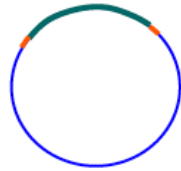


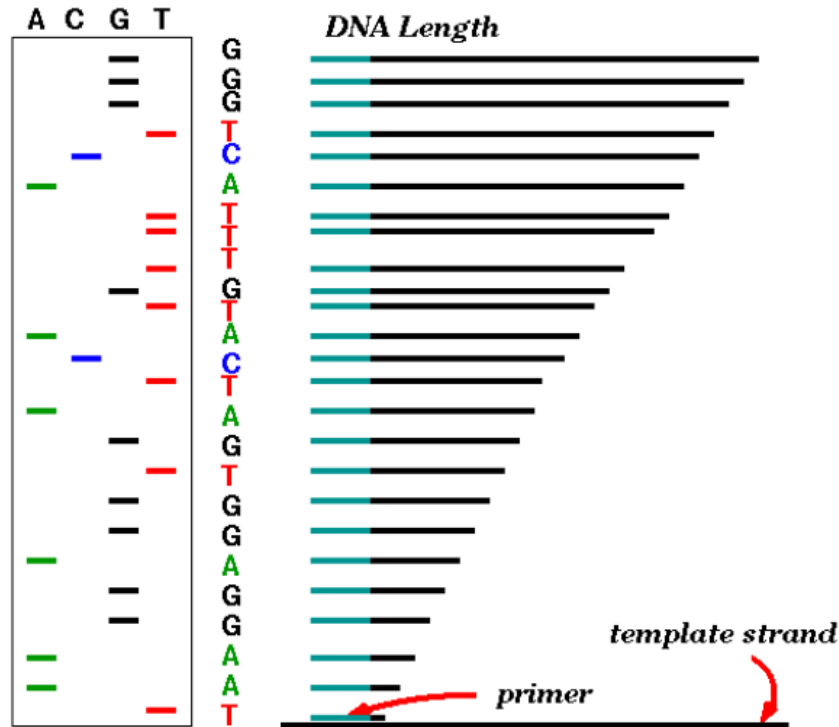
Beyond Laboratory Notebooks: Next Era Biological Data Hurdles for Information Storage, Access and Distribution

Richard Slayden, PhD.
Microbiology, Immunology & Pathology

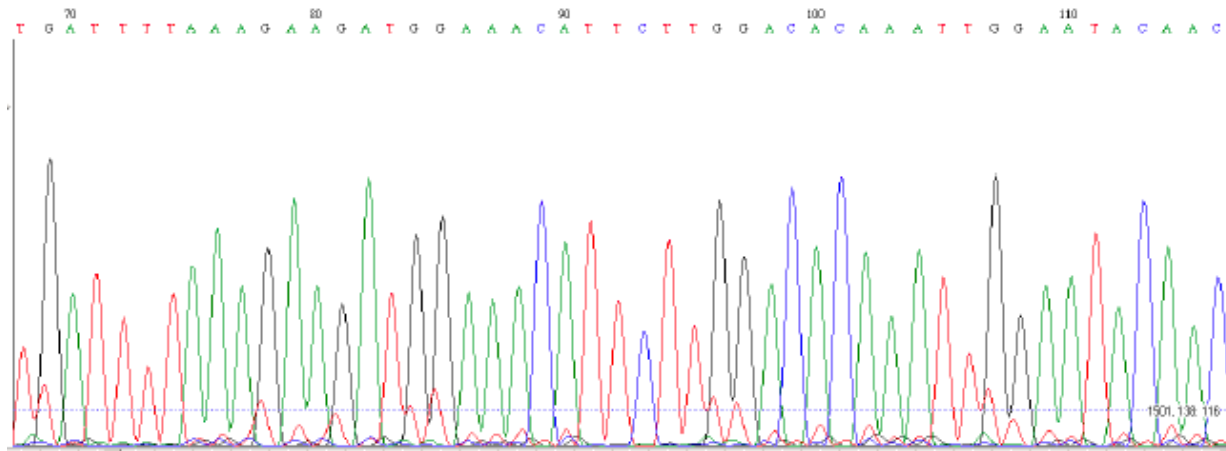
Traditional sequencing: *Sanger Sequencing Process*



