

*Defesa de Mestrado*

# Alinhamento múltiplo de proteínas utilizando algoritmos genéticos

Sérgio Jeferson Rafael Ordine

Orientador: Zanoni Dias



---

# Agenda

---

- ❖ Introdução
- ❖ ALGAe
  - ❖ Melhorias
- ❖ Anubis
- ❖ Considerações Finais

# Introdução

---

# Alinhamento de Sequências

---

- ❖ Arranjo de sequências de caracteres
  - ❖ Maximizar Pontuação
  - ❖ Para sequências biológicas - evidencia similaridades:
    - ❖ Estruturais
    - ❖ Funcionais
    - ❖ Modela a história evolucionária



---

# Alinhamento de Sequências

---

M	P	R	E	D	R	A	T	W	K
M	A	K	L	S	K	Q	Q	K	K
M	A	H	V	A	E	W	K		
M	S	G	A	G	S	K	R	K	

---

# Alinhamento de Sequências

---

M P R E D R  
M A K L S K  
M A H V  
M S G A G

A T W K  
Q Q K K  
A E W K  
S K R K

---

# Alinhamento de Sequências

---

M P R E  
M A K L  
M A  
M S G A

D R A T W K  
S K Q Q K K  
H V A E W K  
G - S K R K



---

# Alinhamento de Sequências

---

M	P	R	E	D	R	A	T	W	K
M	A	K	L	S	K	Q	Q	K	K
-	-	M	A	H	V	A	E	W	K
M	S	G	A	G	-	S	K	R	K



---

# Alinhamento de Sequências

---

- ❖ Alinhamento de 2 sequências
  - ❖ Needleman e Wunsch (1970)
    - ❖ Programação dinâmica
    - ❖  $\theta(mn)$

---

# Alinhamento de Sequências

---

- ❖ Alinhamento de 2 sequências
  - ❖ Semi-Global
  - ❖ Local: Smith e Waterman (1981)

---

# Alinhamento Múltiplo de Sequências

---

- ❖ Multiple Sequence Alignment (MSA)
  - ❖ Conjunto de 3 ou mais sequências
  - ❖ NP-Difícil: Wang e Jiang (1994), Just (2001)
  - ❖ Foco deste trabalho: Sequências de proteínas



---

# MSA

---



Progressivo

---

# MSA

---



Progressivo

---

# MSA

---



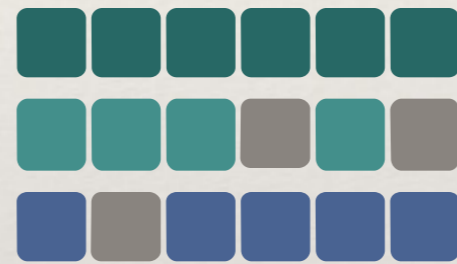
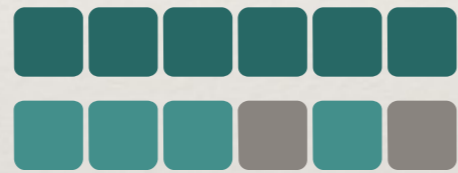
Progressivo



