Sequenceable Event Recorders

Luca Cardelli, University of Oxford

2022-09-27 Applied Systems Biology

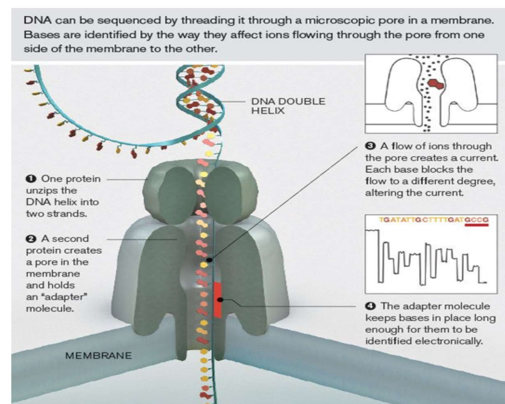
Introduction

- Recording events in DNA

Shipman, S. L., Nivala, J., Macklis, J. D. & Church, G. M. Molecular recordings by directed CRISPR spacer acquisition. *Science* **353**, aaf1175 (2016).

Sheth, R. U., Yim, S. S., Wu, F. L. & Wang, H. H. Multiplex recording of cellular events over time on CRISPR biological tape. *Science* **358**, 1457–1461 (2017).

- High throughput sequencing and synthesis



http://www2.technologyreview.com/news/427677/nanopore-sequencing/

CLONAL GENES
My Genes Project

OVERVIEW > GENE IMPORT > PRICING & SCORE

Change Vector ▾ | + Flanks | Optimize | + Genes | + Custom Vector

#	NAME	SEQUENCE	BP	VECTOR	SCORE	PRICE
1	gene-001	ACTCGACTGACTAGC...	1264	Select Vector ▾	●	$113.76
2	gene-002	ACTCGACTGACTAGC...	1014	Select Vector ▾	●	$91.26
3	gene-003	ACTCGACTGACTAGC...	978	Select Vector ▾	●	$88.02
4	gene-004	ACTCGACTGACTAGC...	848	Select Vector ▾	●	Fix it
5	gene-005	ACTCGACTGACTAGC...	1200	Select Vector ▾	●	$108.00
6	gene-006	ACTCGACTGACTAGC...	1124	Select Vector ▾	●	$101.16
7	gene-007	ACTCGACTGACTAGC...	1200	Select Vector ▾	●	Fix it
8	gene-008	ACTCGACTGACTAGC...	1087	Select Vector ▾	●	$97.83
9	gene-009	ACTCGACTGACTAGC...	1200	Select Vector ▾	●	$108.00

32 GENES · 26,400 BP

All (240) | Easy (24) | Difficult (4) | Error (2) | Not Possible (2)

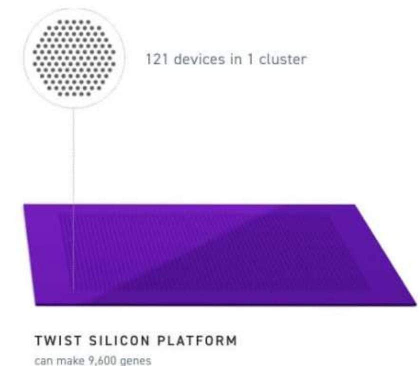
PRICING SUMMARY

NAME	QTY	COST
Easy Genes	24	$2,376.00
Cloning Fee	24	$1,300.00

DELIVERY FORMAT
Plate: 96 Well, Horizontal
Tube Edit

Total $3, 676

Checkout



https://www.twistbioscience.com/technology

Preorder Recorder

- Detecting molecular events is very difficult and very important
- In science we want to know "what's going on?"
- In bioengineering we want to know "what when wrong?"
- We often want to know the *order* of events to help determine *causation*

- We discuss a "preorder recorder" algorithm that reads out the *preorder of first-occurrence* of a set of events in a chemical soup, where an event is the appearance of a DNA/RNA strand in the soup
- These events could be DNA circuit signals, or naturally transcribed RNA, or DNA/RNA transduced in response to e.g. presence of certain proteins

