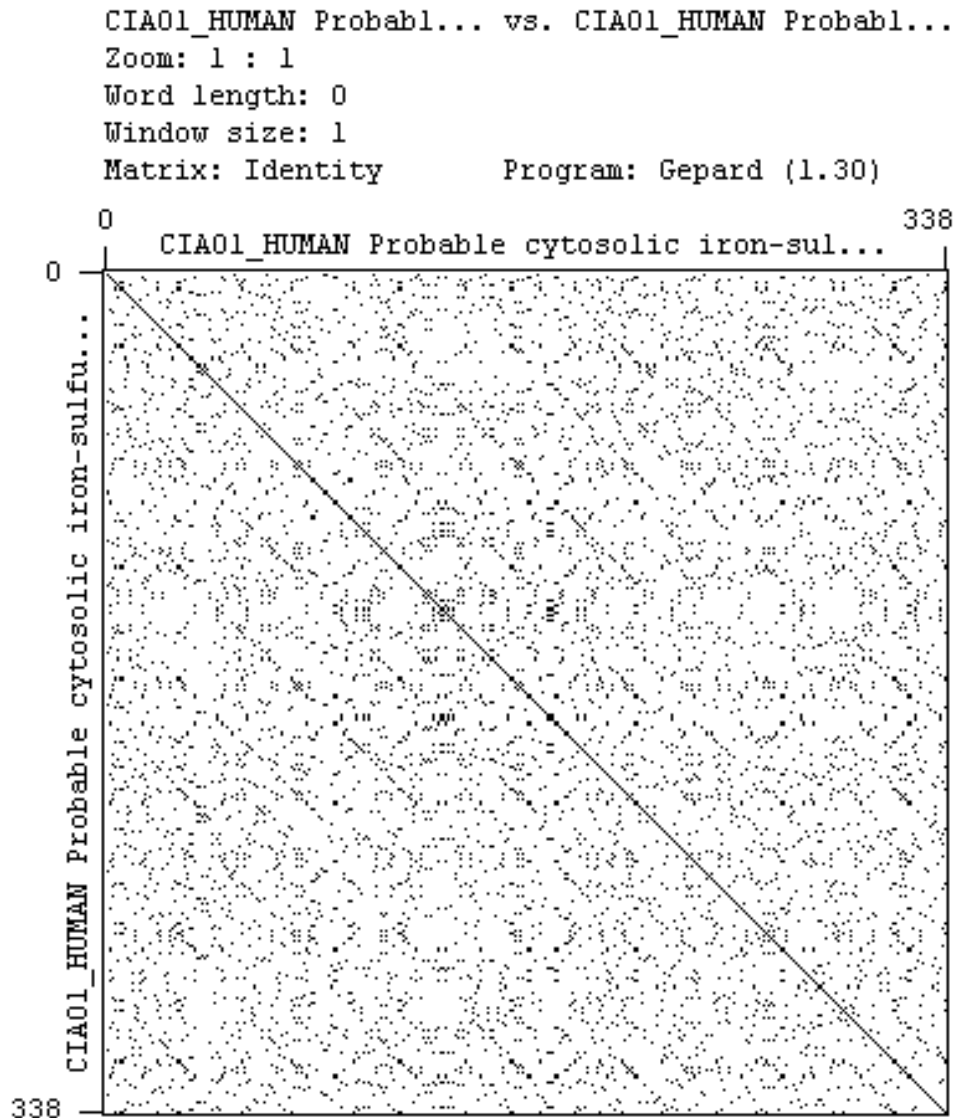


# Dotplots

Anders Krogh

# Simple dotplot

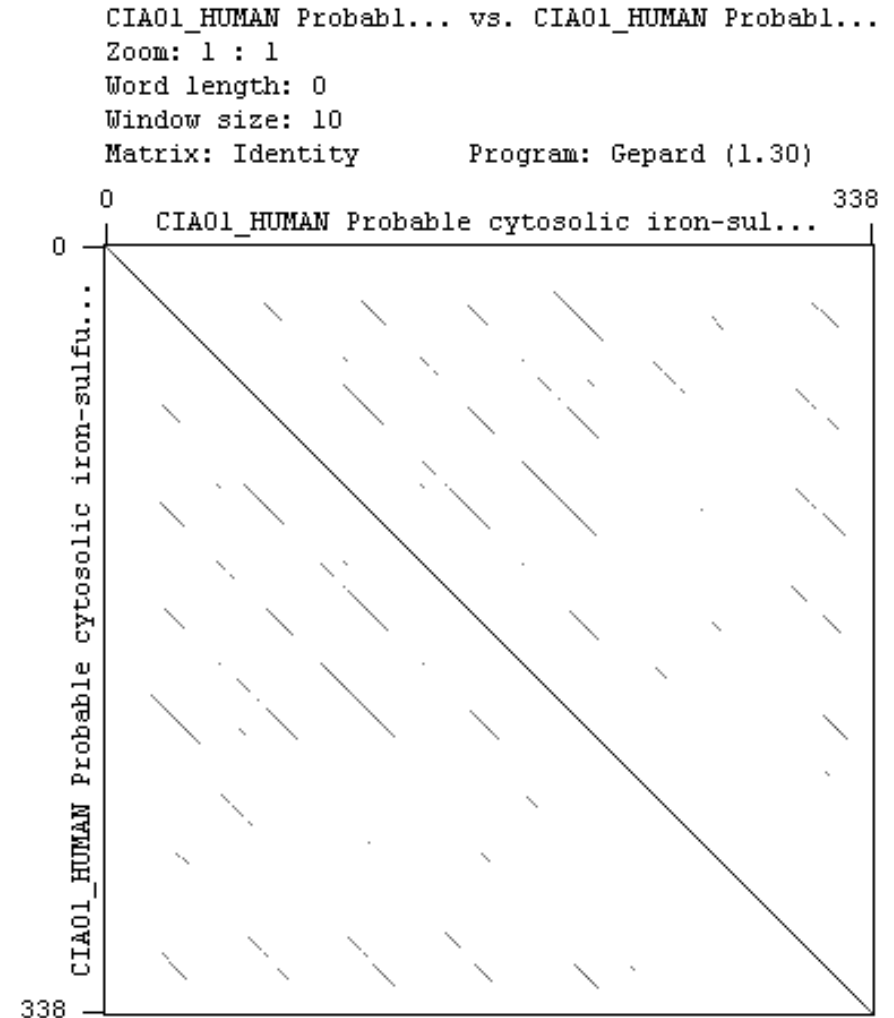


- Make a dot every time the two letters are identical