---

# MSA

---

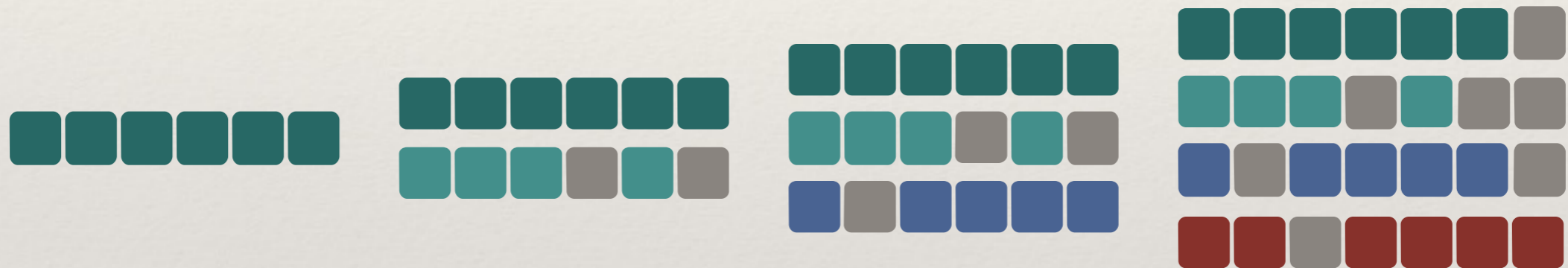


Progressivo

---

# MSA

---

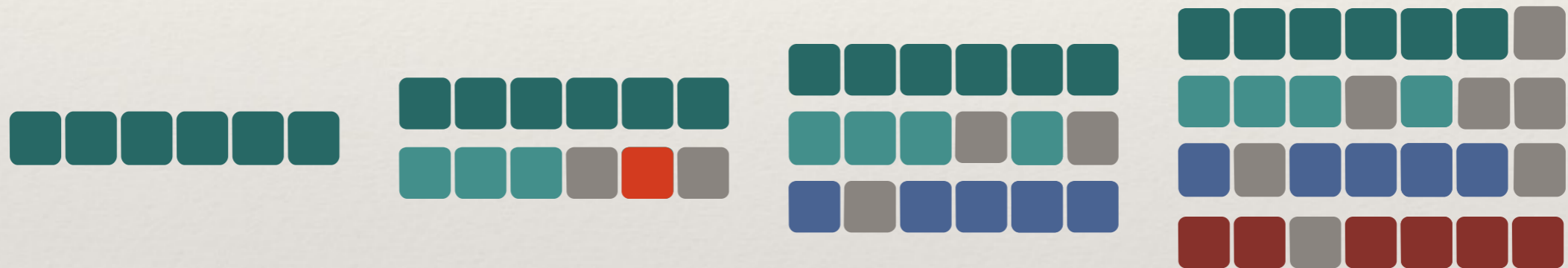


Progressivo

---

# MSA

---



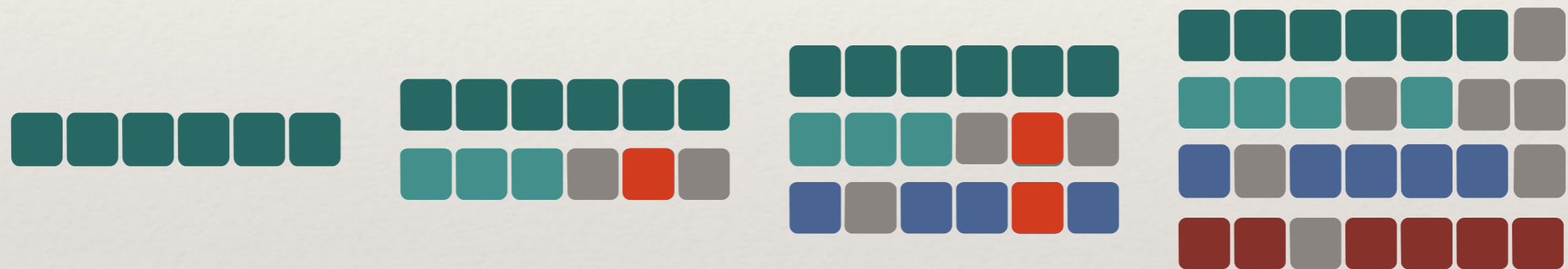
Progressivo



---

# MSA

---

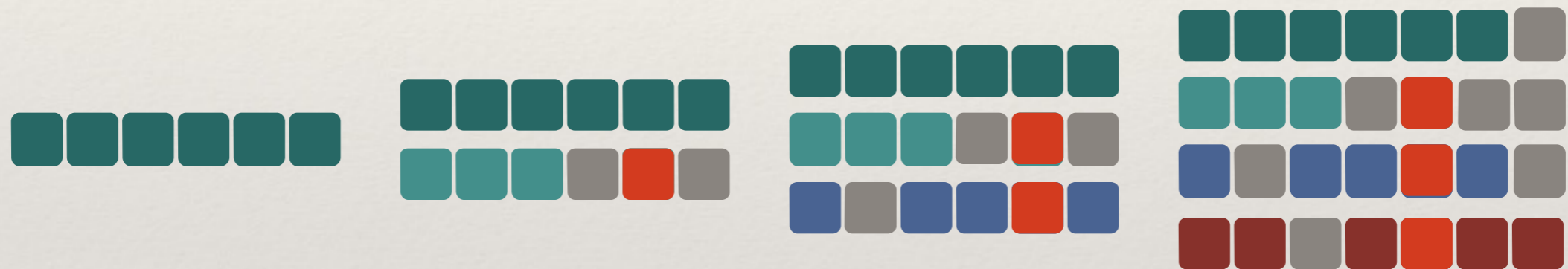


Progressivo

---

# MSA

---



Progressivo

---

# MSA

---



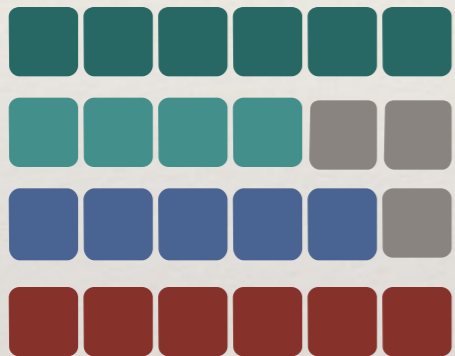
Iterativo



---

# MSA

---

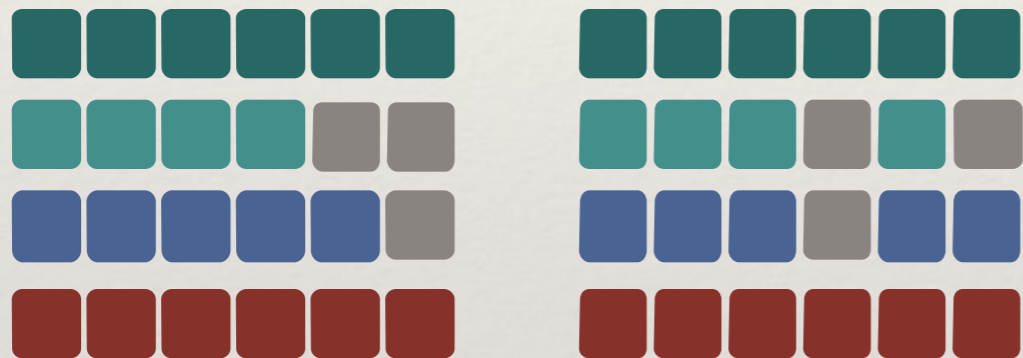


Iterativo

---

# MSA

---

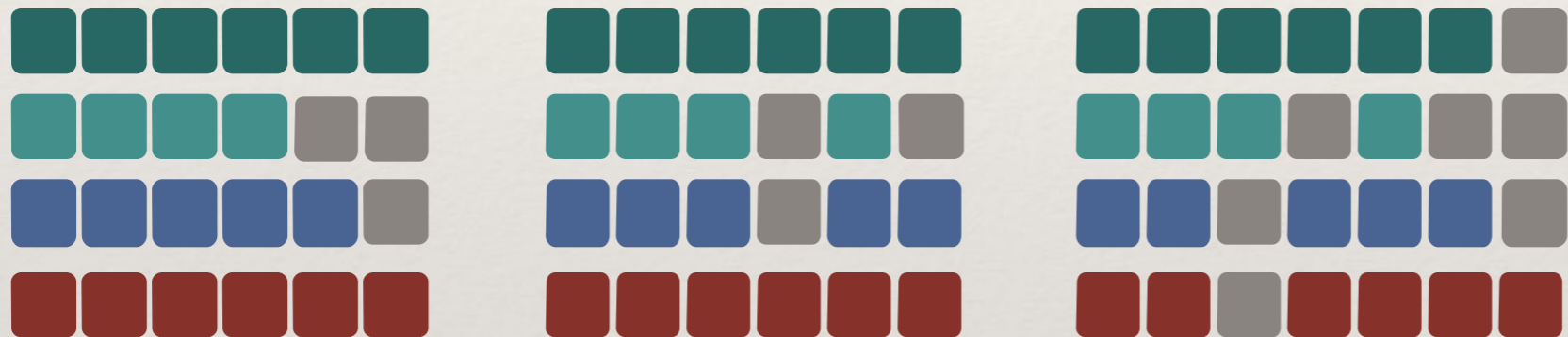


Iterativo

---

# MSA

---



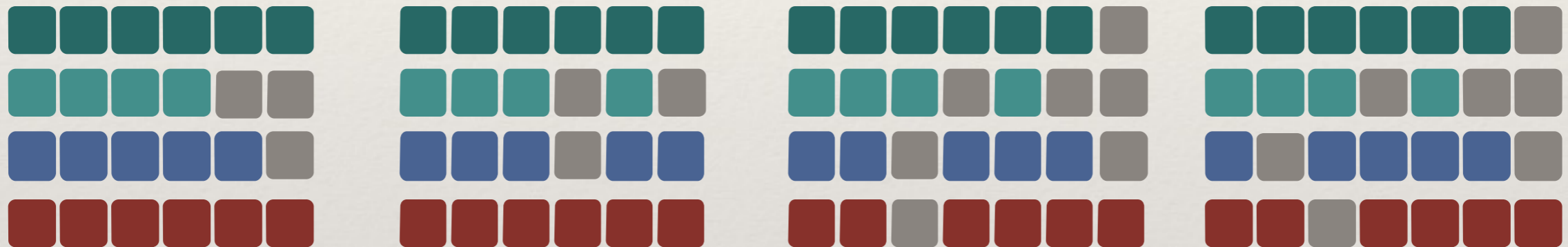
Iterativo



---

# MSA

---

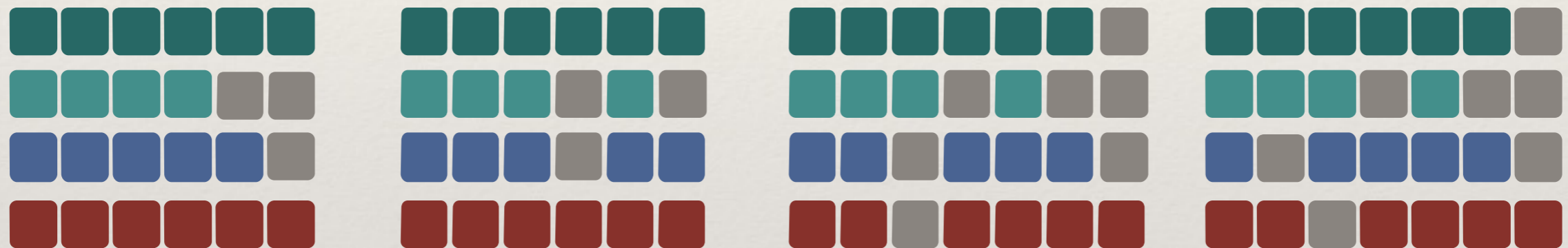


Iterativo

---

# MSA

---



Iterativo

---

# Algoritmos Genéticos

---

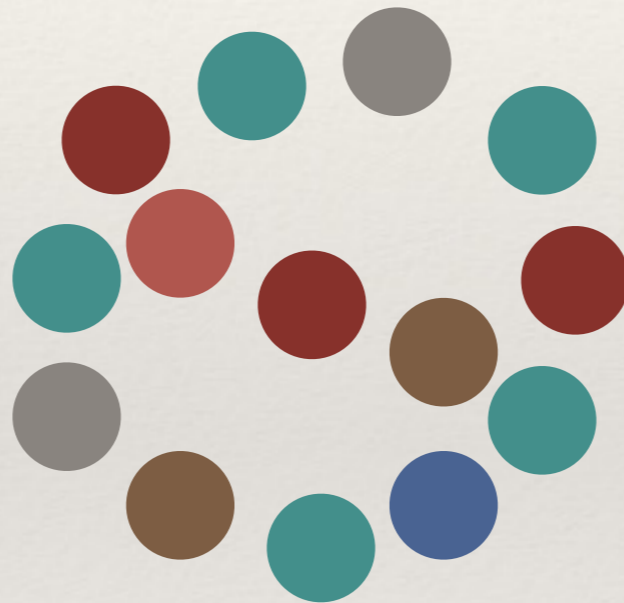
- ❖ Algoritmos de otimização
- ❖ Inspirado na Seleção Natural
- ❖ Holland (1975)
- ❖ SAGA: Notredame e Higgins (1996)
  - ❖ Precursor do uso de Algoritmos Genéticos para MSA



