		+	23885	25201	1317
28	8	X		+	25201	26409	1209		+	25201	26409	1209		+	25201	26409	1209
29	8	X		+	26435	26785	351		+	26435	26785	351		+	26435	26785	351
30	8	X		+	26853	27110	258		+	26790	27110	321		+	26853	27110	258
31	8	X		-	27721	28854	1134		-	27721	28728	1008		-	27721	28728	1008
32	8	X		+	27107	27724	618		+	27107	27724	618		+	27107	27724	618
33	5			-	28914	29399	486		-	28914	29489	576		-	28914	29324	411
34	8	X		+	29498	29755	258		+	29498	29755	258		+	29498	29755	258
35	8	X		+	29752	30150	399		+	29752	30150	399		+	29752	30150	399
36	8	X		+	30213	30749	537		+	30150	30749	600		+	30213	30749	537
37	8	X		+	30749	30931	183		+	30749	30931	183		+	30749	30931	183
38	8	X		+	30928	31278	351		+	30928	31278	351		+	30928	31278	351
39	8	X		+	31275	31424	150		+	31275	31424	150		+	31275	31424	150
40	8	X		+	31424	31774	351		+	31424	31774	351		+	31424	31774	351
41	8	X		+	31758	31910	153		+	31758	31910	153		+	31758	31910	153
42	8	X		+	31907	32992	1086		+	31907	32992	1086		+	31907	32992	1086
43	8	X		+	33010	34413	1404		+	33010	34413	1404		+	33010	34413	1404
44	8	X		+	34416	34949	534		+	34416	34949	534		+	34416	34949	534
45	1		X						+	34952	35080	129					
46	8	X		+	35150	35521	372		+	35150	35521	372		+	35150	35521	372
47	1		X						+	35529	35723	195					
48	8	X		+	35720	36040	321		+	35720	36040	321		+	35720	36040	321
49	8	X		+	36037	36570	534		+	36037	36570	534		+	36037	36570	534
50	8	X		+	36570	37079	510		+	36570	37079	510		+	36570	37079	510

“One” shows
which genes
are identified
by only one
program

Genes	TRNA																	
	TOTAL CALLS	ALL	ONE	PRODIGAL	PRODIGAL	PRODIGAL	PRODIGAL		GLIMMER	GLIMMER	GLIMMER	GLIMMER		METAGENE	METAGENE	METAGENE	METAGENE	
21	8	X		+	20008	20442	435		+	20008	20442	435		+	20008	20442	435	
22	8	X		+	20442	22835	2394		+	20442	22835	2394		+	20442	22835	2394	
23	8	X		+	22850	23170	321		+	22850	23170	321		+	22850	23170	321	
24	8	X		+	23173	23355	183		+	23173	23355	183		+	23173	23355	183	
25	8	X		+	23355	23630	276		+	23355	23630	276		+	23355	23630	276	
26	7			+	23630	23824	195		+	23630	23824	195						
27	8	X		+	23885	25201	1317		+	23885	25201	1317		+	23885	25201	1317	
28	8	X		+	25201	26409	1209		+	25201	26409	1209		+	25201	26409	1209	
29	8	X		+	26435	26785	351		+	26435	26785	351		+	26435	26785	351	
30	8	X		+	26853	27110	258		+	26790	27110	321		+	26853	27110	258	
31	8	X		-	27721	28854	1134		-	27721	28728	1008		-	27721	28728	1008	
32	8	X		+	27107	27724	618		+	27107	27724	618		+	27107	27724	618	
33	5			-	28914	29399	486		-	28914	29489	576		-	28914	29324	411	
34	8	X		+	29498	29755	258		+	29498	29755	258		+	29498	29755	258	
35	8	X		+	29752	30150	399		+	29752	30150	399		+	29752	30150	399	
36	8	X		+	30213	30749	537		+	30150	30749	600		+	30213	30749	537	
37	8	X		+	30749	30931	183		+	30749	30931	183		+	30749	30931	183	
38	8	X		+	30928	31278	351		+	30928	31278	351		+	30928	31278	351	
39	8	X		+	31275	31424	150		+	31275	31424	150		+	31275	31424	150	
40	8	X		+	31424	31774	351		+	31424	31774	351		+	31424	31774	351	
41	8	X		+	31758	31910	153		+	31758	31910	153		+	31758	31910	153	
42	8	X		+	31907	32992	1086		+	31907	32992	1086		+	31907	32992	1086	
43	8	X		+	33010	34413	1404		+	33010	34413	1404		+	33010	34413	1404	
44	8	X		+	34416	34949	534		+	34416	34949	534		+	34416	34949	534	
45	1		X						+	34952	35080	129						
46	8	X		+	35150	35521	372		+	35150	35521	372		+	35150	35521	372	
47	1		X						+	35529	35723	195						
48	8	X		+	35720	36040	321		+	35720	36040	321		+	35720	36040	321	
49	8	X		+	36037	36570	534		+	36037	36570	534		+	36037	36570	534	
50	8	X		+	36570	37079	510		+	36570	37079	510		+	36570	37079	510	

