Viruses in the Era of Genomic

Paolo Amedeo, JCVI
January, 20 2018
OPCUG / PATACS

J. Craig Venter™
INSTITUTE

What is a virus?

- An entity at the boundary between life and something inanimate.
- Egoistic chunk of genetic code.
- A vehicle for evolution and genome plasticity.

  …
- An organism that carries its genetic identity, but needs an host to replicate itself.

Incredible variability in genome size and organization

Genome size: 3,200nt (Hepatitis B virus) – 1,200,000nt (Mimivirus)

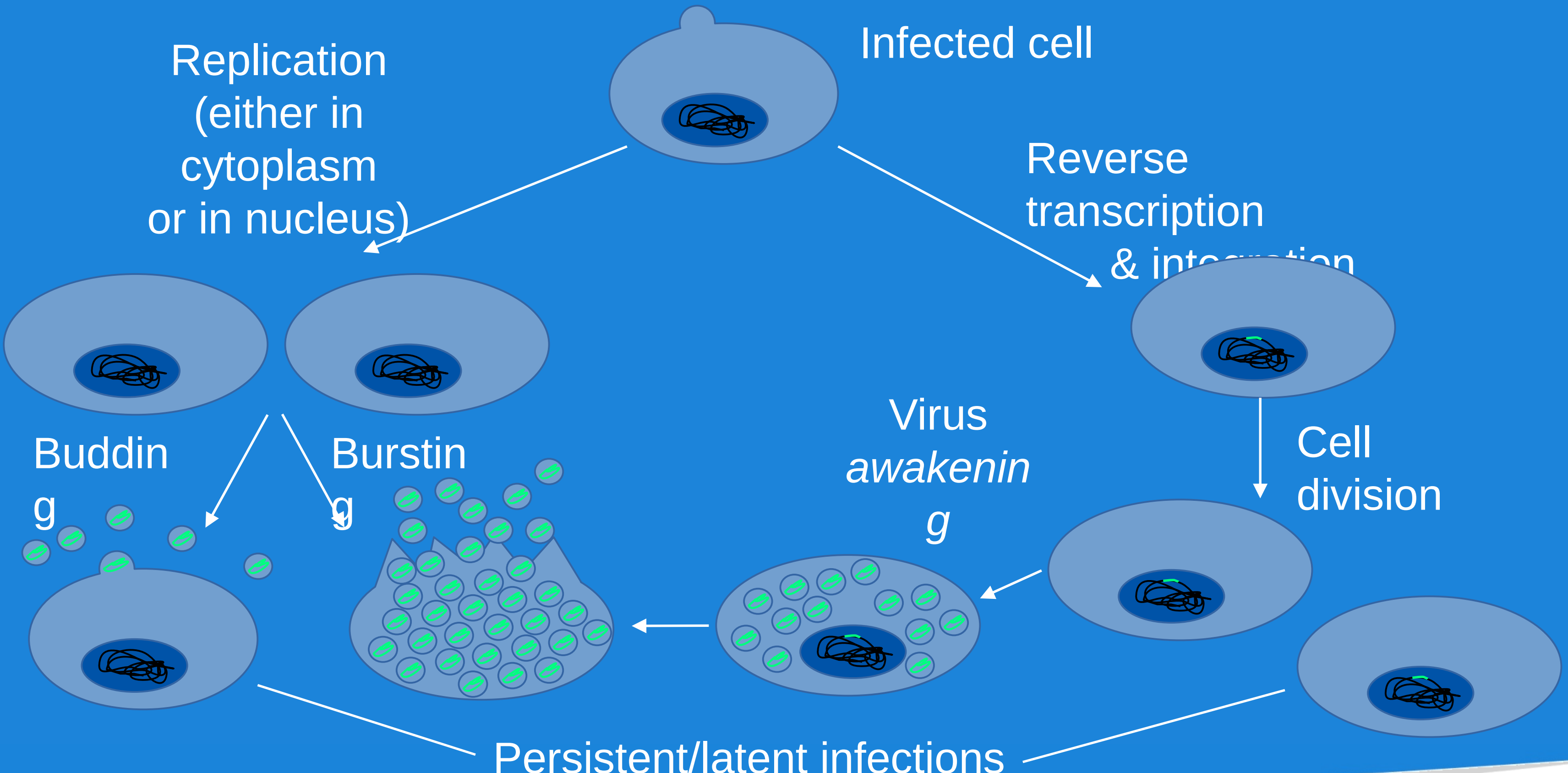
Genome organization: linear, circular, single molecule, multiple molecules.

Nucleic acid: positive, single-strand (SS) RNA, negative SS RNA, double-strand (DS) RNA, SS DNA, DS DNA

Replication: in cytoplasm, in nucleus, insertion in the host genome (retrovirus).

Infectivity: The naked nucleic acid of some viruses in infective, other viruses need the complete viral particle to be able to infect the host cell.

Main replication strategies



Infected cell

Replication (either in cytoplasm or in nucleus)

Reverse transcription & integration

Budding

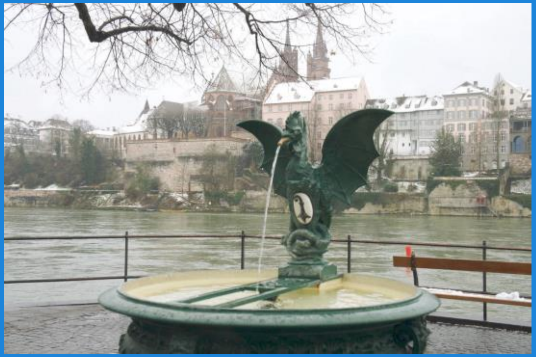
Bursting

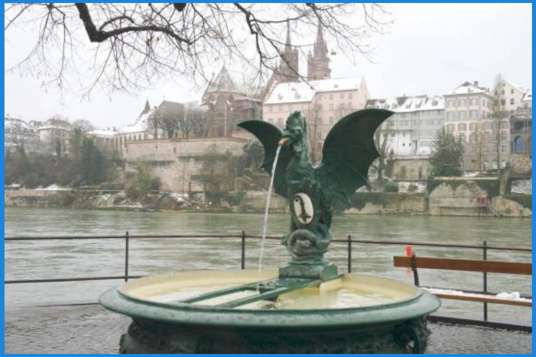
Cell division

Virus *awakening*

Persistent/latent infections

Computer vs. Biology





Computers are very good in recognizing identical objects. But, except for rare cases, identifying similar objects is a rather complex challenge.

In biology, being identical is just like being very similar. And not always what is chemically affine (i.e. similar) looks similar at our eyes.

A refresher of a few concepts of molecular biology

Bases
:

Adenine (A) Thymine (T) only DNA
Cytosine (C) Uracil (U) only RNA
Guanine (G)

Pairing:

A-T (A-U)
C-G

Double-stranded helix (either DNA or RNA) are anti-parallel

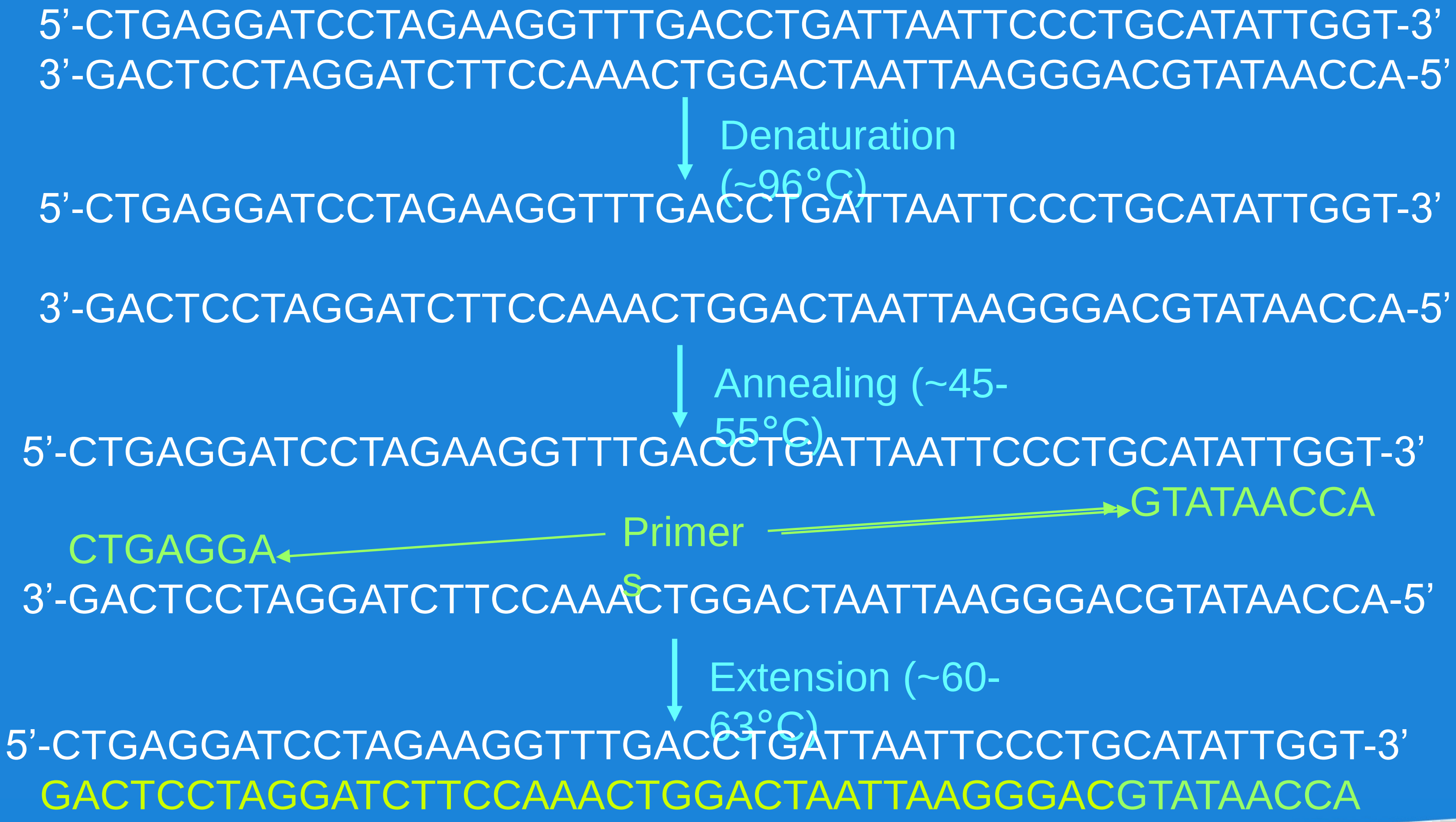
5’-CTGAGGATCCTAGAAGGTTTGACCTGATTAATTCCCTGCATATTGGT-3’
3’-GACTCCTAGGATCTTCCAAACTGGACTAATTAAGGGACGTATAACCA-5’

To replicate DNA or RNA we need a ***primer***, that is a short sequence complementary to the sequence we need to replicate.

ACTCCTAGGA →
CTGAGGATCCTAGAAGGTTTGACCTGATTAATTCCCTGCATATTGGT

DNA and RNA synthesis happens only in direction 5’ → 3’

Polymerase Chain Reaction (PCR)



5’-CTGAGGATCCTAGAAGGTTTGACCTGATTAATTCCCTGCATATTGGT-3’
3’-GACTCCTAGGATCTTCCAAACTGGACTAATTAAGGGACGTATAACCA-5’

Denaturation (~96°C)

5’-CTGAGGATCCTAGAAGGTTTGACCTGATTAATTCCCTGCATATTGGT-3’

3’-GACTCCTAGGATCTTCCAAACTGGACTAATTAAGGGACGTATAACCA-5’

Annealing (~45-55°C)

5’-CTGAGGATCCTAGAAGGTTTGACCTGATTAATTCCCTGCATATTGGT-3’

GTATAACCA

Primers

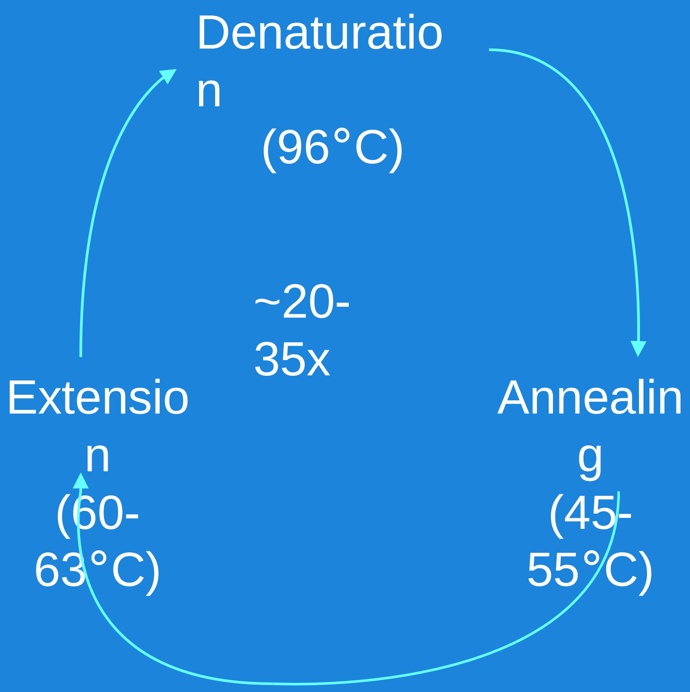
CTGAGGA

3’-GACTCCTAGGATCTTCCAAACTGGACTAATTAAGGGACGTATAACCA-5’

Extension (~60-63°C)

5’-CTGAGGATCCTAGAAGGTTTGACCTGATTAATTCCCTGCATATTGGT-3’
GACTCCTAGGATCTTCCAAACTGGACTAATTAAGGGACGTATAACCA

PCR (Continued)



Denaturation

(96°C)

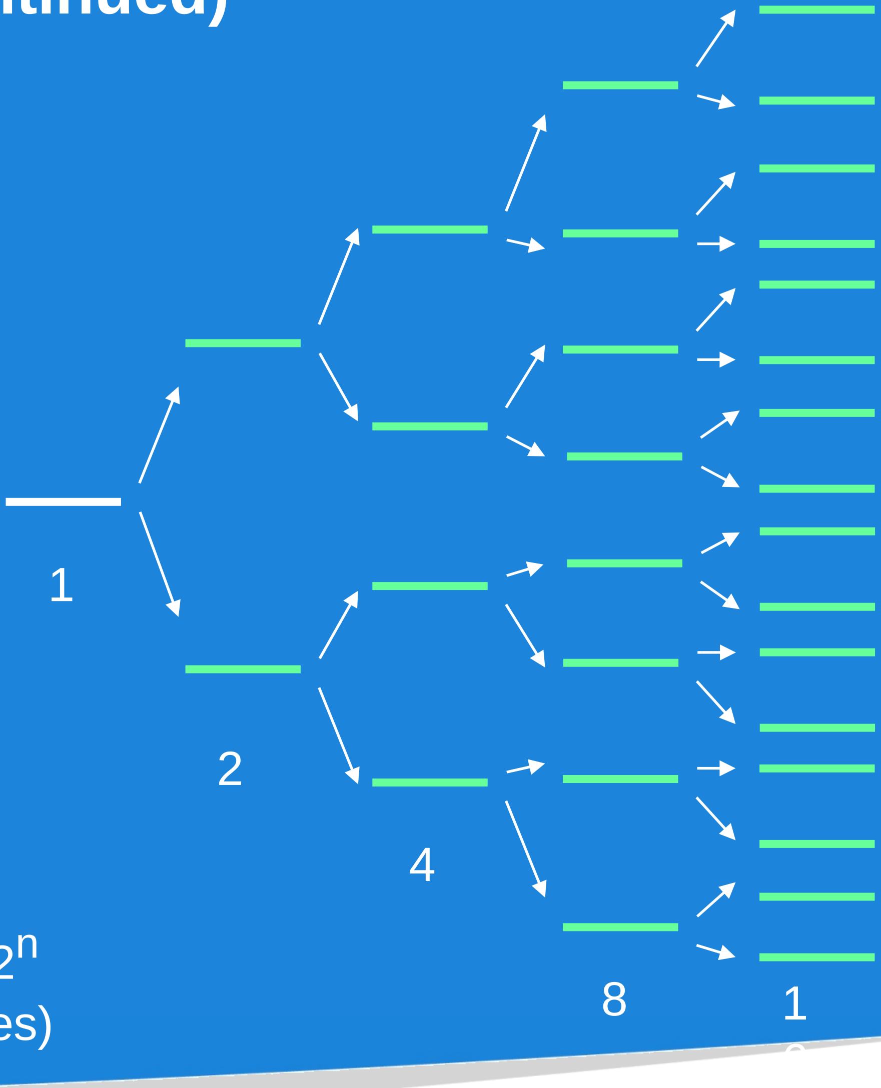
~20-35x

Extension

(60-63°C)