---

# Algoritmos Genéticos

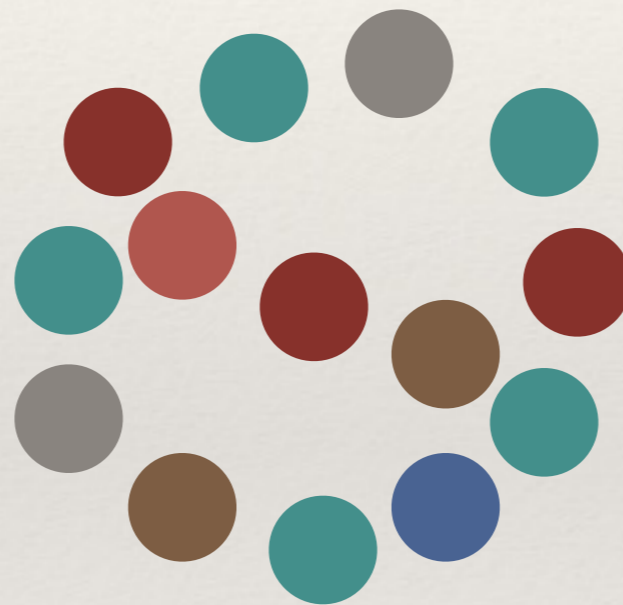
---



---

# Algoritmos Genéticos

---



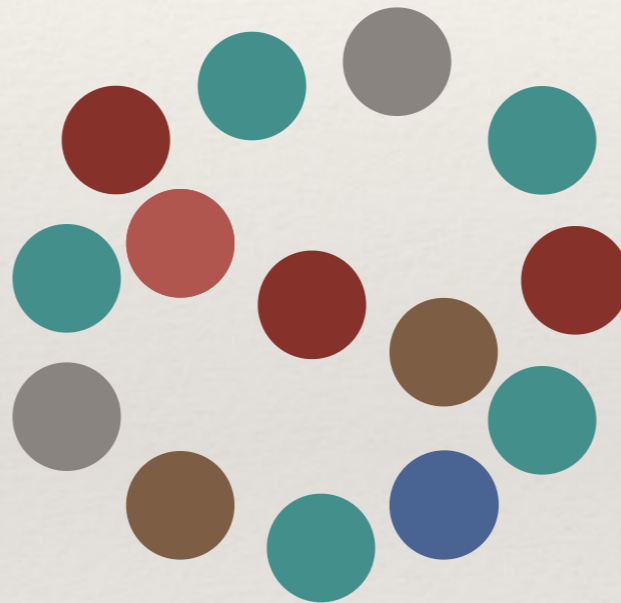
População Inicial

---

# Algoritmos Genéticos

---

- População Inicial



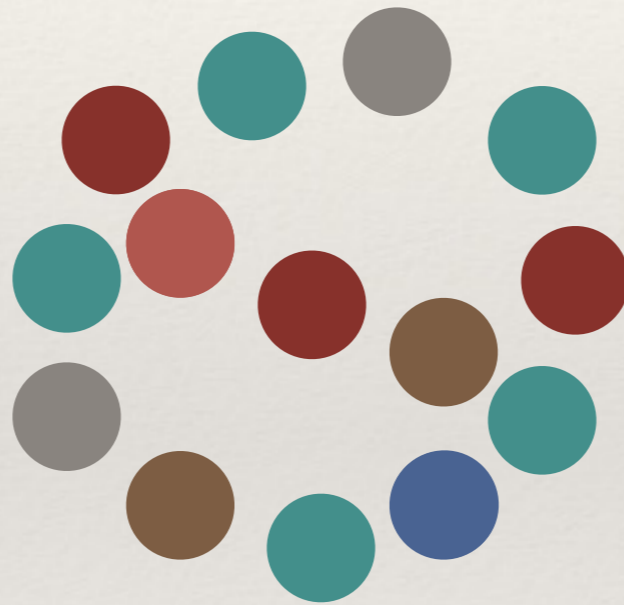


---

# Algoritmos Genéticos

---

- População Inicial



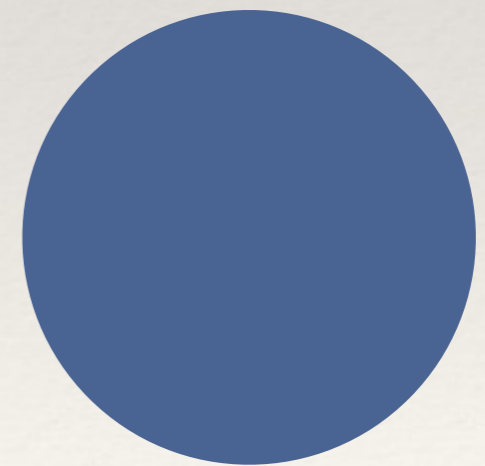
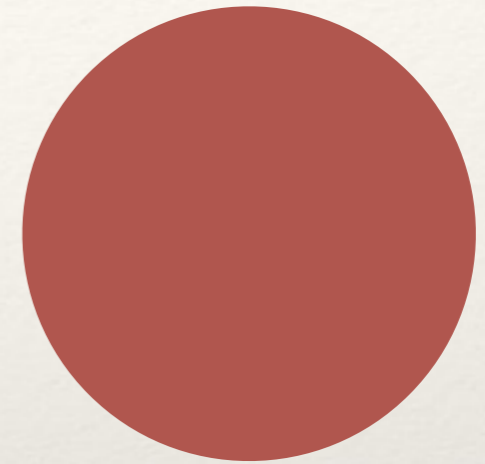
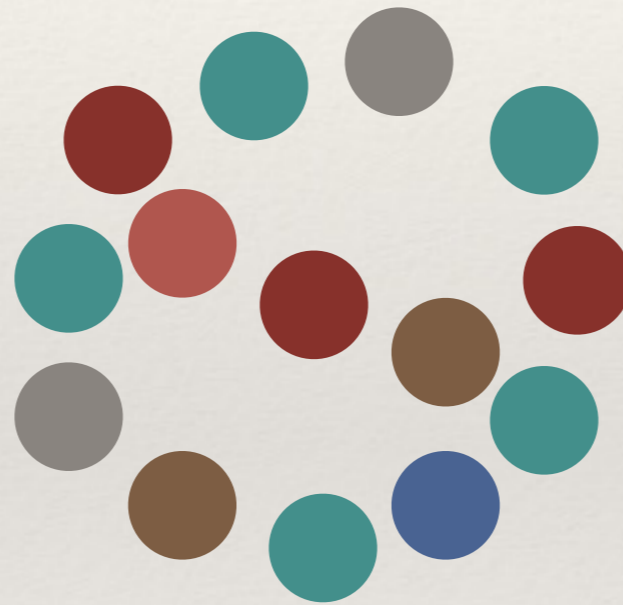
Crossover

---

# Algoritmos Genéticos

---

- População Inicial



Crossover

---

# Algoritmos Genéticos

---

- População Inicial



Crossover

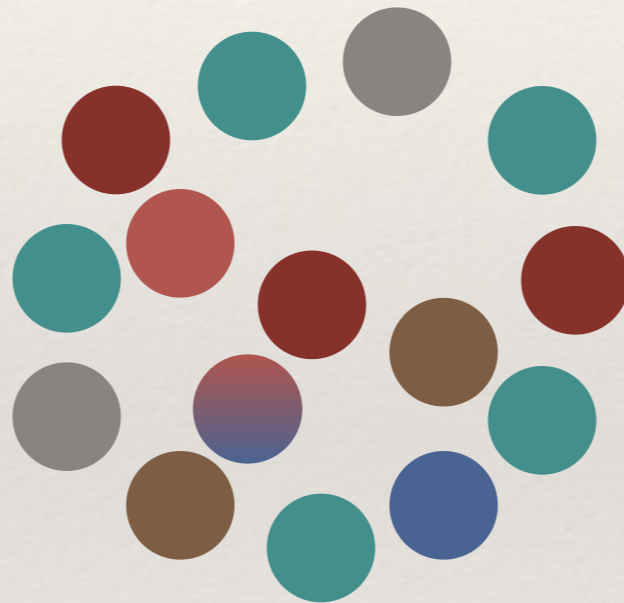


---

# Algoritmos Genéticos

---

- População Inicial
- Crossover

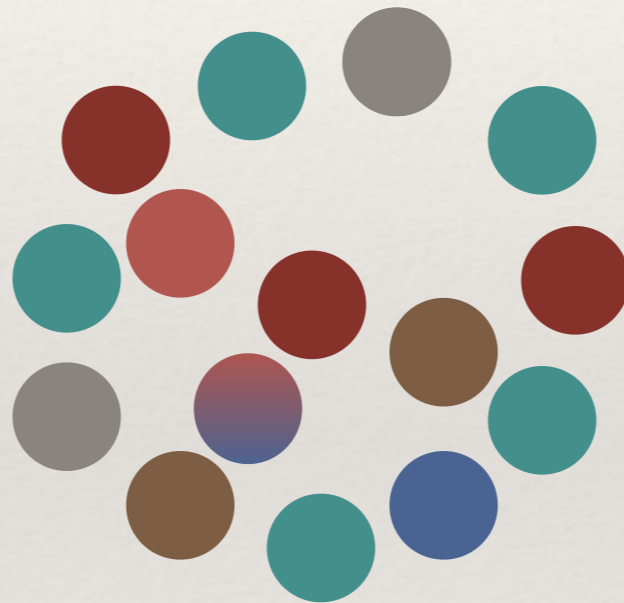


---

# Algoritmos Genéticos

---

- População Inicial
- Crossover



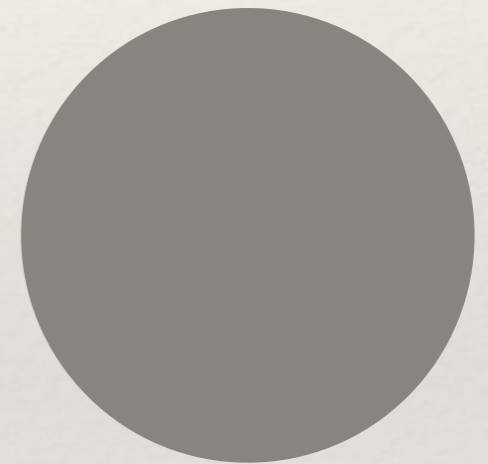
Mutação

---

# Algoritmos Genéticos

---

- População Inicial
- Crossover



Mutação

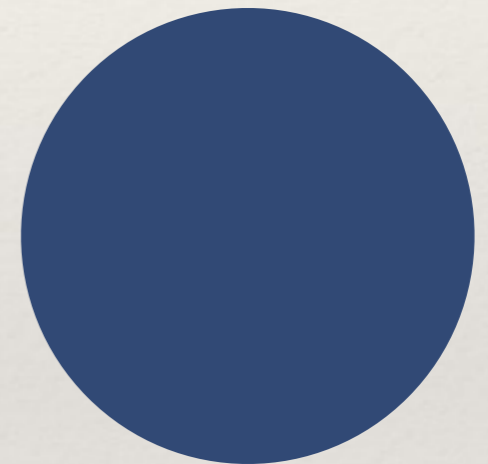
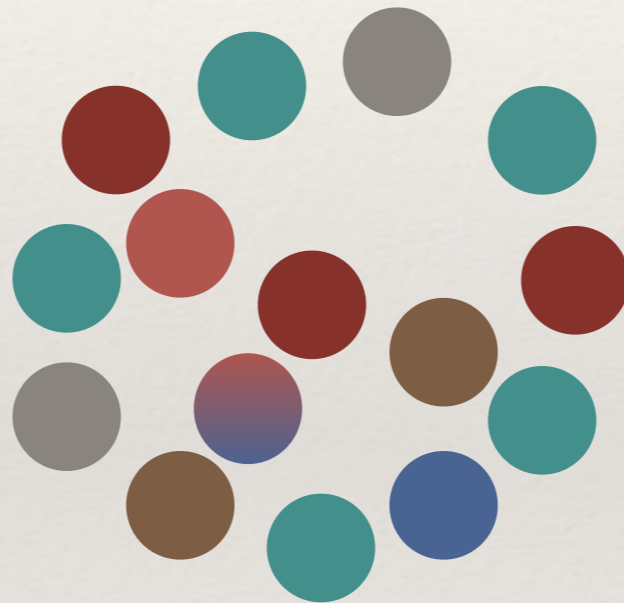


