DNA Encryption: Bioencryption to Store Your Secrets in Living Organisms

About Me

- @JohnDunlap2
- Security Researcher
- Reverse Engineer
- Avid collector of bad software
- Work for GDSSecurity doing code review / RE / Research / pentesting

Let’s be clear

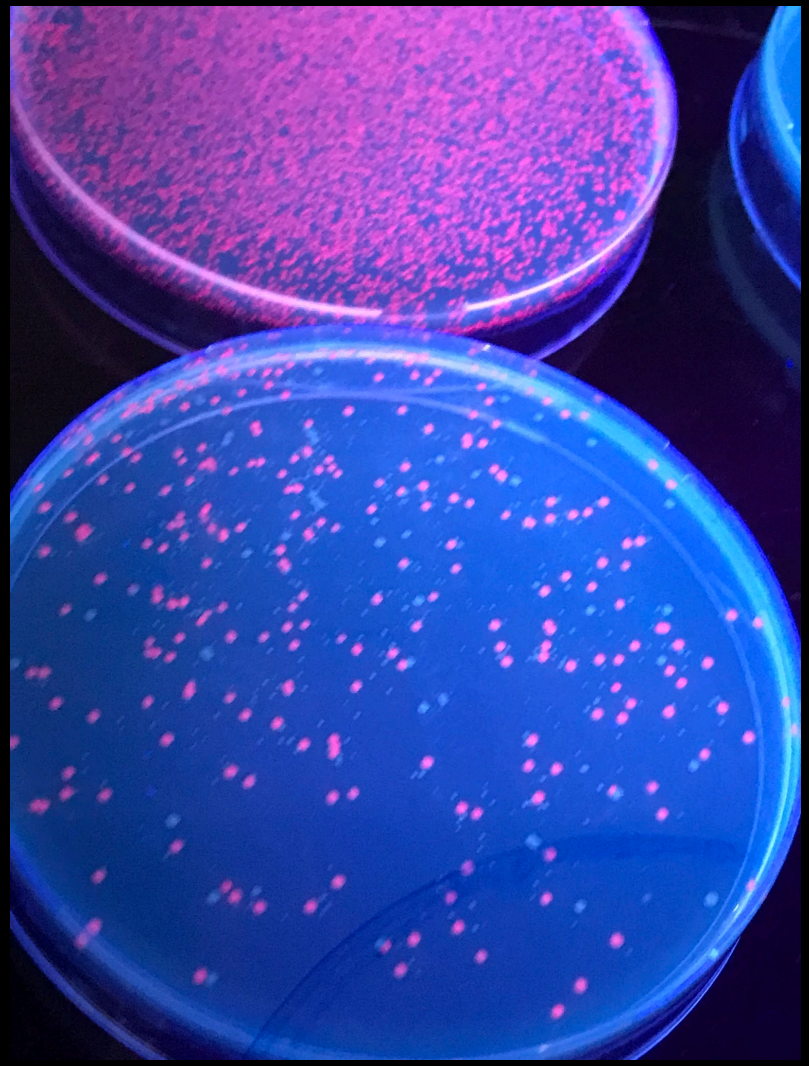
What I am

- Hacker in the traditional sense
- Researcher of hacker history (watch my hope conf talk when it comes out!)
- Programmer
- Curious
- Stubborn
- Perhaps Foolhardy

What I am not

- A biologist
- A geneticist
- A lab professional
- A cryptographer
- Dissuaded from synthesizing DNA of my own design by any of the above facts.

The Impetus for This Talk



Community Biohackerspaces are awesome!

- Please support them
- Be responsible
- Special shout outs to Mike & Will at Genspace for helping me out with all of this

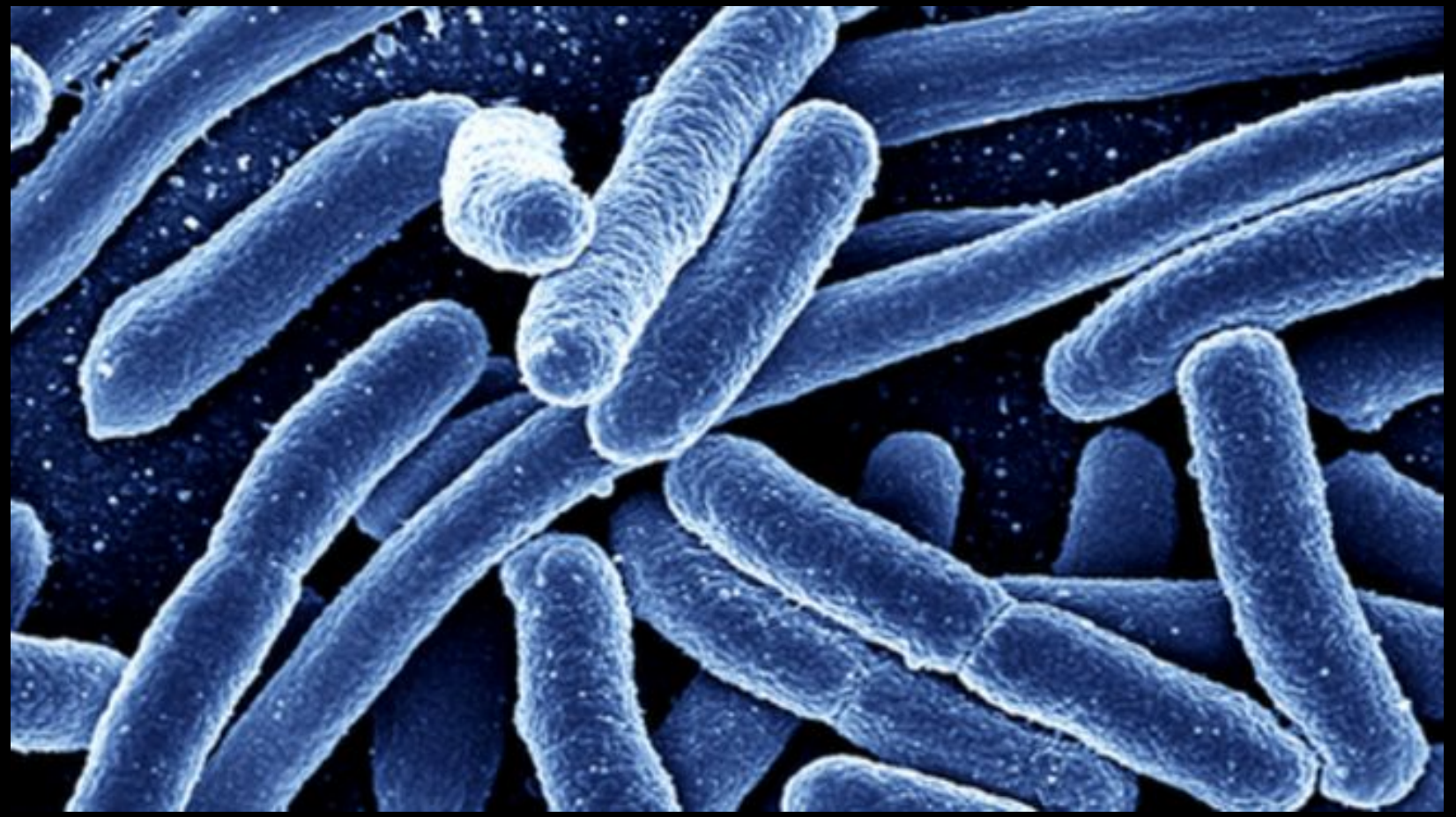
DNA Data Storage

I am going for something more modest.
Hacker Style!