Annealing

(45-55°C)



Max amplification: 2^n

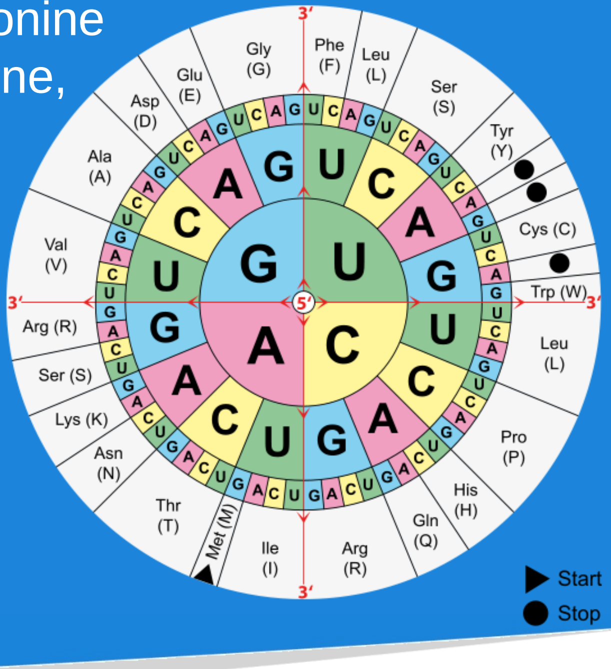
(n = number of cycles)

Translation

RNA is translated into amino acid sequence in groups of three nucleotides, called **codon**.

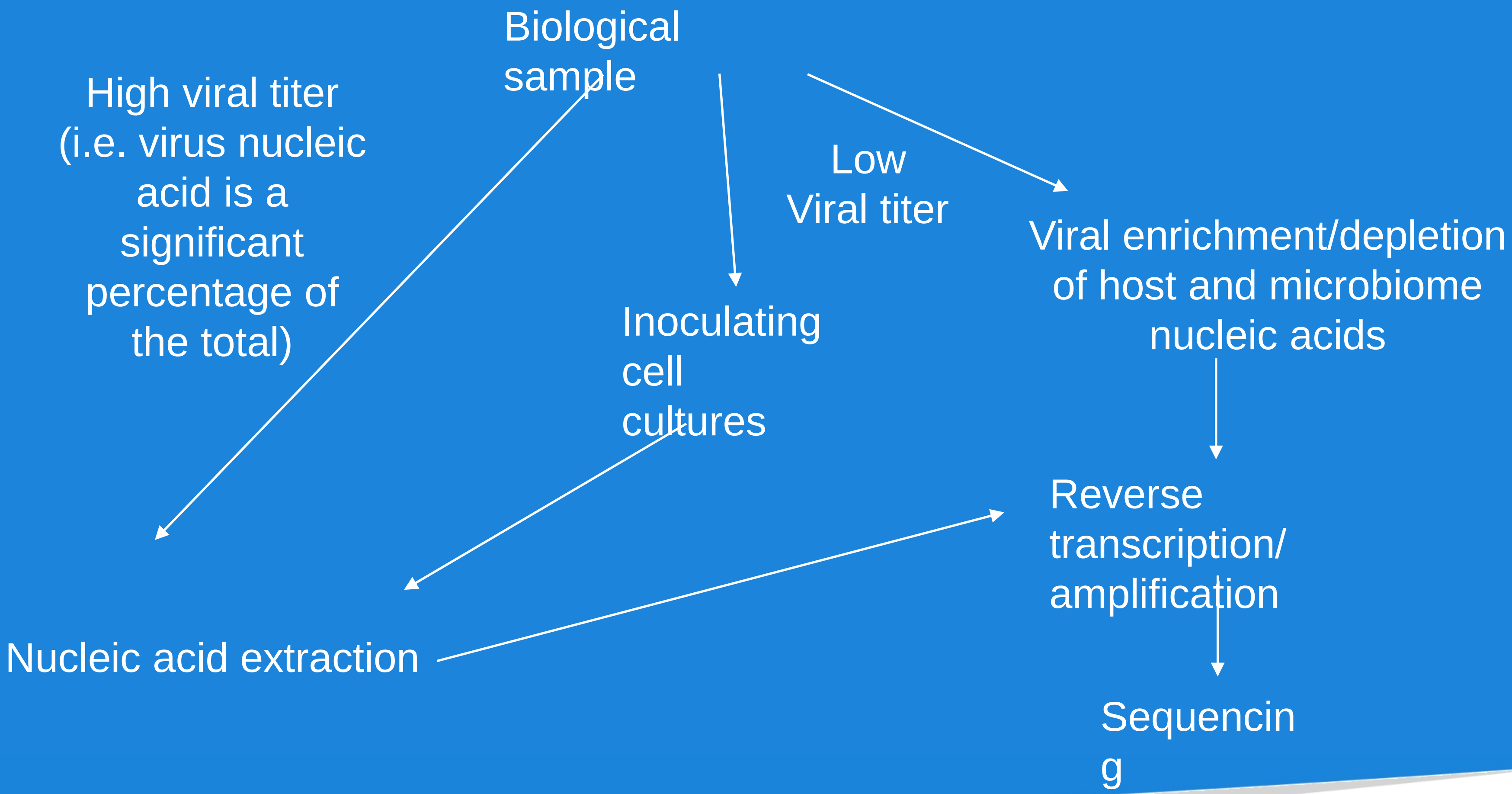
Every protein starts with the amino acid methionine (M) and there are thee triplets that do not codify for any amino acid (stop codons).

Two amino acids are coded by a unique codon (methionine and tryptophan), three are coded by six codons (arginine, leucine, and serine). All the others are coded by anything between two and four codons.
Therefore, two significantly different nucleotide sequences can code for identical or similar protein sequence.



ATG CTG TCG CGG CCG GTG GCG ACG TGC
ATG TTA AGT AGA CCA GTT GCT ACT TGT
Met Leu Ser Arg Pro Val Ala Thr Cys
(100% aa - 55.6% nt identity)

Isolating and sequencing viruses



Biological sample

High viral titer (i.e. virus nucleic acid is a significant percentage of the total)

Low Viral titer

Inoculating cell cultures

Viral enrichment/depletion of host and microbiome nucleic acids

Nucleic acid extraction

Reverse transcription/ amplification

Sequencing

Depletion/enrichment strategies

DNA viruses

RNAse treatment
Size fractionation (remove anything longer than expected genome size)
Subtractive hybridization with host DNA

RNA viruses

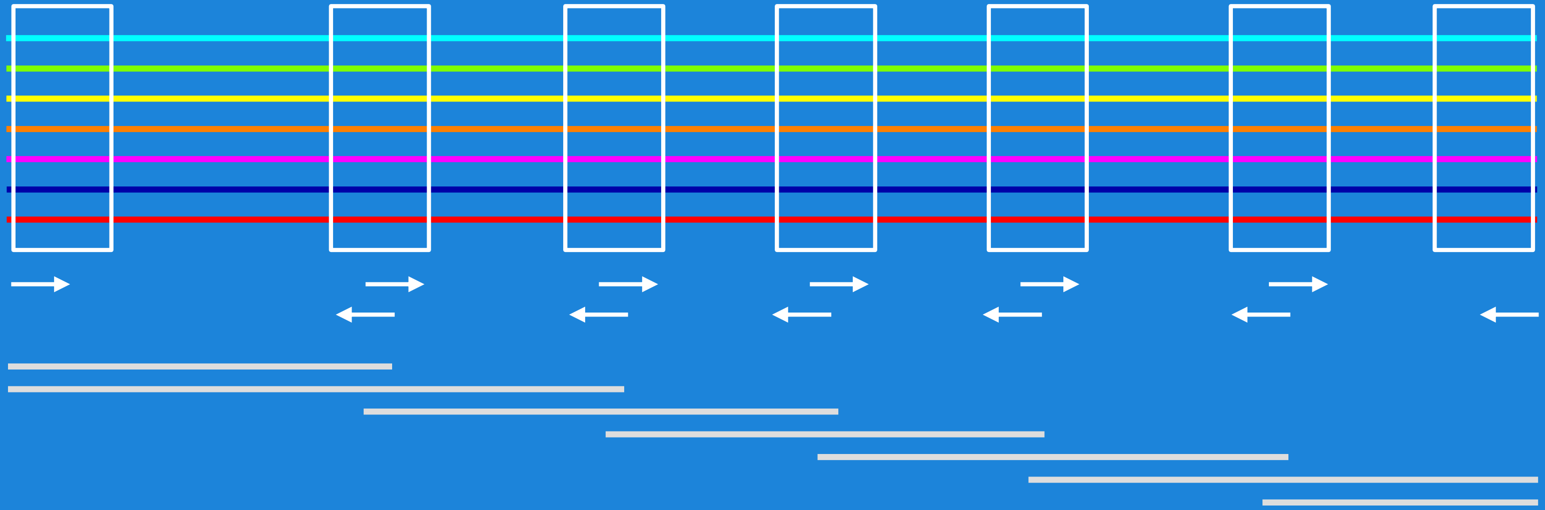
Remove ribosomal RNA
Remove mRNA (poly-A tails, 5’cap)
DNAse treatment
Size fractionation

Each purification treatment comes at the cost of losing part of the viral material (e.g. with a purification step we can go from 0.2% of viral RNA to 30%, but, at the same time we would lose 60% of the initial viral nucleic acids)

When the viral sequence is known...