---

# Algoritmos Genéticos

---

- População Inicial
- Crossover



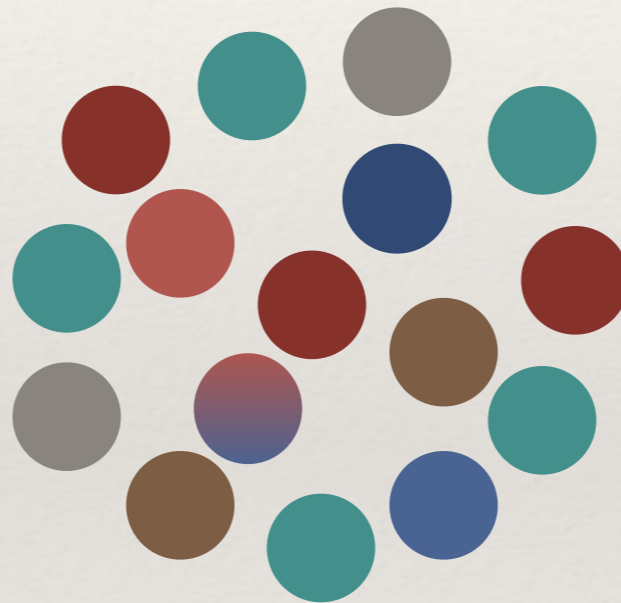
Mutação

---

# Algoritmos Genéticos

---

- População Inicial
- Crossover
- Mutação

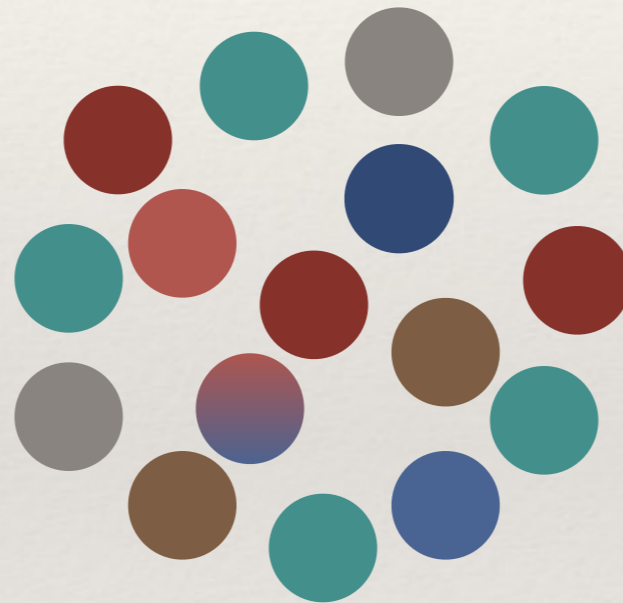


---

# Algoritmos Genéticos

---

- População Inicial
- Crossover
- Mutação



Seleção

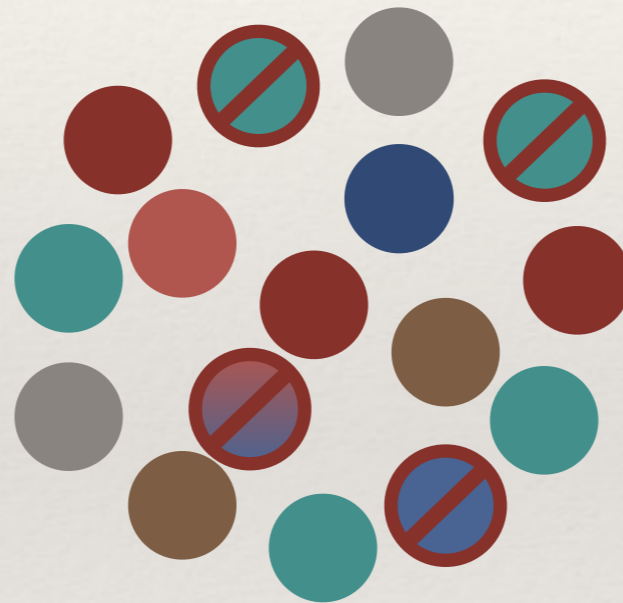


---

# Algoritmos Genéticos

---

- População Inicial
- Crossover
- Mutação



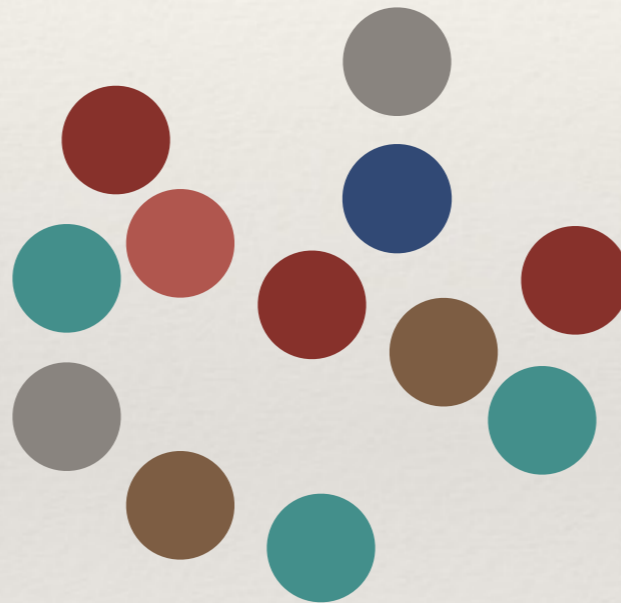
Seleção

---

# Algoritmos Genéticos

---

- População Inicial
- Crossover
- Mutação
- Seleção

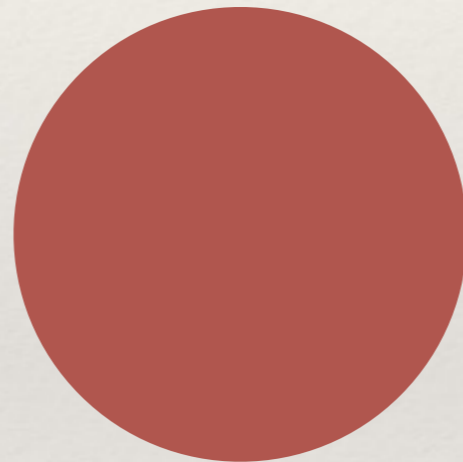


---

# Algoritmos Genéticos

---

- População Inicial
- Crossover
- Mutação
- Seleção





---

# Algoritmos Genéticos

---

- ❖ População Inicial
- ❖ Crossover
- ❖ Mutação
- ❖ Seleção

---

# Algoritmos Genéticos

---

- ❖ População Inicial
- ❖ Crossover
- ❖ Mutação
- ❖ Seleção
- ❖ Função objetivo (*fitness function*)

---

# Algoritmos Genéticos

---

- ❖ Problema
  - ❖ Utilizar algoritmos genéticos para MSA
    - ❖ Diversos parâmetros
    - ❖ Diversas possibilidades para cada parâmetro
    - ❖ Avaliação dos resultados



# ALGae

---

# ALGAe

---

- ❖ Ambiente configurável
  - ❖ Algoritmo genético para resolução de MSA
  - ❖ Java
  - ❖ Log
  - ❖ Testes com algumas variações para cada parâmetro
  - ❖ Foco em proteínas

---

# Parâmetros

---

- ❖ População inicial
- ❖ Mutação
- ❖ Crossover
- ❖ Seleção
- ❖ Função objetivo



---

# População Inicial

---

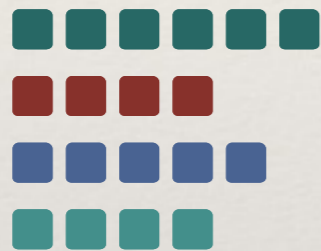
- ❖ Basic Star Alignment
- ❖ Random Star Alignment
- ❖ Super Population
- ❖ Population CR (*Conserved Region*)

---

# População Inicial

---

## Basic Star Alignment Population

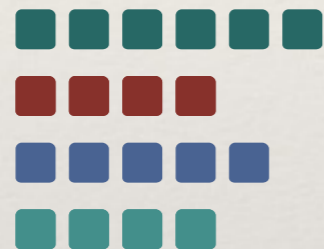
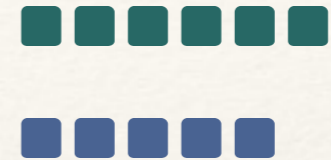


---

# População Inicial

---

**Basic Star Alignment Population**



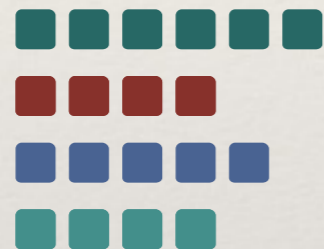


---

# População Inicial

---

**Basic Star Alignment Population**

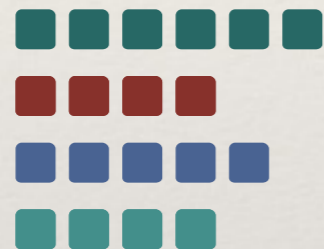
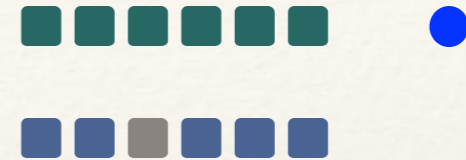


---

# População Inicial

---

**Basic Star Alignment Population**

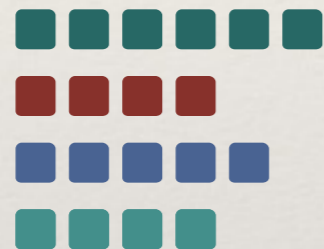


---

# População Inicial

---

**Basic Star Alignment Population**



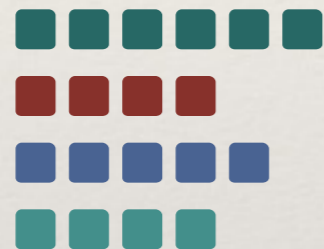


---

# População Inicial

---

**Basic Star Alignment Population**

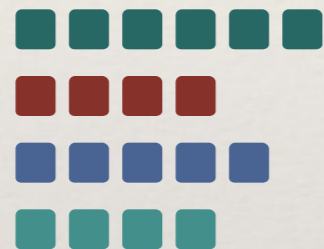
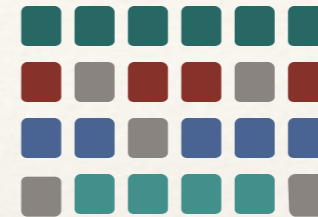


