

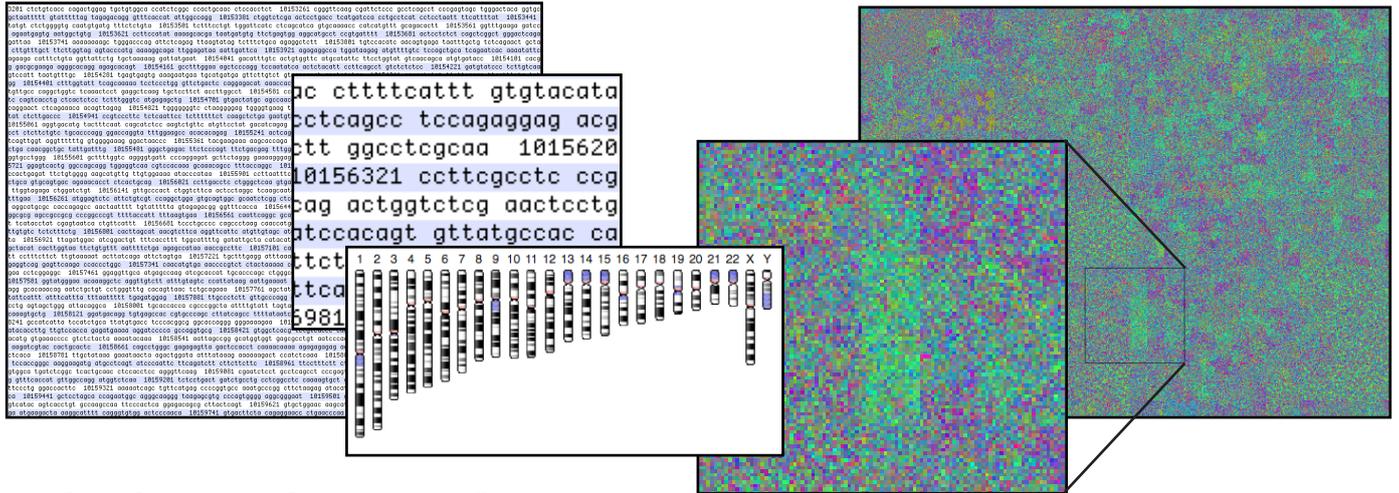
AminoSeeNoEvil



DNA Viewer by Tom Atkinson - aminosee.funk.nz

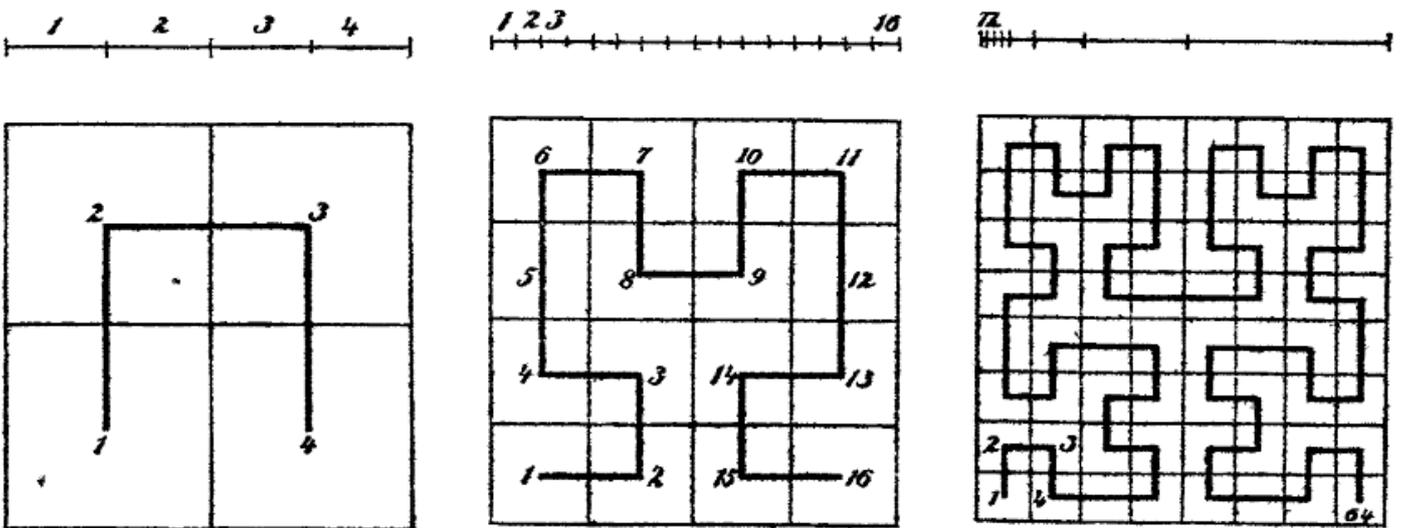


The goal: Visually represent an entire genome graphically
...in one shot



Think Google Maps for DNA: AminoSee allows a full colour stable spatial rendering of our largest molecule, painting each amino acid a unique colour blending many-per-pixel; this clustering concentrates features from the genomes: from large objects like chromosomes, telemers, and centromeres, right down to small pixel level detail mutated genes, and single nucleotide repeats. Zooming out does not destroy (so much) the image as it descends into a blurry 33% grey rainbow smudge like static averaged down.

Why Y? Like the police black and white forensic evidence made via PCR, this is a full-colour version of PCR rendered via computer given a set of digitised DNA file to present in court; while it will show the “hypervariable microsatellites” they are buried in noise; the design goal actually is the cross-species comparisons of ancient humans for my project to understand mutations in chromosome 2 and the Y in the period 340-450 kBP with references to validating the Sumerian account of human historical modification by aliens near the gold mines of Africa, comparing to primate chromosome 2p+2q chosen because oddly while humans have a fused chromosome 2 giving 46 chromosomes whereas all greater primates have 48.