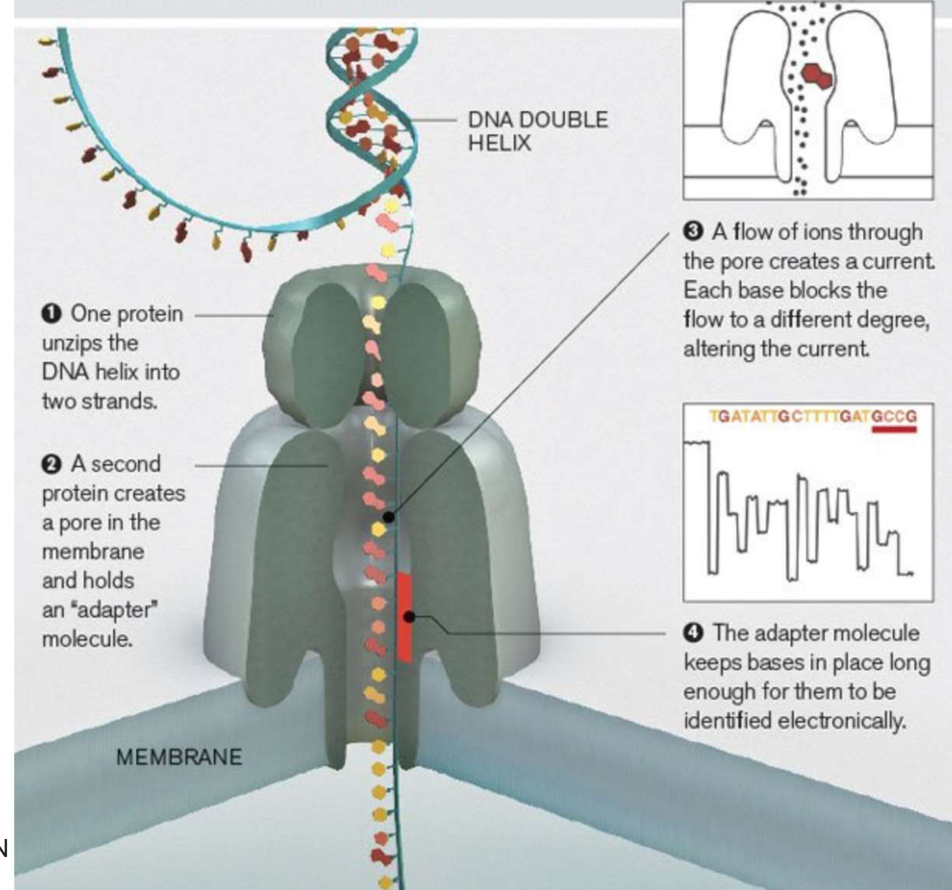
How to Read DNA (Output)

- Fluorescence Readout
  - Limited redout capability: 3/4 "colors" of output.
  - Output can be read continuously over time
- Atomic Force Microscope Readout
  - Detecting shapes and patterns
  - Comprehensive view of the results
- Sequencing Readout
  - At the end of a computation, sequence the strand types left in the soup
  - Output is a multiset of strand types (each with a real-valued concentration)

High Throughput Sequencing

- Sequencing by Synthesis
  - Like Sanger sequencing, but done in parallel on a "lawn" of single strands, removing the fluorophores at each step to carry on.
- Nanopore Sequencing
  - ~ 200 *single* different DNA molecules sequenced in parallel

American astronaut Kate Rubins with a MinION sequencer on the ISS in August 2016.
https://en.wikipedia.org/wiki/Oxford_Nanopore_Technologies



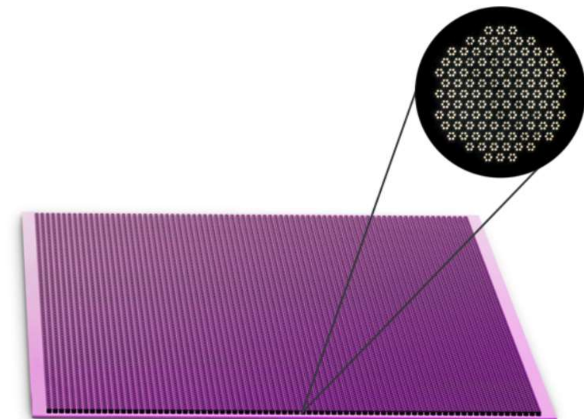
http://www2.technologyreview.com/news/427677/nanopore-sequencing/

How to Write DNA (Gates + Input)

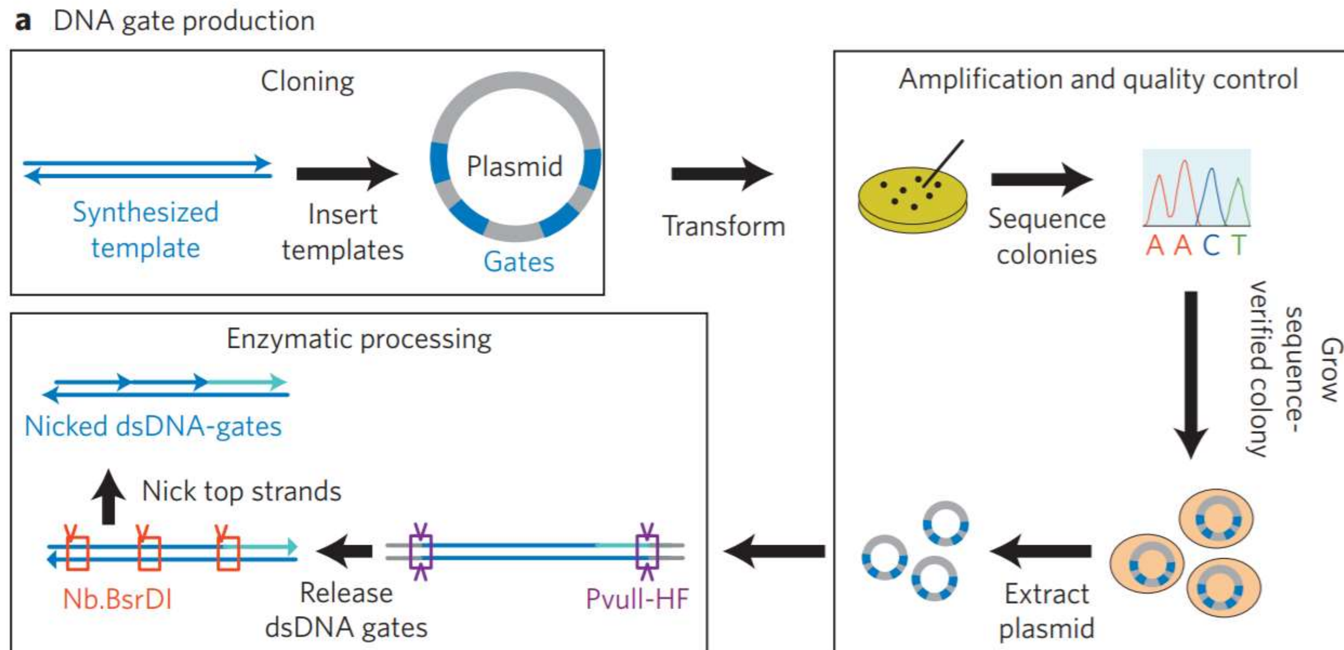
- Synthesizing DNA using silicon microfabrication technology

