Molecular Barcoding for Multiplexing

Now that you can sequence the human genome on a sequencer, what do you do for microbes?

J. Craig Venter™
INSTITUTE

Where we are

- 13:30-14:00 – Primer Design to Amplify Microbial Genomes for Sequencing
- 14:00-14:15 – Primer Design Exercise
- **14:15-14:45 – Molecular Barcoding to Allow Multiplexed NGS**
- 14:45-15:15 – Processing NGS Data – de novo and mapping assembly
- 15:15-15:30 – Break
- 15:30-15:45 – Assembly Exercise
- 15:45-16:15 – Annotation
- 16:15-16:30 – Annotation Exercise
- 16:30-17:00 – Submitting Data to GenBank

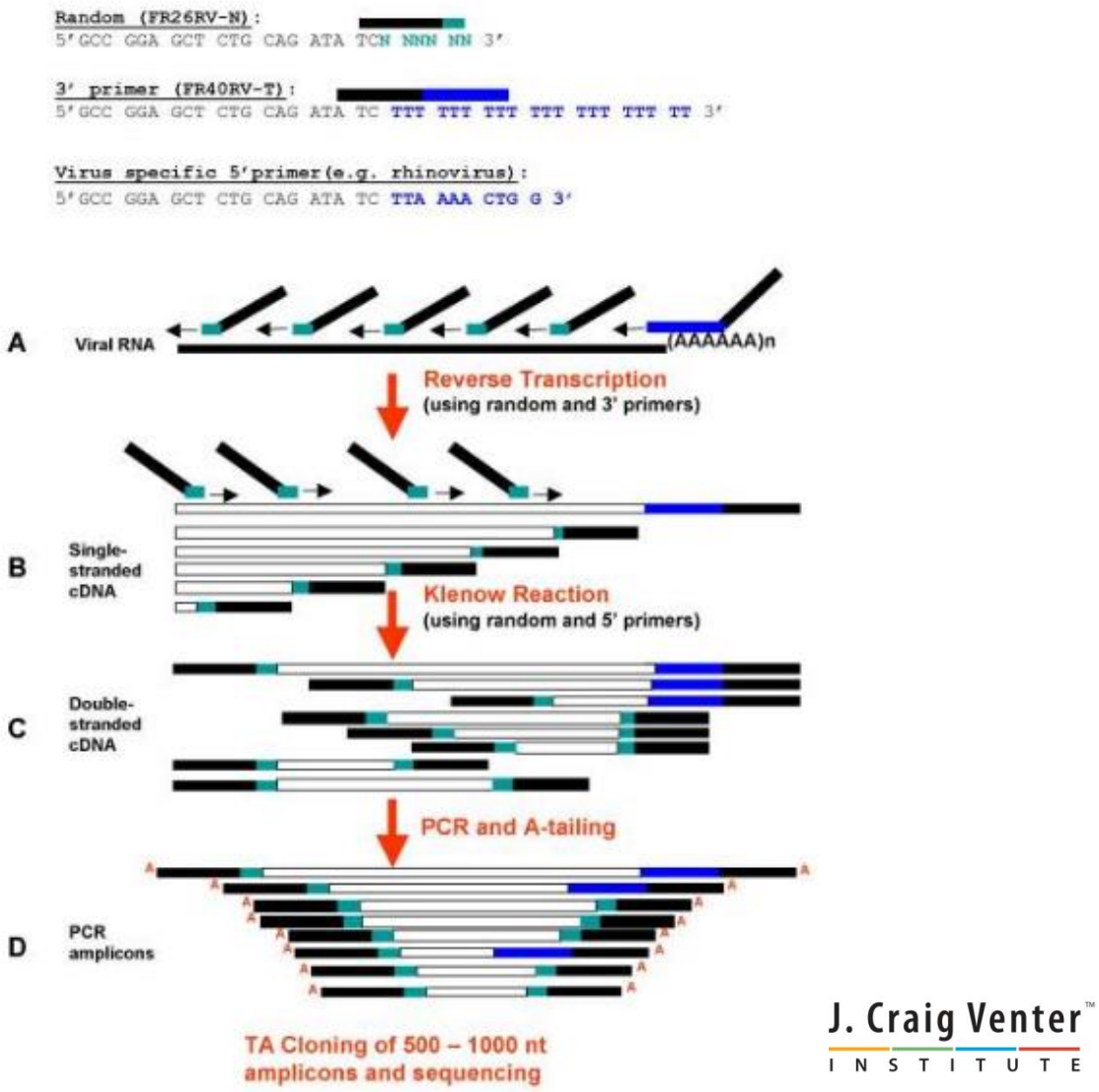
J. Craig Venter™ INSTITUTE

Why add molecular barcodes?