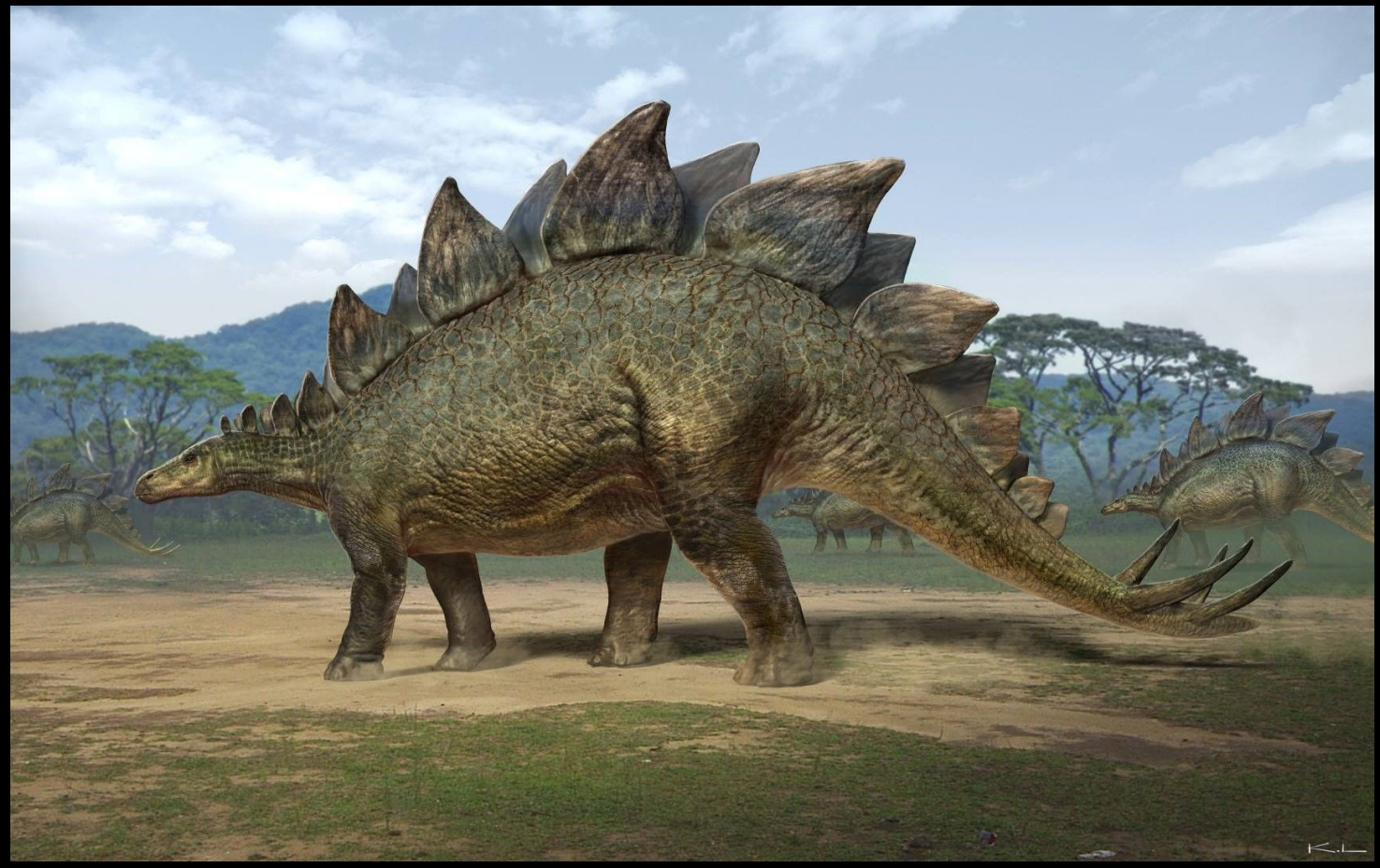
My Dream Attack

But how to get them through?

Biological Encryption



Living Stego



People have been trying this on a research/experimental level for a while

- http://www.csl.sri.com/users/gehani/papers/DIMACS-1999.DNACrypto.pdf Gehani 1999 - good list of early experiments / papers
- https://ieeexplore.ieee.org/document/1205146/ Jie Chen 2003 “DNA Based Biomolecular Cryptography Design”

I am not the first to try (Gehani, 1999)

We will first assemble a large one-time-pad in the form of a DNA strand, which is randomly assembled from short oligonucleotide sequences, then isolated and cloned. These one-time-pads will be assumed to be constructed in secret, and we further assume that specific one-time-pads are shared in advance by both the sender and receiver of the secret message. This assumption requires initial communication of the one-time-pad between sender and receiver, which is facilitated by the compact nature of DNA.

More

The experimental feasibility depends upon the following factors: (i) the size of the lexicon, which is the number of plaintext-ciphertext word-pairs, (ii) the size of each word, (iii) the number of DNA one-time-pads that can be constructed in a synthesis cycle, and (iv) the length of each message that is to be encrypted. If the lexicon used consisted of words of the English language, its size would be in the range of 10,000 to 25,000 word-pairs. If for experimental reasons, a smaller lexicon is required, then the words used could represent a more basic set such as ASCII characters, resulting in a lexicon size of 128. The implicit tradeoff is that this would increase message length. We estimate that in a single cloning procedure, we can produce 10^6 to 10^8 different one-time-pad DNA sequences. Choice of word encodings must guarantee an acceptable Hamming distance between sequences such that the fidelity of annealing is maximized. When generating sequences that will represent words, the space of all possibilities is much larger than the set that is actually needed for the implementation of the words in the lexicon. We also note that if the lexicon is to be split among multiple DNA one-time-pads, then care should be taken during pad construction to prevent a single word from being mapped to multiple targets.

DNA Chips

If long-PCR with high fidelity enzymes introduces errors and the data in question is from an electronic source, we can pre-process it using suitable error-correction coding. If instead we are dealing with a wet database, the DNA one-time-pad's size can be restricted. This is done by splitting the single long one-time-pad into multiple shorter one-time-pads. In this case each cipher word would be modified to include a subsequence prefix that would denote which shorter one-time-

pad should be used for its decryption. This increases the difficulty of cloning the entire set of pads.

Stego was proposed back in 1999 at least

Steganography using DNA is appealling due to its simplicity. One method proposed involves taking "plaintext" input DNA strands, tagging each with "secret key" strands, and then hiding them among random "distracter" strands. The plaintext is retrieved by hybridization with the complement of the secret key strands. It has been postulated that in the absence of knowledge of the secret key, it would be necessary to examine all the strands including the distracters to retrieve the plaintext. Based on the likely difference in entropy of the distracters and the plaintext, we argue that the message can be retrieved without the key.

2003, Jie Chen

A.1. Construction of DNA Cipher Words

There are a number of possible methodologies for construction of cipher words used for the cryptosystems. One methodology is the random assembly of one-time-pads in solution. We view such methods less favorably due to the difficulty of achieving both full coverage and yet still avoiding possible conflicts by repetition of plaintext and/or cipher words. Our favorable method is to employ a DNA chip technology [11], [12]. Such DNA chips are currently commercially available; and chemical methods for construction of custom variants are well developed. The DNA chip has an array of immobilized DNA strands, so that multiple copies of a single sequence are grouped together in a microscopic array. These microscopic arrays of the DNA chip are carbon nanotube (CNT) probe addressable. There is known technology for growing distinct DNA sequences at each site of the array. DNA synthesis can be conducted in parallel. Therefore, the number of sequences prepared far exceeds the number of chemical reactions required. For preparation of oligonucleotides of length L, the $4L$ sequences are synthesized in $4n$ chemical reactions. For example, the 65,000 sequences of length 8 require 32 synthesis cycles. The plaintext and cipher pairs can be constructed so that there is a nearly unique word mapping between plaintext and cipher pairs. These resulting cipher word, plaintext word pairs can be assembled together in random order on a long DNA strand by a number of known methods, e.g.,

DNA Chip


AFFYMETRIX
@52048200653457112008403657070173
GeneChip®
Genome-Wide
Human SNP Array 5.0
EpiC813
PBC3
P/N: 520482
Lot #: 4036570
Exp. Date: 11/2008
For Research Use Only

Lots of crypto papers but not many reports of experiments…

My Idea: Do it small. Do it simple.

- Don’t have the lab, the funding, the team of grad students for microarrays
- I’m going to cook some cryptography up in an E.Coli culture in a way that’s replicable on a DIYish scale
- Edit a plasmid via restriction enzyme digestion.
- Sequence the culture and see if I can get the plain text back out.
- Giggle madly?

Oligos?

Oligonucleotides are short DNA or RNA molecules, oligomers, that have a wide range of applications in genetic testing, research, and forensics. Commonly made in the laboratory by solid-phase chemical synthesis, these small bits of nucleic acids can be manufactured as single-stranded molecules with any user-specified sequence, and so are vital for artificial gene synthesis, polymerase chain reaction (PCR), DNA sequencing, library construction and as molecular probes. In nature, oligonucleotides are usually found as small RNA molecules that function in the regulation of gene expression (e.g. microRNA), or are degradation intermediates derived from the breakdown of larger nucleic acid molecules.

