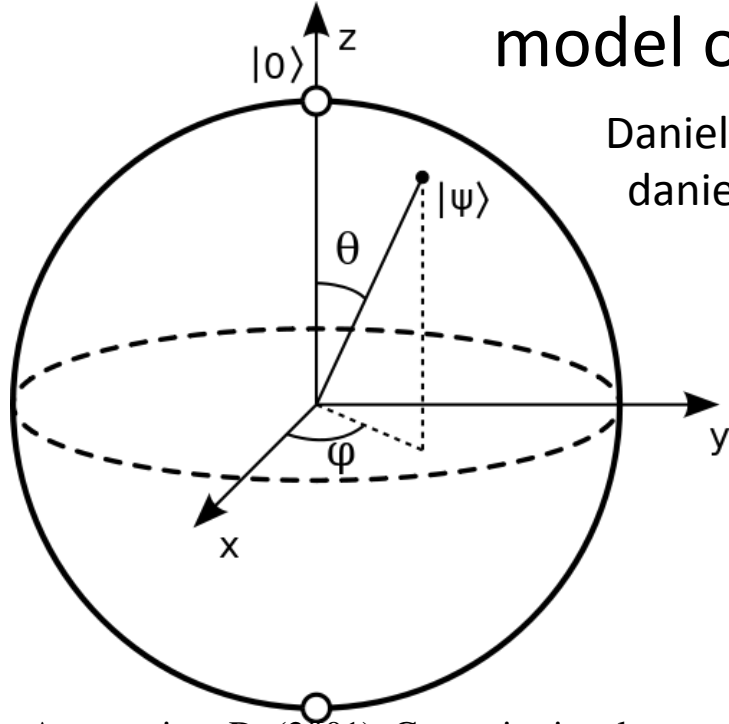
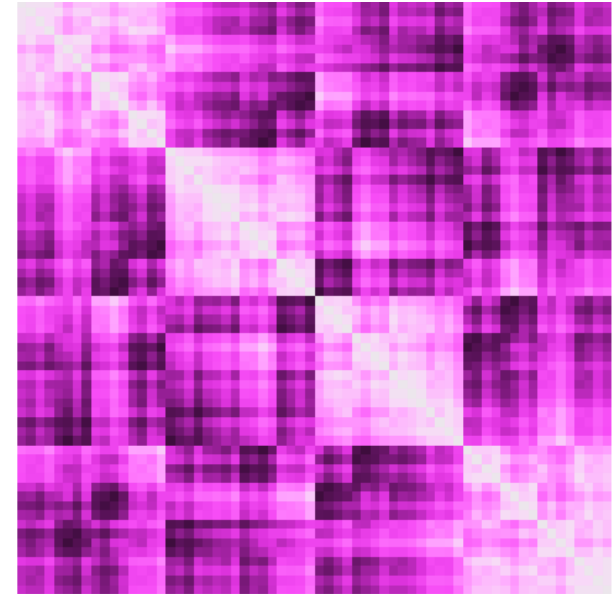


Genetic Phase Angle Distance

computing pairwise codon distance using a 3-dimensional model of the genetic code.

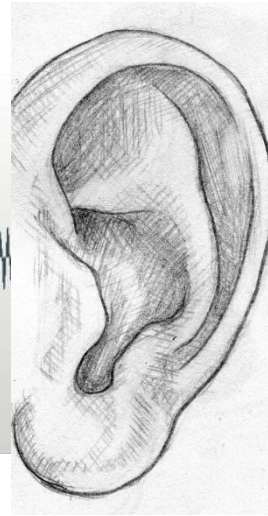
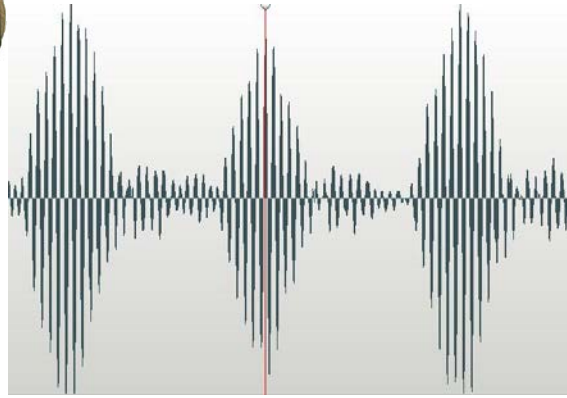


Daniel K. Pratt, 15 April 2013
daniel.k.pratt@gmail.com



- Anastassiou, D. (2001). Genomic signal processing. *IEEE Signal Processing Magazine*, 18(4), 8–20.
- Aragon-Camarasa, G., Aragon-Gonzalez, G., Aragon, J. L., & Rodriguez-Andrade, M. A. (2008). Clifford Algebra with Mathematica. Retrieved from <http://arxiv.org/pdf/0810.2412>
- Cristea, P. D. (2005). Representation and analysis of DNA sequences. In E. R. Dougherty (Ed.), *EURASIP book series on signal processing and communications v. 2. Genomic signal processing and statistics*. New York, N.Y: Hindawi Pub. Corp.
- Zhang, H., Zhu, C., Peng, Q., & Chen, J. (2006). Using geometric algebra for 3D linear transformations. *Computing in Science & Engineering*, 8(3), 68–75.