- Goal is to sequence as many genomes as possible, to the required depth of coverage, as economically as possible
- All NGS platforms have their own methods for molecular barcoding, but are often expensive
- JCVI developed a method for adding our own barcodes, less expensive, but has some drawbacks

Sequence Independent Single Primer Amplification - SISPA

- Early SISPA research at JCVI was developed to amplify viral genome sequences for downstream TA cloning and Sanger sequencing
- The method has been adapted to allow molecular barcoding of samples by designing 100's of unique sequences that have the following characteristics:
  - Tm and sequence composition is compatible with SISPA protocol
  - Compatible with NextGen library construction and sequencing protocols
  - Distinguishable from one another, allowing up to 10% sequencing error

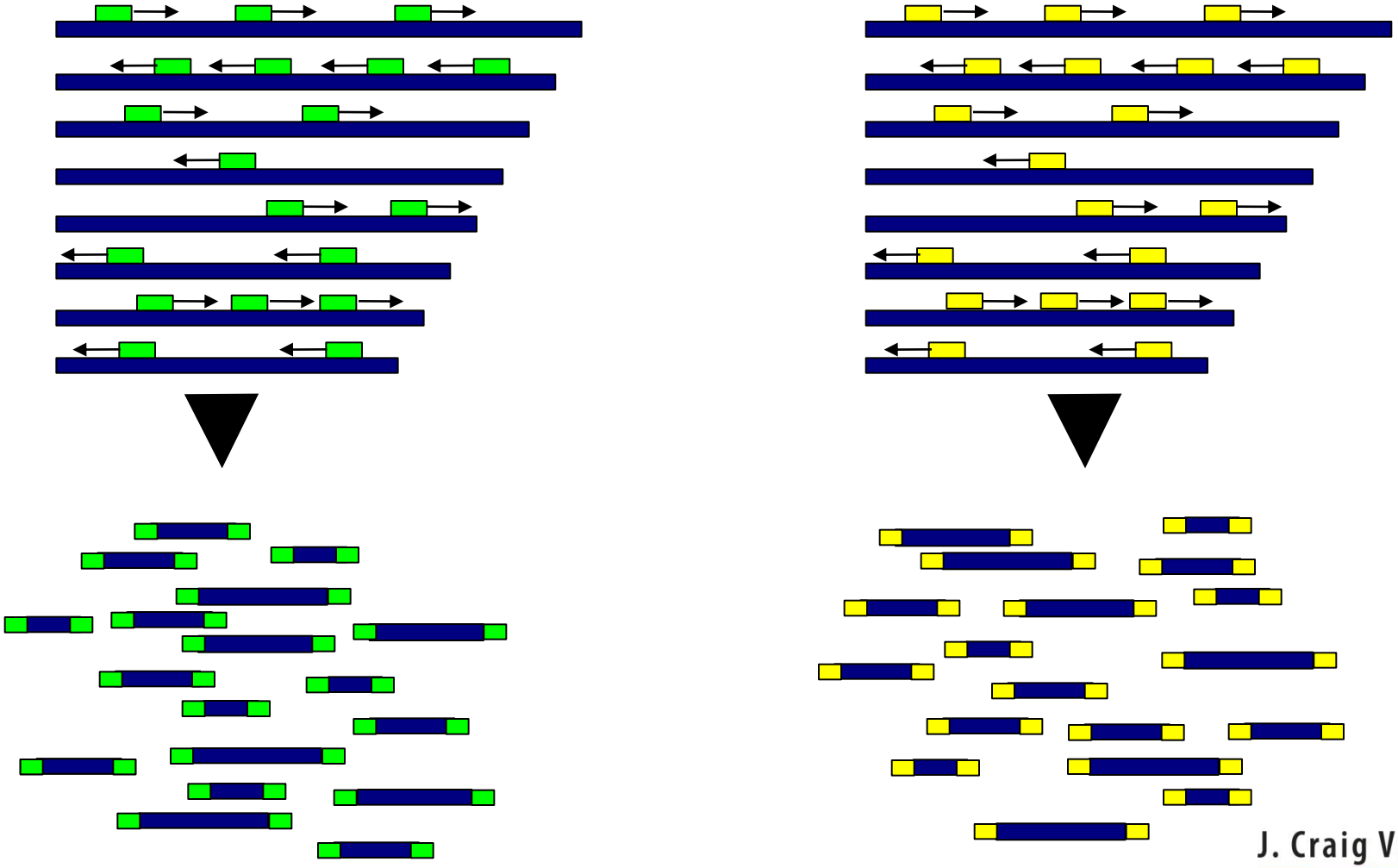


Djikeng A, Halpin R, Kuzmickas R, Depasse J, Feldblyum J, Sengamalay N, Afonso C, Zhang X, Anderson NG, Ghedin E, Spiro DJ. Viral genome sequencing by random priming methods. BMC Genomics. 2008 Jan 7;9:5.