Amino Acid	Colour Lookup	Hue	RGB	
Ochre	AminoSee converts plain-text files containing DNA or	0°	255,128,128	
Glutamic acid	RNA code to an image by first turning it into a series of	16°	255,162,128	
Aspartic acid	colours blended many-per-pixel, then wraps the invisibly	31°	255,193,128	
Amber	thin 1-dimensional coil into clumpy 2D and 3D projec-	47°	255,227,128	
Cysteine	tions using space filling curve provides resolution inde-	63°	249,255,128	
Glycine	pendent: genes can bunch in blobs rather than spread	78°	217,255,128	
Alanine	out and always maintain their proximity.	94°	183,255,128	
Methionine	[AAA] [CAT] [GAT]	110°	149,255,128	
Valine	becomes	125°	128,255,138	
Leucine	[Lysine] [Histidine] [Aspartic acid]	141°	128,255,172	
Isoleucine		157°	128,255,206	
Phenylalanine		172°	128,255,238	
Tryptophan		188°	128,238,255	
Serine		203°	128,206,255	
Threonine		219°	128,172,255	
Opal		240°	128,128,255	
Glutamine		250°	149,128,255	
Asparagine		266°	183,128,255	
Tyrosine		282°	217,128,255	
Arginine		297°	249,128,255	
Lysine		The exact colour values are (not yet final) are intended	313°	255,128,227
Histidine		to be the actual colour of the compound, like Opal and	329°	255,128,193
Start Codons		Ochre. Colder blues for water solubles, yellow for sulphur,		
Stop Codons		reds for the acids.		
Proline			344°	255,128,162

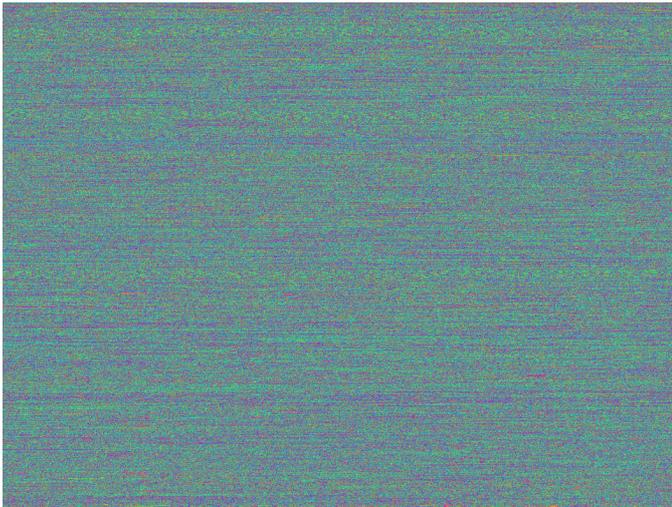
Two passes: 1D linear then 2D Hilbert curve