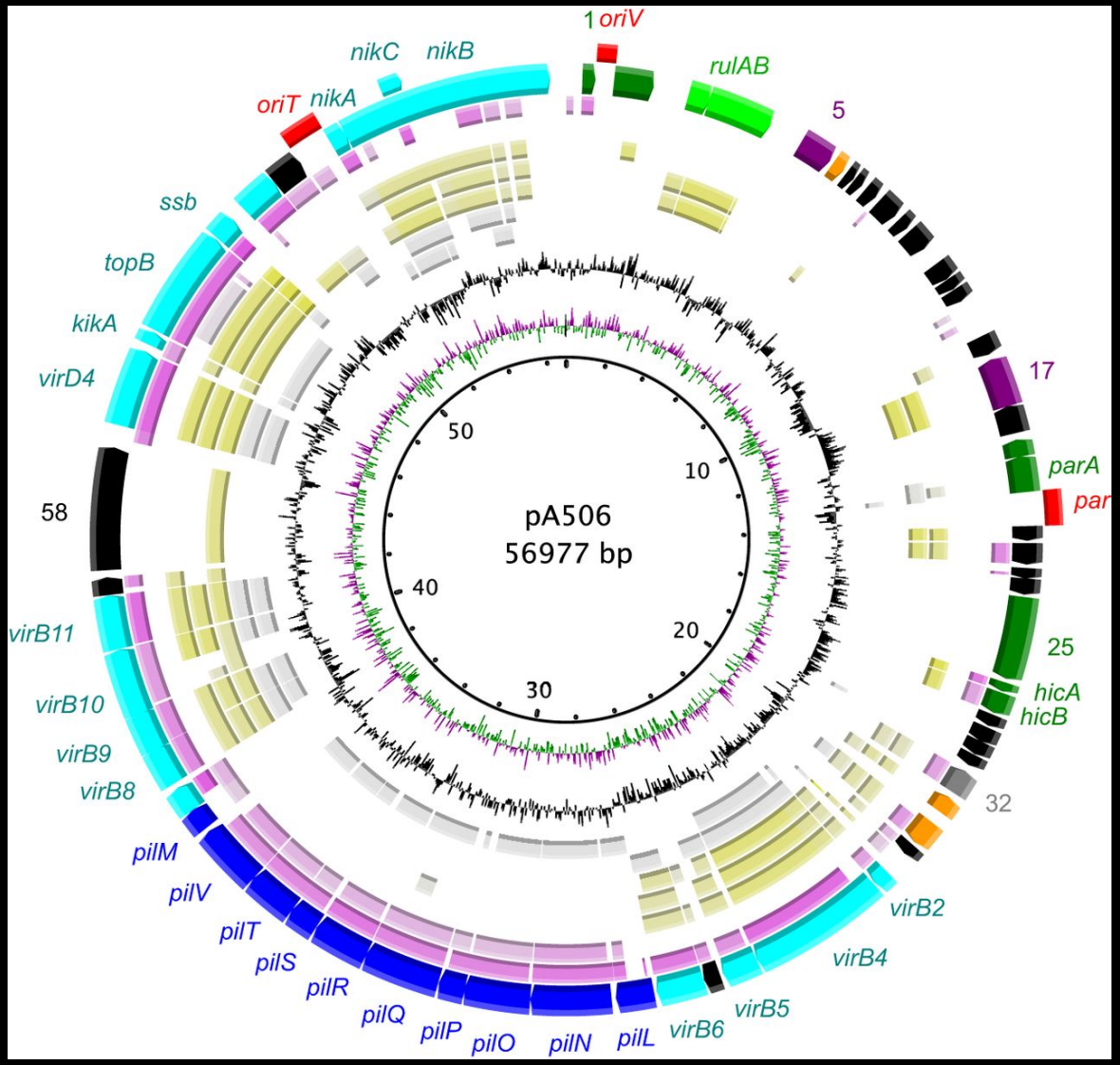
Biosyn has a FAQ on this

The term “oligonucleotide” or “oligo” usually refers to a synthetic laboratory-made DNA or RNA strand. Oligonucleotides are used in biochemistry, biology, molecular diagnostics, genomics, and other molecular biology experiments. Almost all applications using oligos involve synthesizing the complementary strand of a targeted, naturally occurring, strand of nucleic acid sequence.
In a hybridization experiment, the synthesized oligo will bind to the targeted sequence according to classic Watson-Crick base-pairing rules to form a double stranded nucleic acid molecule. The oligo base sequence is fully determined by the targeted sequence moiety of the duplex.
Since the length of an oligonucleotide is directly proportional to its synthesis cost, and inversely proportional to its efficacy, or the ability of the oligo to produce a desired or intended result, the desired oligo length is usually only a fraction of the targeted sequence’s length. Therefore, when

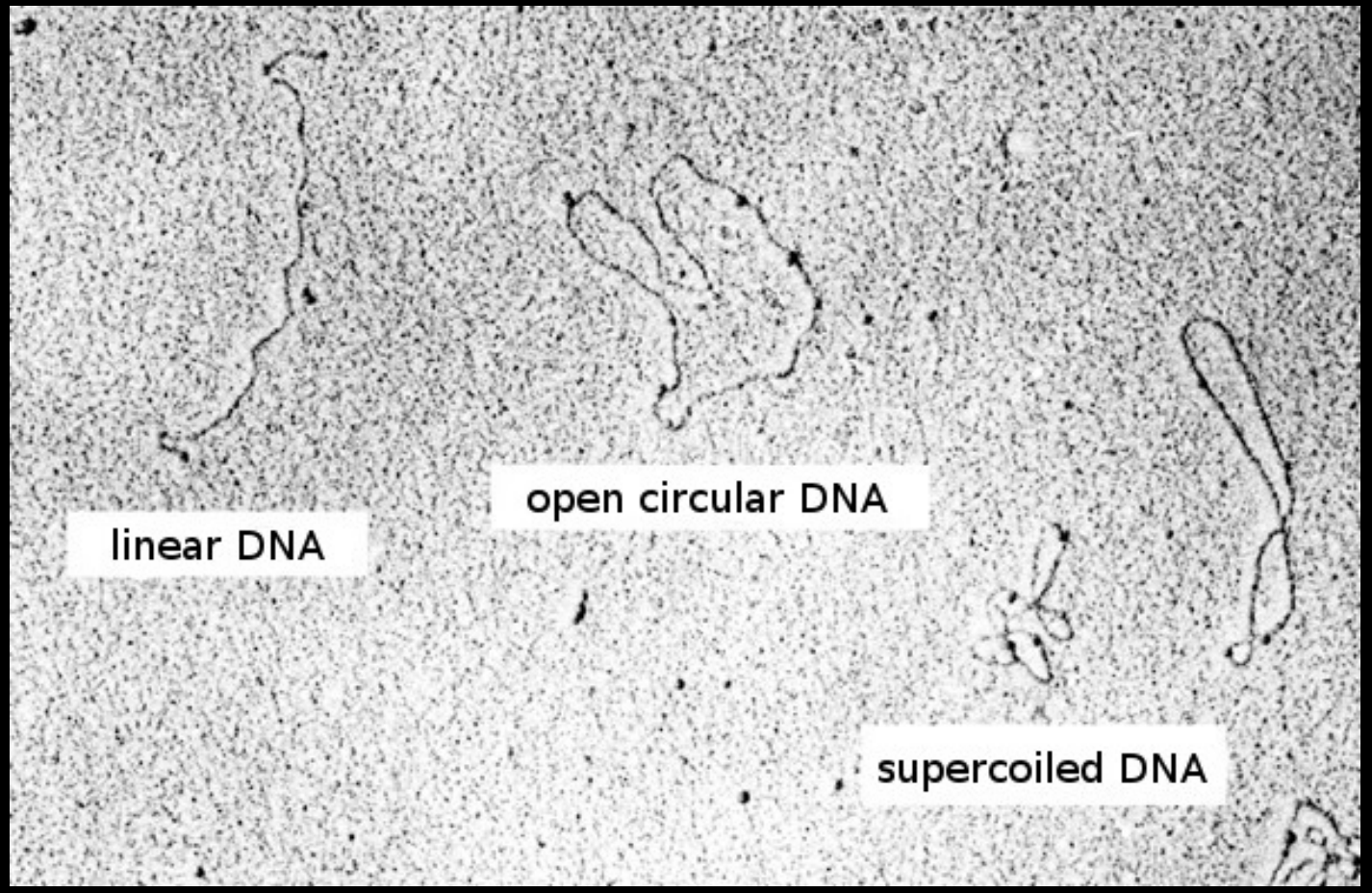
Plasmid

A **plasmid** is a small DNA molecule within a cell that is physically separated from a chromosomal DNA and can replicate independently. They are most commonly found as small circular, double-stranded DNA molecules in bacteria; however, plasmids are sometimes present in archaea and eukaryotic organisms. In nature, plasmids often carry genes that may benefit the survival of the organism, for example antibiotic resistance. While the chromosomes are big and contain all the essential genetic information for living under normal conditions, plasmids usually are very small and contain only additional genes that may be useful to the organism under certain situations or particular conditions. Artificial plasmids are widely used as vectors