J. Craig Venter™ INSTITUTE

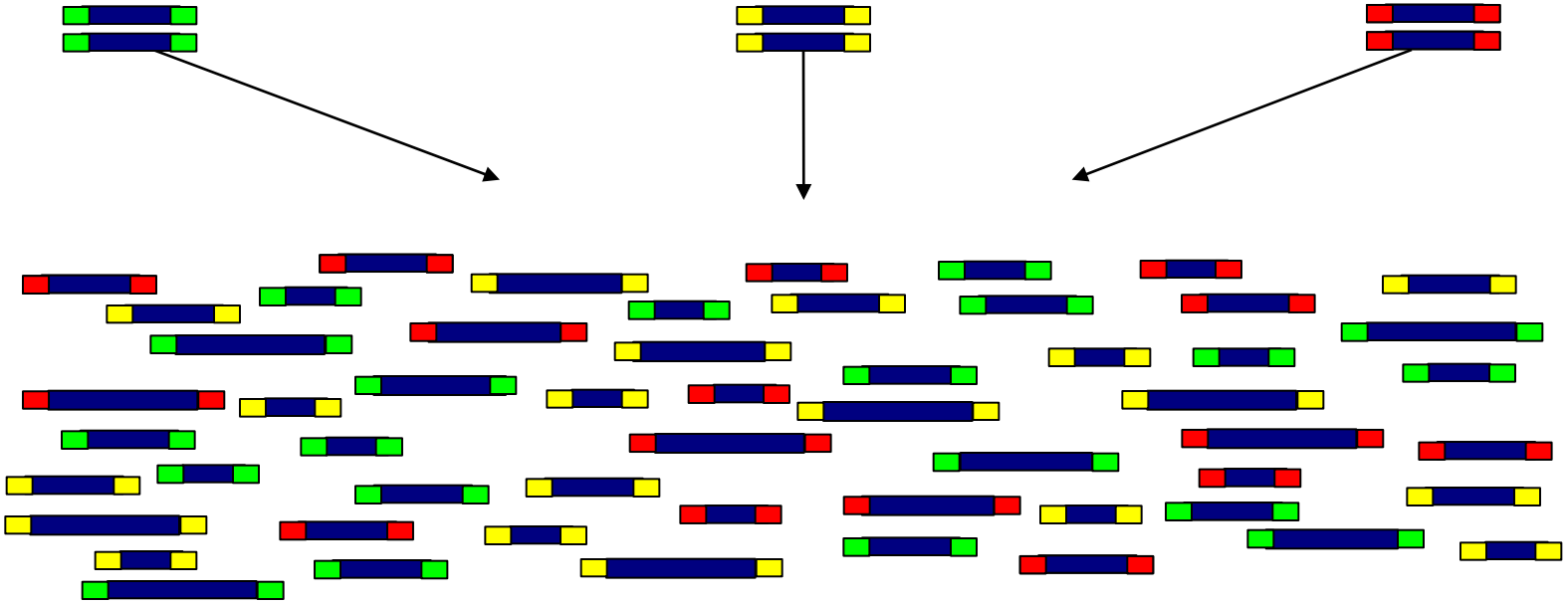
Bar-Coding and Genome Amplification

Sequence Independent Single Primer Amplification (SISPA)



J. Craig Venter™ INSTITUTE

Sample Pooling/NextGen Sequencing



J. Craig Venter™
INSTITUTE

Our typical SISPA Protocol

- For SISPA Barcoding
  - For every ~15kb of viral genome – give sample two SISPA barcodes - examples
    - Norovirus – 7.5kb – two unique SISPA barcode
    - Influenza Virus – 15kb – two unique SISPA barcodes
    - Coronavirus – 30kb – four unique SISPA barcodes
- Normalize and Pool SISPA products
  - 192 normalized SISPA products go into our typical pool
- Illumina Libraries and Pooling into Lanes
  - Each SISPA product pool is used to build an indexed Illumina PE library or Roche/454 Rapid Library
  - On Illumina HiSeq 2000, we have been pooling 3 of these libraries per lane