Twist Bioscience developed a proprietary semiconductor-based synthetic DNA manufacturing process featuring a high-throughput silicon platform that allows us to miniaturize the chemistry necessary for DNA synthesis. This miniaturization allows us to reduce the reaction volumes by a factor of 1,000,000 while increasing throughput by a factor of 1,000, enabling the synthesis of 9,600 genes on a single silicon chip at full scale. Traditional synthesis methods produce a single gene in the same physical space using a 96-well plate.

=> DNA Storage



Cloning



- Higher quality than ordinary synthesis, and supporting very long strands (very complex "gates").

The Pace of Biotechnology

Sequencing costs (USD / Mb)

Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP) Available at: www.genome.gov/sequencingcostsdata.

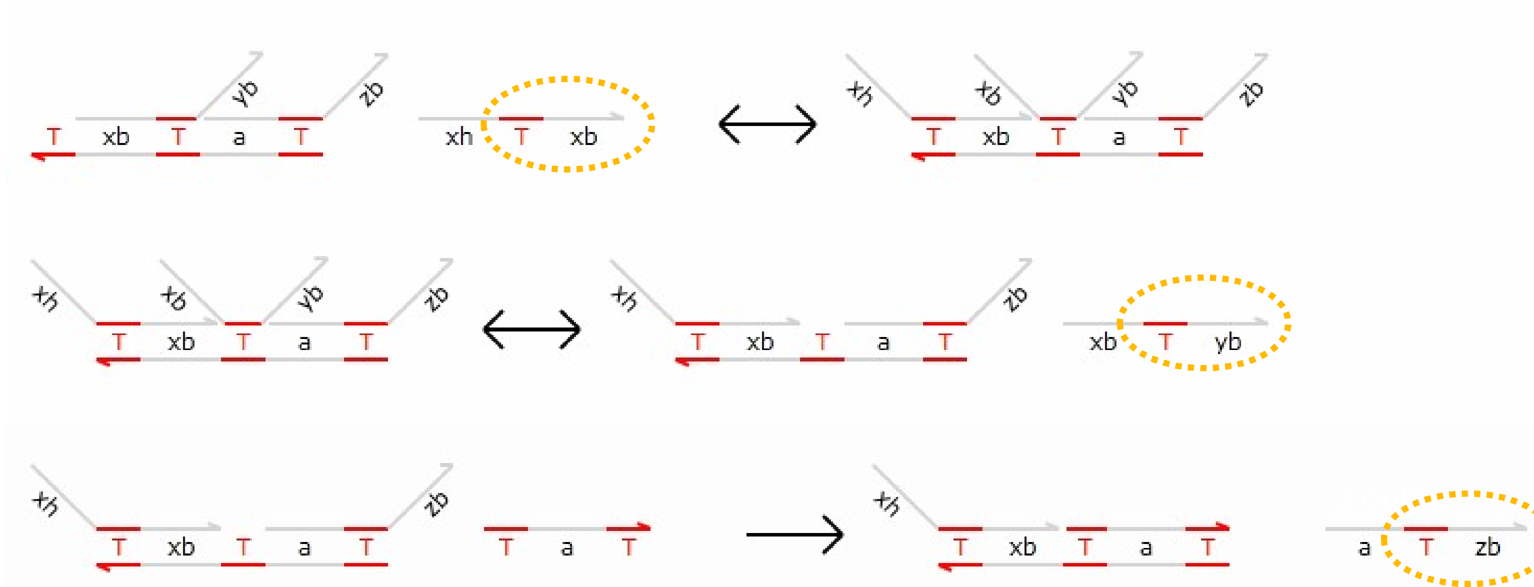
How can we take full advantage of all this, for DNA-based algorithms?