(and there is not too much sequence variation)

We can design PCR primers over conserved regions and amplify directly the virus



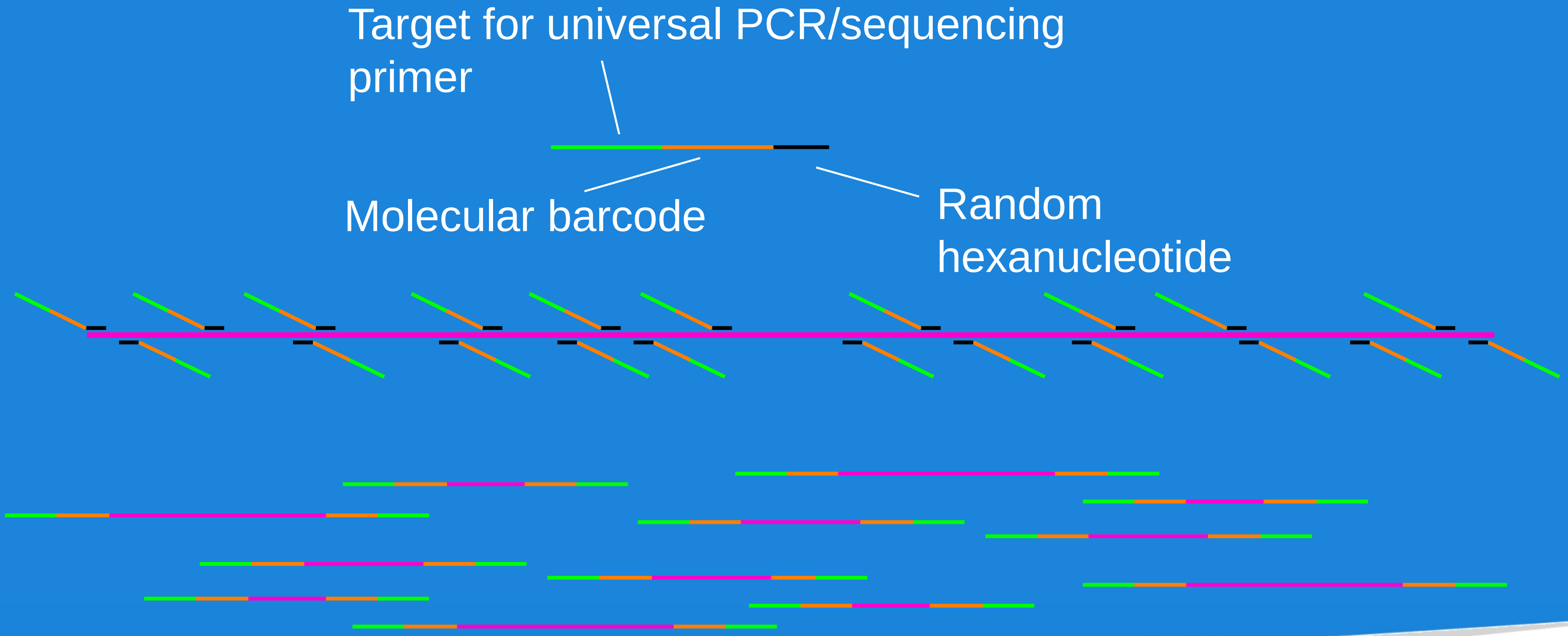
- Align known sequences and identify conserved regions
- Design primers around the conserved regions
- Amplify via PCR overlapping fragments of the genome

When the viral sequence is unknown...

(or the virus has too much variation at the nucleotide level)

We need to resort to sequence-independent amplification strategies

Sequence-Independent Single Primer Amplification (SISPA)



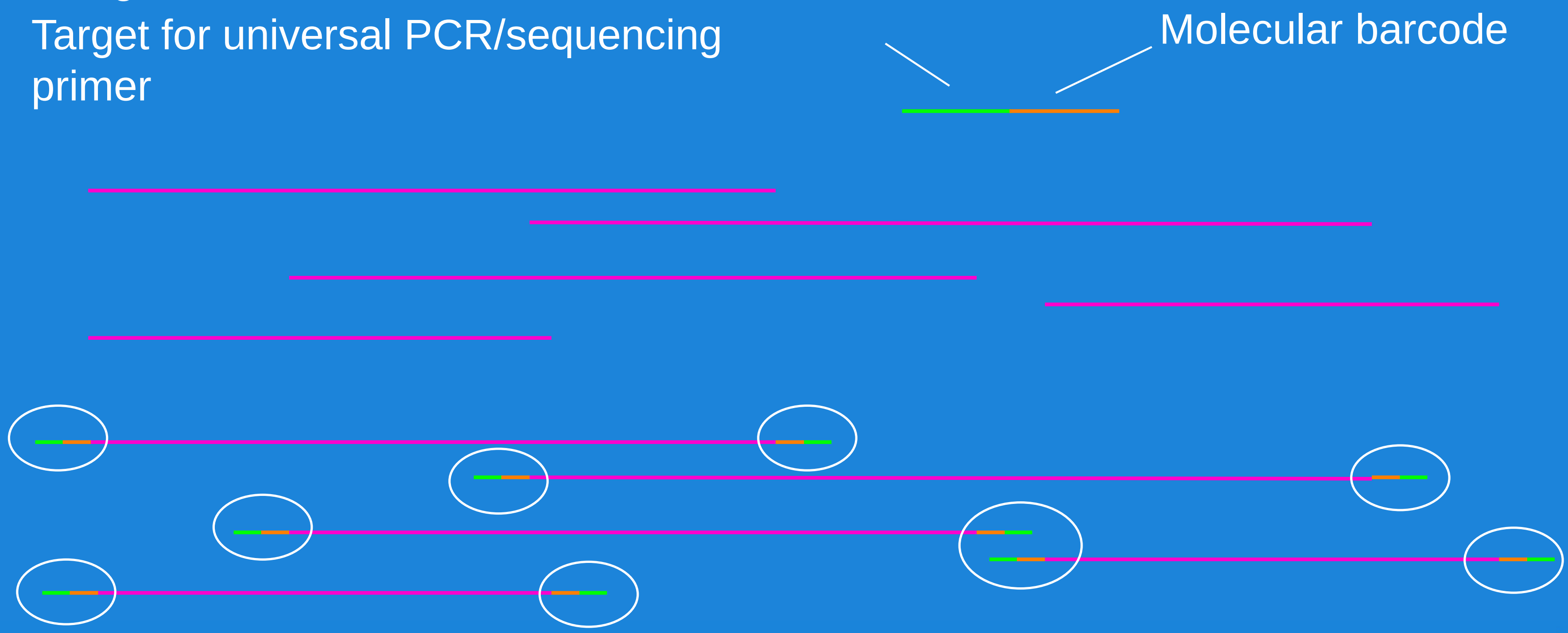
Target for universal PCR/sequencing primer

Molecular barcode

Random hexanucleotide

When the viral sequence is unknown...
(continued)