The first “linear” pass establishes the total number of codons, produces the images on the left, a simple map from left to right, top to bottom, the way a television scans each pixel. Structures like genes are hard to see and spread out visually horizontally, splitting for each scanline. Codons per pixel ≈ 10.3 , meaning 7,149,024 codons are displayed. Final hilbert map shrinks this to 27.3 codons per pixel, during anti-aliasing. After projection through the Hilbert curve, the various genetic objects are much more visible.

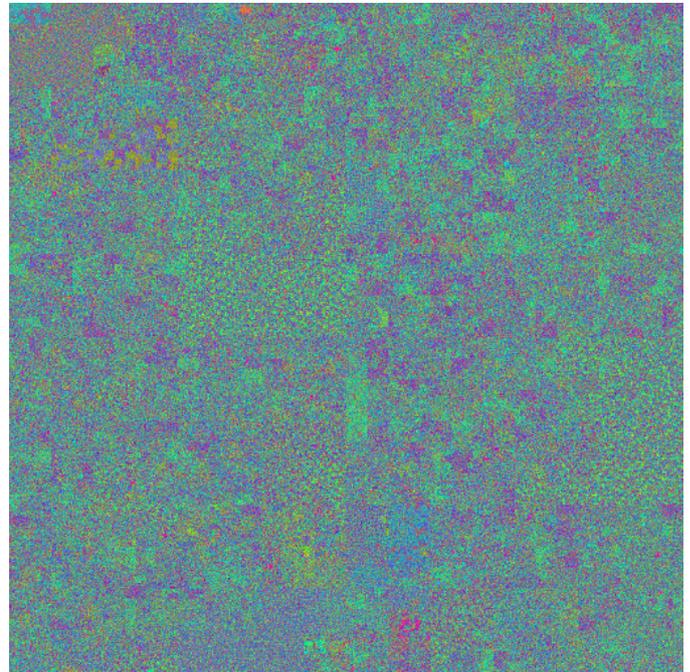
Final image pixel count is power of 4 (64, 256, 1024, 4096) etc with the highest resolution being 2048 x 2048 before Node experiences a call stack overflow (yeah, I broke node*),

* reproduce this condition with high values of m:

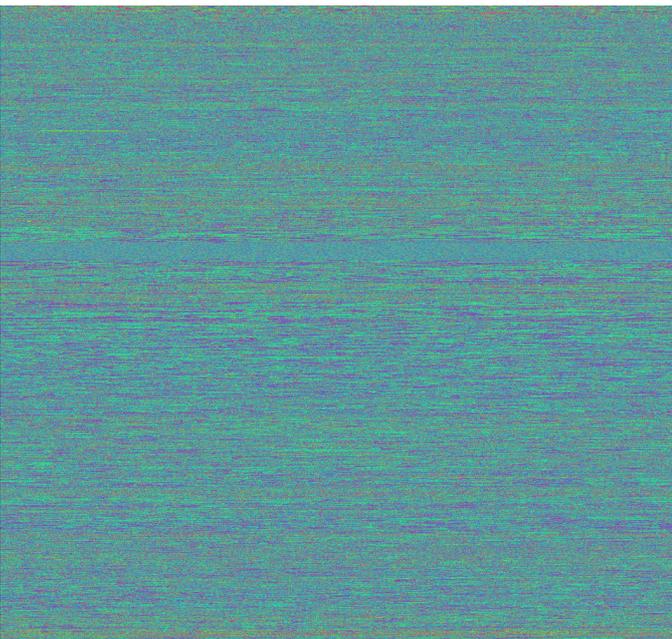
“aminosee -m 9 --test”