- This plot is made with gepard (like all the rest)

# Windowing

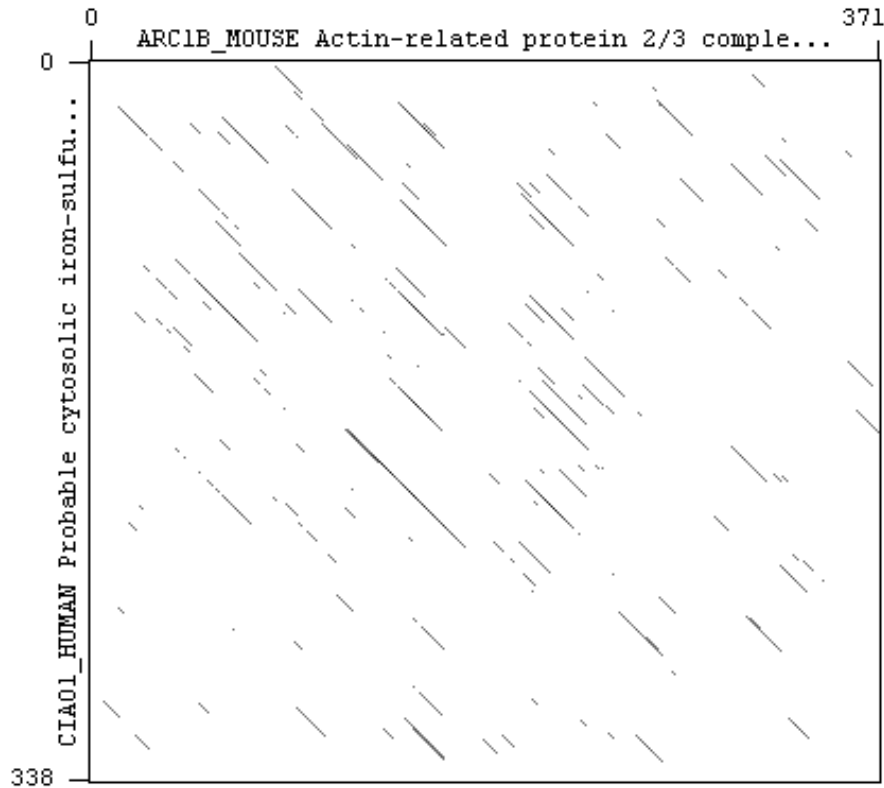
....WSPCGNYLASASFDATTCI....  
...WSPCGNYDLGSADFDDTTIWK..