Many DNA strand displacement computational schemes are "Universal"

- 4-domain, 3-domain, 2-domain, split-domain …
- Can be used to systematically compile arbitrary finite chemical reaction networks to DNA molecules that exhibit (approximately) the same kinetics.
- But not all can be written by cloning and read by sequencing.

A Typical 3-domain Scheme

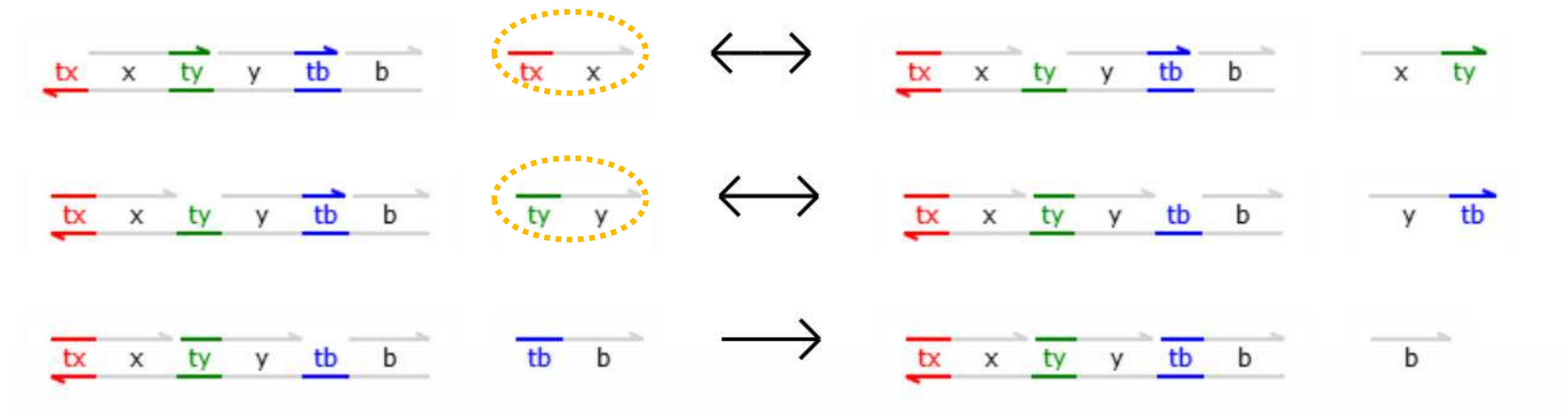
2-input "join"   x + y -> z



Non-clonable, non-sequenceable
because of the secondary structures

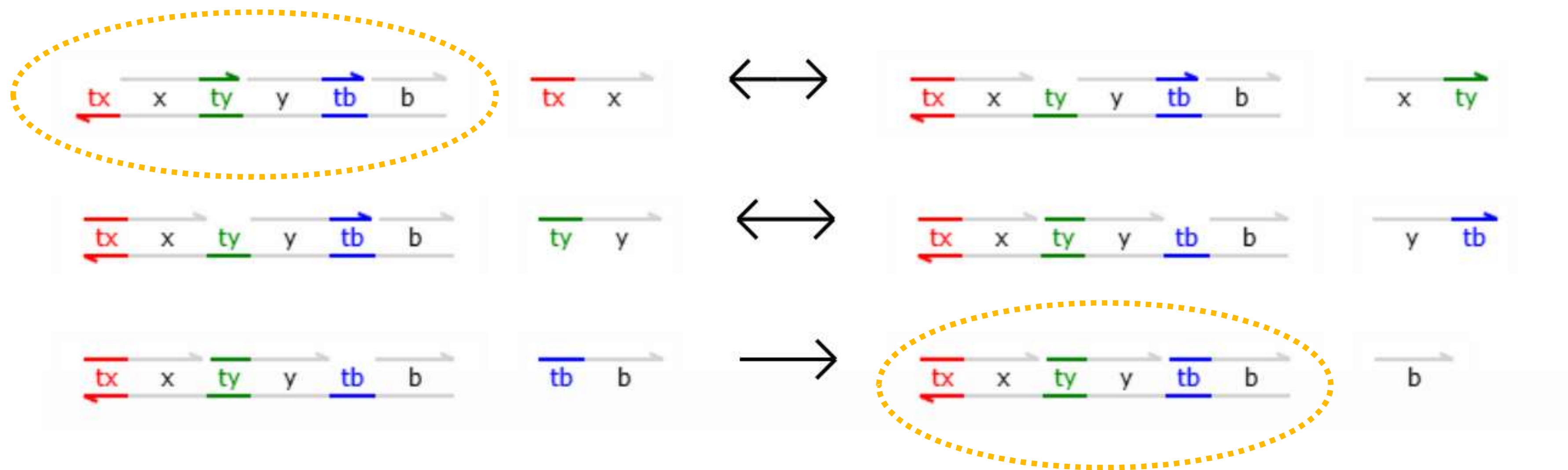
A 2-domain Scheme

2-input "join" x + y -> Ø



Clonable but not Sequenceable

Sequencing (of double strands) must be preceded by *polymerase extension* (to remove single-stranded gaps) and *ligation* (to remove nicks)

