

Genome 540 Discussion

Jan. 15th 2026



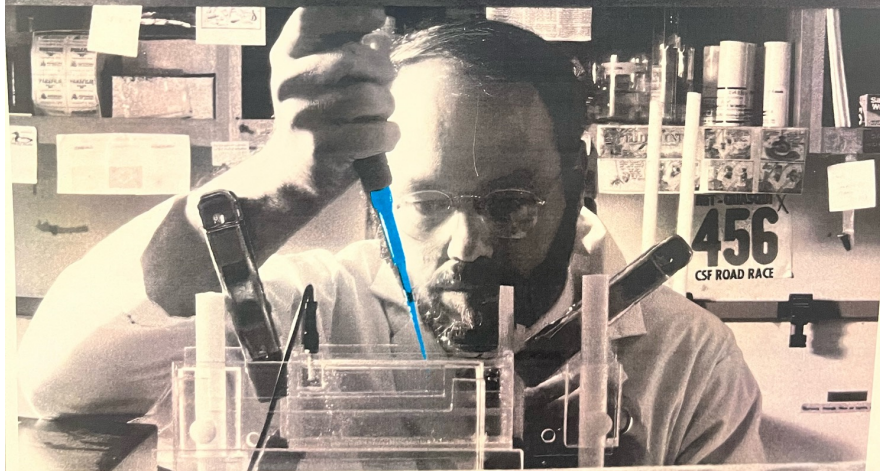
AN UNCOMMON PRODUCTIONS DOCUMENTARY
NARRATED BY MARK RUFFALO

COGS SCREENING EVENT
JANUARY 15TH @ 3:30PM
FOEGE AUDITORIUM
SOCIAL HOUR TO FOLLOW

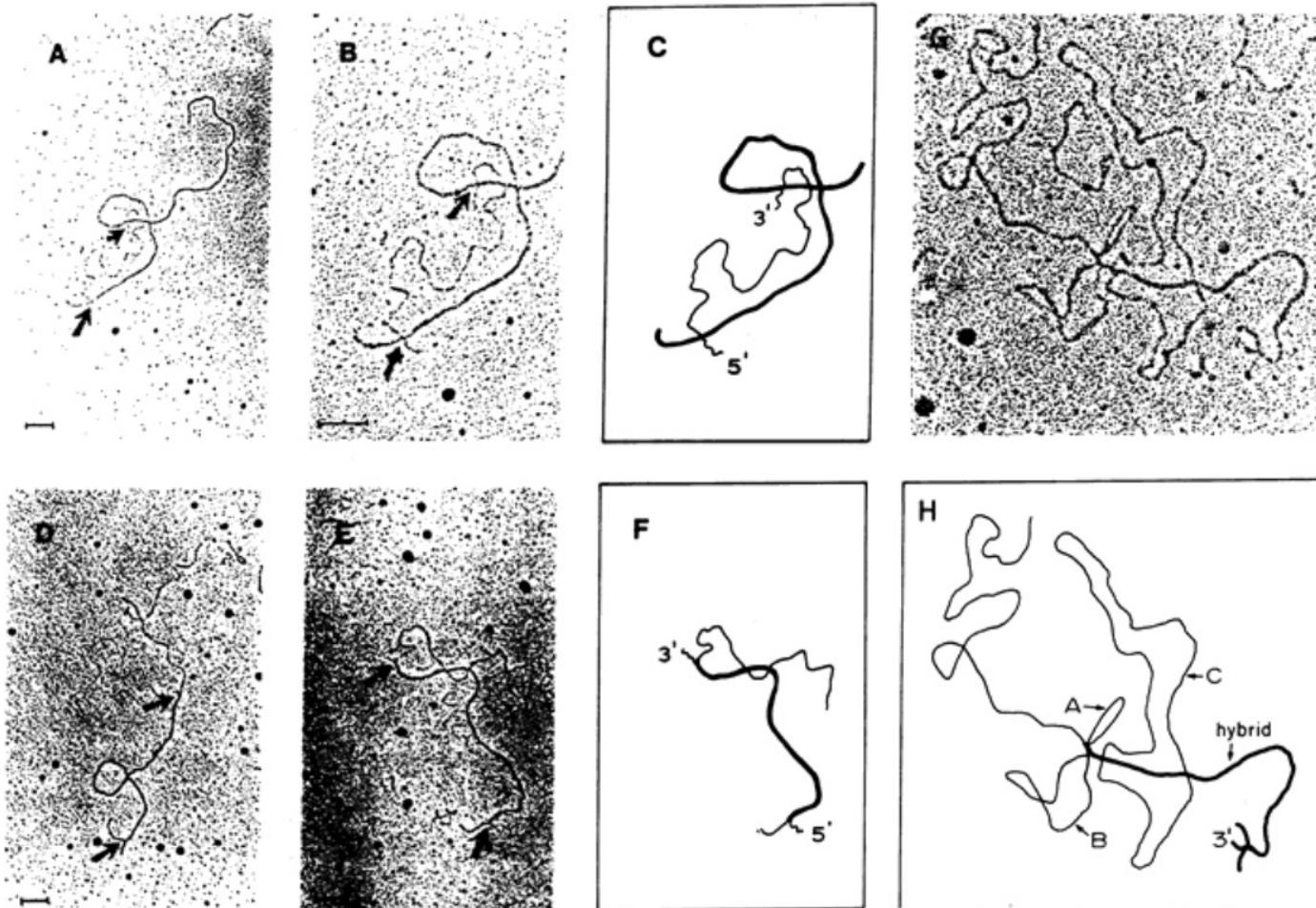


RSVP HERE

DIRECTED BY BILL HANEY



CRACKING THE CODE:
PHIL SHARP & THE BIOTECH REVOLUTION



HW1 Bits and Bobs

Can use the `time`
command if your
program is a binary

Assignment: GS 540 HW1
Name: Karl Young
Email: youngkar@uw.edu
Language: Rust
Runtime: 8.02 sec

```
hex ~/G/0/genome 540/ 🦀 time ./target/release/hw1 > young_hw1.txt
```

Executed in	8.02 secs	fish	external
usr time	7.06 secs	0.17 millis	7.06 secs
sys time	0.96 secs	1.48 millis	0.96 secs

HW1 Bits and Bobs

report "real time",
not usr or sys time

Assignment: GS 540 HW1
Name: Karl Young
Email: youngkar@uw.edu
Language: Rust
Runtime: 8.02 sec

```
hex ~/G/0/genome 540/ 🦑 time ./target/release/hw1 > young_hw1.txt
```

Executed in	8.02 secs	fish	external
usr time	7.06 secs	0.17 millis	7.06 secs
sys time	0.96 secs	1.48 millis	0.96 secs

HW1 Bits and Bobs

Otherwise, most languages have convenient ways time things

Assignment: GS 540 HW1
Name: Karl Young
Email: youngkar@uw.edu
Language: Rust
Runtime: 8.02 sec

```
>>> import time
>>> start = time.perf_counter()
>>> # code you need to time here
>>> end = time.perf_counter()
>>> print(f"Runtime: {end - start:.2f} seconds")
Runtime: 12.26 seconds
>>> █
```

HW1 Bits and Bobs

Make sure you're getting the same numbers from the example datasets,

“non-alphabetic” characters does *not* include spaces

“*” means any nucleotide

be careful with upper/lower case.

```
Fasta 1: CP001872.fna
Non-alphabetic characters: 851
>gi|284930242|gb|CP001872.1| Mycoplasma gallisepticum str. R(high)
, complete genome
*=1012027
A=349322
C=159094
G=159365
T=344246
N=0

Fasta 2: CP003913.fna
Non-alphabetic characters: 681
>gi|440453185|gb|CP003913.1| Mycoplasma pneumoniae M129-B7, comple
te genome
*=816373
A=249201
C=162924
G=163697
T=240551
N=0
```

HW1 Bits and Bobs

For each suffix in in sequence 1:

Find the longest match

Add 1 to the bucket