- Add identical letters in window



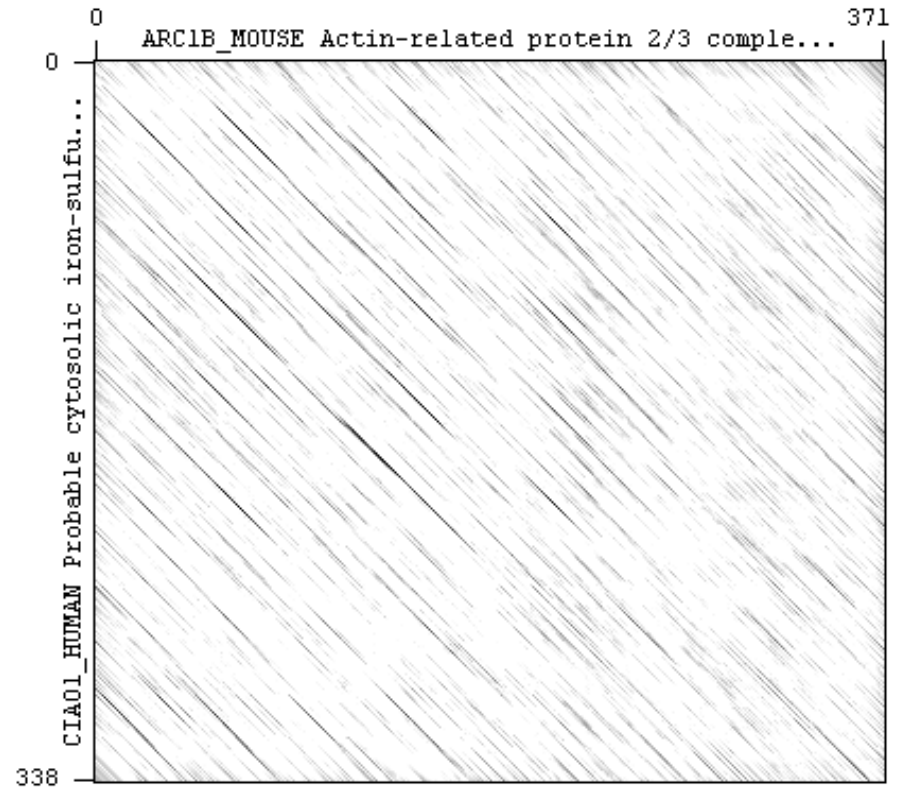
# Substitution matrix and gray scale

ARC1B\_MOUSE Actin-rel... vs. CIA01\_HUMAN Probable ...  
Zoom: 1 : 1  
Word length: 0  
Window size: 20  
Matrix: Identity      Program: Gepard (1.30)



Mouse protein vs human protein, win. 20

ARC1B\_MOUSE Actin-rel... vs. CIA01\_HUMAN Probable ...  
Zoom: 1 : 1  
Word length: 0  
Window size: 25  
Matrix: BLOSUM62      Program: Gepard (1.30)