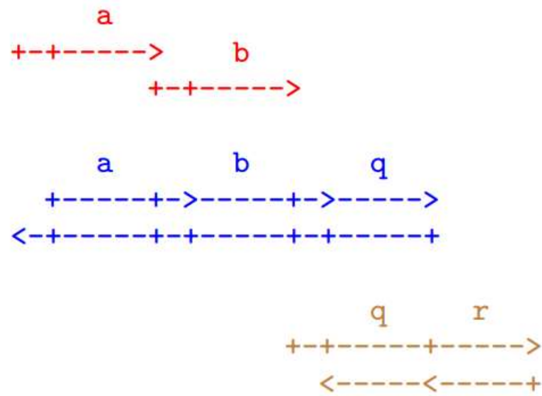


Then input and output look the same under sequencing

Sequenceable Join gate

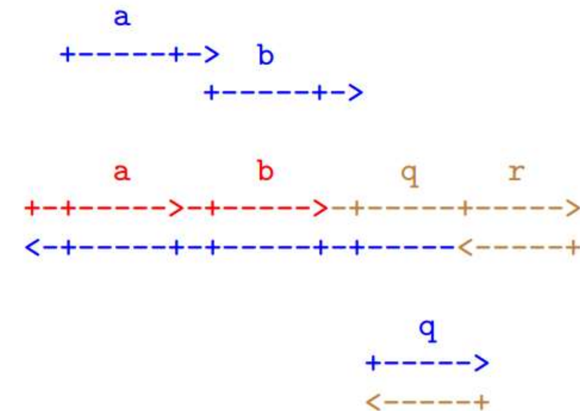
A 2-input join gate, join(a,b):

[Georg Seelig & Yuan-Jyue Chen]



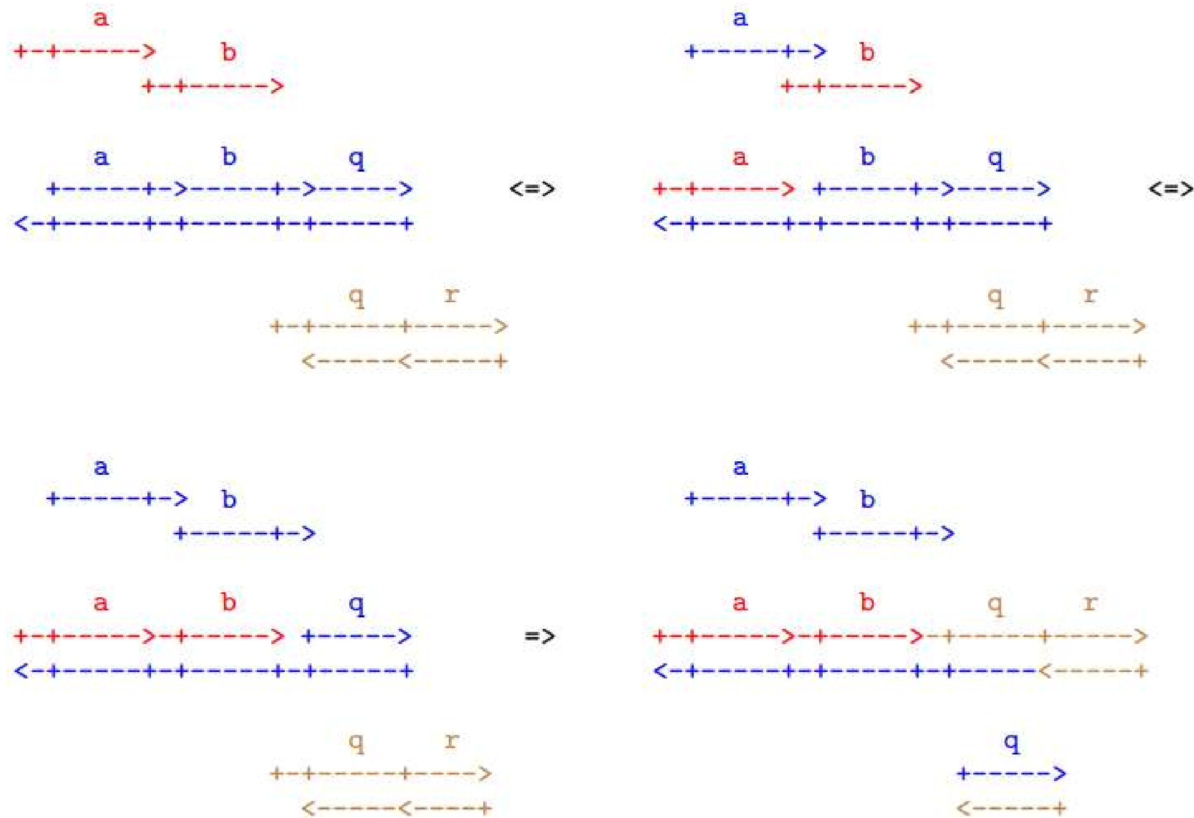
Two-domain gate architecture [L.Cardelli 2013]
based on double stranded DNA (no secondary structure)
hence gates can be sequenced by standard means

if a, b are present *together*, then after full activation:



an “abqr+q” read (after ligation) reveals there was activation of join(a,b), hence both a and b occurred. Otherwise, we would read “abq+qr”.