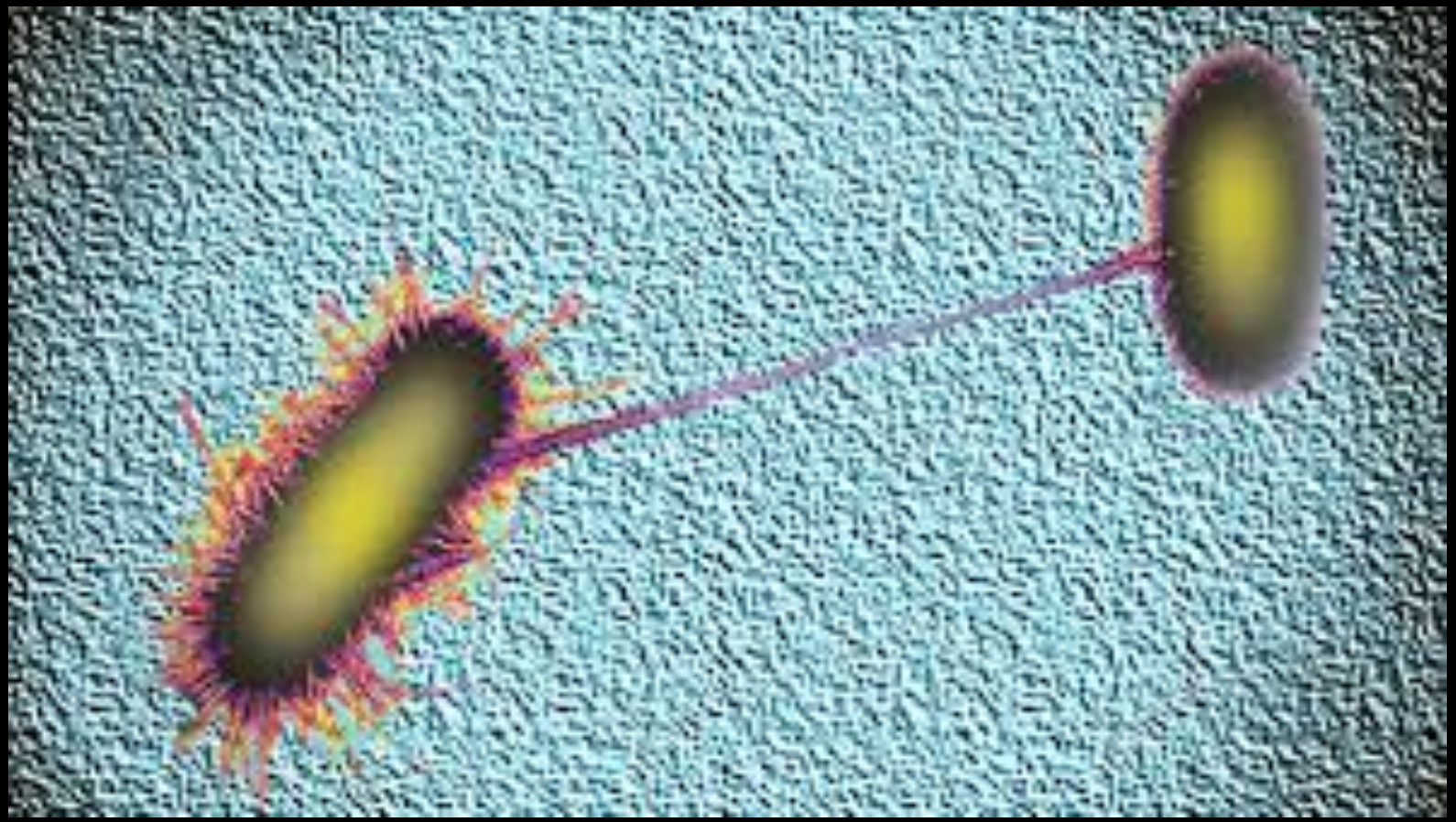
Think DNA Donut



SEM Plasmid



Bacteria sharing plasmids via conjugation



Step 1: Pick an Amenable Model Organism

Step 2: Pick an amenable encryption strategy

- Requirements include: Small size, low complexity, high integrity, low cost
- Low cost
- Low cost
- Something I can do in a community lab

Constraints of Affordable Synthesis

- Size – anything above about 60 nucleotides gets much more expensive and complex
- No repeating patterns
- No self compliments / palindromes

Algorithms That Wont Work

- AES/DES – too big
- Any kind of public key crypto – I don’t want to synthesize a bunch of oligos to implement key sharing. Also too long.

Professional cryptographers are coming up with a way to do those though

- People are coming up with some novel methods of hiding keys in DNA
- Attempts at all modern forms of cryptography, at least on the theoretical level have been proposed.
- I haven’t seen anyone document doing it in a lab

One Time Pad

- It can be small
- It can be perfect
- I don’t want to implement the key sharing in the DNA due to complexity and expense. We will assume the conspirators have little padbooks like back in the day.
- The sheer data density of DNA gives an attractive level of obscurity even if we know that we shouldn’t rely on that.

Step 3: Design Message Encoding Options

Nucleotides

- We want to convert ascii strings into ACGT
- How we do this will determine the density of data we can encode

Base 4

Value	Acid
0	Adenine
1	Cytosine
2	Guanine
3	Thymine

Side note: I had no idea people called b4 integers quats. This is amusing to me.

No, it doesn't. Let's see you design a 16-quat full adder that takes fewer transistors or less die area than an 32-bit full adder.
Base 3 or higher are a lose for implementing logic. Base 4 is useful in some kinds of memory, and this has been done by Intel since

	Binary	4 Quat	8 Quat	12 Quat	Quaternary
Complete Errors	8	6	129	213	302
Small Errors	3	174	166	188	199

Byte Width

- Depending on how big our bytes are we can encode different quantities of information
- I prefer a four *quat* bytes (qytes? Crumb? Quites?)
- 4 quat bytes can represent the entire ascii table

Comparison

Byte Size	Max Int	Max Message Size for 32 bp	It can represent…
4	256	8	Ints, full ascii table
3	64	10	Most alphabets with some wiggle room for other stuff
2	16	16	Viable subset of English letters

Step 4: Design a primer

- I want to do PCR so the rest of the oligo will help with the primer
- 20 nt primer binding site (random)
- 20 nt reverse primer (compliment)
- Primer Binding site -> payload -> reverse primer
- The lab I contacted me assured this would work with their in house plasmid and universal primer.
- They also gave me a lot of help and advice. They are very nice people.
- Literally taught me all of this stuff.

IDT has a good explanation

When do researchers use random (degenerate, wobble, mixed) base oligos

Mixed bases are used in primers to bind to templates that contain variability or a mixture of sequences at the primer binding sites. Mixed bases can also be used to create diversity in clone libraries and in site directed mutagenesis. IDT offers random base oligos. Order them by using an upper case letter N or other IUPAC-IUB symbol (see table below). IDT offers two

Repeating Sequences (Bioinformatics wiki)

Sequences with short repeating units

The two most common terms used for sequences containing short repeating units are *simple sequence repeat (SSR)* and *microsatellite*.

SSRs are composed of short (1 to 5 bp), tandemly repeating units that are exact in identity and repetition. Although the elongation of SSR tracts may be due to more than one mechanism [1], much is thought to be the result of slip-strand replication errors. In the process of nascent strand formation, reannealing can occur. And when the strands contain repetitive elements, such as with SSR tracts, the annealing can be imperfect, leading to the addition of the same elements. The errors become permanent when an additional round of replication occurs before they are discovered by repair enzymes [2][3].

Some biochemists use the descriptor *polymeric* and modifications thereon in order to describe a repeat more precisely, e.g.:

- *homopolymer* or *heteropolymer*
- *poly()*
  - *dU/dY*
  - *dU.dY*
  - *dU:dY*
  - *poly(dU).poly(dY)*
- *polypurine.polypyrimidine*

Self Compliments are Bad

- Some sequencing techniques will generate significant errors when you do this

Biosoft explains this better than I do

6. Primer Secondary Structures: Presence of the primer secondary structures produced by intermolecular or intramolecular interactions can lead to poor or no yield of the product. They adversely affect primer template annealing and thus the amplification. They greatly reduce the availability of primers to the reaction.

i) Hairpins: It is formed by intramolecular interaction within the primer and should be avoided. Optimally a 3' end hairpin with a ΔG of -2 kcal/mol and an internal hairpin with a ΔG of -3 kcal/mol is tolerated generally.

ΔG definition: The Gibbs Free Energy G is the measure of the amount of work that can be extracted from a process operating at a constant pressure. It is the measure of the spontaneity of the reaction. The stability of hairpin is commonly represented by its ΔG value, the energy required to break the secondary structure. Larger negative value for ΔG indicates stable, undesirable hairpins. Presence of hairpins at the 3' end most adversely affects the reaction.

$$\Delta G = \Delta H - T\Delta S$$

Temperature

3. Primer Annealing Temperature: The primer melting temperature is the estimate of the DNA-DNA hybrid stability and critical in determining the annealing temperature. Too high T_a will produce insufficient primer-template hybridization resulting in low PCR product yield. Too low T_a may possibly lead to non-specific products caused by a high number of base pair mismatches,. Mismatch tolerance is found to have the strongest influence on PCR specificity.

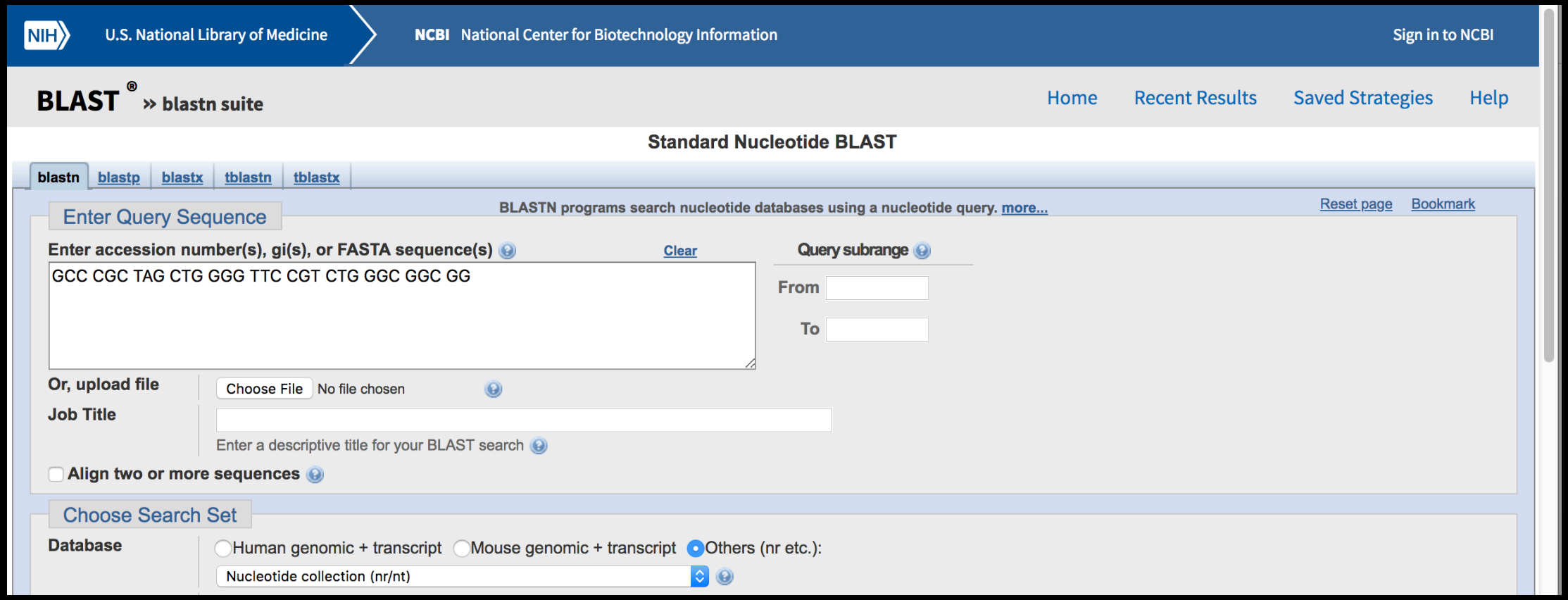
$$T_a = 0.3 \times T_m(\text{primer}) + 0.7\ T_m\ (\text{product}) - 14.9$$