End-ligation of *adapters* and fragmentation of the genome



Molecular barcodes & multiplexing

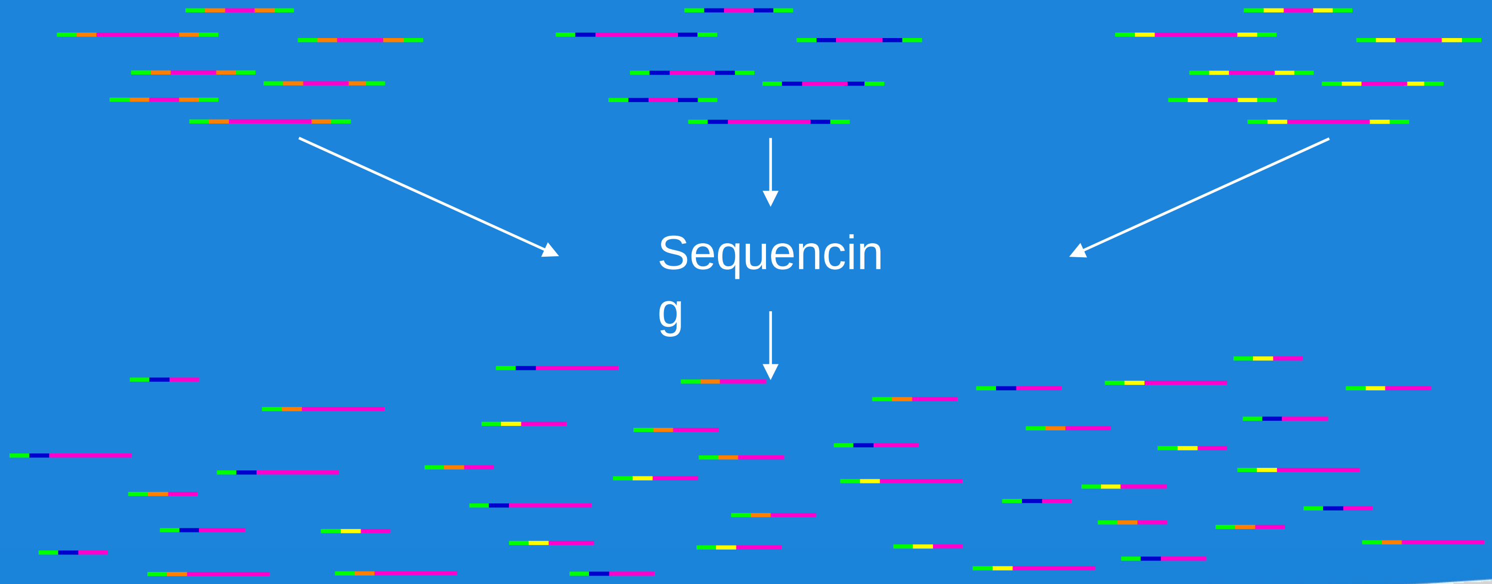
Modern sequencing technologies produce an enormous amount of sequences for each individual sequencing reaction, a few order of magnitude more than what it is needed to cover a viral genome.

The cost per sequencing reed is very small. However, the cost per sequencing run is rather high.

It is therefore necessary to combine multiple samples into a single sequencing reaction (i.e. multiplexing).

Molecular barcodes & multiplexing
(continued)

By attaching at each individual sample an unique artificial sequence (**molecular barcode**), it is possible to pool together many samples into a single sequencing reaction and tease the reads apart after sequencing.



Why sequencing viral genomes?

- To understand better their diversity and their biology.
- To follow their rapid evolution and to try to predict which variant will cause the next epidemic.
- To understand the dynamics of infection and viral spread.
- To try to understand which mutations confer resistance to antiviral drugs.
- To find new potential targets for drugs and antibodies (vaccine development).

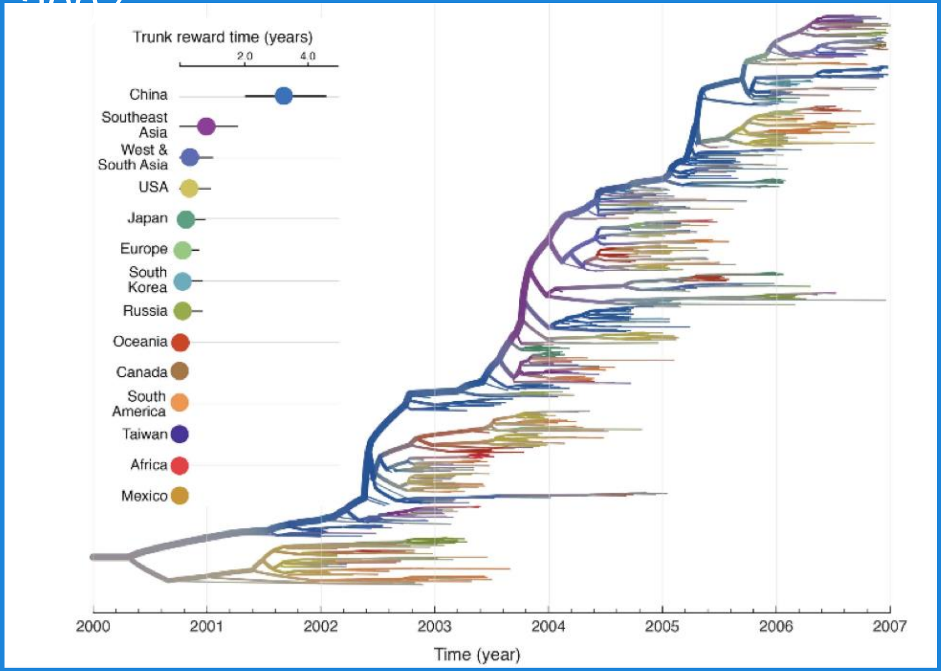
...

Understanding viral evolution

Especially RNA viruses like flu evolve very rapidly.

Many “branches” become dead ends either because of environmental factors (e.g. they evolve at the end of a flu season and do not reach a critical mass to survive till the next season, etc.), or because the mutations cause a loss of virulence or fitness.

Evolution of Influenza A H3N2 2000-2007

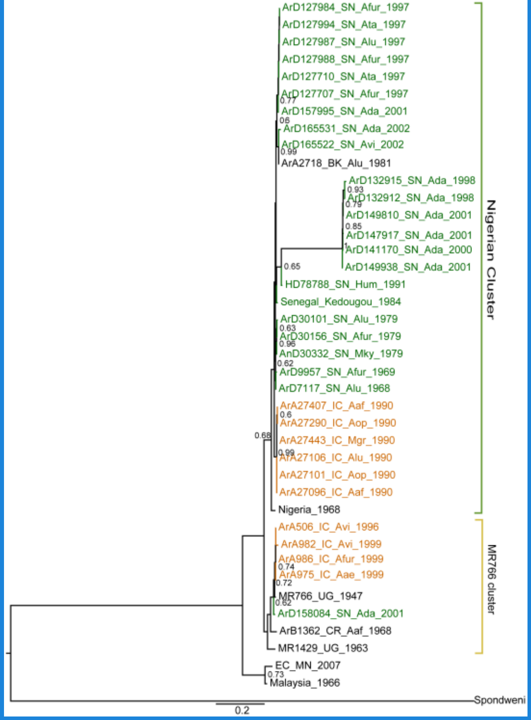


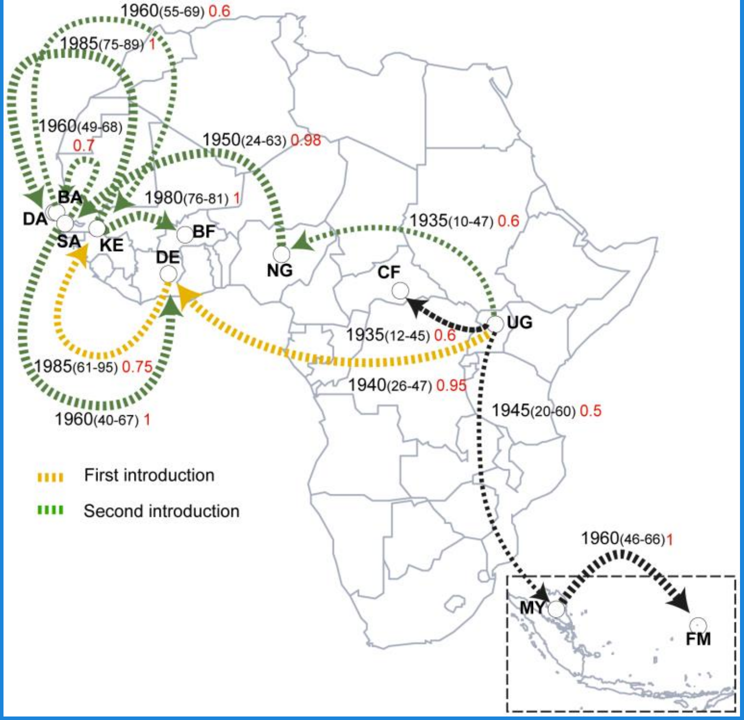
From Lemey P, Rambaut A, Bedford T, Faria N, Bielejec F, et al. PLoS Pathog. 2014 Feb 20;10(2):e1003932

Understanding the dynamics of the spread of a virus (phylogeography)

By combining the date when each sample has been isolated with the geographical location and comparing the differences between the samples we can reconstruct the path of the infection and interpolate dates.