Join Gate activation steps



Sequence the soup: an "abqr" read indicates that both "a" and "b" were present.

What we can use

- Technologies to write (synthesize) whole sets of DNA strands in parallel
- Technologies to read (sequence) whole sets of DNA strands in parallel
- An architecture to do computation on DNA strands and produce sequenceable results
- Hence … highly concurrent algorithms!

Coincidence Recorder

Goal: determine which pairs of a set of events were present *together* in the pot.

Algorithm:
At the beginning, add *all* the pairs join(x,y) for x,y in Events.
At the end, sequence the whole pot.
End.

N.B. join(x,x) tells us if x was ever present.

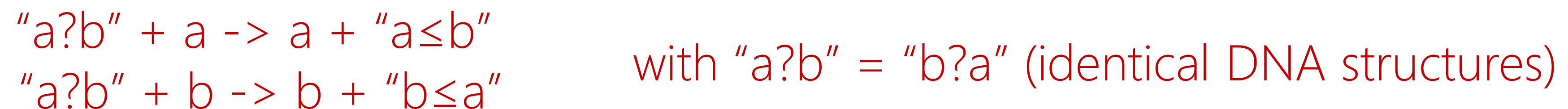
N^2 algorithm: great, we make "good use" of high-throughput synthesis and sequencing!
It uses no power when events are not present (it does not record *timing*, only *coincidence*).

Choice gate Specification

A *choice* gate is a two-input gate denoted $a?b$ between input events a and b. As an abstract operator it is symmetric: $a?b = b?a$. Its desired behavior is as follows:

- If a arrives no later than b, then $a?b$ produces a distinct result that we indicate $a \leq b$ or equivalently $b \geq a$.
- If b arrives no later than a, then $a?b$ produces a distinct result that we indicate $b \leq a$ or equivalently $a \geq b$.
- If a and b arrive together, then $a?b$ produces a result that we indicate $a \sim b$ or equivalently $b \sim a$. (This is in practice an equal mixture of $a \leq b$ and $b \leq a$, or an unequal mixture if they arrive slightly offset.)
- As a special case, if a ever arrives, then $a?a$ produces a result $a \sim a$.

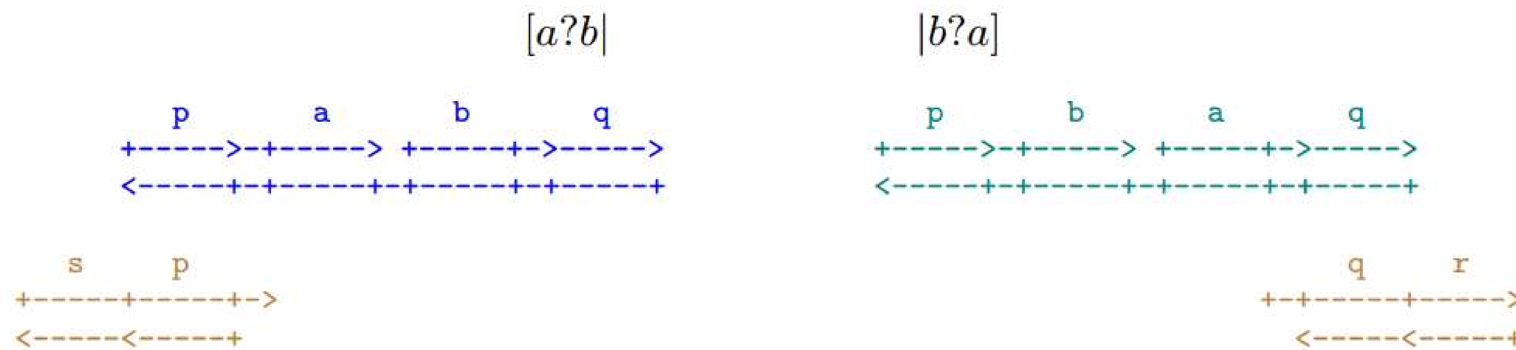
That is, we want to implement the CRN:



And we want the output to be readable by sequencing

Sequenceable Choice gate

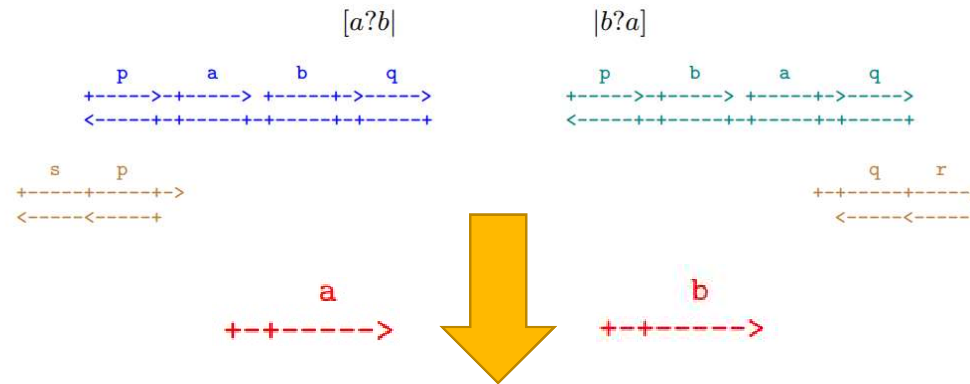
$$a?b = [a?b| + |b?a] = [b?a| + |a?b] = b?a.$$