---

# População Inicial

---

**Basic Star Alignment Population**

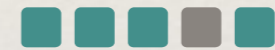
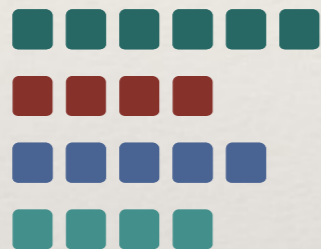


---

# População Inicial

---

## Basic Star Alignment Population



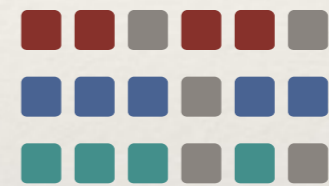
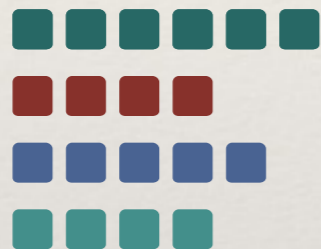


---

# População Inicial

---

## Basic Star Alignment Population

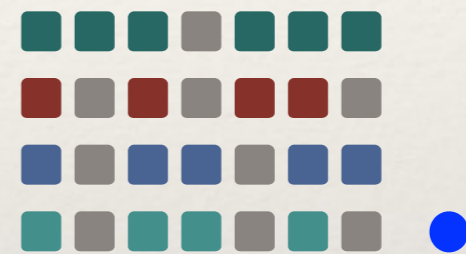
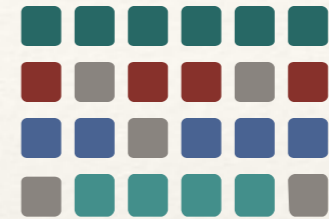
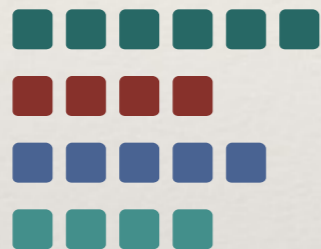


---

# População Inicial

---

## Basic Star Alignment Population

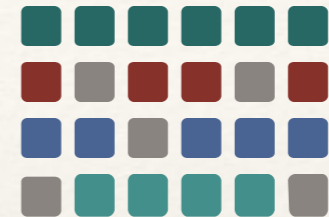
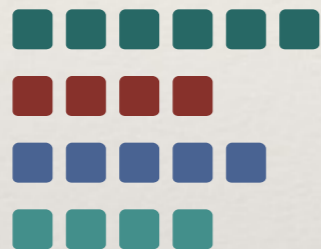


---

# População Inicial

---

## Basic Star Alignment Population



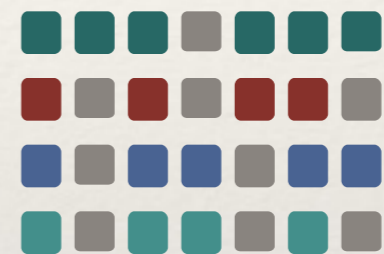
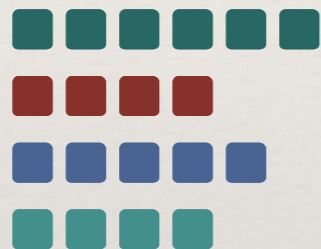


---

# População Inicial

---

## Basic Star Alignment Population

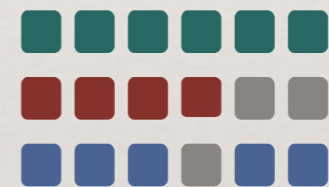
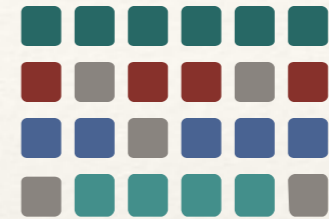
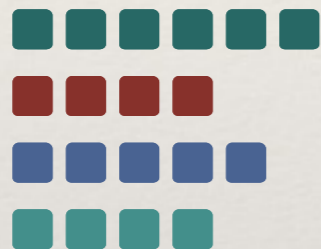


---

# População Inicial

---

## Basic Star Alignment Population

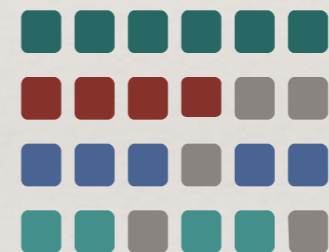
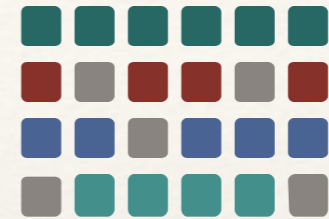
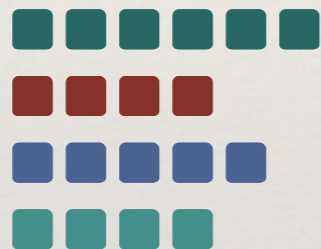


---

# População Inicial

---

## Basic Star Alignment Population



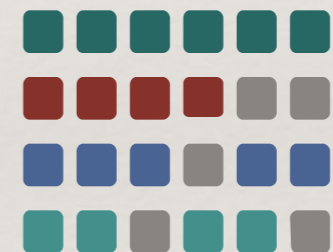
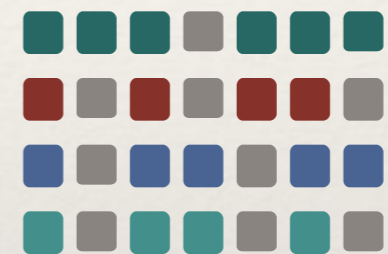
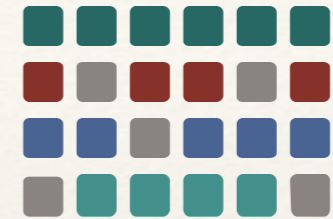
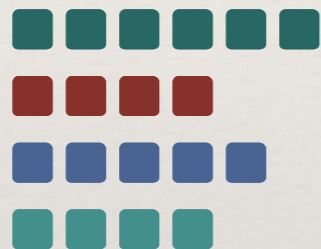


---

# População Inicial

---

## Basic Star Alignment Population

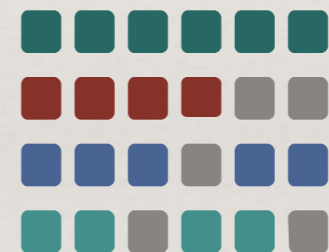
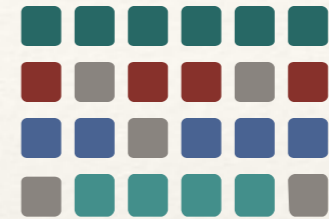
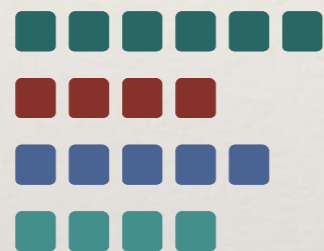


---

# População Inicial

---

## Basic Star Alignment Population

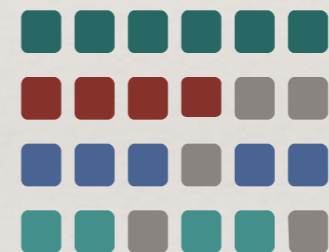
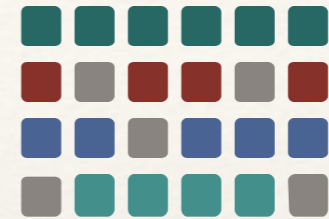
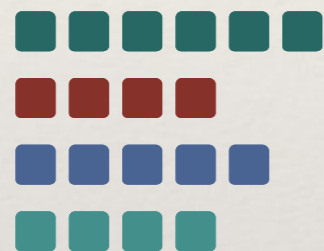


---

# População Inicial

---

## Basic Star Alignment Population



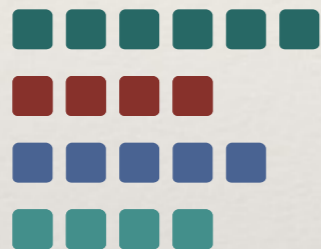


---

# População Inicial

---

## Random Star Alignment Population

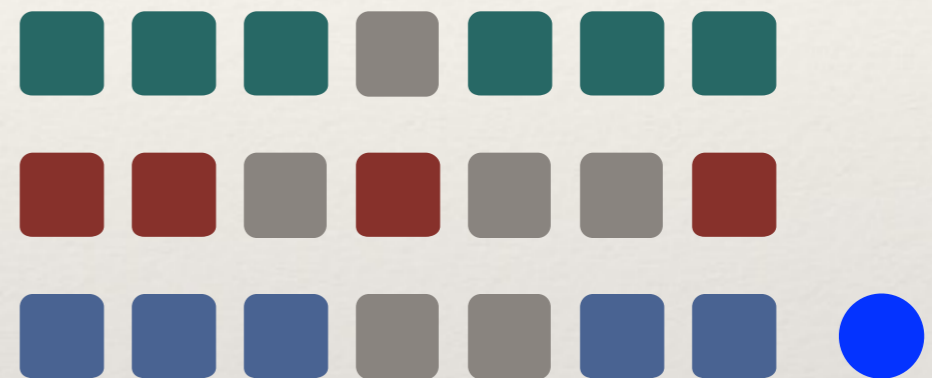
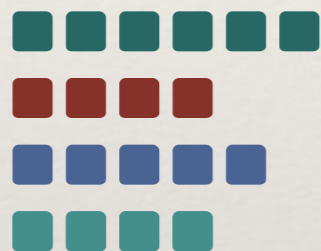