Tracking the evolution and the movements of Zika

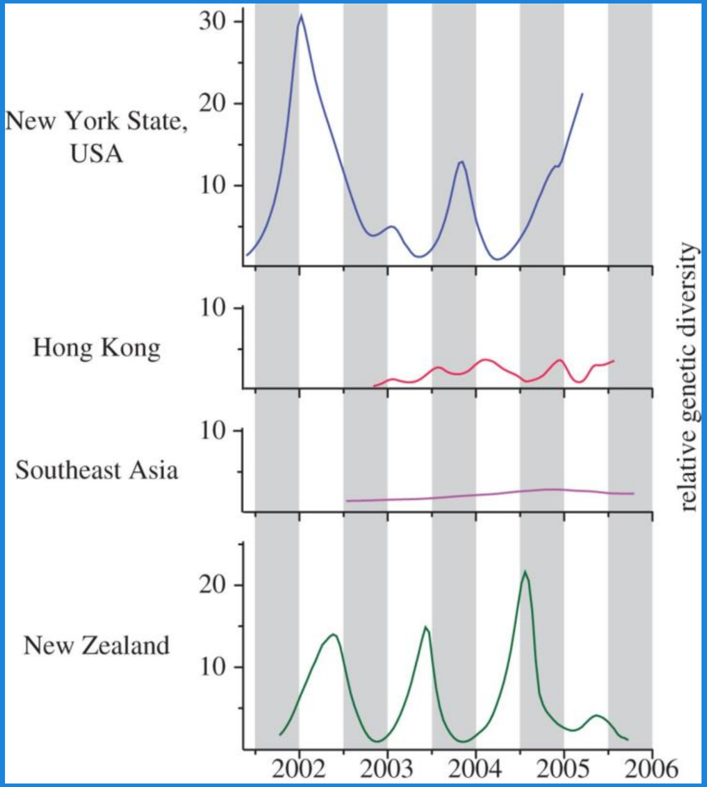




O. Faye, C.C.M. Freire, A. Iamarino, et al. PLoS Negl Trop Dis. 2014 Jan 9;8(1):e2636

Distinguish between endemic area and areas where a virus is newly-introduced each epidemic season

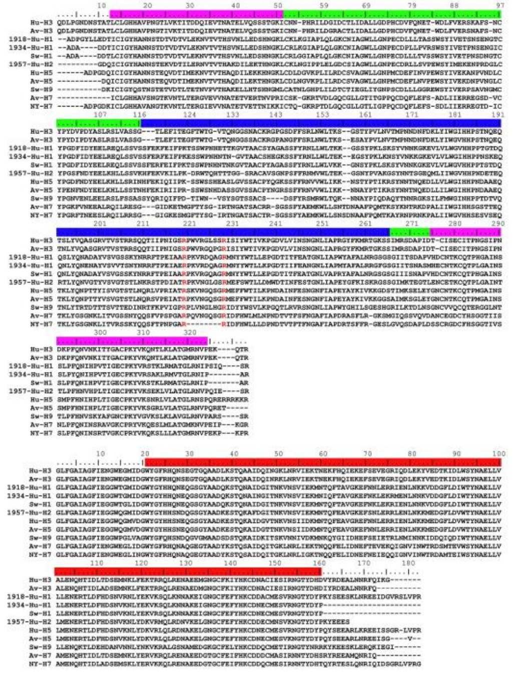
By analyzing the geographic distribution and the isolation year of the samples in each main branch of a phylogenetic tree it is possible to distinguish between virus that are endemic in the area and viruses that have been introduced from other areas and lack an endemic primary host.

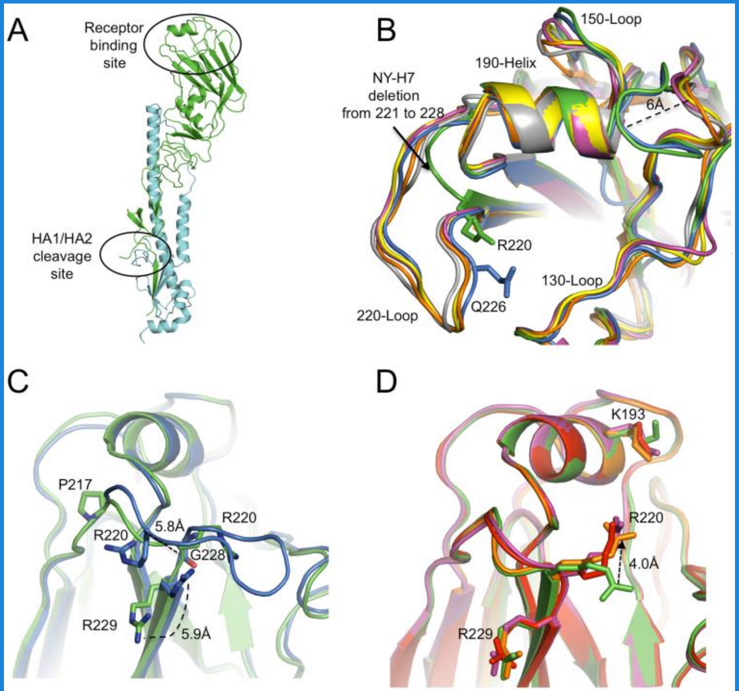


Viboud C, Nelson MI, Tan Y, Holmes EC.
Philos Trans R Soc Lond B Biol Sci. 2013 Feb 4;368(1614):20120199

Sequences & protein crystal structures

By having both the crystal structure of a protein and the sequences of several variants of it, it is possible to identify the regions that are most important for the activity (much less variable) and the ones most targeted by the host immune system (most variable).





Yang H, Chen LM, Carney PJ, et al PLoS Pathog. 2010 Sep 2;6(9):e1001081

Challenge: Create a new Influenza A live attenuated vaccine strain in a week using synthetic biology

(Biomedical Advanced Research and Development Authority - BARDA)

Terms:

JCVI would receive a sample with an unknown strain of Influenza A and, in at most seven days, it would deliver to the vaccine manufacture plan the DNA corresponding to the new vaccine strain.

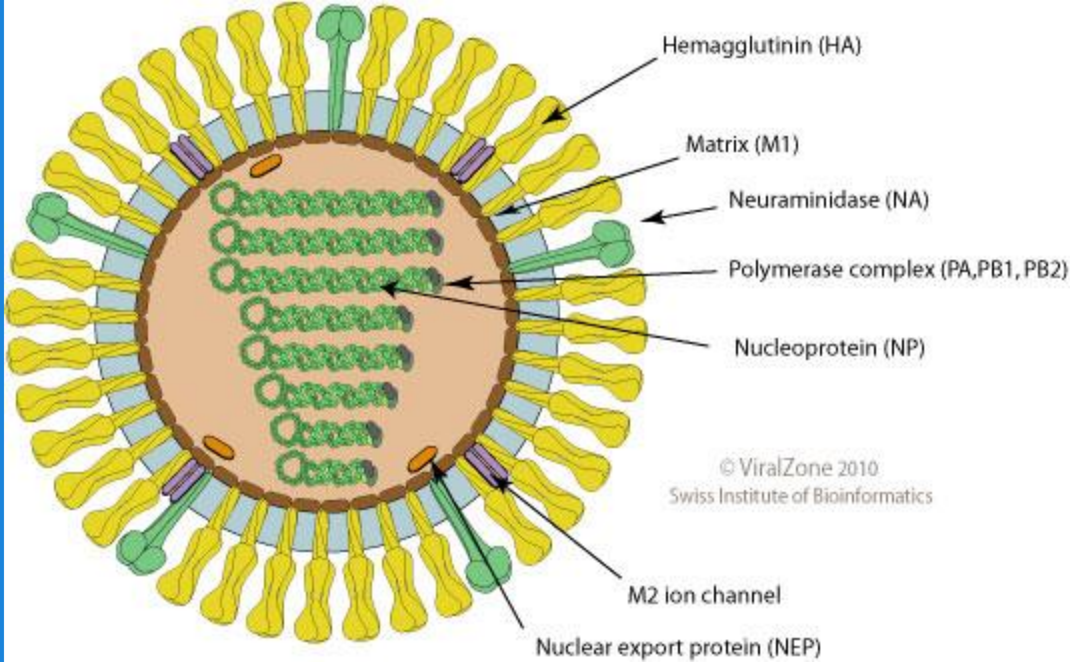
The DNA would be freshly synthesized and assembled, not a single piece could be amplified from the original sample, so that the sequencing of the unknown strain and the synthesis of the new vaccine could happen thousands of miles apart.