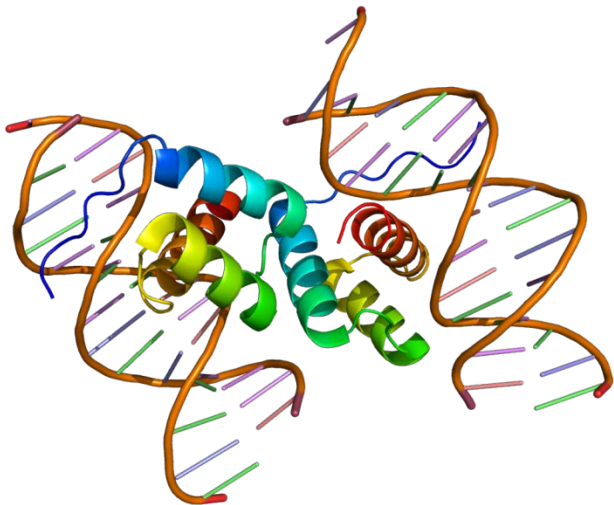
“imaginary space”



Linear
Form

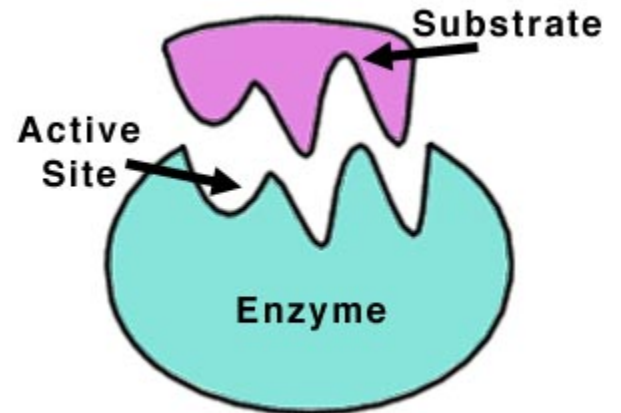
nonsequential, simultaneous interaction of **ALL** components

NonLinear
Function



???

“imaginary space”
(Anastassiou, 2001)



Nucleotides are mapped into a complex vector space,
According to three biochemical properties.

e1 axis: 2/3-hydrogen bonds

e2 axis: Amino/Keto

e3 axis: Purine/Pyrimidine

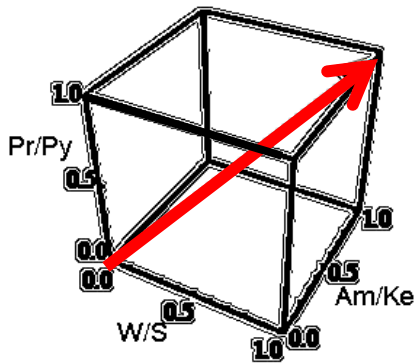
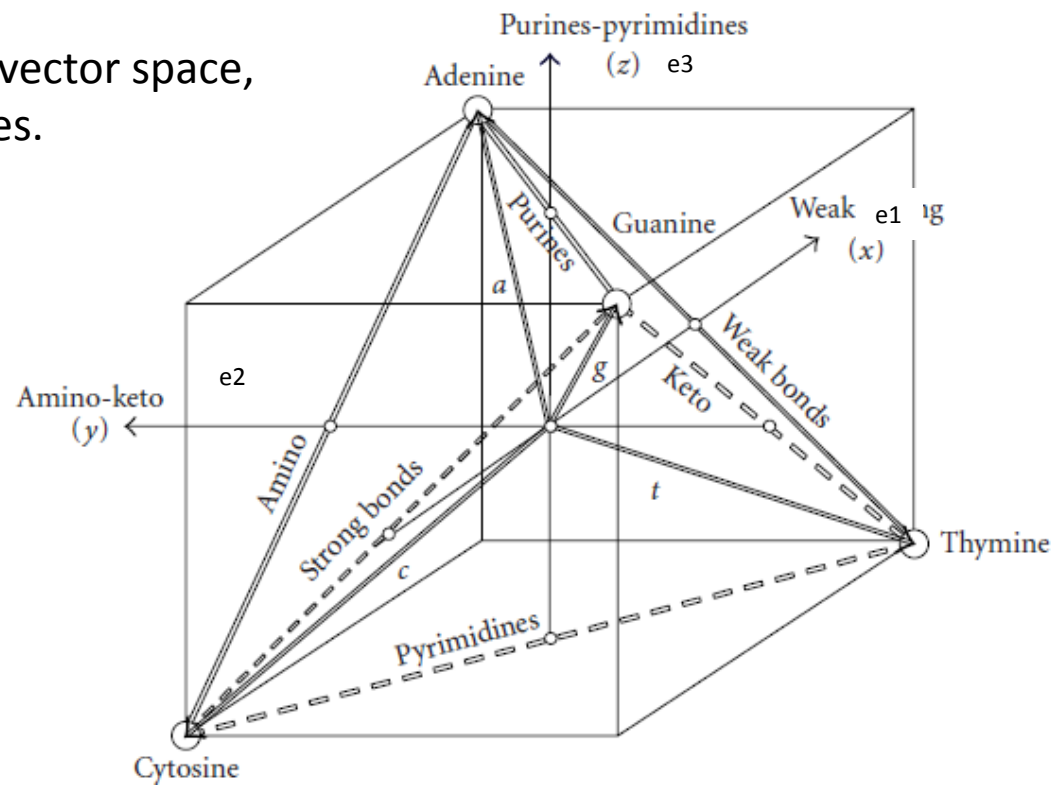
From (Cristea, 2005)

$$\vec{a} = \vec{i} + \vec{j} + \vec{k},$$

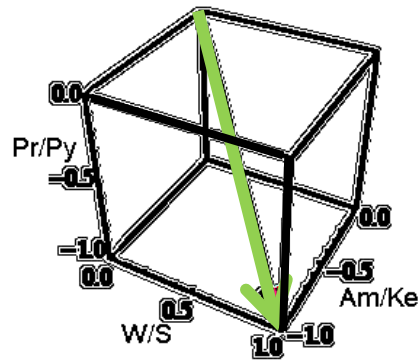
$$\vec{c} = -\vec{i} + \vec{j} - \vec{k},$$

$$\vec{g} = -\vec{i} - \vec{j} + \vec{k},$$

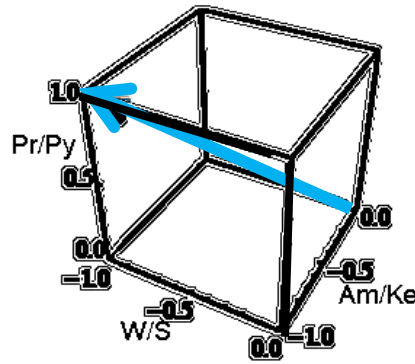
$$\vec{t} = \vec{i} - \vec{j} - \vec{k}.$$