(also clonable)

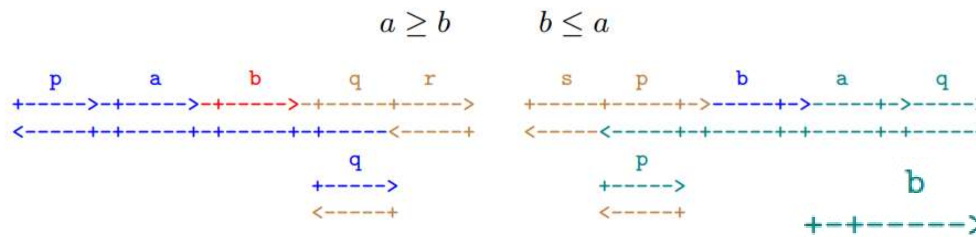
Sequenceable Choice gate outcomes

$a?b$



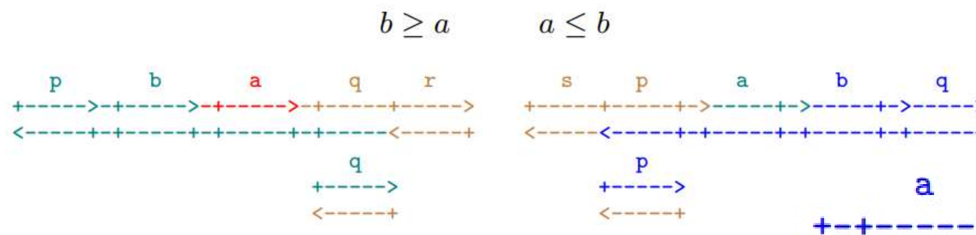
Sequencing pattern:

If b arrives first:



pabqr + spbaq

If a arrives first:



pbaqr + spabq

Preorder Recorder

Goal: Record the preorder of first arrivals of a set of events that occur in a pot.

Algorithm:
At the beginning, add *all* the pairs x?y, for x,y in Events.
At the end, sequence the whole pot and reconstruct the preorder by transitive reduction.
End.

E.g.: Events = {a,b,c}

gates	structures	after '−c'	after '−b'
$a?a$	$[a?a\|\ \ \|a?a]$	$[a?a\|\ \ \|a?a]$	$[a?a\|\ \ \|a?a]$
$b?b$	$[b?b\|\ \ \|b?b]$	$[b?b\|\ \ \|b?b]$	$b \geq b\ \ b \leq b$
$c?c$	$[c?c\|\ \ \|c?c]$	$c \geq c\ \ c \leq c$	$c \geq c\ \ c \leq c$
$a?b$	$[a?b\|\ \ \|b?a]$	$[a?b\|\ \ \|b?a]$	$a \geq b\ \ b \leq a$
$a?c$	$[a?c\|\ \ \|c?a]$	$a \geq c\ \ c \leq a$	$a \geq c\ \ c \leq a$
$b?c$	$[b?c\|\ \ \|c?b]$	$b \geq c\ \ c \leq b$	$b \geq c\ \ c \leq b$

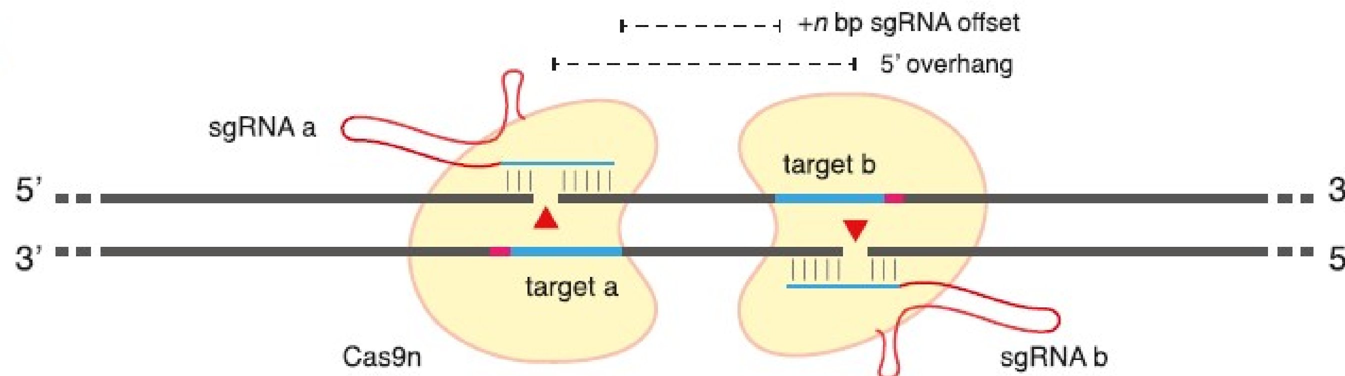
That's a definite $c < b$, because we observe $c \leq b$ but not $b \leq c$. Moreover, we do not observe $a \leq a$ which means that a never arrived. If we were to observe $c \leq b$ and $b \leq c$, then we would deduce that c, b arrived together, up to our time resolution.