where,

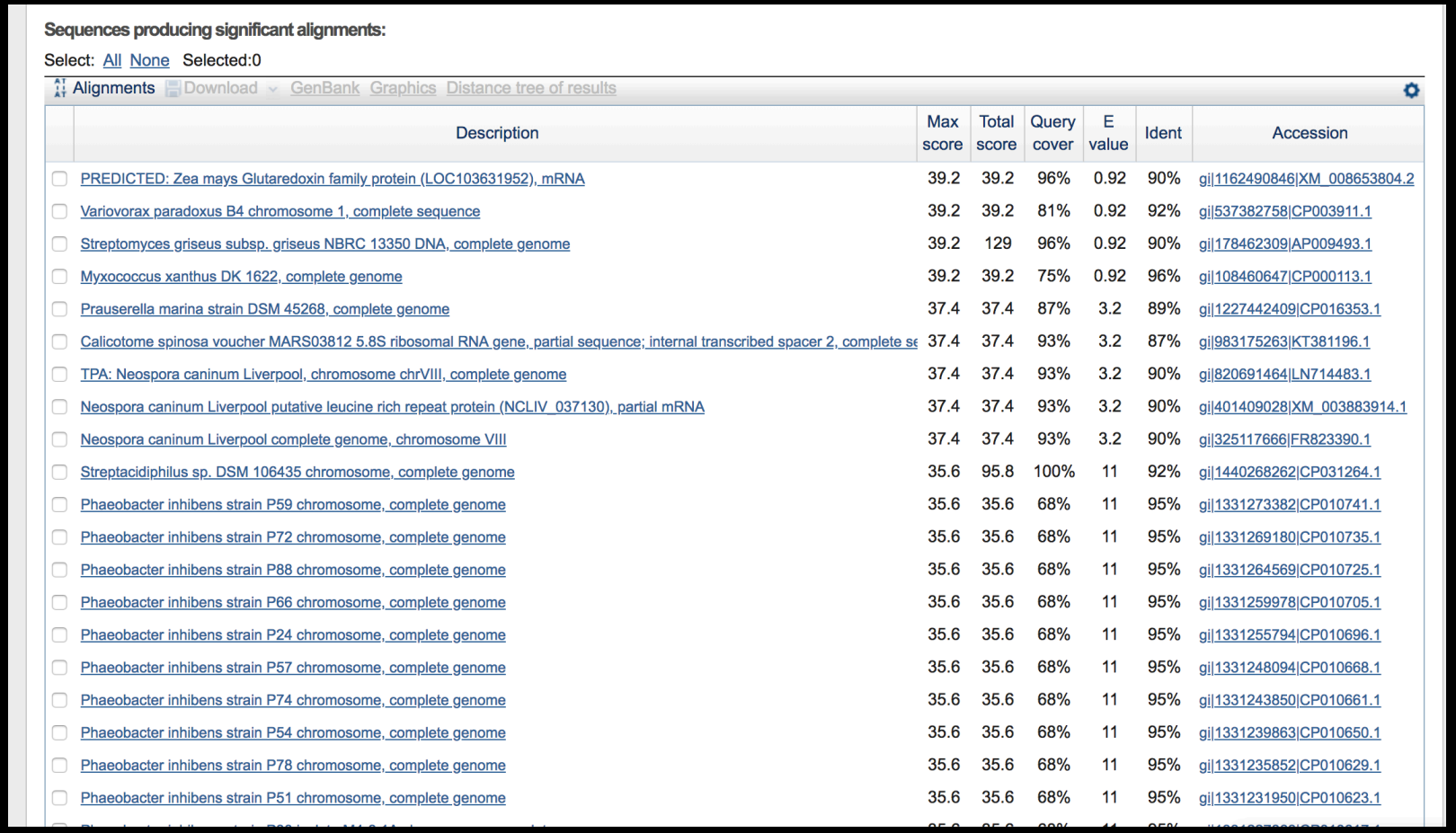
T_m(primer) = Melting Temperature of the primers

T_m(product) = Melting temperature of the product

BLAST


NIH
U.S. National Library of Medicine
NCBI National Center for Biotechnology Information
Sign in to NCBI
BLAST ® » blastn suite
Home
Recent Results
Saved Strategies
Help
Standard Nucleotide BLAST
blastn
blastp
blastx
tblastn
tblastx
BLASTN programs search nucleotide databases using a nucleotide query. more...
Reset page
Bookmark
Enter Query Sequence
Enter accession number(s), gi(s), or FASTA sequence(s)
Clear
GCC CGC TAG CTG GGG TTC CGT CTG GGC GGC GG
Query subrange
From
To
Or, upload file
Choose File
No file chosen
Job Title
Enter a descriptive title for your BLAST search
Align two or more sequences
Choose Search Set
Database
Human genomic + transcript
Mouse genomic + transcript
Others (nr etc.):
Nucleotide collection (nr/nt)

Stuff BLAST finds



Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenBank Graphics Distance tree of results

Description	Max score	Total score	Query cover	E value	Ident	Accession
PREDICTED: Zea mays Glutaredoxin family protein (LOC103631952), mRNA	39.2	39.2	96%	0.92	90%	gi|1162490846|XM_008653804.2
Variovorax paradoxus B4 chromosome 1, complete sequence	39.2	39.2	81%	0.92	92%	gi|537382758|CP003911.1
Streptomyces griseus subsp. griseus NBRC 13350 DNA, complete genome	39.2	129	96%	0.92	90%	gi|178462309|AP009493.1
Myxococcus xanthus DK 1622, complete genome	39.2	39.2	75%	0.92	96%	gi|108460647|CP000113.1
Prauserella marina strain DSM 45268, complete genome	37.4	37.4	87%	3.2	89%	gi|1227442409|CP016353.1
Calicotome spinosa voucher MARS03812 5.8S ribosomal RNA gene, partial sequence; internal transcribed spacer 2, complete se	37.4	37.4	93%	3.2	87%	gi|983175263|KT381196.1
TPA: Neospora caninum Liverpool, chromosome chrVIII, complete genome	37.4	37.4	93%	3.2	90%	gi|820691464|LN714483.1
Neospora caninum Liverpool putative leucine rich repeat protein (NCLIV_037130), partial mRNA	37.4	37.4	93%	3.2	90%	gi|401409028|XM_003883914.1
Neospora caninum Liverpool complete genome, chromosome VIII	37.4	37.4	93%	3.2	90%	gi|325117666|FR823390.1
Streptacidiphilus sp. DSM 106435 chromosome, complete genome	35.6	95.8	100%	11	92%	gi|1440268262|CP031264.1
Phaeobacter inhibens strain P59 chromosome, complete genome	35.6	35.6	68%	11	95%	gi|1331273382|CP010741.1
Phaeobacter inhibens strain P72 chromosome, complete genome	35.6	35.6	68%	11	95%	gi|1331269180|CP010735.1
Phaeobacter inhibens strain P88 chromosome, complete genome	35.6	35.6	68%	11	95%	gi|1331264569|CP010725.1
Phaeobacter inhibens strain P66 chromosome, complete genome	35.6	35.6	68%	11	95%	gi|1331259978|CP010705.1
Phaeobacter inhibens strain P24 chromosome, complete genome	35.6	35.6	68%	11	95%	gi|1331255794|CP010696.1
Phaeobacter inhibens strain P57 chromosome, complete genome	35.6	35.6	68%	11	95%	gi|1331248094|CP010668.1
Phaeobacter inhibens strain P74 chromosome, complete genome	35.6	35.6	68%	11	95%	gi|1331243850|CP010661.1
Phaeobacter inhibens strain P54 chromosome, complete genome	35.6	35.6	68%	11	95%	gi|1331239863|CP010650.1
Phaeobacter inhibens strain P78 chromosome, complete genome	35.6	35.6	68%	11	95%	gi|1331235852|CP010629.1
Phaeobacter inhibens strain P51 chromosome, complete genome	35.6	35.6	68%	11	95%	gi|1331231950|CP010623.1

I’m a programmer so I automated

- Converts ASCII into one time pad
- Fill in any blanks
- Base 4 it
- Add the one time pad
- Configurable byte size
- Decrypts as well

You can grab my script on github

- https://github.com/MrSynAckSter/CryptoPlasmid

Small PoC

- We will convert some 1337 $p34k into DNA, because that amuses me.
- A commercial synthesis company actually already made this 1337 infected DNA. There were no complaints over my s1ck w4r3z
- The tool doesn’t do some of the more robust structural checks on the DNA (self compliment, temperature, etc…) – if people are interested I may implement some of that.