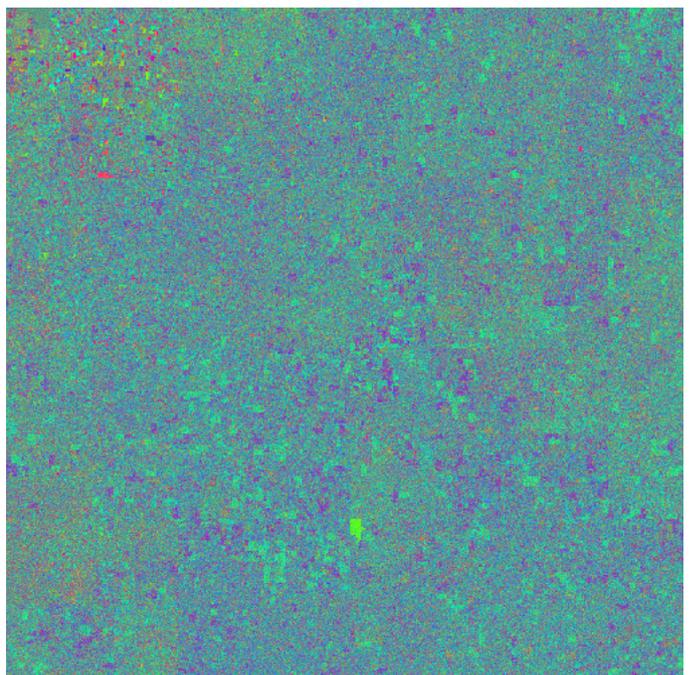
Linear map 960 × 723 pixels
Chimp Y chromosome
~10 DNA codons per pixel