corresponding to that length

What does it mean for match
length 1 to have 1 entry? What is
happening at match length 7?

Match Length Histogram:

1	1
2	1
3	1
4	1
5	1
6	1
7	696
8	21780
9	139804
10	299679
11	292645
12	160266
13	62582
14	21008
15	6701
16	2217
17	840
18	352
19	231
20	176

HW1 Bits and Bobs

“Number of match strings” only counts unique strings...

But! Each unique string might occur in multiple places.

```
119 2
120 2
121 2
122 2
```

```
The longest match length: 122
Number of match strings: 1
```

```
Match string: GTCGGGTAAATTCGTCCTTGAATGGTGAACCATCTCTTGACTGTCTCGGCTATA
GACTCGGTGAAATCCAGGTACGGGTGAAGACACCCGTTAGGCGCAACGGGACGGAAAGACCCC
Description: This sequence comes from [look up entry in .gbff annotation
file using the position information below]
```

```
Fasta: CP001872.fna
Position: 338240
Strand: forward
```

```
Fasta: CP001872.fna
Position: 82469
Strand: forward
```

```
Fasta: CP003913.fna
Position: 122005
Strand: forward
```

HW1 Bits and Bobs

Just copy-paste your source code here.

Don't need build instructions.

Program:

```
int main() {  
    do_analysis();  
    return 0;  
}
```

HW1 Bits and Bobs

Run your program on the provided example files

You can use ``diff`` to check differences between the template and your output.

Many IDEs also have visual ways to compare files.

```
hex ~/G/0/g/h/example/  diff hw1_template.txt hw1_test.txt
2,5c2,5
< Name: Conor Camplisson
< Email: concamp@uw.edu
< Language: C++
< Runtime: 8.08 sec
---
> Name: Karl Young
> Email: youngkar@uw.edu
> Language: Rust
> Runtime: xx sec
158c158
< Position: 338240
---
> Position: 82469
162c162
< Position: 82469
---
> Position: 338240
171,175d170
<
< int main() {
<     do_analysis();
<     return 0;
< }
```

HW1 Bits and Bobs

Thinking ahead? HW2

Simulate some random sequences, save as a fasta files.

Run HW1 *on the simulated sequences!*

Can you structure your HW1 program such that it's easy to swap files in and out?

HW1 Bits and Bobs

Please! Put your name and the assignment number in the file you submit!

Something like:

young_hw1_2026.txt

Submission is just an email to Phil and myself.