---

# População Inicial

---

## Random Star Alignment Population

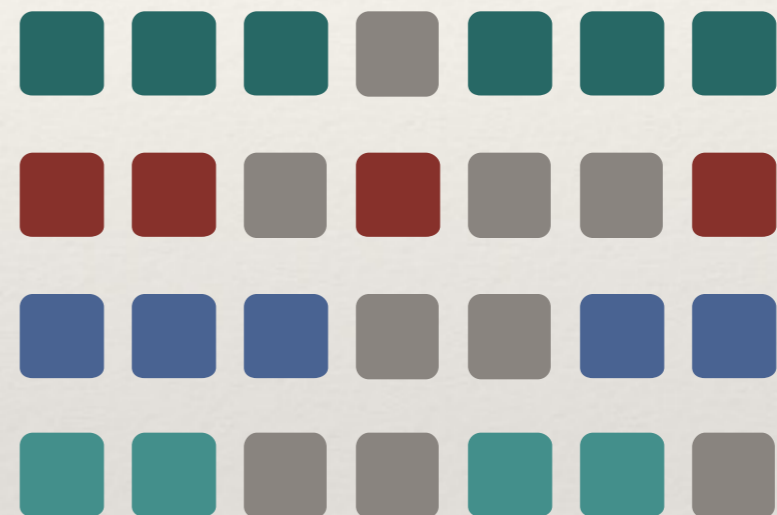
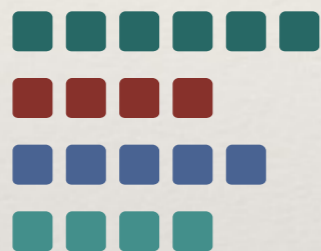


---

# População Inicial

---

## Random Star Alignment Population





---

# População Inicial

---

- ❖ Super Population
  - ❖ União das populações dos métodos anteriores
  - ❖ População maior que o limite populacional
    - ❖ Extinção em massa na primeira geração

---

# População Inicial

---

- ❖ PopulationCR
  - ❖ Criado sobre o método Super Population
  - ❖ Alinhamento baseado em regiões conservadas
    - ❖ Par a par
    - ❖ Progressivo - Sequências com maior similaridade

---

# População Inicial

---

- ❖ Abordagem A1
  - ❖ Alinhamento local
  - ❖ Recursivo - prefixo e sufixo



---

# População Inicial

---

## PopulationCR

	A	V	P	E	A	S	S	M	P	I	T	L	N	R	S	M	N	V	A	L	Q	
W	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
A	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
S	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
K	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Y	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
M	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
P	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
A	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
R	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
D	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
L	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■