Final Hilbert map 512 × 512 pixels
Possibly centromere is green in centre

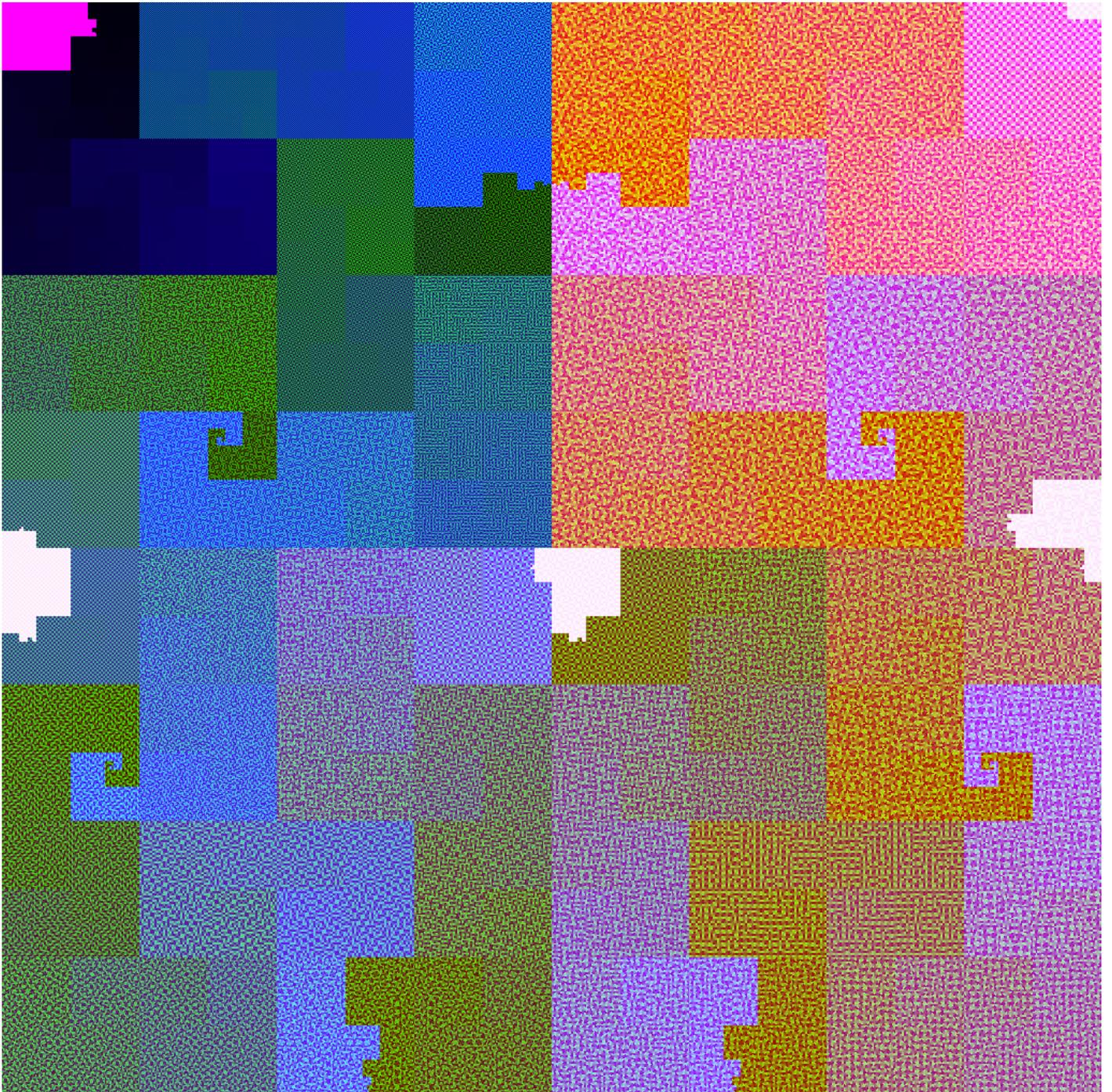


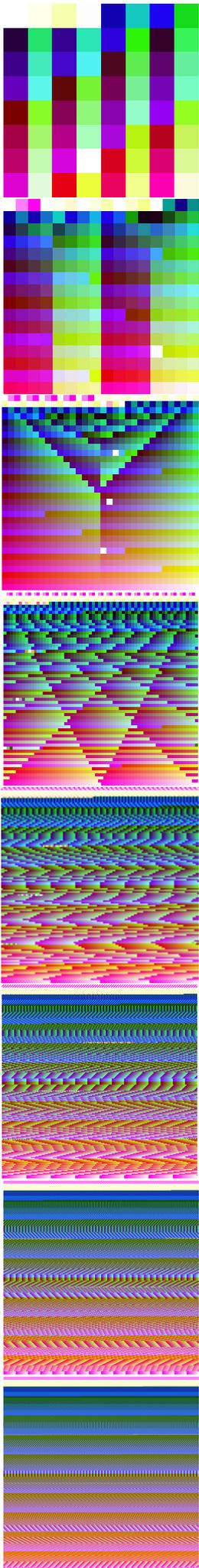
Linear map 960 × 1013 pixels
Chimp X chromosome
~50 DNA codons per pixel



Hilbert map 512 × 512 pixels
Chromosomal structures are much more clearly visible, retaining relative proximity during zoom

