

A

k-mer sequence

A G C C T

T G C G A

Binary Encode

0 2 1 1 3

3 2 1 2 0

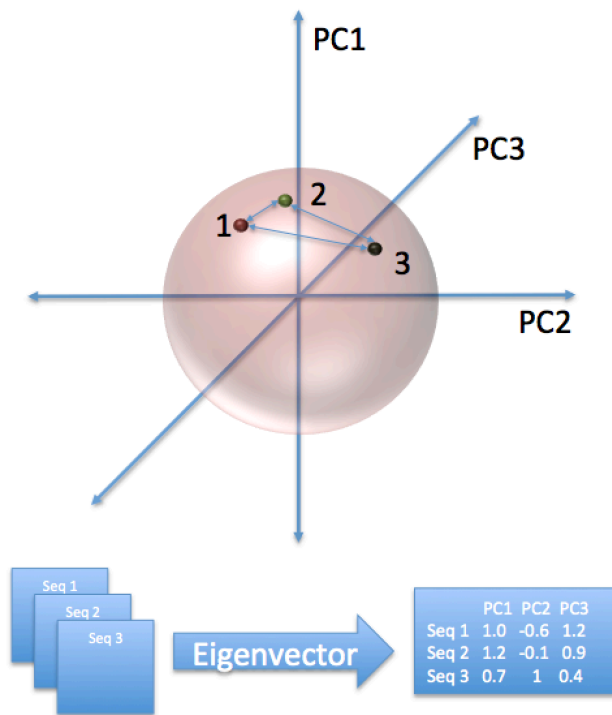
Decimal Encode

151

920

B

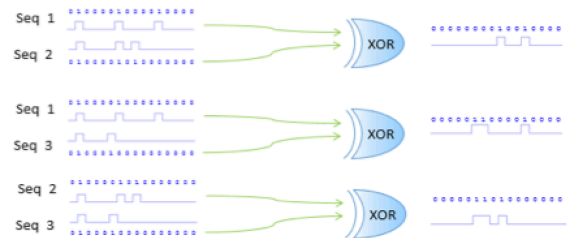
Linear Algebra  
(Orthogonal transformation)



Calculation of pairwise Euclidian distances  
Between points in Eigen space

Discrete Mathematics  
(Boolean transformation)

	ACGT	
Seq 1	AC	AA AC AG AT CA CC CG CT GA GC GG GT TA TC TG TT
	CG	O 1 O O O O 1 O O O O 1 O O O O
	GT	
	GA	
Seq 2	ACGA	AA AC AG AT CA CC CG CT GA GC GG GT TA TC TG TT
	AC	O 1 O O O O 1 O 1 O O O O O O O
	CG	
	GA	
Seq 3	ACCC	AA AC AG AT CA CC CG CT GA GC GG GT TA TC TG TT
	AC	O 1 O O O 1 O O O O O O O O O
	CC	
	CC	



Calculation of pairwise differences (XOR)  
using Boolean matrices

C

Invariant pairwise distance calculation for each genome  
Boolean (-logBoolean)

Seq 1 / Seq 2 : 0.9031

Seq 1 / Seq 3 : 0.7270

Seq 2 / Seq 3: 0.7270