Row shading is proportional to how many programs identify a gene (darker = more programs, white = only one program)

DNA strand
(+ or -)

Gene start (stop
if strand "-")

Gene stop (start
if strand is "-")

Gene length

Phage Commander - untitled*

File

	Genes																	
	TRNA																	
	PRODIGAL	PRODIGAL	PRODIGAL	PRODIGAL		GLIMMER	GLIMMER	GLIMMER	GLIMMER		METAGENE	METAGENE	METAGENE	METAGENE		HEURISTIC	HEURISTIC	HEURISTIC
21	+	20008	20442	435		+	20008	20442	435		+	20008	20442	435		+	20008	20442
22	+	20442	22835	2394		+	20442	22835	2394		+	20442	22835	2394		+	20442	22835
23	+	22850	23170	321		+	22850	23170	321		+	22850	23170	321		+	22850	23170
24	+	23173	23355	183		+	23173	23355	183		+	23173	23355	183		+	23173	23355
25	+	23355	23630	276		+	23355	23630	276		+	23355	23630	276		+	23355	23630
26	+	23630	23824	195		+	23630	23824	195							+	23630	23824
27	+	23885	25201	1317		+	23885	25201	1317		+	23885	25201	1317		+	23885	25201
28	+	25201	26409	1209		+	25201	26409	1209		+	25201	26409	1209		+	25201	26409
29	+	26435	26785	351		+	26435	26785	351		+	26435	26785	351		+	26435	26785
30	+	26853	27110	258		+	26790	27110	321		+	26853	27110	258		+	26853	27110
31	-	27721	28854	1134		-	27721	28728	1008		-	27721	28728	1008		-	27721	28728
32	+	27107	27724	618		+	27107	27724	618		+	27107	27724	618		+	27107	27724
33	-	28914	29399	486		-	28914	29489	576		-	28914	29324	411				
34	+	29498	29755	258		+	29498	29755	258		+	29498	29755	258		+	29498	29755
35	+	29752	30150	399		+	29752	30150	399		+	29752	30150	399		+	29752	30150
36	+	30213	30749	537		+	30150	30749	600		+	30213	30749	537		+	30402	30749
37	+	30749	30931	183		+	30749	30931	183		+	30749	30931	183		+	30749	30931
38	+	30928	31278	351		+	30928	31278	351		+	30928	31278	351		+	30928	31278
39	+	31275	31424	150		+	31275	31424	150		+	31275	31424	150		+	31275	31424
40	+	31424	31774	351		+	31424	31774	351		+	31424	31774	351		+	31424	31774
41	+	31758	31910	153		+	31758	31910	153		+	31758	31910	153		+	31758	31910
42	+	31907	32992	1086		+	31907	32992	1086		+	31907	32992	1086		+	31907	32992
43	+	33010	34413	1404		+	33010	34413	1404		+	33010	34413	1404		+	32989	34413
44	+	34416	34949	534		+	34416	34949	534		+	34416	34949	534		+	34416	34949
45						+	34952	35080	129									
46	+	35150	35521	372		+	35150	35521	372		+	35150	35521	372		+	35150	35521
47						+	35529	35723	195									
48	+	35720	36040	321		+	35720	36040	321		+	35720	36040	321		+	35720	36040
49	+	36037	36570	534		+	36037	36570	534		+	36037	36570	534		+	36037	36570
50	+	36570	37079	510		+	36570	37079	510		+	36570	37079	510		+	36570	37079