Test Pattern



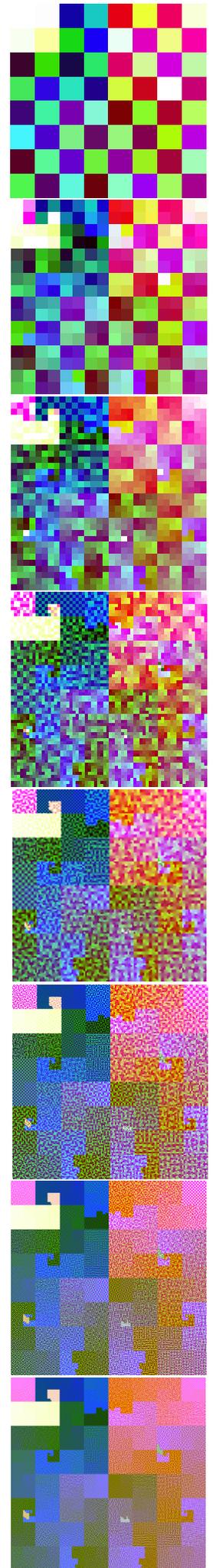
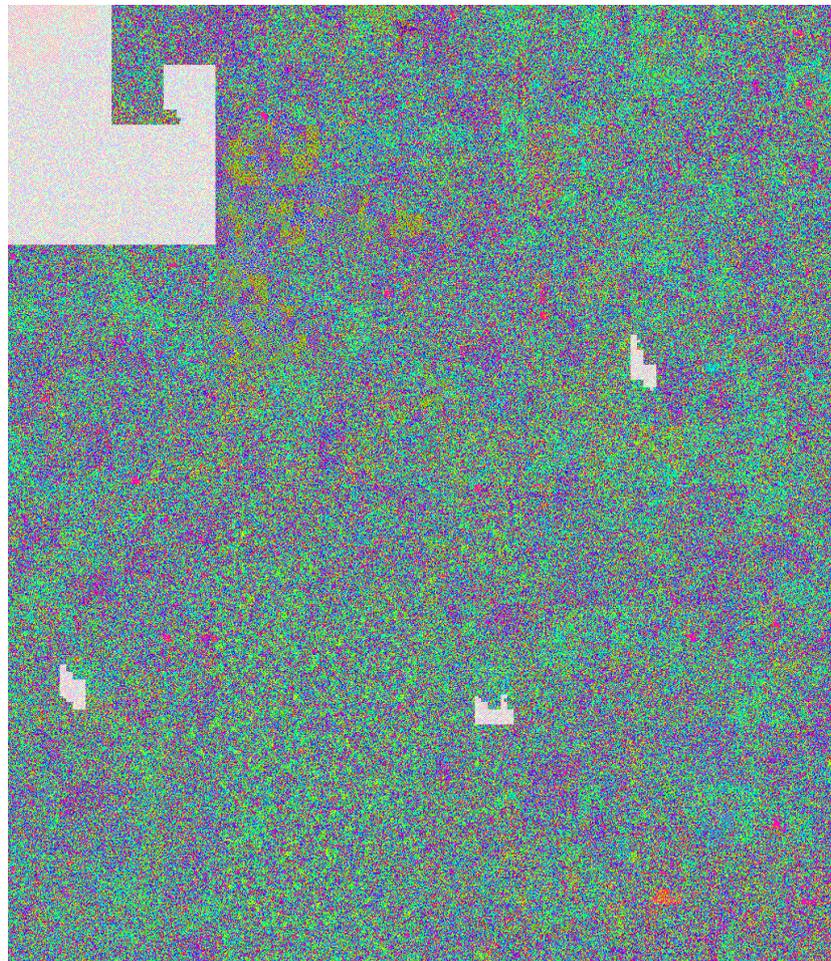