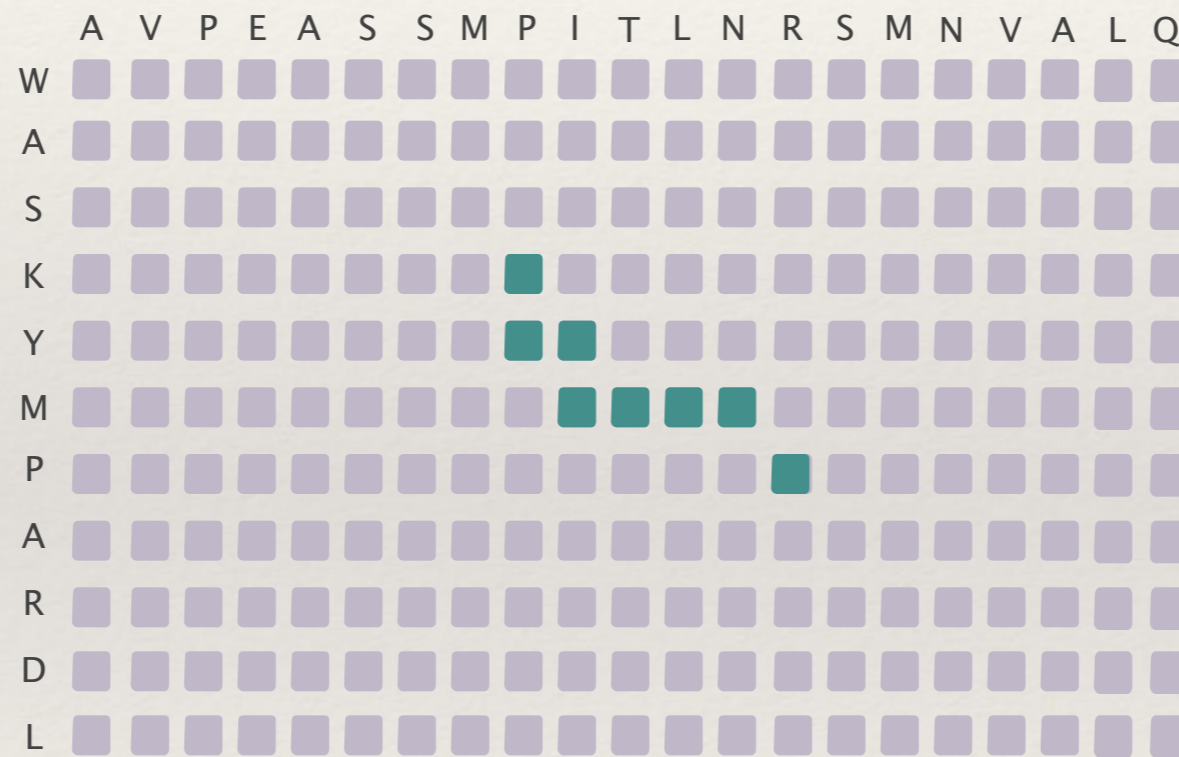
## Abordagem A1

---

# População Inicial

---

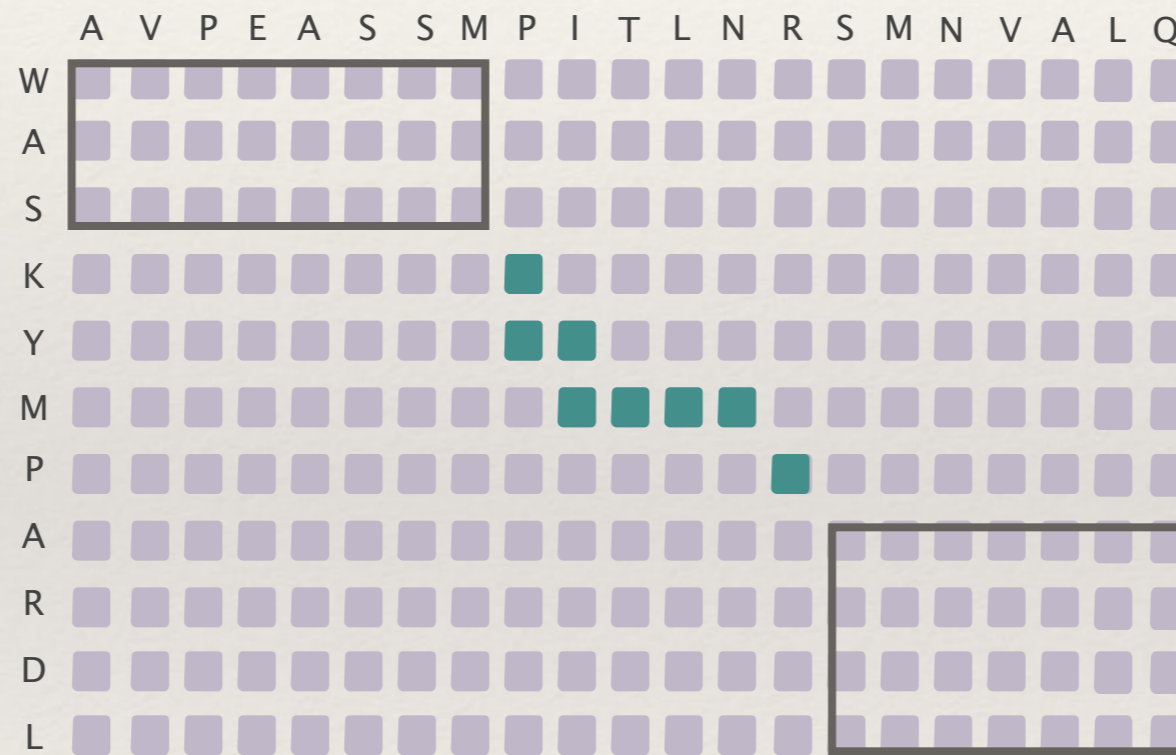
## PopulationCR



## Abordagem A1

# População Inicial

## PopulationCR

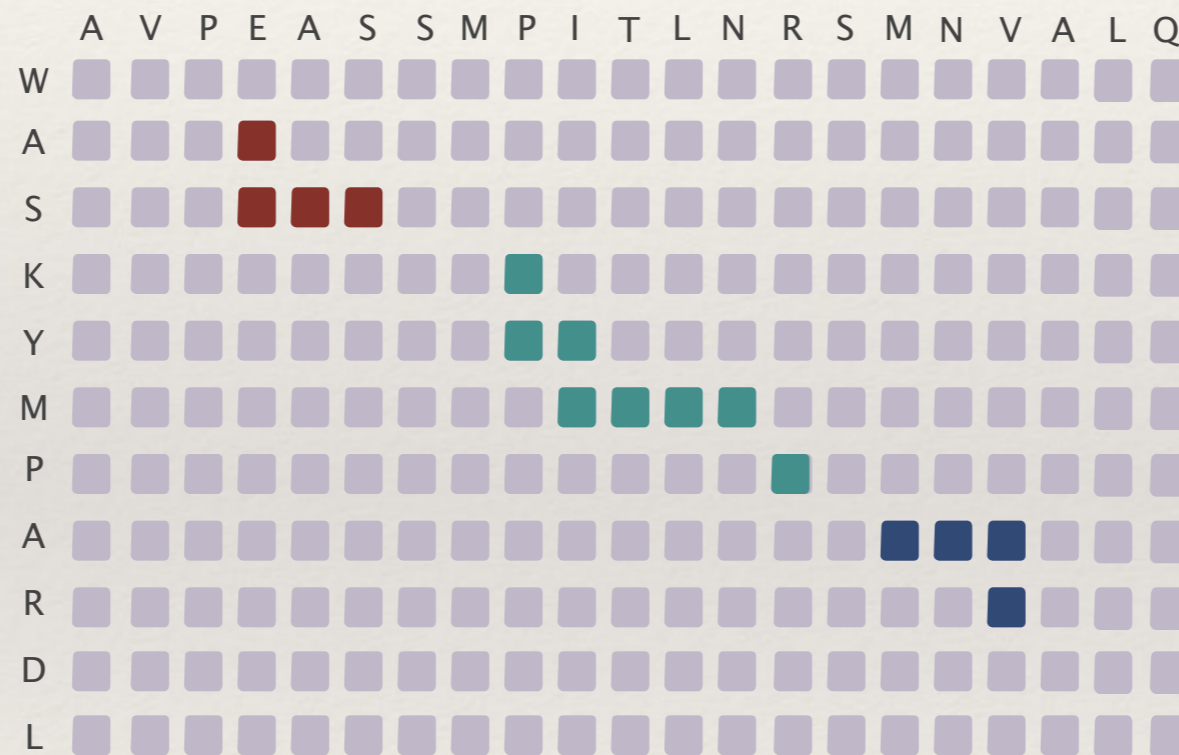


## Abordagem A1



# População Inicial

## PopulationCR

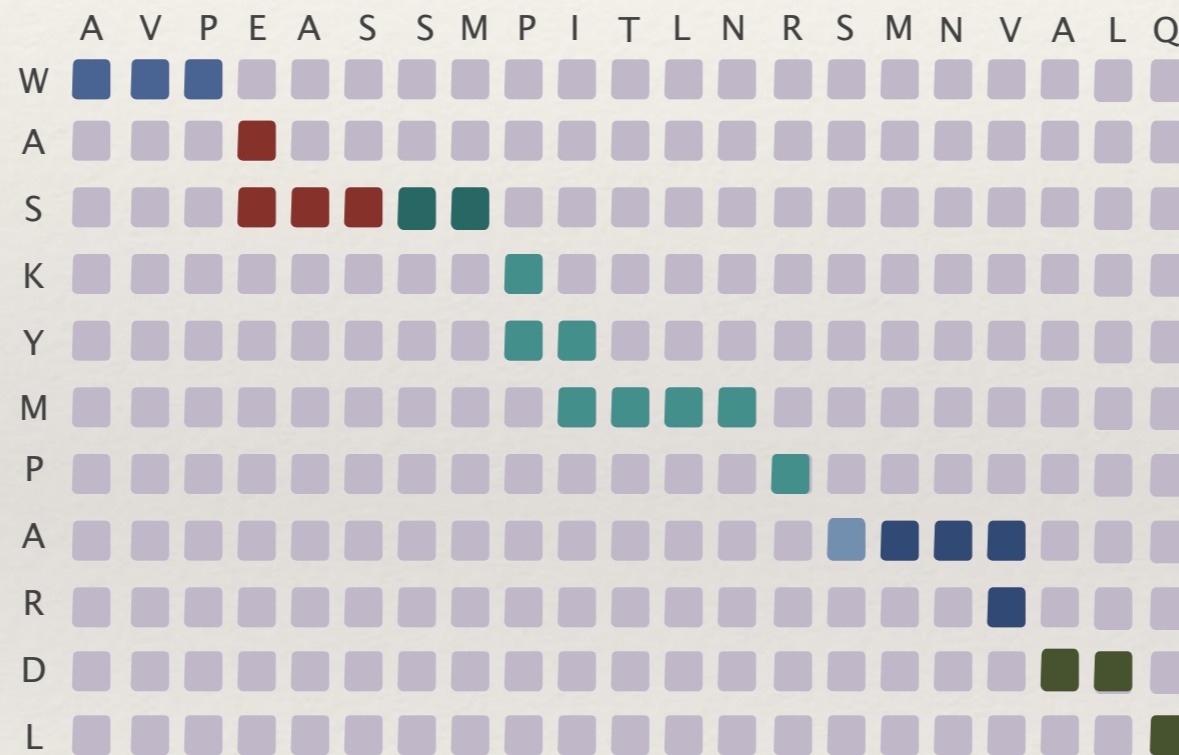


## Abordagem A1



# População Inicial

## PopulationCR



## Abordagem A1



---

# População Inicial

---

- ❖ Abordagem A2
  - ❖ Blocos idênticos
  - ❖ Janela deslizante
  - ❖ Grafo orientado de blocos consistentes
    - ❖ Caminho máximo

---

# População Inicial

---

- ❖ Abordagem A2
  - ❖ Caminho máximo
    - ❖ Grafo acíclico
    - ❖ Ordenação topológica

---

# População Inicial

---

**PopulationCR**

A D K T T Y A D K T T  
K T T A D K T T A K T T

**Abordagem A2**



---

# População Inicial

---

**PopulationCR**



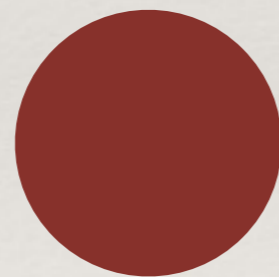
**Abordagem A2**

---

# População Inicial

---

**PopulationCR**



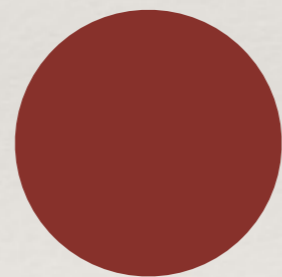
**Abordagem A2**

---

# População Inicial

---

**PopulationCR**



**Abordagem A2**



---

# População Inicial

---

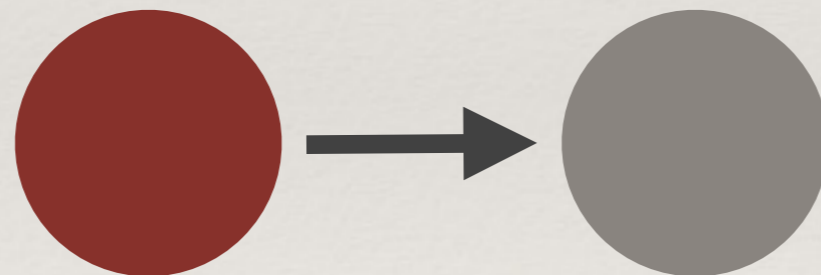
**PopulationCR**



**Abordagem A2**

# População Inicial

**PopulationCR**



**Abordagem A2**

---

# População Inicial

---

**PopulationCR**

A D K T T Y A D K T T  
K T T A D K T T A K T T

**Abordagem A2**



---

# População Inicial

---

**PopulationCR**



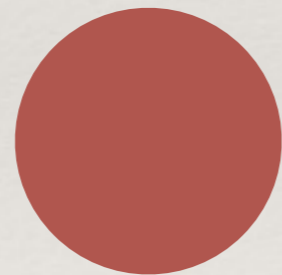
**Abordagem A2**

---

# População Inicial

---

**PopulationCR**



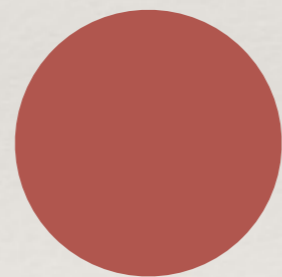
**Abordagem A2**

---

# População Inicial

---

**PopulationCR**



**Abordagem A2**



# População Inicial

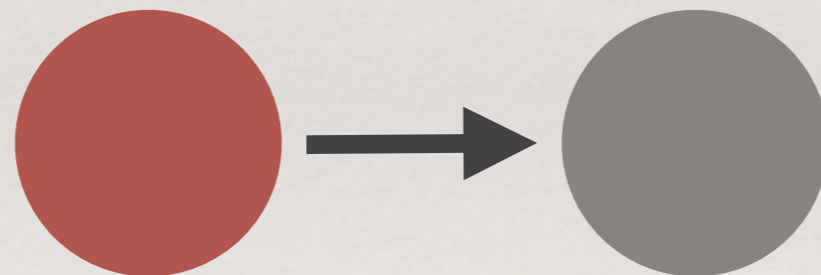
**PopulationCR**



**Abordagem A2**

# População Inicial

**PopulationCR**



**Abordagem A2**

---

# População Inicial

---

**PopulationCR**

A D K T T Y A D K T T  
K T T A D K T T A K T T

**Abordagem A2**



---

# População Inicial

---

**PopulationCR**

A D K T T Y A D K T T  
K T T A D K T T A K T T

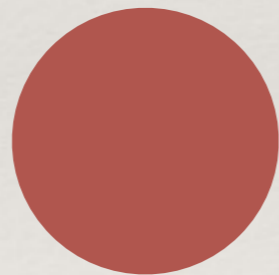
**Abordagem A2**

---

# População Inicial

---

**PopulationCR**



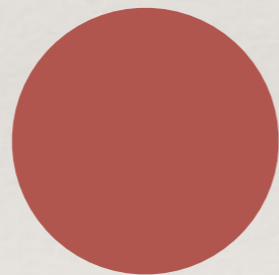
**Abordagem A2**

---

# População Inicial

---

**PopulationCR**



**Abordagem A2**



---

# População Inicial

---

**PopulationCR**



**Abordagem A2**

---

# População Inicial

---

**PopulationCR**

A D K T T Y A D K T T  
K T T A D K T T A K T T

**Abordagem A2**

---

# População Inicial

---

**PopulationCR**

A D K T T Y A D K T T  
K T T A D K T T A K T T

**Abordagem A2**

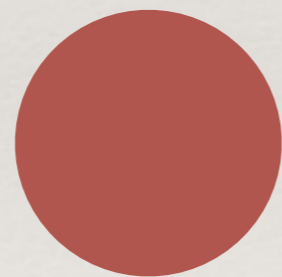


---

# População Inicial

---

**PopulationCR**



**Abordagem A2**

---

# População Inicial

---

**PopulationCR**



**Abordagem A2**

---

# População Inicial

---

**PopulationCR**



**Abordagem A2**



---

# População Inicial

---

- ❖ Abordagem A2
  - ❖ Janela deslizante ( $k = 5$ )