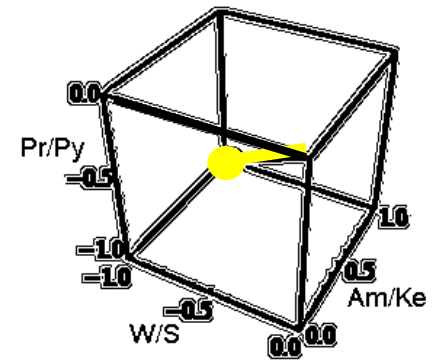
$$A1 = e1 + e2 + e3$$



$$T1 = e1 - e2 - e3$$



$$G1 = -e1 - e2 + e3$$

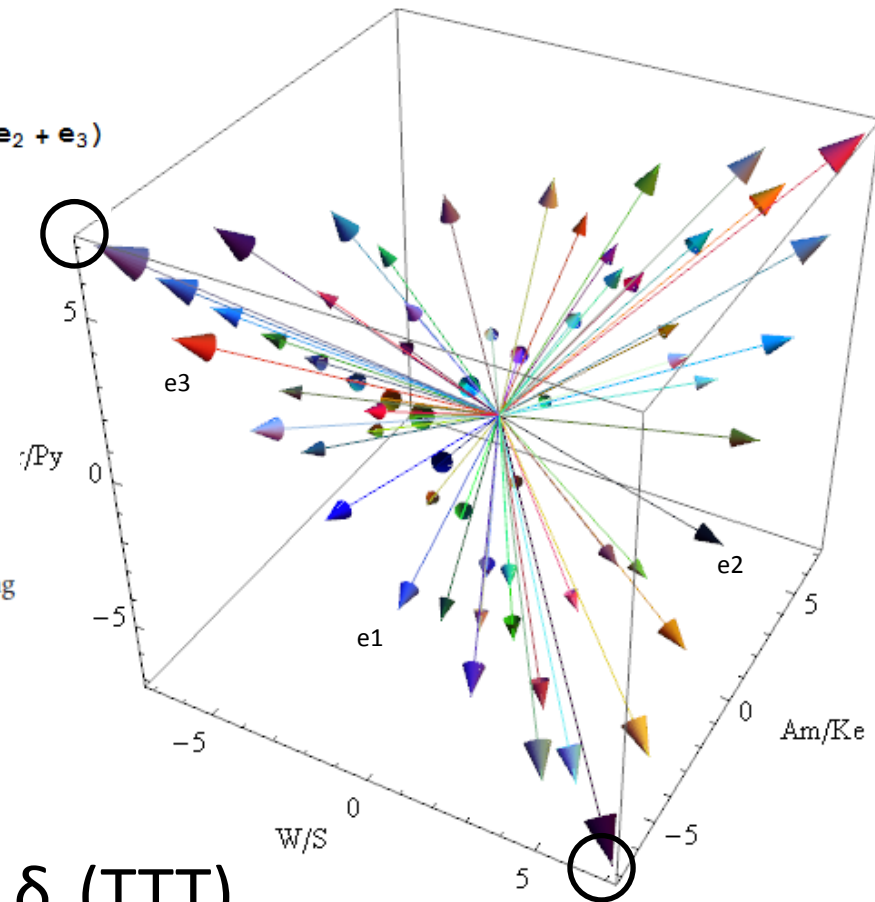
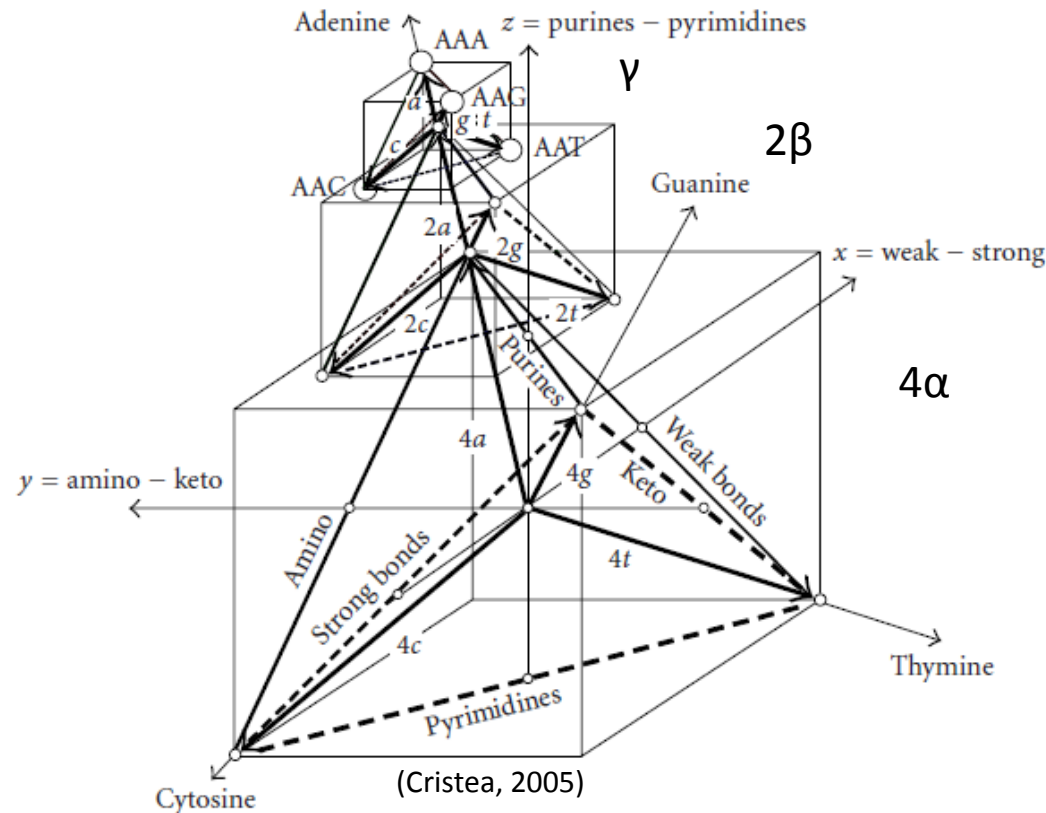