1. Start at primer
2. Grow DNA chain
3. Include dideoxynucleoside (modified a, c, g, t)
4. Stops reaction at all possible points
5. Separate products with length, using gel electrophoresis



source: robotics.stanford.edu/~serafim/cs262/Spring2003/Slides/Lecture9.ppt



Background of Next Generation Sequencing: Impact

2001 First human genome

sequence draft: ~ 13 years and 300 million US\$

Technology Review

May 2005: ~ 6 month and 20 to 30 million US\$

The Scientist

(Vol. 20,2 p.67) 454: ~ 1 month and 900 000 US\$ (1x coverage)

The Scientist

(Vol. 20,2 p.67) Solexa: ~ 6 month and 50 000 US\$ (15x coverage)

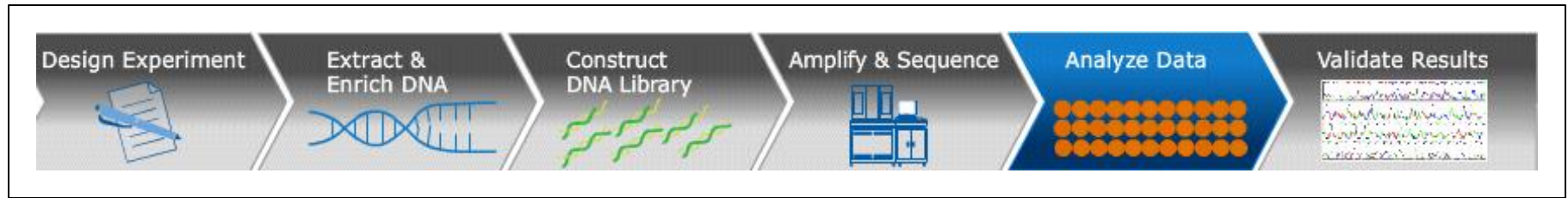
Published literature using AB SOLiD

SOLiD sequencer: 14 days and 20 000 US\$ (~10x coverage)

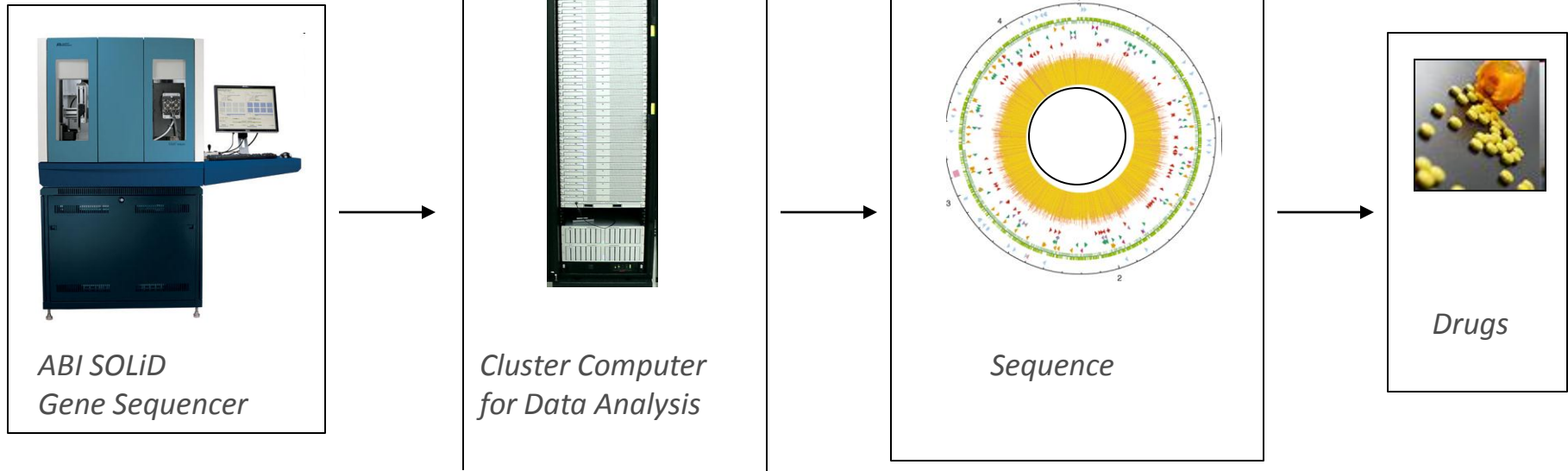
Next Generation Sequencing-*Infectious Diseases Research Center*