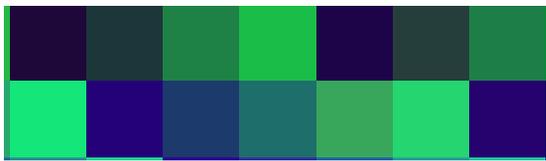
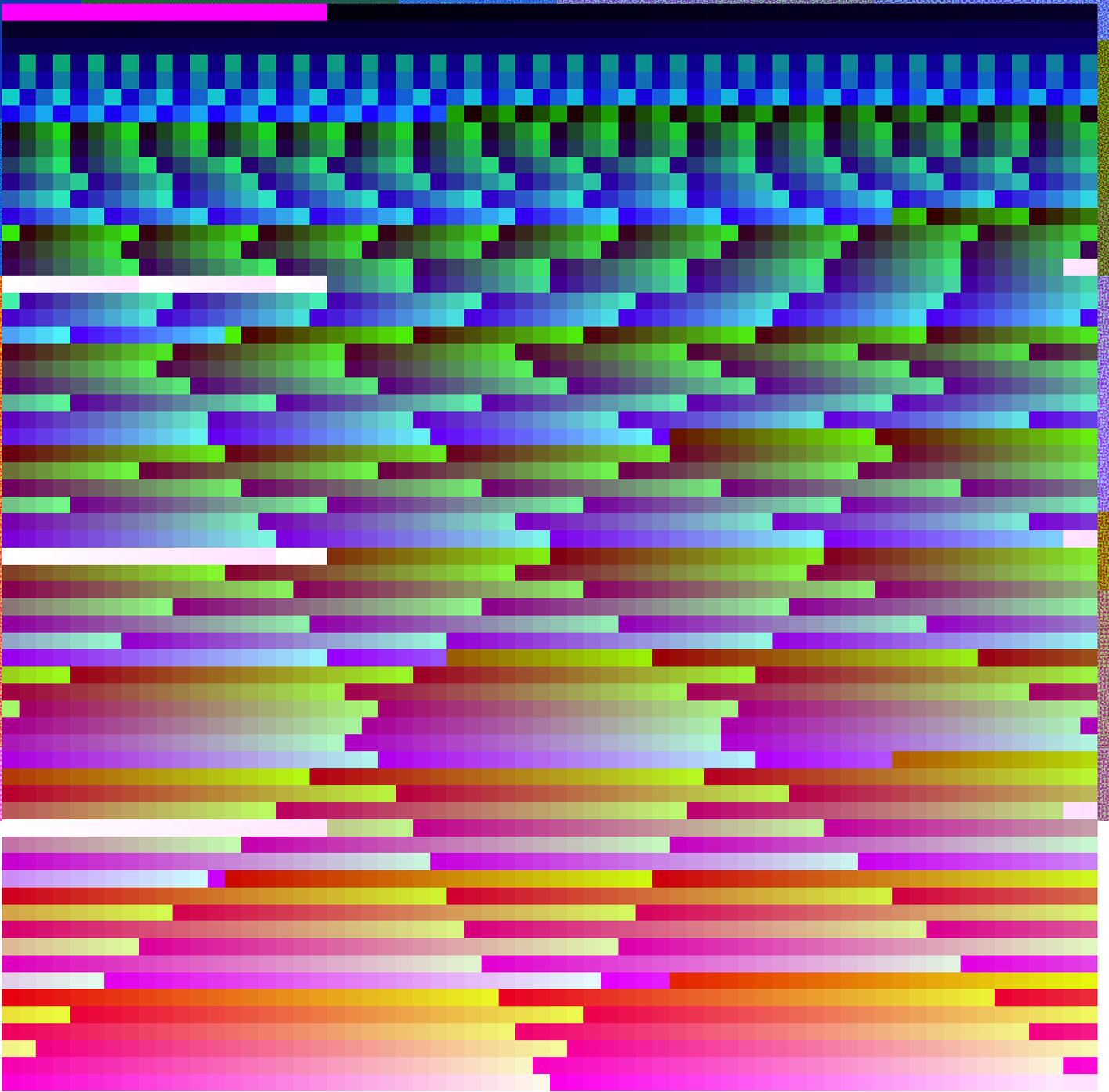
DNA The human genome is 3.2 billion base pairs of DNA coiled up in each cell, which if uncoiled would be 2 meters in length. But even if you did somehow organise it to lay it out onscreen, what order would you put the sequence and what resolution?

Both sets of test pattern images are compromised of the same sequence of colours, but notice how the set on the right has much more visual stability than those on the left?

By plotting the sequence using a Hilbert on the right, parts of the image that were previous close, remain close to their neighbours.

Chimp Chromosome Y "Clint" The white spiral registration marks can be disabled.



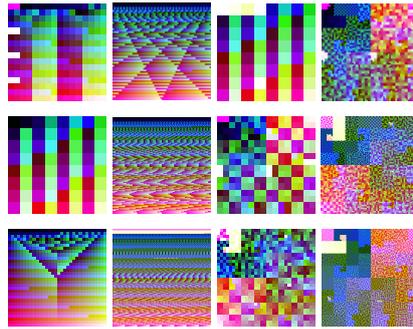


5 pixel ramp



32 pixel ramp

Hilbert Curve Calibration Images



The calibration sequence is 1-dimensional like DNA.