$$C1 = -e1 + e2 - e3$$

As represented in Mathematica using 'clifford.m' (Aragon, 2008; Zhang, 2006);
Constructed by (Pratt, 2013)

64 codons map to 64 unique vector positions, resulting in a tetrahedral genetic code.
GENETIC CODE MAPPING: $\delta_n(\text{codon}) = 4\alpha + 2\beta + \gamma$; where α =first, β =second, γ =third

$$\begin{aligned} &= 4(\mathbf{G}) + 2(\mathbf{G}) + \mathbf{G} \\ \delta_1(\mathbf{GGG}) &= 4(-\mathbf{e}_1 - \mathbf{e}_2 + \mathbf{e}_3) + 2(-\mathbf{e}_1 - \mathbf{e}_2 + \mathbf{e}_3) + (-\mathbf{e}_1 - \mathbf{e}_2 + \mathbf{e}_3) \\ &= -7\mathbf{e}_1 - 7\mathbf{e}_2 + 7\mathbf{e}_3 \end{aligned}$$

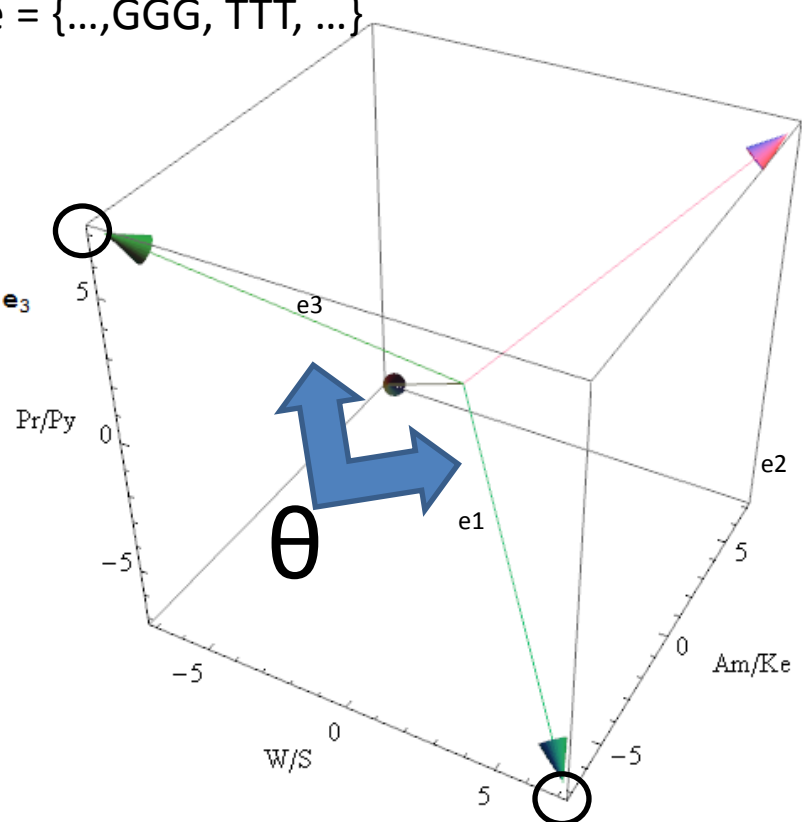


$$\begin{aligned} \delta_2(\mathbf{TTT}) &= 4(\mathbf{T}) + 2(\mathbf{T}) + \mathbf{T} \\ &= 4(\mathbf{e}_1 - \mathbf{e}_2 - \mathbf{e}_3) + 2(\mathbf{e}_1 - \mathbf{e}_2 - \mathbf{e}_3) + (\mathbf{e}_1 - \mathbf{e}_2 - \mathbf{e}_3) \\ &= 7\mathbf{e}_1 - 7\mathbf{e}_2 - 7\mathbf{e}_3 \end{aligned}$$