The Key

- One time pad.
- Same size as message.
- User Specified.
- You'll have plenty of time between ordering oligos to figure out the next key.

The Params

- Plaintext: 1337f347 - because we accomplished something cool.
- Key: cryptkid – because we are using a script to become a genetic engineering script kiddy

Help / Args

```
DNA encryption for plasmids with Python

optional arguments:
  -h, --help  show this help message and exit
  -k KEY      Key string. Must be equal size to your final payload. Defaults
              to the key from the original talk
  -f KF       Key file. Loads a key from a file.
  -e PT       specify string to encrypt
  -i PF       file to encrypt
  -l PL       Optional. Length of payload. 32 bp is assumed. This tool does
              not account for bp larger than ~60.
  -b BSZ      byte-size. Specifies the size of the quaternary byte. Defaults
              to 4 quats. For values < 4 sensible ascii table offsets are
              assumed.
  -d DEC      Specify string for decryption
  -df DF      specify file for decryption
  -x DX       double check success by decrypting
```

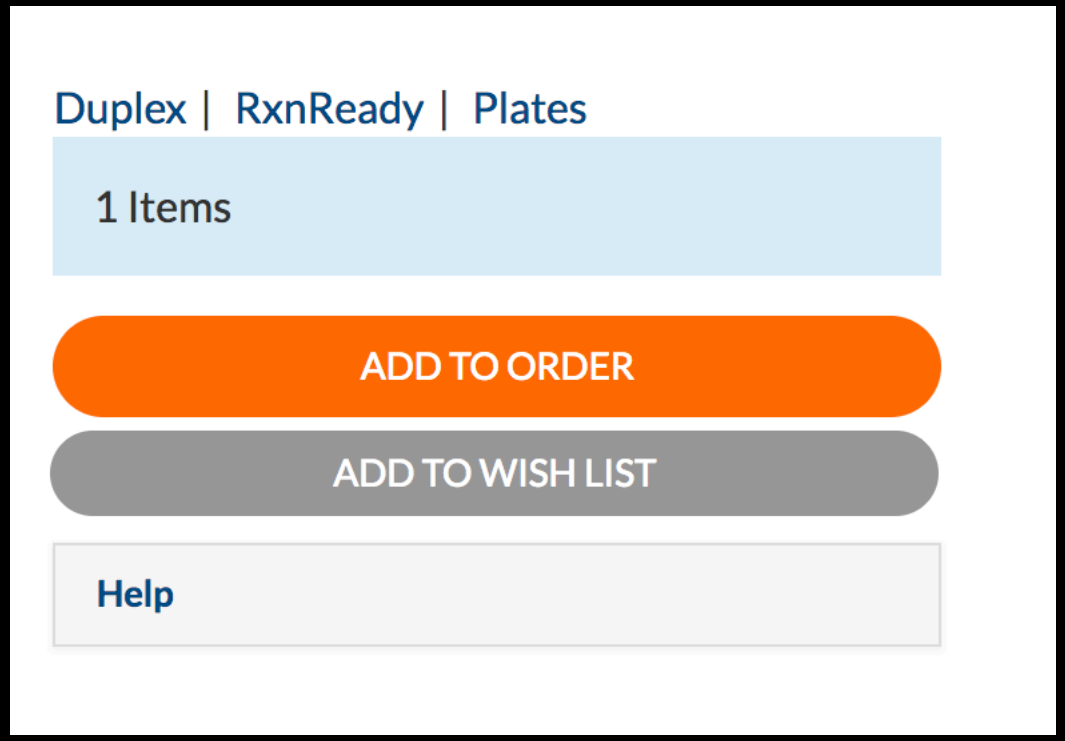
Running

```
Johns-MacBook-Pro:CryptoPlasmid jdunlap$ python cryptoplasmid.py -e 1337f347 -k cryptkid
Final String for Synthesis
GCCCGCTAGCTGGGGTTCCGGGTAGGCGGCGG
```

Decrypting

```
Johns-MacBook-Pro:CryptoPlasmid jdunlap$ python cryptoplasmid.py -df GCCCGCTAGCTGGGGTTCCGGGTAGGCGGCG
G -k cryptkid
decrypting payload
1337f347
```

Step 5: Ordering DNA for Synthesis



You can’t actually just do this

- Synthesis providers typically only sell to labs
- You will have to develop a relationship with a lab in good standing with a synthesis provider in order to buy the DNA
- The scientific industry is actually very responsible when it comes to checking up on sketchy people trying to order DNA / Lab equipment / chemicals
- On a positive note the ordering forms often have nifty validation suites to help make sure your DNA is viable

It would be pretty tough to do the synthesis on your own

- I hear the chemicals are pretty nasty and it’s not easy.
- Commercial synthesis is cheap enough that it’s not worth the trouble/danger/expense/equipment
- Wasn’t really an option for me.

Step 6: Editing E. Coli

- Use a restriction enzyme to edit a plasmid
- Basically the original DNA editing technique
- Essentially snip the plasmid leaving some “sticky ends” which our synthetic DNA may attach to
- Exactly like the class I took. Recreate intro to gene hacking experiment with our encrypted ASCII data integrated into the plasmid

A Vague High Level Lab Strategy

- https://www.genscript.com/restriction-digestion-protocol.html has a wonderful summary of what you would do. It's a good site to read layman accessible versions of this stuff
- I ended up asking a lot of questions and needing a lot of help. Having a lab buddy who knows what they are doing is a good thing.

Lab pics

- Stop for pictures here if time permits

Muh DNA!

IDT
INTEGRATED DNA TECHNOLOGIES
1710 COMMERCIAL PARK
CORALVILLE, IOWA 52241
(319) 626-8400

PACKING LIST

Package# 36810957

Packaging Date: 03-Aug-2018

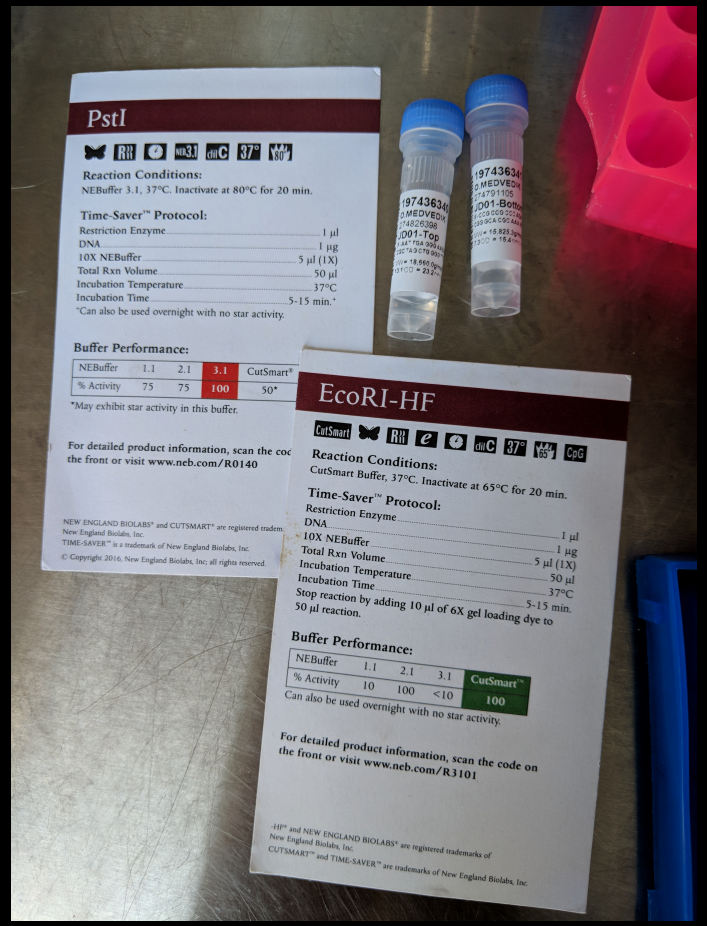
OLIVER MEDVEDIK
GENSPACE
140 32ND ST
STE. 108
BROOKLYN, NY 11232
USA
(516) 200-6537

Cust # 265729 Sales Order # 15314891 P.O. # Credit Card Order Date 01-Aug-2018

Qty	Ref #	Product	Description	Purification
1	197436340	25 nmole DNA Oligo	JD01-Top	Standard Desalting
1	197436341	25 nmole DNA Oligo	JD01-Bottom	Standard Desalting
2		Sales Order Total		
2		Package Total		

Page 1 of 1

THE BUILDING BLOCKS OF LIFE!