Sequencing genomes and transcripts for drug discovery

NextGen workflow

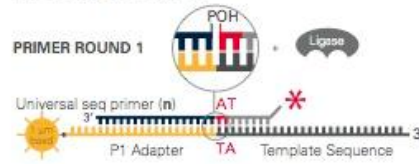


NextGen technology

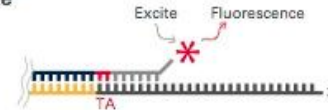


Ligation Based Sequencing

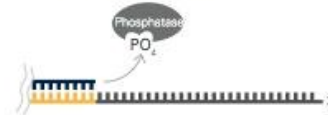
1. Prime and Ligate



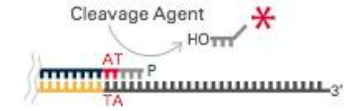
2. Image



3. Cap Unextended Strands



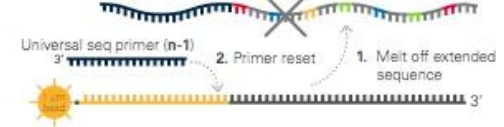
4. Cleave off Fluor



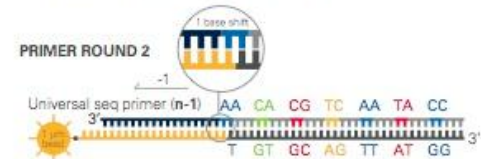
5. Repeat steps 1-4 to Extend Sequence



6. Primer Reset



7. Repeat steps 1-5 with new primer



8. Repeat Reset with , n-2, n-3, n-4 primers

	Read Position	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35			
1	Universal seq primer (n) 3'		•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•		
2	Universal seq primer (n-1) 3'			•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	
3	Universal seq primer (n-2) 3'				•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	
4	Universal seq primer (n-3) 3'					•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
5	Universal seq primer (n-4) 3'						•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•

• Indicates positions of interrogation

Ligation Cycle: 1 (blue), 2 (orange), 3 (green), 4 (red), 5 (purple)

Background of Next Generation Sequencing: Sequencing software

The screenshot displays the Applied Biosystems SOLiD 3.0 software interface, titled "MICHILLE - SOLiD - 3.0 - User: lab_user". The browser address bar shows the URL: https://www3.appliedbiosystems.com/cms/groups/portal/documents/web_content/cms_059386.jpg.

The interface is divided into two main panels for Flow Cell 1 and Flow Cell 2.

Flow Cell 1 (Stopped):

- Run Name: MICHILLE_20080929
- Created by: lab_user
- Sample Slide: 1 Sample -in- 8_spot_mask_sf
- Protocol: SOLiD v3.0_v7.3_1x, R3 + F3, 25 + 25 bases
- Pre-Scan: R3, F3
- Status: Stopped
- Buttons: Start Run, Load Flowcell..., Clear Flowcell...

Flow Cell 2 (Running):