Sequence = {...,GGG, TTT, ...}

$\delta_1(\text{GGG})$

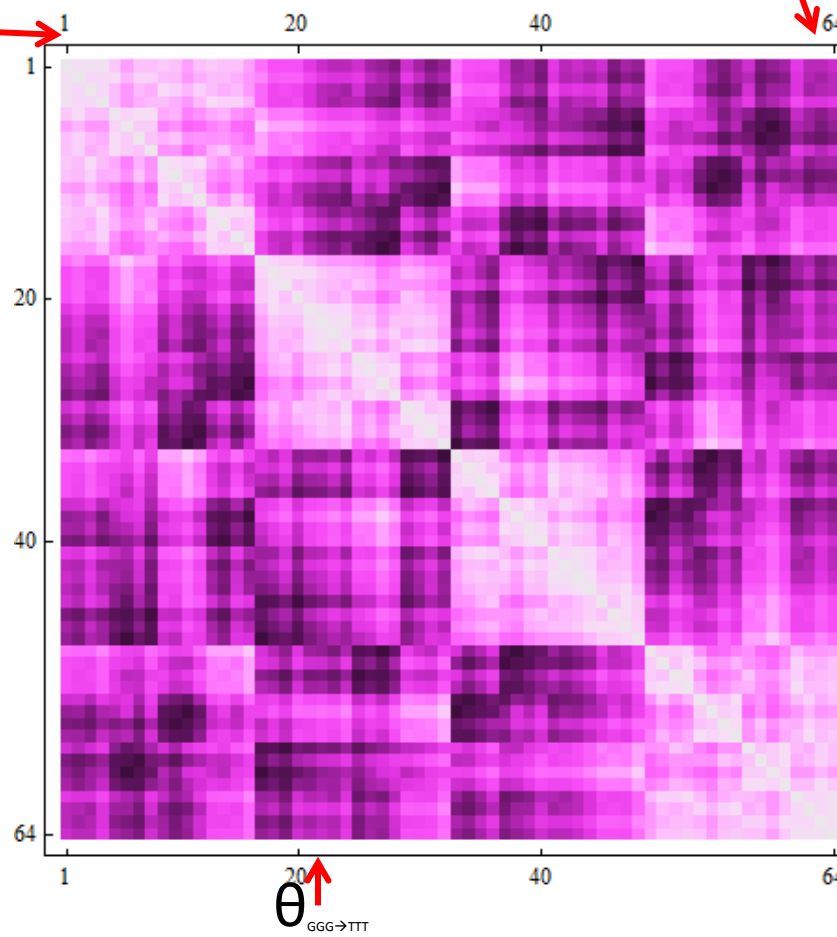
$= -7 \mathbf{e}_1 - 7 \mathbf{e}_2 + 7 \mathbf{e}_3$



$\delta_2(\text{TTT}) = 7 \mathbf{e}_1 - 7 \mathbf{e}_2 - 7 \mathbf{e}_3$

θ

- = VectorAngle[δ_1, δ_2]
- = VectorAngle[GGG, TTT]
- = VectorAngle[$-7 \mathbf{e}_1 - 7 \mathbf{e}_2 + 7 \mathbf{e}_3, 7 \mathbf{e}_1 - 7 \mathbf{e}_2 - 7 \mathbf{e}_3$]
- = ArcCos[$-\frac{1}{3}$] = 1.9106332362490186\`



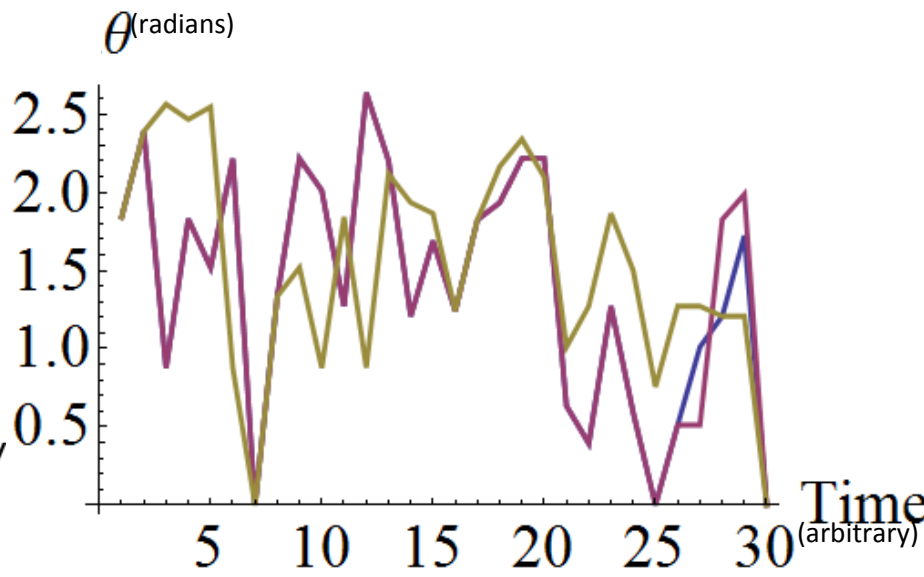
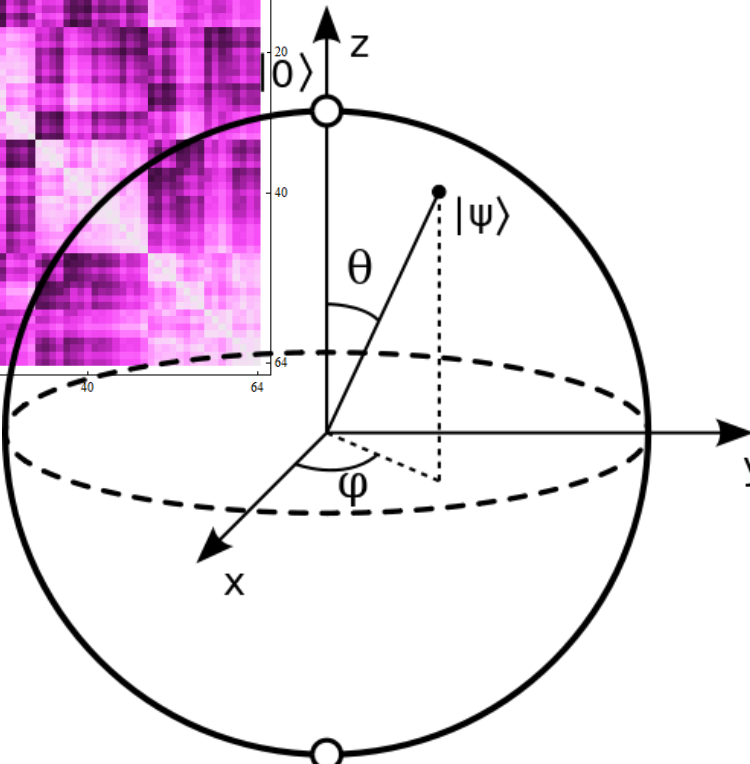
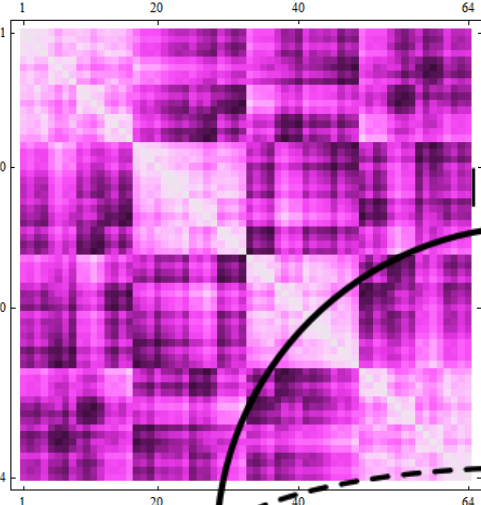
A matrix of all possible values of θ (small $\theta \rightarrow$ light, large $\theta \rightarrow$ dark).