Some background about Influenza A virus

The genome is divided in eight segments, each coding for one or more proteins.

The surface of the viral particle is coated essentially with two proteins, essential for the virus to enter the host cell.

These two proteins, Hemagglutinin (HA) and Neuraminidase (NA) are the main targets of our immune system.



© ViralZone 2010
Swiss Institute of Bioinformatics

A brief background about vaccines

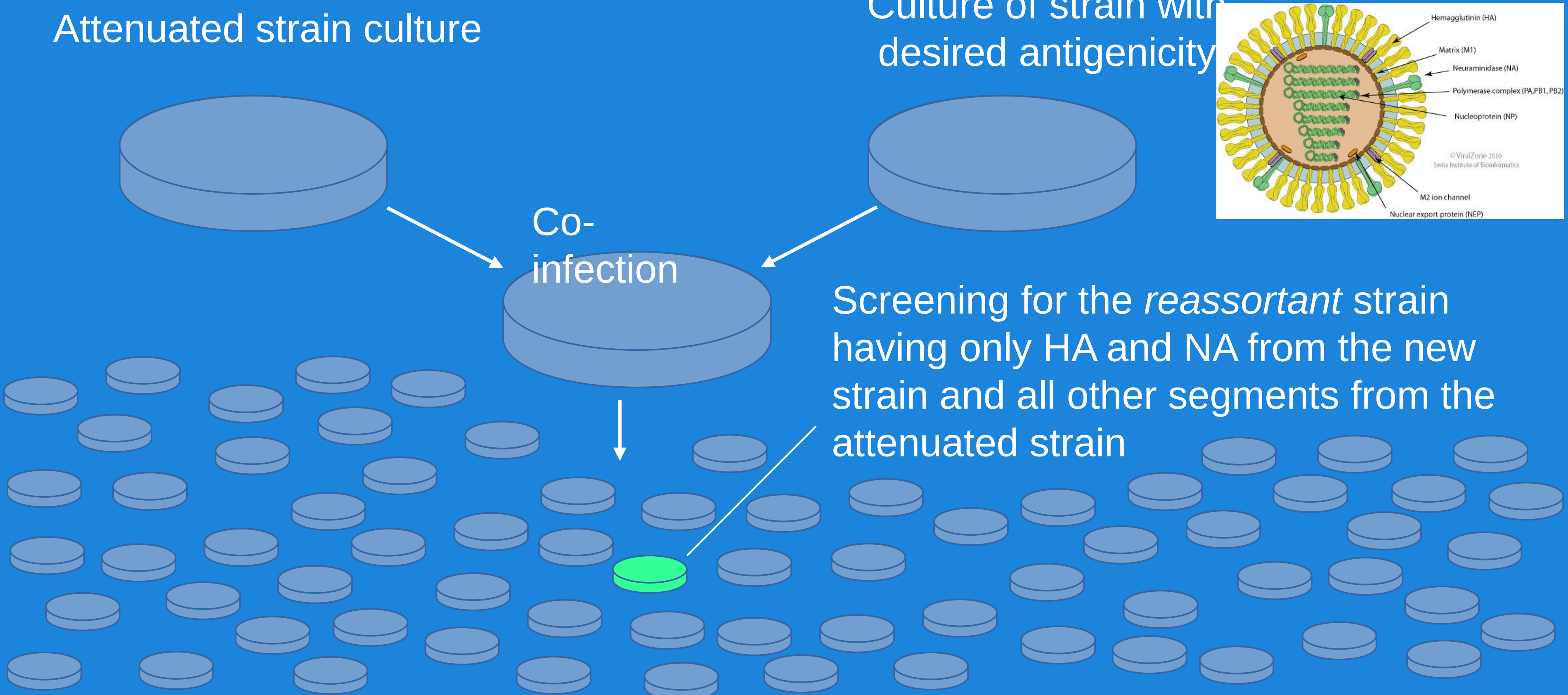
There are essentially two classes of vaccines: live, attenuated viruses and inactivated viruses.

A live, attenuated vaccine is essentially a virus with several mutations that allow it still to replicate in our body, but not to cause a disease. That is, once it triggers the immune response, it is easily cleared by our immune system.

A inactivate vaccine is instead composed by the actual virus that has been partially destroyed (all the genetic material removed). A vaccine dose consists of a considerable amount of viral fragments, equivalent to the amount of viral material circulating in our bloodstream during a viral infection.
Many inactivated vaccines require one or more booster doses, in order to be active.

Both type of vaccines will cause us to “feel sick” for a couple of days: it is our immune system reacting to the new antigens.

A brief background in flu vaccine manufacturing



Attenuated strain culture

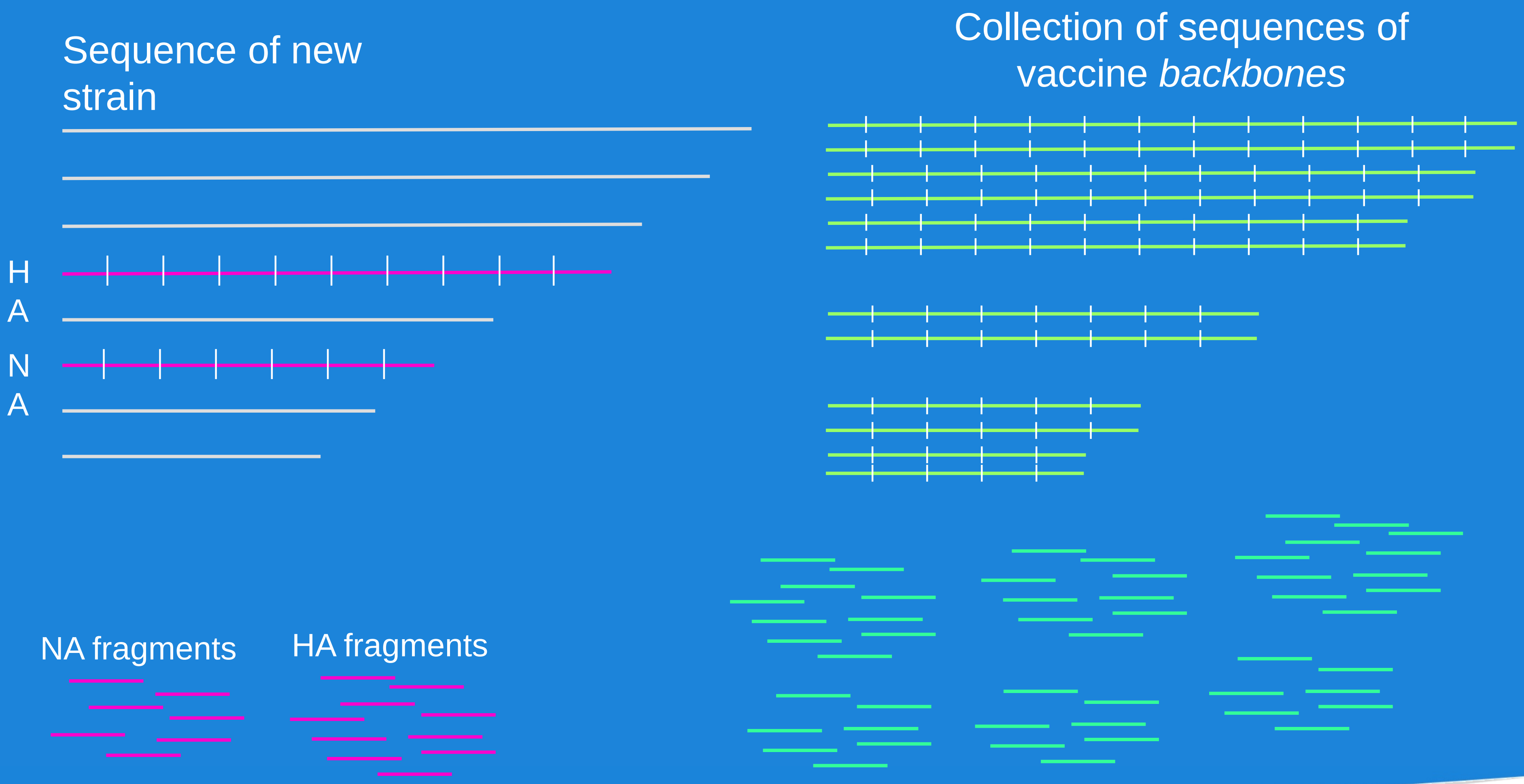
Culture of strain with desired antigenicity

Co-infection

Screening for the *reassortant* strain having only HA and NA from the new strain and all other segments from the attenuated strain

~3-4 months, several thousands of cultures to be carefully screened, before starting the actual vaccine production.