Gene starts chosen by the majority of programs shown in black or white font

Gene starts chosen by a minority of programs are shown in different color font (orange, green, yellow, violet)

tRNA genes
Identified by Aragorn
in “TRNA” tab

Phage Commander - C:/Users/Administrator/Dropbox/209X/Hoot Aragorn.gq

File

Genes		TRNA					
	TOTAL CALLS	ALL	ONE	ARAGORN	ARAGORN	ARAGORN	ARAGORN
1	1	X		+	5403	5475	73
2	1	X		+	5480	5553	74
3	1	X		+	5595	5667	73

Total calls is always 1
since Aragorn is the only
program for identifying
tRNA genes

Saving your work

- Phage Commander output can be saved as .gq file for later use
- **File menu** → **Save As**
- **File menu** → **Load** (for loading .gq files)

Exporting Phage Commander Results

- To export as excel spreadsheet, select:
 - **File** → **Export** → **Excel**
- To export as GenBank (.gb) format file, select:
 - **File** → **Export** → **GenBank**

Exporting Phage Commander results in .gb format

Export genes identified by at least one program (i.e. all genes, least stringent)

Export genes identified by at least the set number of programs

Export only those genes identified by all programs (most stringent)

The screenshot shows the 'Export to Genbank' dialog box. It has two main sections: 'Export genes chosen by:' and 'Export gene starts chosen by:'. The 'Export genes chosen by:' section has three radio button options: 'Any program (Export all genes)', 'At least 8 programs', and 'All programs'. The 'At least' option has a dropdown menu showing '8'. There is a checkbox for 'Export tRNA called by Aragorn'. The 'Export gene starts chosen by:' section has three radio button options: 'Majority rule' (which is selected), 'Longest open reading frame (called by program)', and 'Specific program'. Below this is a 'Program:' dropdown menu with 'Prodigal' selected. At the bottom, there are 'Export' and 'Cancel' buttons, and a 'Save as...' button next to a text input field.

Finally, press "Export"

Press "Save as" and enter desired filename

Export gene starts chosen by majority rule

Export gene starts that generate the longest open reading frame (but are chosen by at least 1 program)

Export gene starts chosen by a specific program (e.g. Prodigal – if program does not call a particular gene, default to majority rule)

Export Gene settings

- “Any program” will export all genes identified by at least one program – this is the least conservative setting (I use this one most of the time)
- “At least” will export those genes identified by a number of programs greater than the threshold (e.g. genes identified by 2 or more programs. Use “0” as the threshold and this setting to export all genes identified)
- “All programs” will export only those genes chosen by all programs used (this is the most conservative setting, will export the fewest genes)
- “Export tRNAs called by Aragorn” will export the tRNA genes identified by Aragorn

Which gene starts to export?

- “Majority rule” – for each gene, export the start chosen by the majority of programs (in the event of tie, export the start that produces the longer of the two reading frames)
- “Longest open reading frame” – for each gene, export the start that produces the longest possible reading frame (provided it is chosen by at least one program)
- “Specific program” – export the gene starts chosen by a particular program (in the event that a gene is not identified by the chosen program, default to majority rule)