(Pratt, 2013)

Phase distance θ is measured as the angle between sequentially neighboring codon vectors.

The coding sequences of the first exon of β -globin gene of eleven different species.

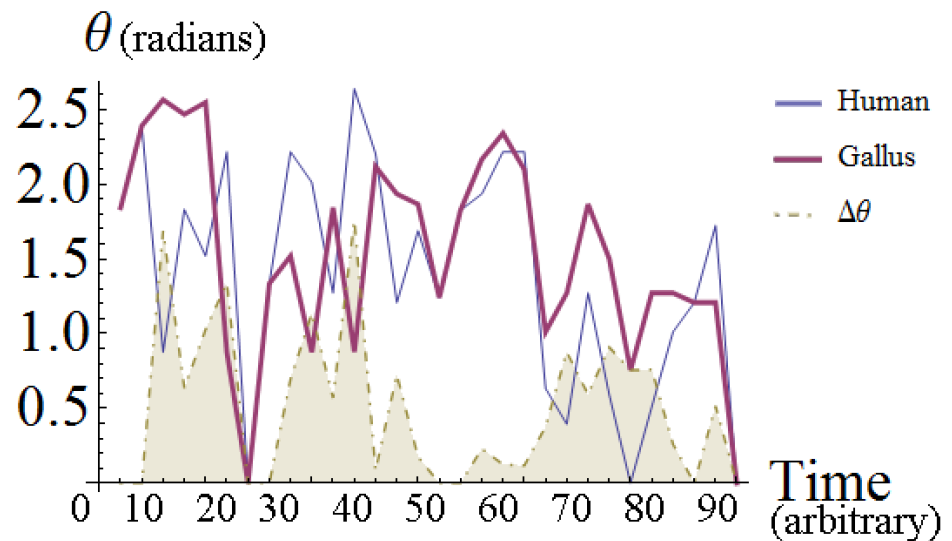
Species	Coding sequence (N=30)	Alignment made in MEGA5 software using MUSCLE (codon) algorithm
1. 1-Human	ATGGTGCACCTGACTTCCTGAGGAGAAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAAACGTGGATGAAATTTGGTGGTGAAGGCCCTGGGCAG	
2. 2-Chimpanzee	ATGGTGCACCTGACTTCCTGAGGAGAAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAAACGTGGATGAAATTTGGTGGTGAAGGCCCTGGGCAG	
3. 3-Goat	ATG-----CTGACTGCTGAGGAGAAAGGCTGCCGTTACTGCCCTGTGGGGCAAGGTGAAAGTGGATGAAATTTGGTGGTGAAGGCCCTGGGCAG	
4. 4-Bovine	ATG-----CTGACTGCTGAGGAGAAAGGCTGCCGTTACTGCCCTTTTGGGGCAAGGTGAAAGTGGATGAAATTTGGTGGTGAAGGCCCTGGGCAG	
5. 5-Gallus	ATGGTGCACCTGGAATTCCTGAGGAGAAAGCTTCATCACCGGCCCTCTGGGGCAAGGTCAATGTGGCCGAAATGTGGGGCCGAAAGCCCTGGCC--	
6. 6-Mouse	ATGGTGCACCTGACTGATGCTGAGAAAGGCTGCTGCTCTGTCCTGTGGGGAAAGGTGAACTCCGATGAAATTTGGTGGTGAAGGCCCTGGGCAG	
7. 7-Rat	ATGGTGCACCTAACTGATGCTGAGAAAGGCTACTGTTAGTGGCCCTGTGGGGAAAGGTGAAACCCTGATTAATGTGGCCCTGAGAGGCCCTGGGC--	
8. 8-Gorilla	ATGGTGCACCTGACTTCCTGAGGAGAAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAAACGTGGATGAAATTTGGTGGTGAAGGCCCTGGGCAG	
9. 9-Rabbit	ATGGTGCATCTGTTCCAGTGAAGGAGAAAGTCTGCCGGTCACTGCCCTGTGGGGCAAGGTGAAATGTGGAAAGAAATTTGGTGGTGAAGGCCCTGGGC--	
10. 10-Opossum	ATGGTGCACCTGACTTTCCTGAGGAGAAAGAACTGCAATCACCTACCAATCTGGTCTAAGGTTGACCAAGGACTGGTGGTGAAGGCCCTTGGCAG	
11. 11-Lemur	ATGACTTTTGCCTGAGTGCCTGAGGAGAAATGCTCAATGTCACCTCTCTGTTGGGGCAAGGTGGATGTAGAGAAAGTTTGGTGGCGAGAGGCCCTGGGCAG	



Phase angle θ sequence maps for species: Human (blue), Chimpanzee (red), and Gallus (gold)

Angular difference $\Delta\theta$ is computed as the absolute value of the difference in phase θ between species under investigation, at equivalent (aligned) genetic loci.