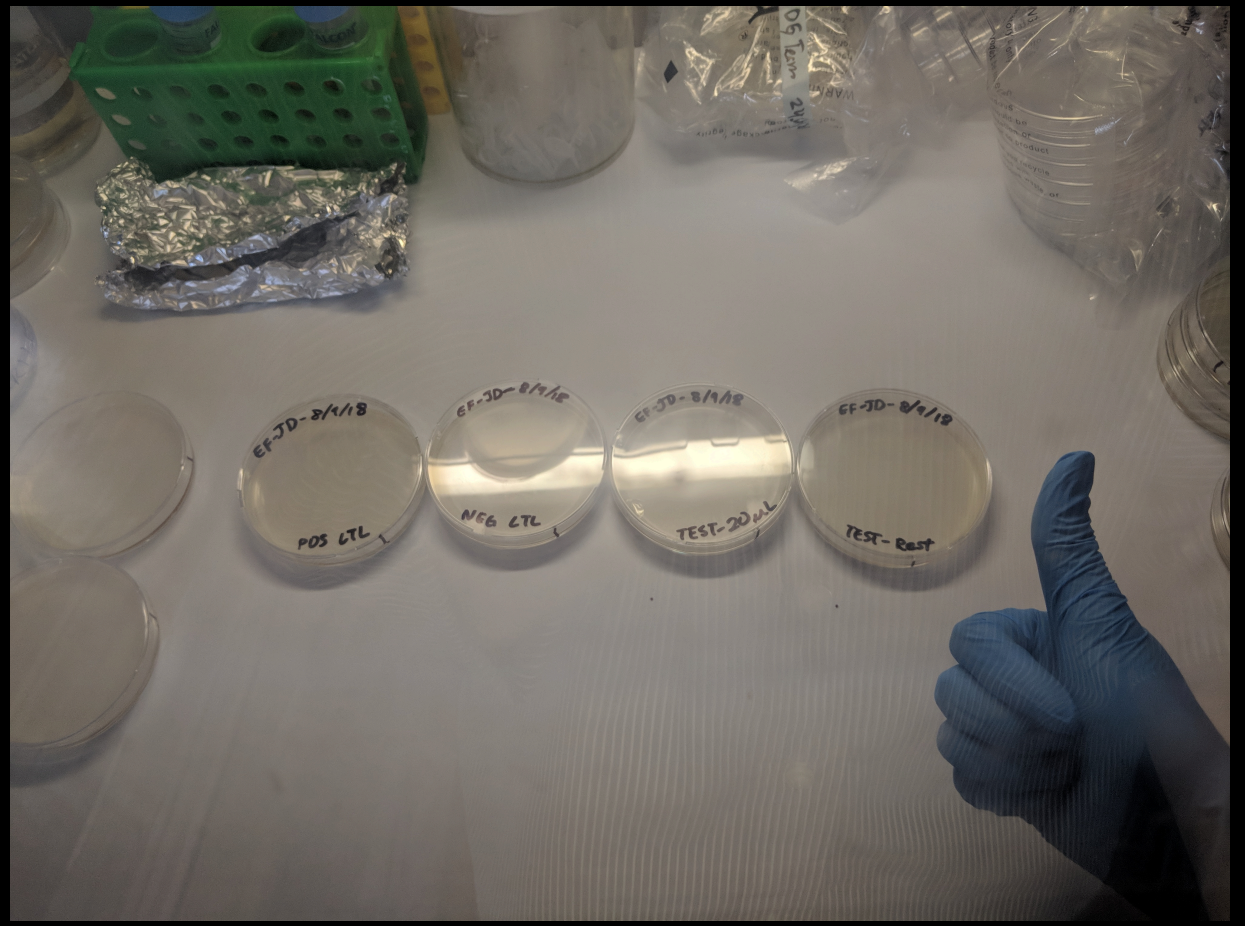
PCR Machine


1 Hld
1 Hld
80.0
37.0
20:00
30:00
Start
Store
Print
More
Cancel
F1
F2
F3
F4
F5
POWER
STOP
1
2
3
4
5
6
7
8
9
ENTER
0
CE
GeneAmp®
PCR System 9700

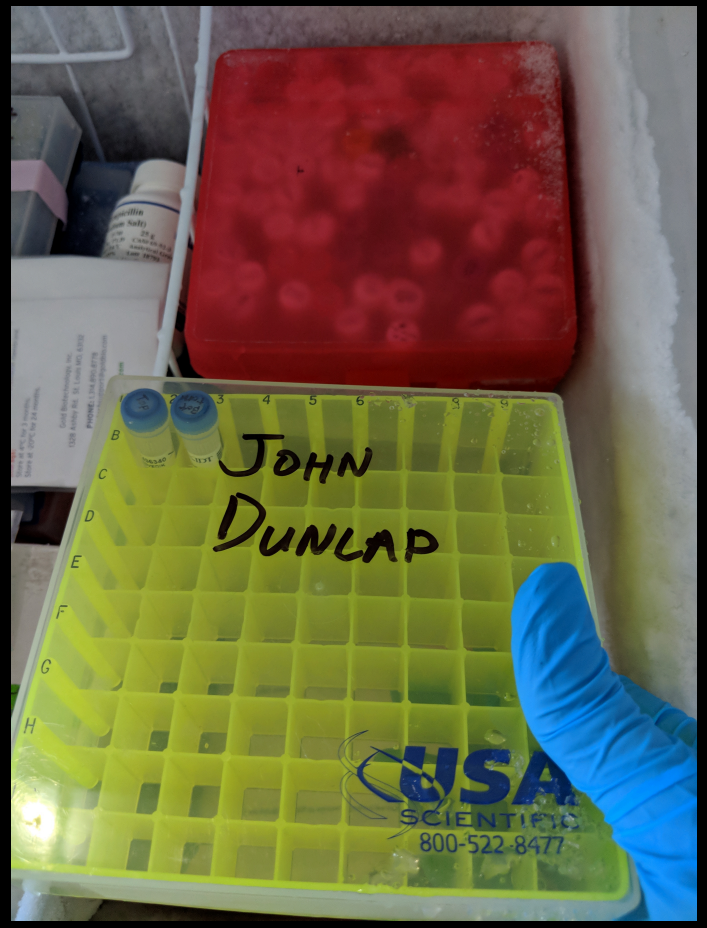
More PCR


T4 DNA Ligase
M0202S
20,000 units
400,000 cohesive end units/ml
Exp: 12/18
RECOMBINANT
Store at –20°C
Lot: 1201612
Description: Catalyzes the formation of a phosphodiester bond between juxtaposed 5´ phosphate and 3´ hydroxyl termini in duplex DNA or RNA. This enzyme will join blunt end and cohesive end termini as well as repair single-stranded nicks in duplex DNA, RNA or DNA/RNA hybrids (1).
Source: Purified from E. coli C600 pci857 pPLc28 lig8 (2)
Applications:
• Cloning of restriction fragments (3)
• Joining linkers and adapters to blunt-ended DNA
1 Hld
16.0
30:00
Start
Store
Print
F1
F2
F3
F4
F5
POWER
STOP
GeneAmp
PCR System 9700
ENTER
CE