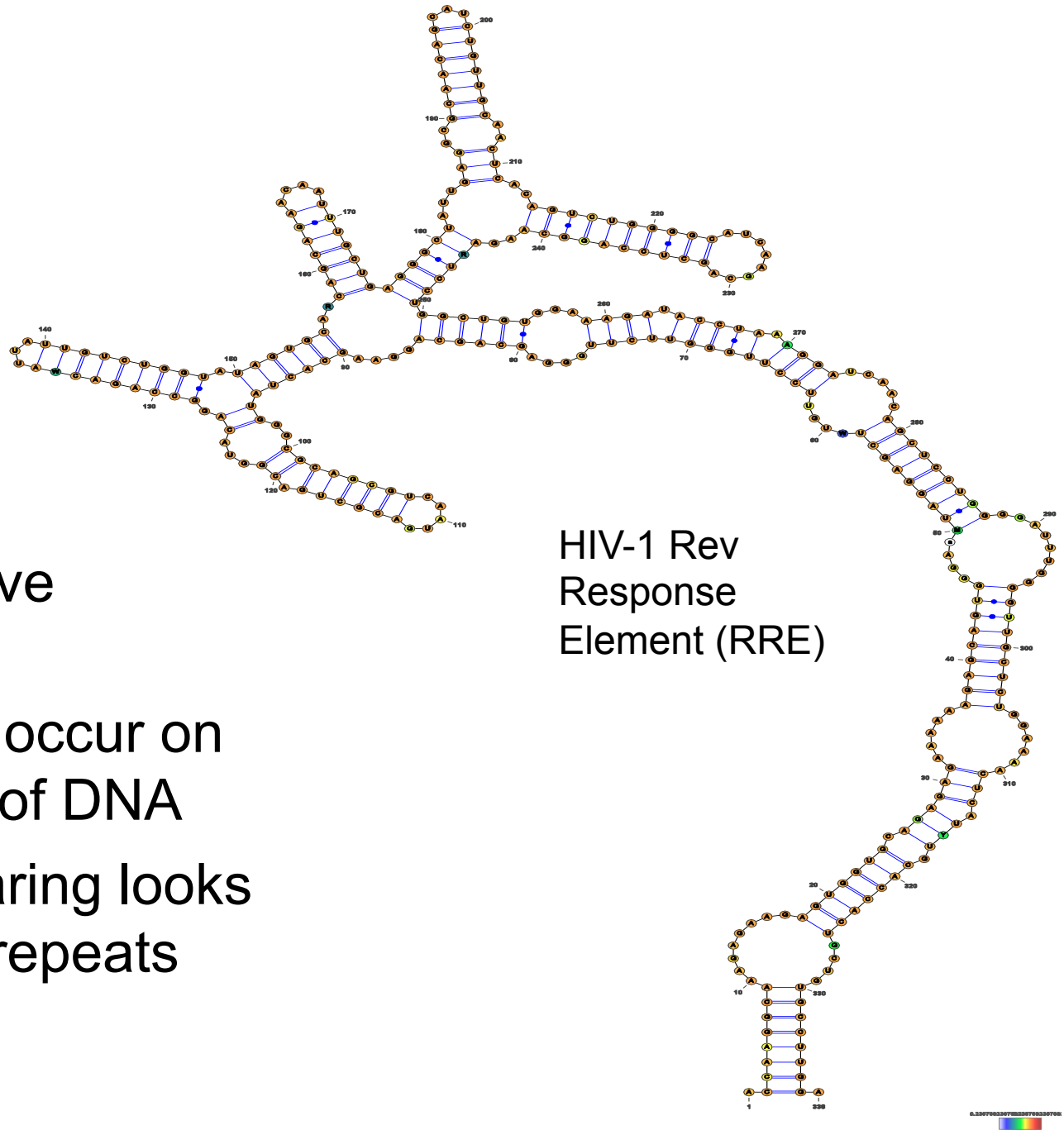
BLOSUM62 and win. size 25

# BLOSUM62 matrix

[illegible]