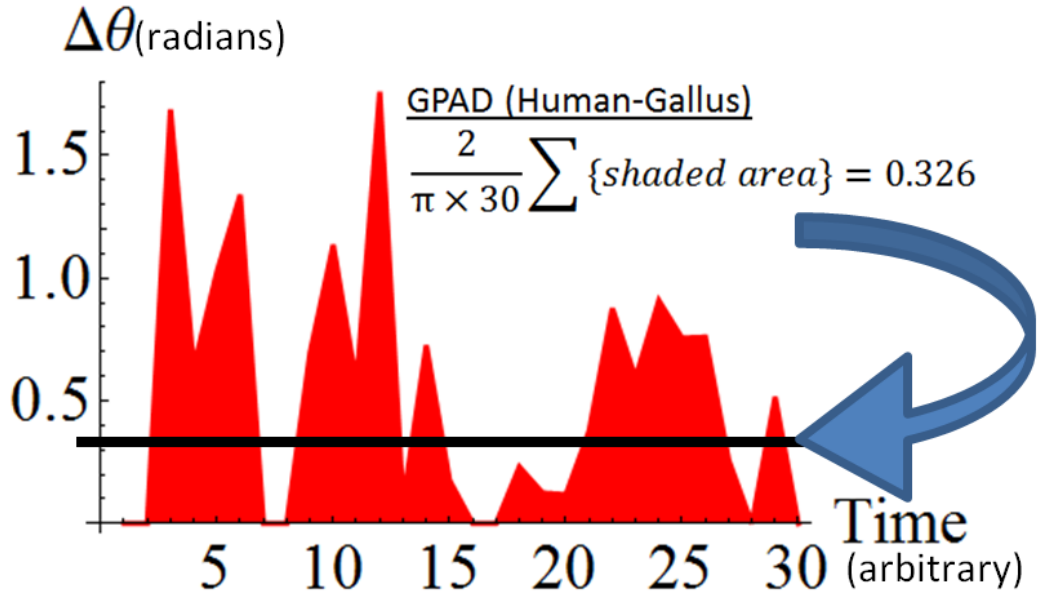
shaded : $\Delta\theta_{HG} = | \Delta\theta(\text{Human}) - \Delta\theta(\text{Gallus}) |$

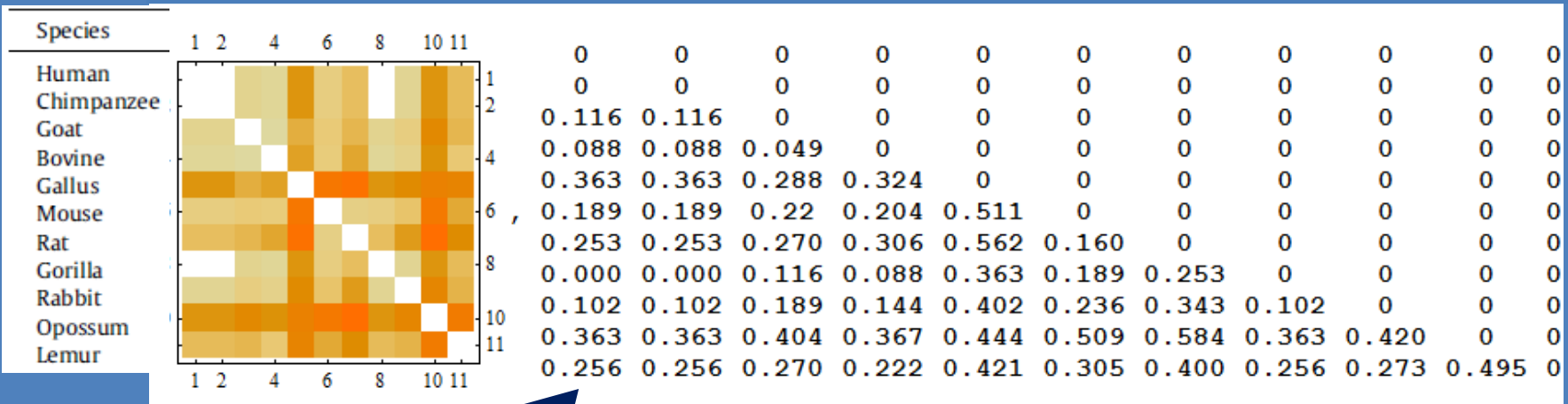
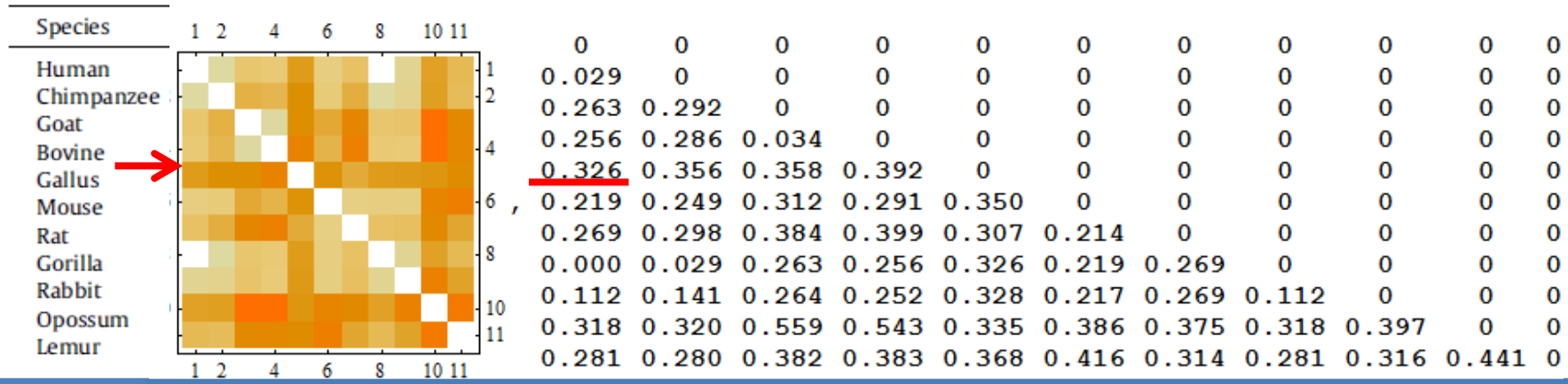