Slowly rising red over the course from 0% at top to 100% at bottom

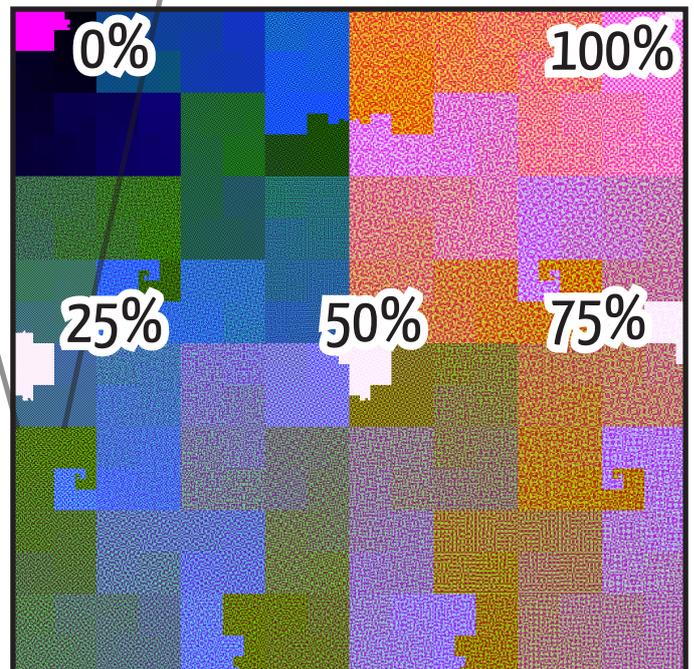
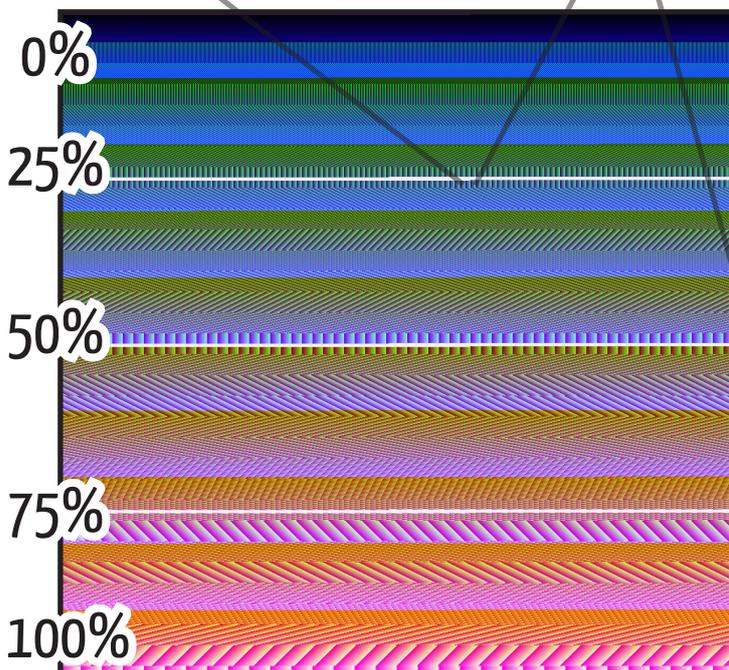
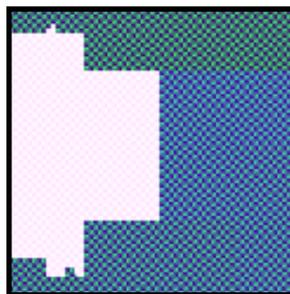
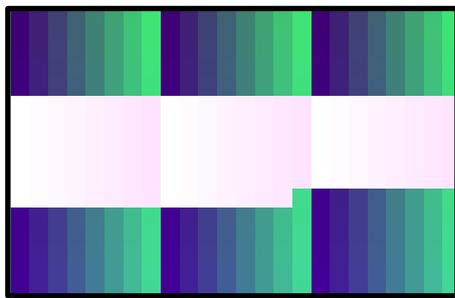
Blue rises from black to full 10 times peaking at 10%, 20%, 30%

Green steps through 32 patterns in sync with the 64 chess squares

4% of image is inverted at 1/4, 1/3, 3/4 positions

Final specification is still being finalised at this time.

- “aminosee --test” ;# output a series of 8 calibration images
- “aminosee --test --reg” ;# output images with registration marks
- “aminosee --test -m 9” ;# try to output 4096x4096 calibration breaking node



Entire Chimpanzee Genome

Eucalyptus



Octopus

Human Chromosome 2