The synthetic biology way



Sequence of new strain

Collection of sequences of vaccine *backbones*

HA

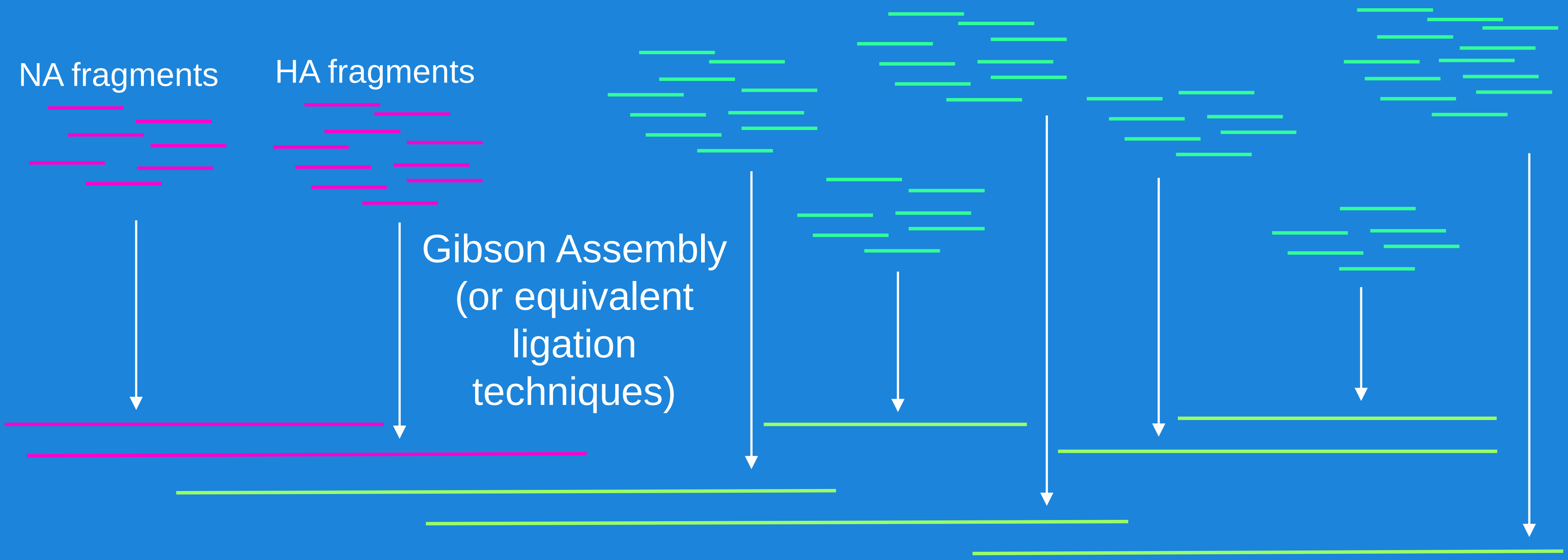
NA

NA fragments

HA fragments

J. Craig Venter™ INSTITUTE

The synthetic biology way
(continued)



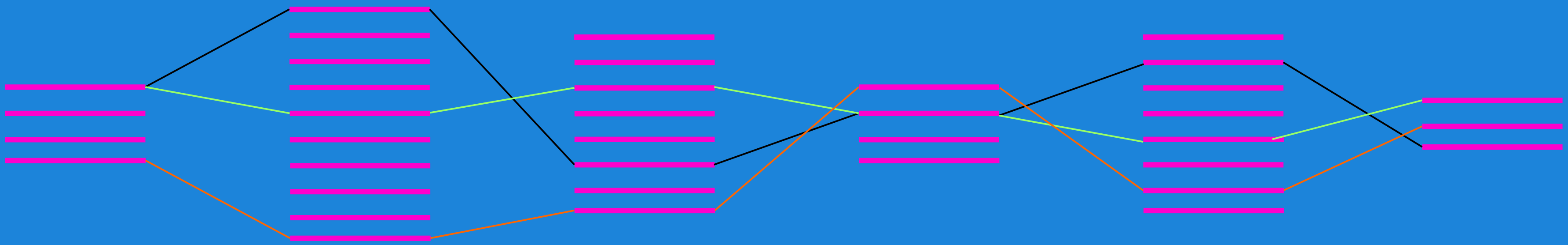
The resulting synthetic segments are verified by sequencing and shipped overnight to the vaccine manufacturing plan.
Total time: 1 week

Synthetic flu vaccine preparedness program

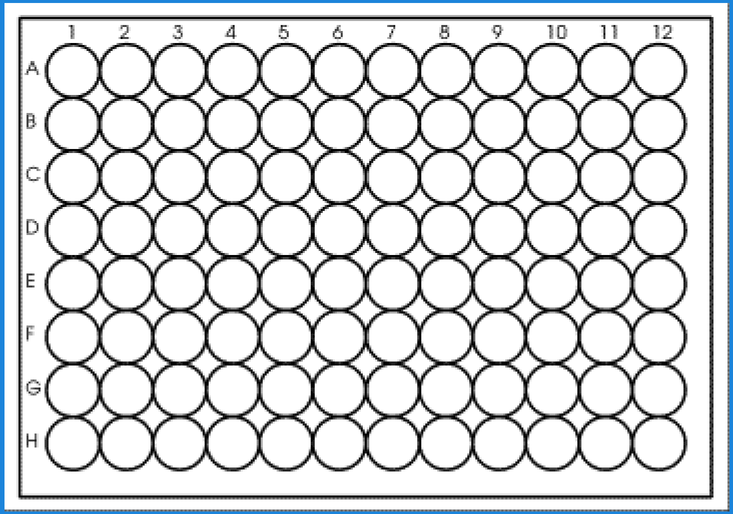
Goal:
Create a collection of synthetic flu fragments sufficient to manufacture in real time any flu A vaccine strain ever created and the software that will produce the correspondent program for a liquid-handling robot to assemble Gibson Assembly reactions picking fragments from the stock microtiter plates.

Create a minimal set of fragments and maps describing which fragments to use for assembling each strain

Several regions of each segment is common to many strains. The first step is to identify a non-redundant set of fragments that cover all the diversity and can be used to reconstitute each of the vaccine strains.



Organize the fragments in microtiter plates so that a reaction to assemble any of the segments would use the least number of plates



Each well in the plate is identified by alphanumeric coordinates (letters for rows, number for columns) and the synthetic DNA fragments were stored in a specific order to optimize the process with the liquid-handling robot.

J. Craig Venter™ INSTITUTE

Proof of concept

To prove the system we were assigned a certain number of vaccine strains that would need to be assembled in a very limited amount of time, sequence-verified and sent to the vaccine-manufacturing plan.

Bottom line:
Despite these technological advances, the current vaccines are still produced mostly using millions of chicken eggs and traditional methods, mostly due to the costs and challenges of having the new methods validated and approved by the regulatory agencies.

Acknowledgements

JCVI

J. Craig Venter
Karen Nelson
Brett Pickett
Gene Tan
Alan Durbin
Torrey Williams
Mark Novotny
Pilar Viedma
Fortuna Arumeni
Nadia Fedorova
Lihui Hu
James Christensen
Richard Isom
John Miller

JCVI Alumni

Tim Stockwell
David Wentworth
Reed Shabman
Suman Das
Becky Halpin
Brian Bishop
Danny Katzel
Mark Williams
Susmita Shrivastava
Seth Schobel
Karla Stuker
Yi Tan
Meghan Shilts

Questions?

J. Craig Venter™
INSTITUTE