Cas9 Join gate

Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity

F. Ann Ran,[1,2,3,4,5,11] Patrick D. Hsu,[1,2,3,4,5,11] Chie-Yu Lin,[1,2,3,4,6] Jonathan S. Gootenberg,[1,2,3,4] Silvana Konermann,[1,2,3,4] Alexandro E. Trevino,[1] David A. Scott,[1,2,3,4] Azusa Inoue,[7,8,9,10] Shogo Matoba,[7,8,9,10] Yi Zhang,[7,8,9,10] and Feng Zhang[1,2,3,4,*]

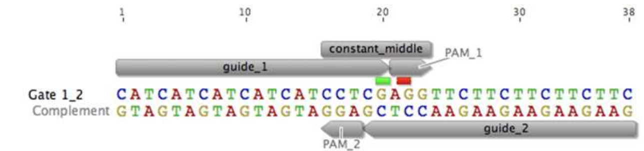
- The target DNA is (completely) cut only if both gRNAs are present at the same time (DNA repair will otherwise fix single nicks):

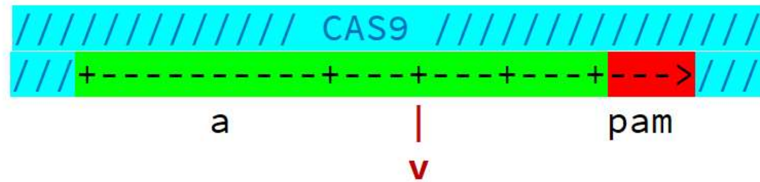


- This scheme is sufficient to run a coincidence recorder, assuming single nicks are being repaired.

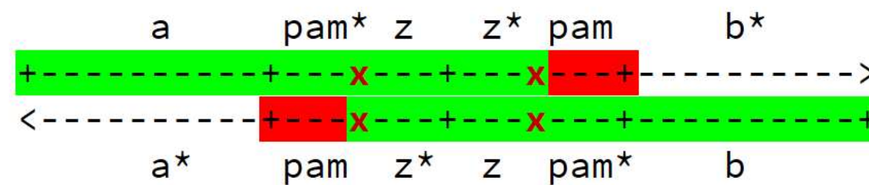
Cas9 Choice gate

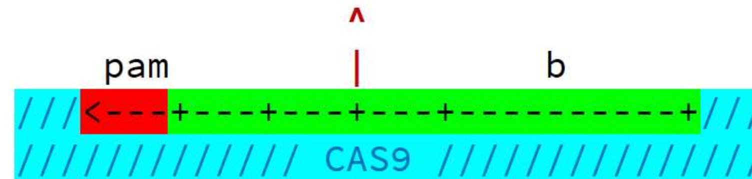
[Jeff Nivala]

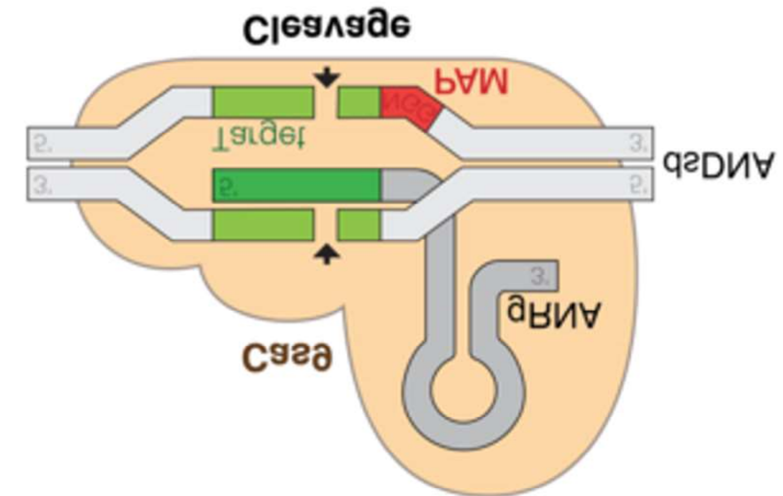




a?b







The two cuts are distinguishable by sequencing.

N.B. the events must be in the form of gRNA, so they can be picked up by Cas9

Conclusions

- Technological advances
  - High-throughput synthesis and sequencing

- Provide new readout opportunities
  - Reading and writing n^2 elements feasibly

- Which can inspire a new class of parallel algorithms
  - Coincidence Recorder, Preorder Recorder, ... ???

Sequenceable Event Recorders. Luca Cardelli.

- ©2021 In E. Winfree, N. Jonoska (Eds.): Visions of DNA Nanotechnology at 40 for the Next 40. Accepted 2021-11-02. Springer (to appear).
- ©2021 [PDF] ArXiv 2021-12-04: http://arxiv.org/abs/2105.15039.