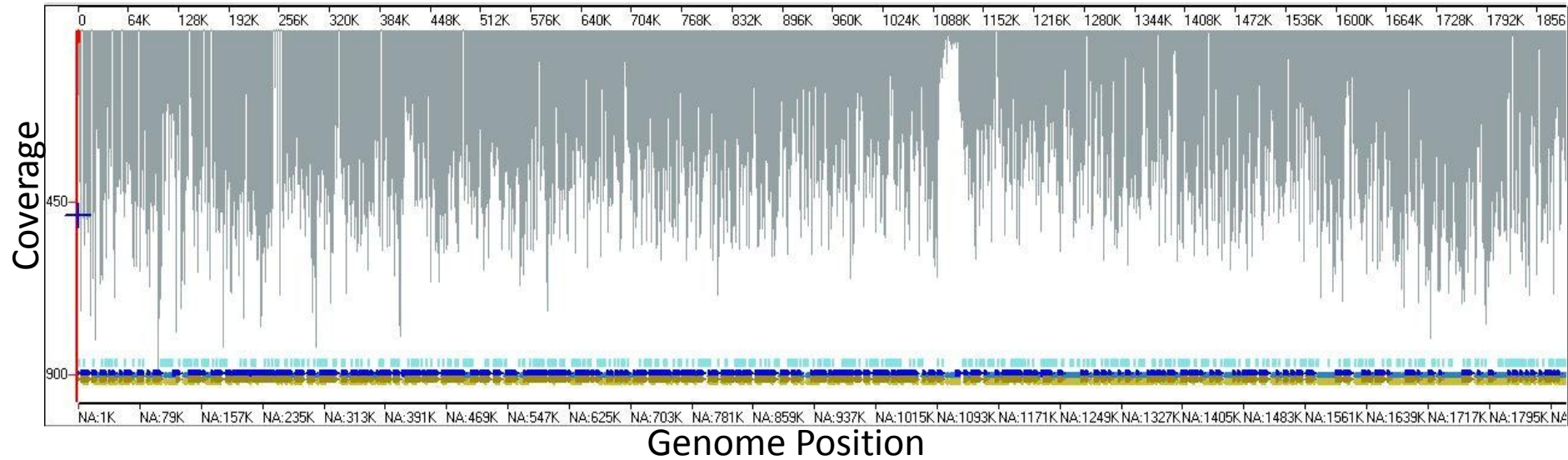
- Run Name: MICHILLE_20080925_syn_2
- Created by: lab_user
- Sample Slide: 1 Sample -in- 1_spot_mask_sf
- Protocol: SOLiD v3.0_v7.3_2x, F3, 25 bases
- Pre-Scan: F3
- Process: F3 - Primer 2 - Ligation 3 (Ligate-Dar, Dark Ligat, Phosphate, SOLiD 25, Scan Slide, Cleave Lig)
- Status: Running
- Buttons: Pause Run..., Stop Run

System Status (Left Sidebar):

- Prime Buffer Line: Priming
- Chiller: 4.0 C, Cooling
- Lamp: 99 hrs, On
- Reset..., Turn Off
- Door: Closed-Locked, Unlock Doors

The bottom status bar indicates "Imaging is Running", "Fluidics is Online", and "146M of 297M".

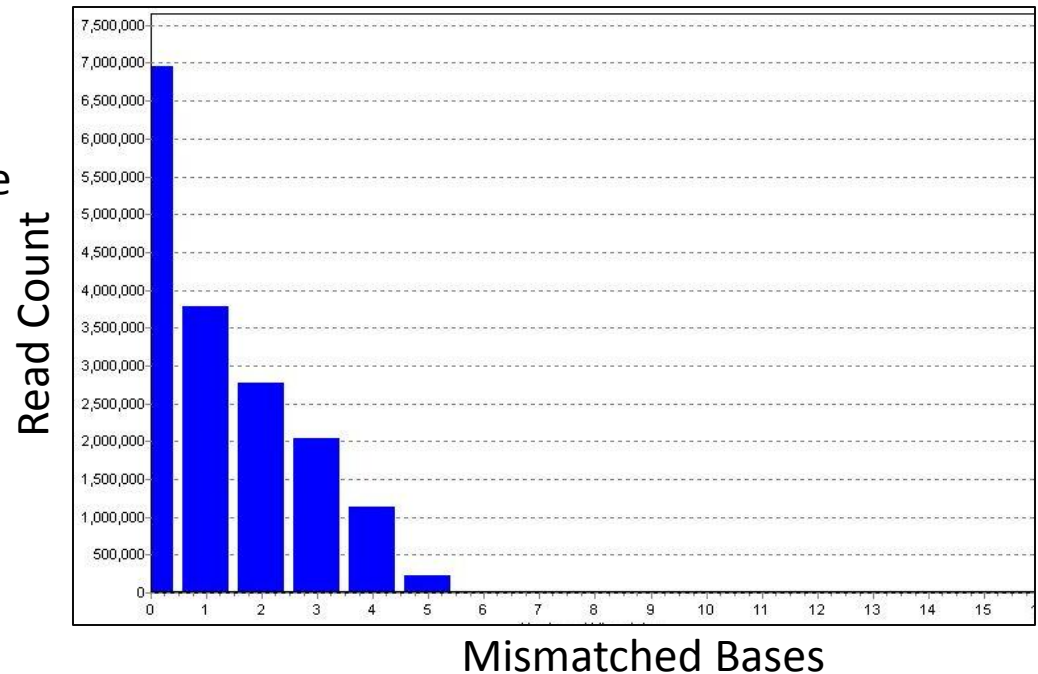
Background of Next Generation Sequencing: Whole Genome sequencing