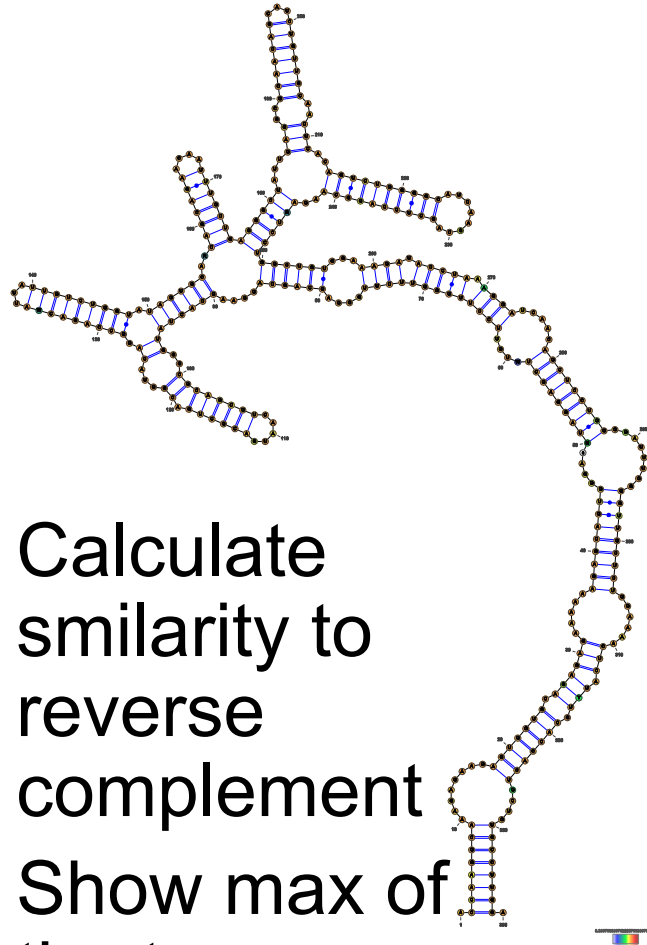
# Inverted repeats etc.

- Genomes have inversions
- Repeats can occur on both strands of DNA
- RNA base-pairing looks like inverted repeats