My plates

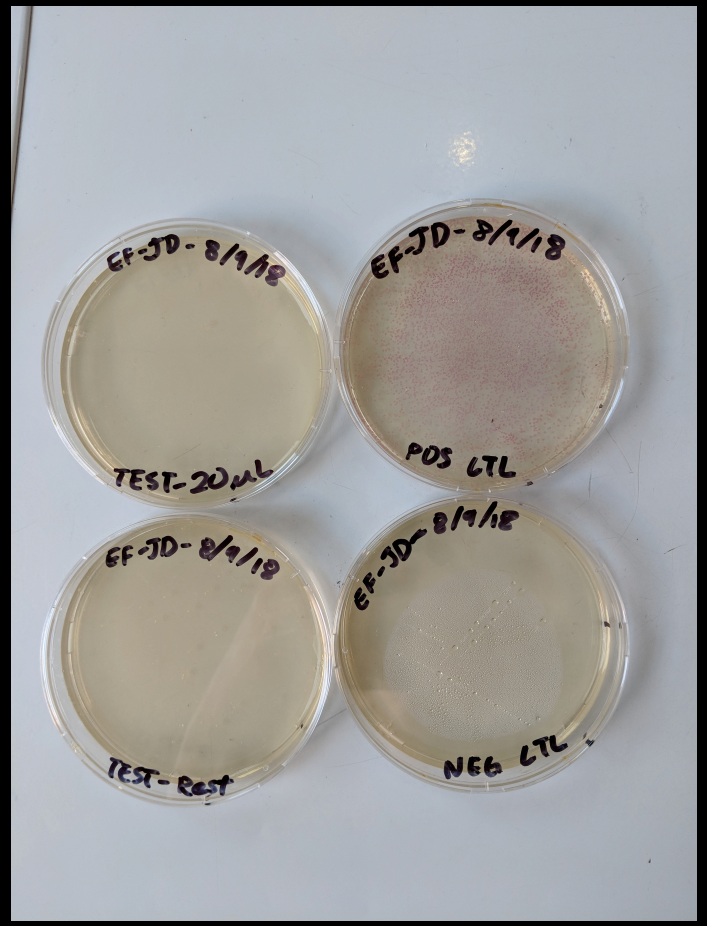


Holding the tubes

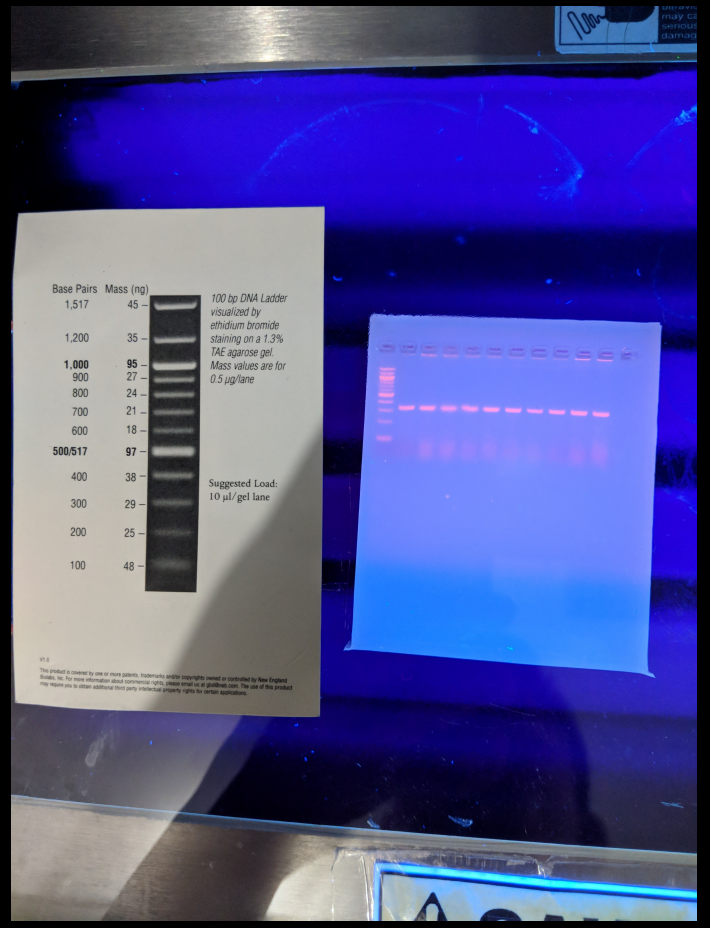


Getting ready to PCR

More plates


EF-JD- 8/9/18
TEST- 20 µL
EF-JD- 8/9/18
POS LTL
EF-JD- 8/9/18
TEST- Rest
EF-JD- 8/9/18
NEG LTL

PCR Gels


Base Pairs
Mass (ng)
1,517
45
1,200
35
1,000
95
900
27
800
24
700
21
600
18
500/517
97
400
38
300
29
200
25
100
48
100 bp DNA Ladder visualized by ethidium bromide staining on a 1.3% TAE agarose gel. Mass values are for 0.5 µg/lane
Suggested Load:
10 µl/gel lane

Genscript's Protocol

- Thaw all reagents on ice.
- Assemble reaction mix into 50 µL volume in a microfuge tube.
- Add reagents in following order: water, buffer, BSA, DNA template, restriction enzyme.
- Gently mix by tapping tube. Briefly centrifuge to settle tube contents.
- Prepare negative control reaction without template DNA.
- Prepare positive control reaction with template of known cutting site corresponding to the restriction enzyme of choice.
- Typical Incubation time and temperature is 37 C° for 1 hour, though time and temperature will vary depending on restriction enzyme used.
- Restriction enzymes are typically inactivated by incubation at high temperature. Incubation time and temperature is 65C° for 20 min, though time and temperature will vary depending on restriction enzyme used.
- Analyze the results of your PCR reaction via gel electrophoresis.

There a lot of variations on the instructions

- http://2009.igem.org/wiki/images/1/1d/PKU_DNA_digestion_protocol.pdf
- https://www.addgene.org/protocols/restriction-digest/ - has a nice video of their protocol in action
- https://www.thermofisher.com/us/en/home/life-science/cloning/restriction-enzyme-digestion-and-ligation.html
- http://www.methodbook.net/dna/restrdig.html
- Procedures are likely to vary a bit based on your setup

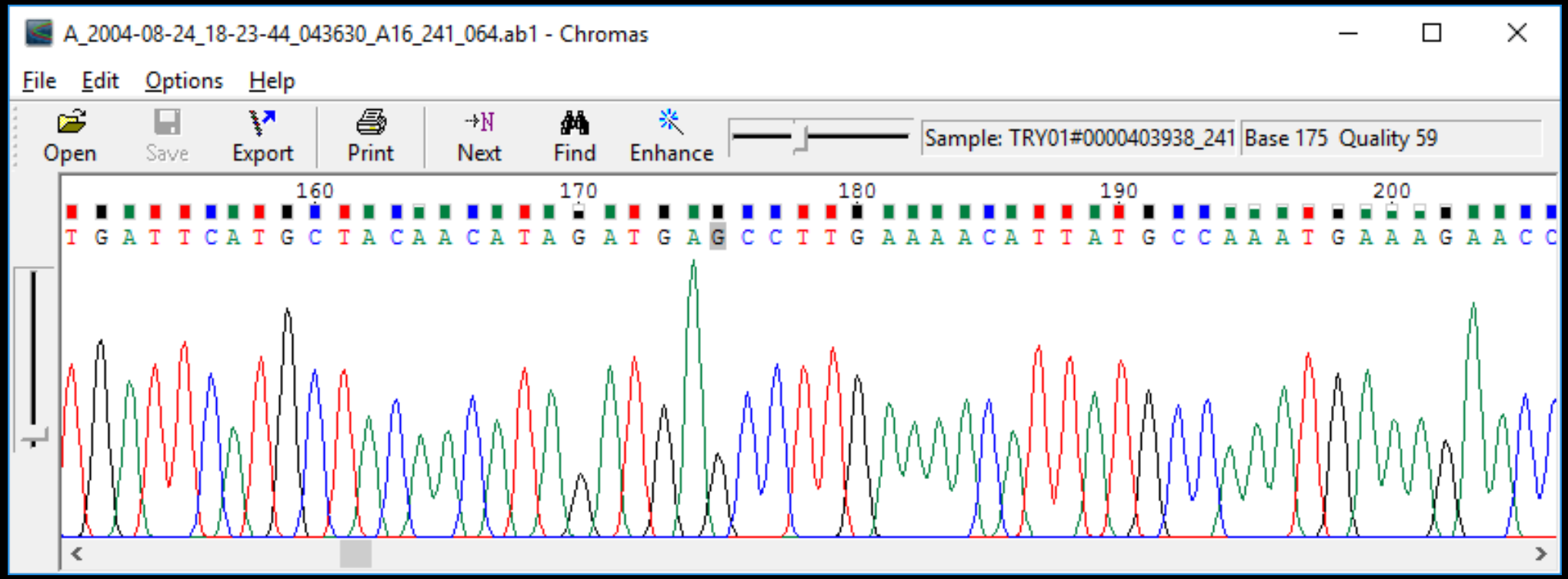
Step 7: Sequencing

- Again, having a relationship with a legit lab is going to make this way easier
- With my skillset/tools using a service was the only option
- Costs much more money as you ask for more sequencing
- Doing more prep/purification will lower the cost
- Sending them a raw bacterial colony will cost a bit more
- Roughly a few hundred dollars, give or take

Step 8: Interpret Results

- The lab will send you a big list of nts
- Things may have gone horribly wrong
- You may have to try again

Chromatograph Readings



Now you can take your secret data anywhere (that's cool with E.Coli)

Costs

Item	Cost (per 32 bytes of data)
Lab Membership	~$300+ (ongoing)
Synthesis	$40
Lab expendables (enzymes etc…)	$200 (to get set up)
Sequencing	$300 (could be less in some places)
Total	~$800

Benefits

- No one will ever guess it’s in the bacteria.
- Not an obvious storage medium
- No one will ask you for key escrow on your pet bacteria
- It would take a very dedicated effort to back door this one
- The joy of birthing synthetic life
- Lots of practice pipetting

Downsides

- Pretty expensive
- Slow
- Key exchange sucks
- Hard to say if your plasmid will survive for long in successive generations of bacteria. Likely will not offer a biological benefit. Natural selection may quickly degrade/remove it.
- Can’t really do it without the involvement of several labs
- There will be a lot of un-stealthy records/invoices.
- Then again I can imagine organizations with an interest in stealth building labs for this sort of thing from scratch.

Cost per byte is a little high compared to on chip AES

My Aspirations for the Future

- New model organisms
- Try it with CRISPR
- More elaborate algorithms
- A viable method to package the living colony for stealthy transport
- Some longish term bacterial husbandry to determine how well this bare bones method survives in the colony after successive generations
- Error correction code implementation
- Angering the lab people less with decreased begging for help

Yeast?

I Hope You Had Fun

- If you are well informed molecular biologist, I am so sorry. This must be like when I watch CSI cyber.
- The biohacking community is very supportive and you should get involved with it even if you don’t think you have the background
- If you are a computer person there are a lot of fun bioinformatics projects to get involved with
- If you are thinking of starting a community lab, please do. The world needs more.
- Feel free to hit me up on github or @JohnDunlap2 if you have suggestions or questions about my project

Happy Biohacking!