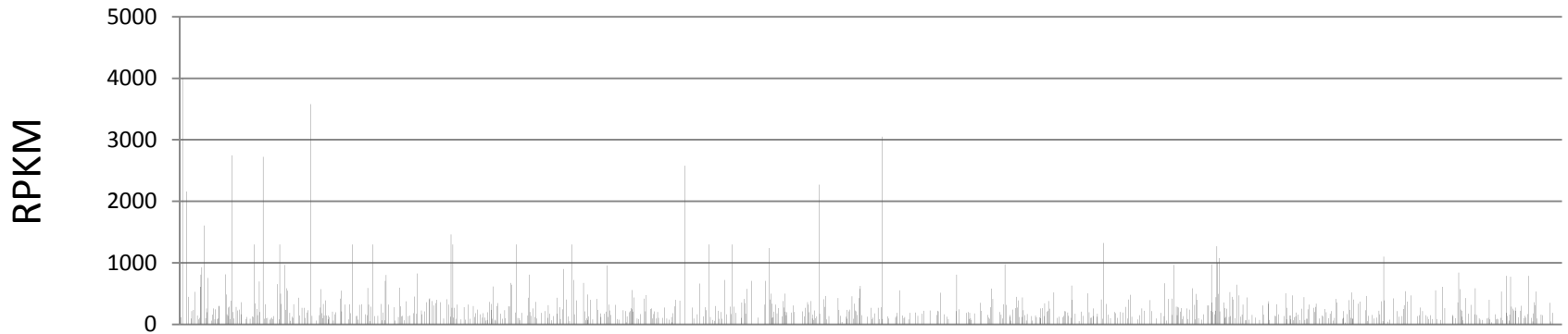
-1624 SNPs

-74% Within Proposed Coding Sequence

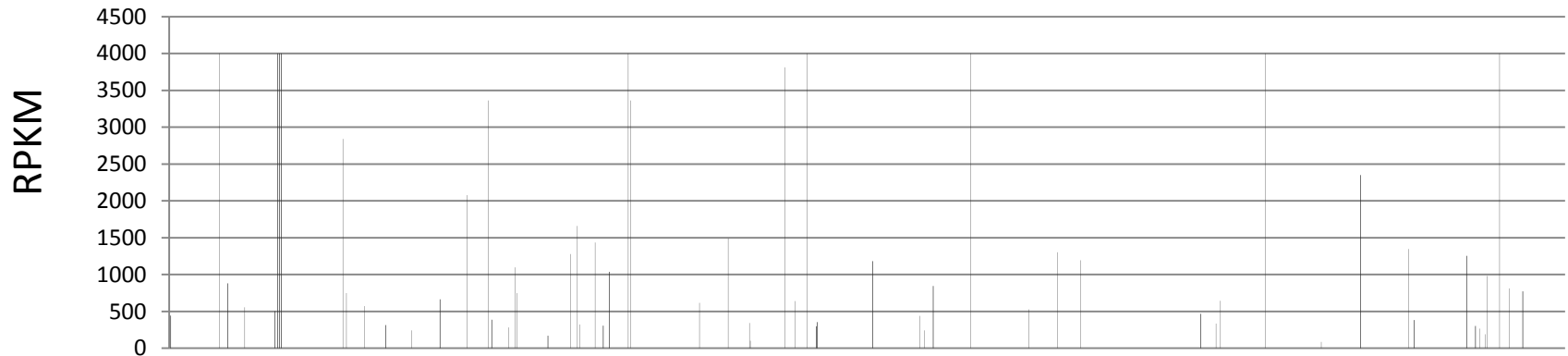
-1419 Non-synonymous SNPs



Background of Next Generation Sequencing: Unbiased Molecular Detection

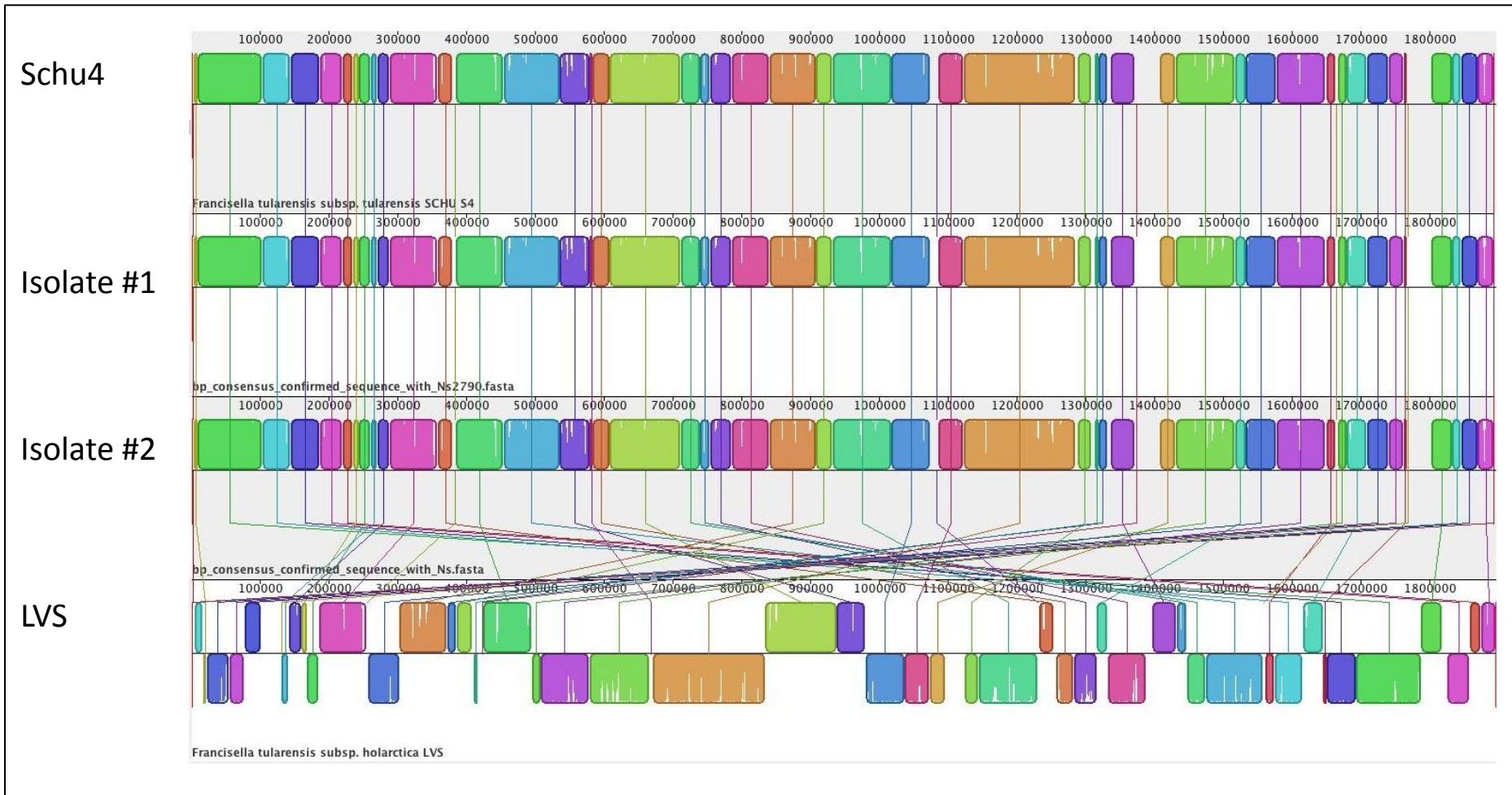


Annotated open reading frames



Non-annotated open reading frames

Advanced Analysis of Next Generation Sequencing



Current Data management & Storage Issues

Data output:

- ✓ *One sequencing run: 2-8 Tb*
- ✓ *Single sequencing run: 4-14 days*
- ✓ *Instrument able to perform two sequencing runs as a single time*

Envisioned Needs:

1. *Data Storage*
2. *Movement of data (bandwidth)*
3. *Access to data files*
4. *Distribution of data files*
5. *Bioinformatics software*