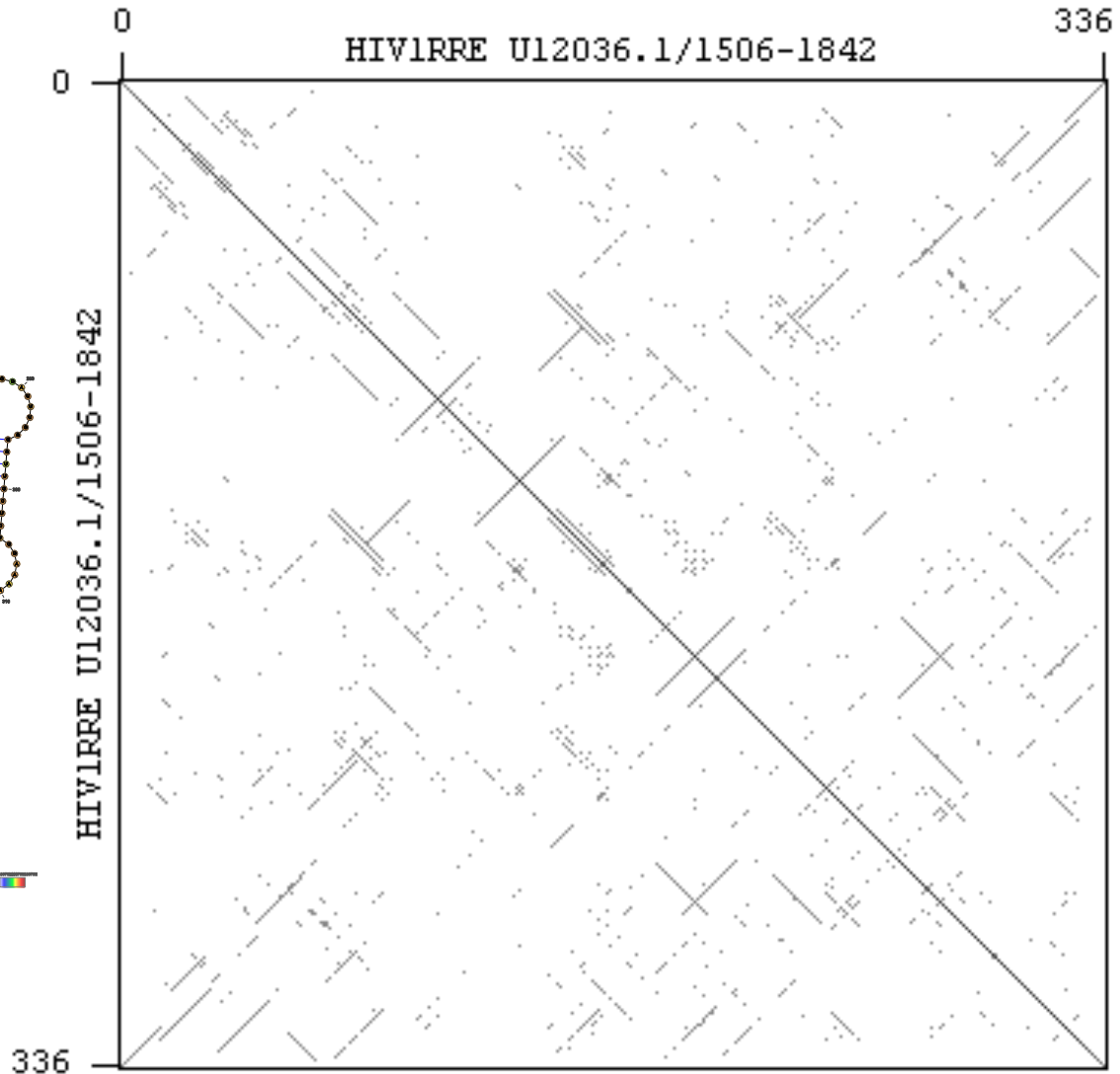


# RNA dotplot

HIV1RRE U12036.1/1... vs. HIV1RRE U12036.1/1...  
Zoom: 1 : 1  
Word length: 0  
Window size: 20  
Matrix: DNA  
GC ratio seq1: 0.5074  
GC ratio seq2: 0.5074  
Program: Gepard (1.30)