Genetic Phase Angle Distance

$$\frac{2}{\pi \times N} \sum \{ \Delta\theta_n, \dots, \Delta\theta_{N-1}, \Delta\theta_N \}$$

(Pratt,2013)



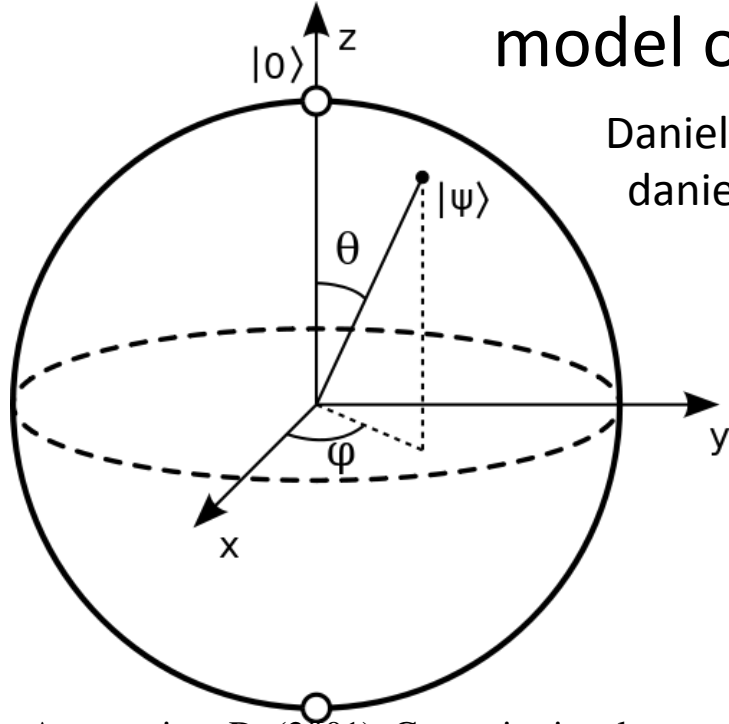


	1	2	3	4	5	6	7	8	9	10	11
1. 1-Human		0.000	0.040	0.034	0.085	0.054	0.063	0.000	0.038	0.087	0.068
2. 2-Chimpanzee	0.000		0.040	0.034	0.085	0.054	0.063	0.000	0.038	0.087	0.068
3. 3-Goat	0.116	0.116		0.024	0.075	0.058	0.068	0.040	0.051	0.094	0.071
4. 4-Bovine	0.088	0.088	0.049		0.079	0.054	0.074	0.034	0.044	0.087	0.060
5. 5-Gallus	0.363	0.363	0.288	0.324		0.109	0.125	0.085	0.090	0.099	0.089
6. 6-Mouse	0.189	0.189	0.220	0.204	0.511		0.050	0.054	0.063	0.107	0.068
7. 7-Rat	0.253	0.253	0.270	0.306	0.562	0.160		0.063	0.084	0.122	0.088
8. 8-Gorilla	0.000	0.000	0.116	0.088	0.363	0.189	0.253		0.038	0.087	0.068
9. 9-Rabbit	0.102	0.102	0.189	0.144	0.402	0.236	0.343	0.102		0.094	0.069
10. 10-Opossum	0.363	0.363	0.404	0.367	0.444	0.509	0.584	0.363	0.420		0.112
11. 11-Lemur	0.256	0.256	0.270	0.222	0.421	0.305	0.400	0.256	0.273	0.495	

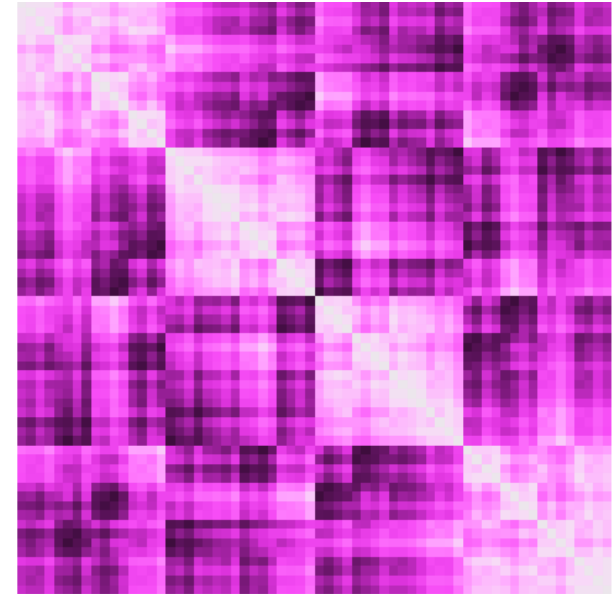
For comparison, Tamura-3 distance measure computed using MEGA5

Genetic Phase Angle Distance

computing pairwise codon distance using a 3-dimensional model of the genetic code.



Daniel K. Pratt, 15 April 2013
daniel.k.pratt@gmail.com



- Anastassiou, D. (2001). Genomic signal processing. *IEEE Signal Processing Magazine*, 18(4), 8–20.
- Aragon-Camarasa, G., Aragon-Gonzalez, G., Aragon, J. L., & Rodriguez-Andrade, M. A. (2008). Clifford Algebra with Mathematica. Retrieved from <http://arxiv.org/pdf/0810.2412>
- Cristea, P. D. (2005). Representation and analysis of DNA sequences. In E. R. Dougherty (Ed.), *EURASIP book series on signal processing and communications v. 2. Genomic signal processing and statistics*. New York, N.Y: Hindawi Pub. Corp.
- Zhang, H., Zhu, C., Peng, Q., & Chen, J. (2006). Using geometric algebra for 3D linear transformations. *Computing in Science & Engineering*, 8(3), 68–75.