Performance on HiSeq 2000

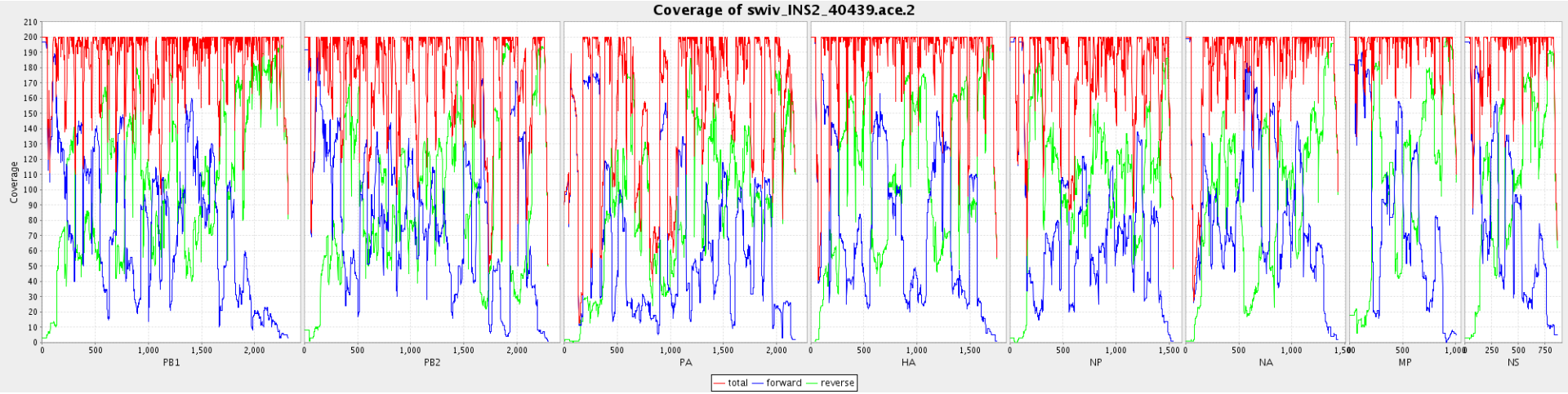
LANE	Library Name	Yield (Mbases)	% PF	# Reads	% of >= Q30 Bases (PF)	Mean Quality Score (PF)	
349	SISPA20120406-PE-IL15-01	4,622	63.85	71,665,850	62.67	26.87	
349	SISPA20120407-PE-IL16-01	4,174	62.64	65,965,754	62.38	26.82	
349	SISPA20120408-PE-IL23-01	3,623	63.03	56,901,646	66.01	27.81	
350	SISPA20120416-PE-IL24-01	3,457	72.38	47,295,656	69.11	28.89	
350	SISPA20120417-PE-IL26-01	5,697	70.94	79,503,452	68.83	28.81	
350	SISPA20120418-PE-IL27-01	6,500	71.62	89,859,024	69	28.87	
339	SISPA20120326RESUB-PE-IL10-01	4,628	65.31	70,163,402	68.44	28.67	
339	SISPA20120328-PE-IL12-02	9,831	73.75	131,980,380	72.82	29.91	
		5,317	***67.9***	***76,666,896***	***67.4***	***28.3***	***avg***
		2,088	***4.7***	***25,847,948***	***3.5***	***1.1***	***stdev***

PhiX spike in lane

LANE	Library Name	Yield (Mbases)	% PF	# Reads	% of >= Q30 Bases (PF)	Mean Quality Score (PF)	
352	20120515SISPA21-PE-IL29-01	4,532	73.56	60,999,882	69.62	29.03	
352	SISPA20120425-PE-IL28-01	3,164	72.44	43,244,082	70.02	29.1	
352	SISPA20120601-PE-IL30-01	3,117	72.23	42,725,366	67.33	28.3	
352	PhiX spike in + unknown	4,413	93.44	46,765,556	88.37	34.59	
		3,604	***72.7***	***48,989,777***	***69.0***	***28.8***	***avg***
		804	***0.7***	***10,404,289***	***1.5***	***0.4***	***stdev***

J. Craig Venter™ INSTITUTE

Sequencing Coverage of mRT-PCR-amplified, SISPA-barcoded Influenza Sample



NOTE: We limit depth of coverage to 200x max so that assembly files remain at reasonable size, and other analyses can run in under 2 hours.

J. Craig Venter™ INSTITUTE

SISPA Software

- **JCVI Sequence BARcode Designer** – to design SISPA barcodes, http://sourceforge.net/projects/jcvibard/
- **JCVI DNA Barcode Deconvolution** – to demultiplex reads based on barcodes used in a NGS run, http://sourceforge.net/projects/deconvolver/

NGS Vendor Barcoding

- Roche/454 – Multiplex Identifier (MID) in adaptors
- Illumina – Nextera tagmentation and Illumina Index reads
- LifeTech Ion PGM – IonXpress barcoding
- All of these are added as part of individual library construction, and demultiplexed by vendors' software