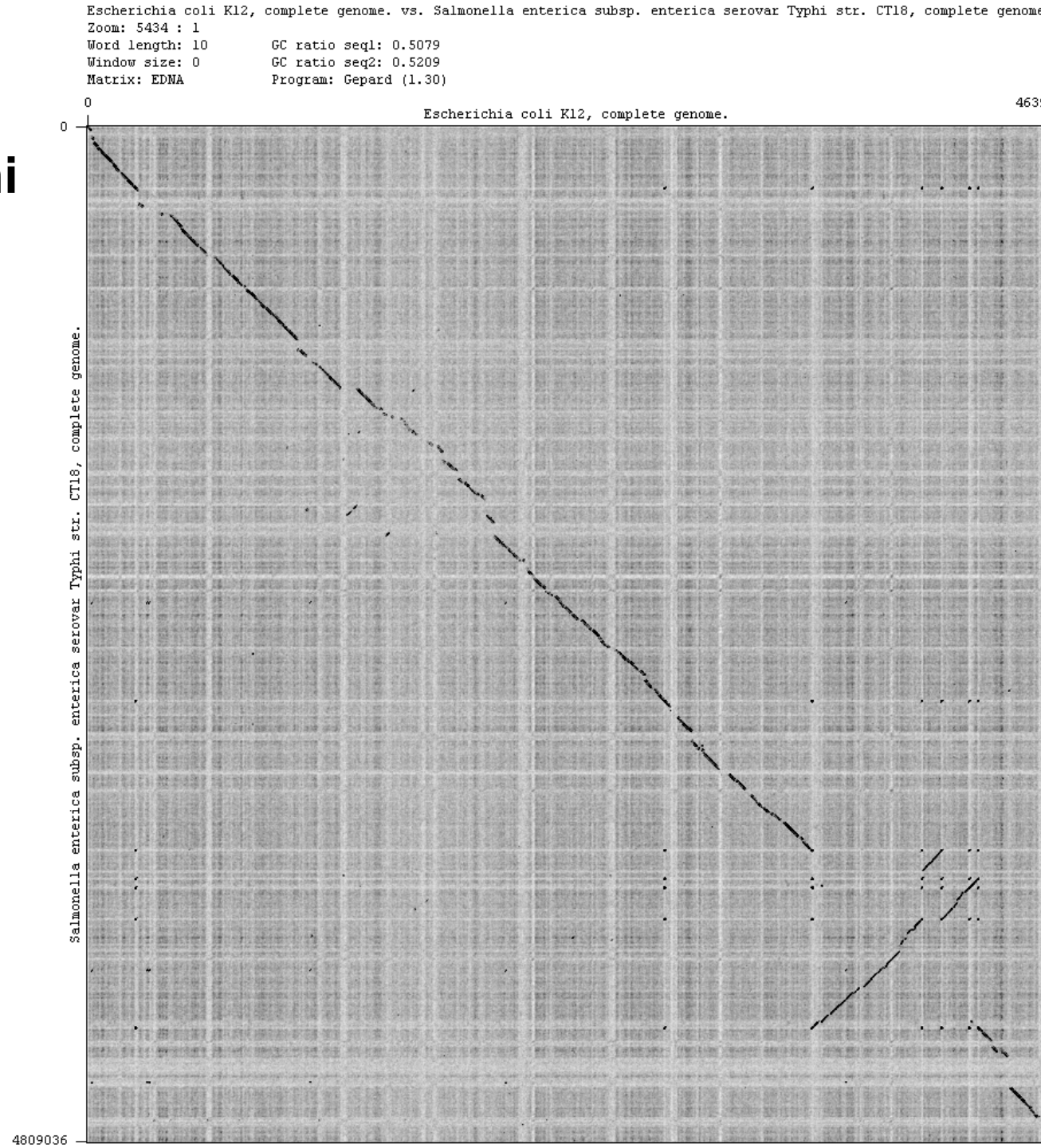
- Calculate similarity to reverse complement
- Show max of the two scores



# Coli vs Salmonella Typhi

The two are almost co-linear

One major region is inverted





# Computing dotplots

- $2 \times 10^{13}$  points in the genome comparison!
- Smart ways to make fast computations
- Dotter  
<http://sonnhammer.sbc.su.se/Dotter.html>  
is fast
- gepard <http://cube.univie.ac.at/gepard> is faster

# Pairwise alignment

# Examples

(a)

HBA_HUMAN	GSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKL
	G+ +VK+HGKKV A++++AH+D++ +++++LS+LH KL
HBB_HUMAN	GNPKVKAHGKKVLGAFSDGLAHLNFKGTFTLSELHCDKL

(b)

HBA_HUMAN	GSAQVKGHGKKVADALTNAVAHV---D---DMPNALSALSDLHAHKL
	++ ++++H+ KV + +A ++ +L+ L+++H+ K
LGB2_LUPLU	NNPELQAHAGKVFCLKVYEAAIQLQVTGVVVTDATLKNLGSVHVSKG

(c)

HBA_HUMAN	GSAQVKGHGKKVADALTNAVAHVDDMPNALSALSD----LHAHKL
	GS+ + G + +D L ++ H+ D+ A +AL D ++AH+
F11G11.2	GSGYLVGDSLTFVDLL--VAQHTADLLAANAALLDEFPPQFKAHQE

# BLOSUM62 – again

[illegible]