---

# População Inicial

---

- ❖ Abordagem A2a
  - ❖ Alfabetos comprimidos
  - ❖ Janela deslizante ( $k = 12$ )

---

# População Inicial

---

- ❖ PopulationCR
  - ❖ Comparação das abordagens utilizadas
    - ❖ BAliBASE 3
    - ❖ Abordagem A1



---

# Mutação

---

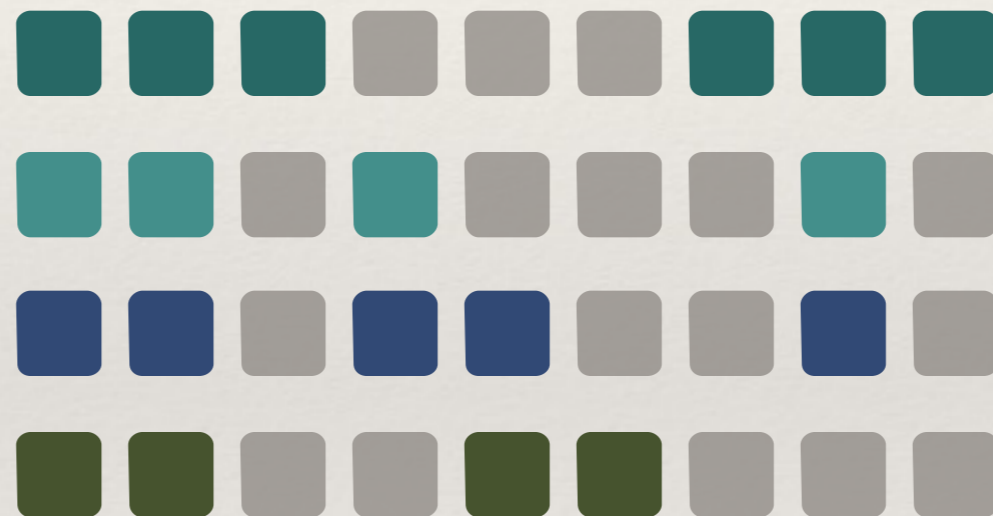
- ❖ Simple Mutation
- ❖ Shift Gap Block Mutation
- ❖ Change Gap Block Mutation

---

# Mutação

---

## Simple Mutation

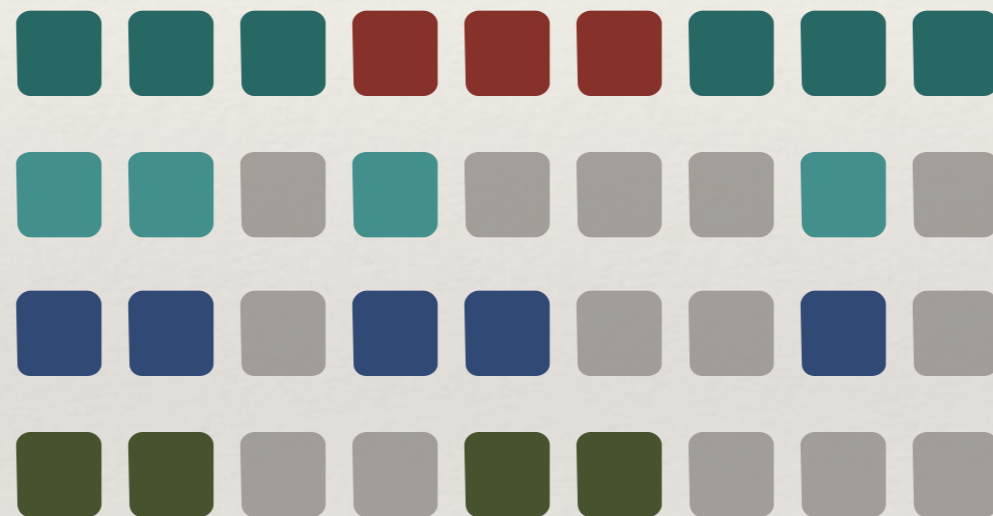


---

# Mutação

---

## Simple Mutation



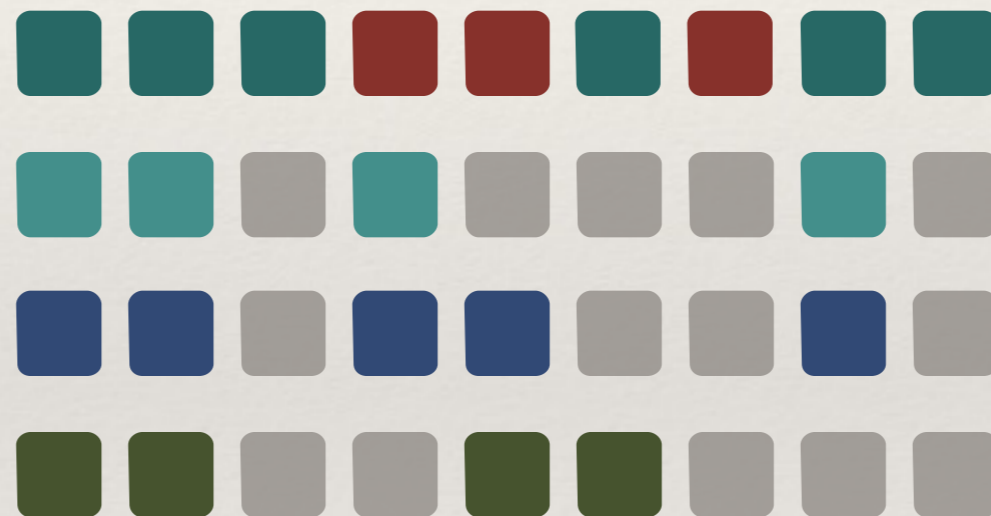


---

# Mutação

---

## Simple Mutation

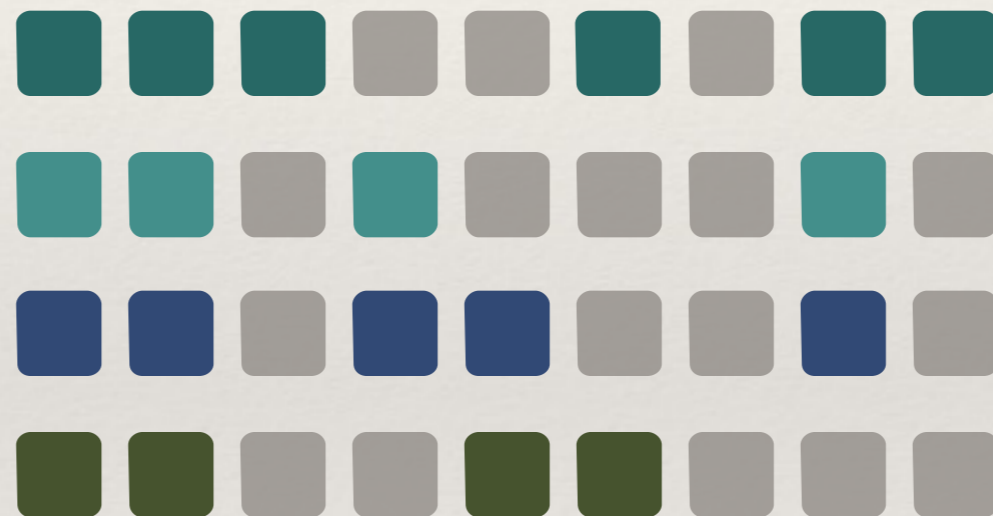


---

# Mutação

---

## Simple Mutation

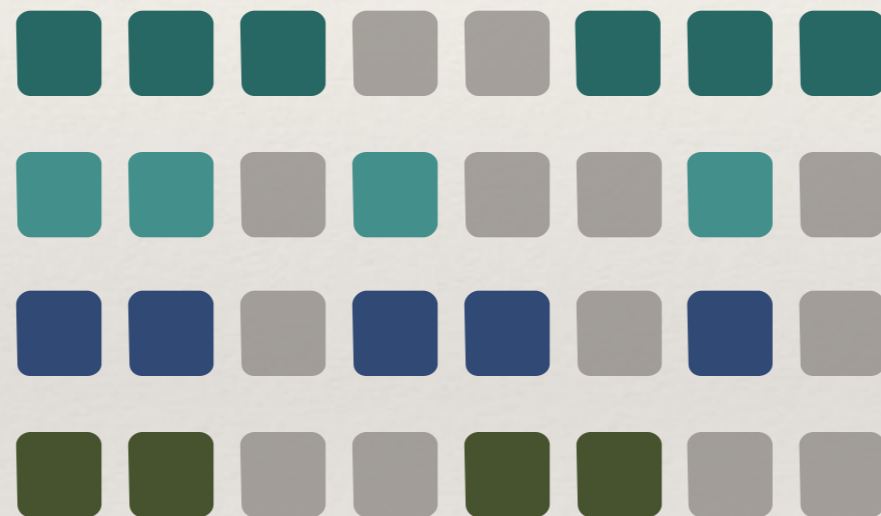


---

# Mutação

---

## Simple Mutation



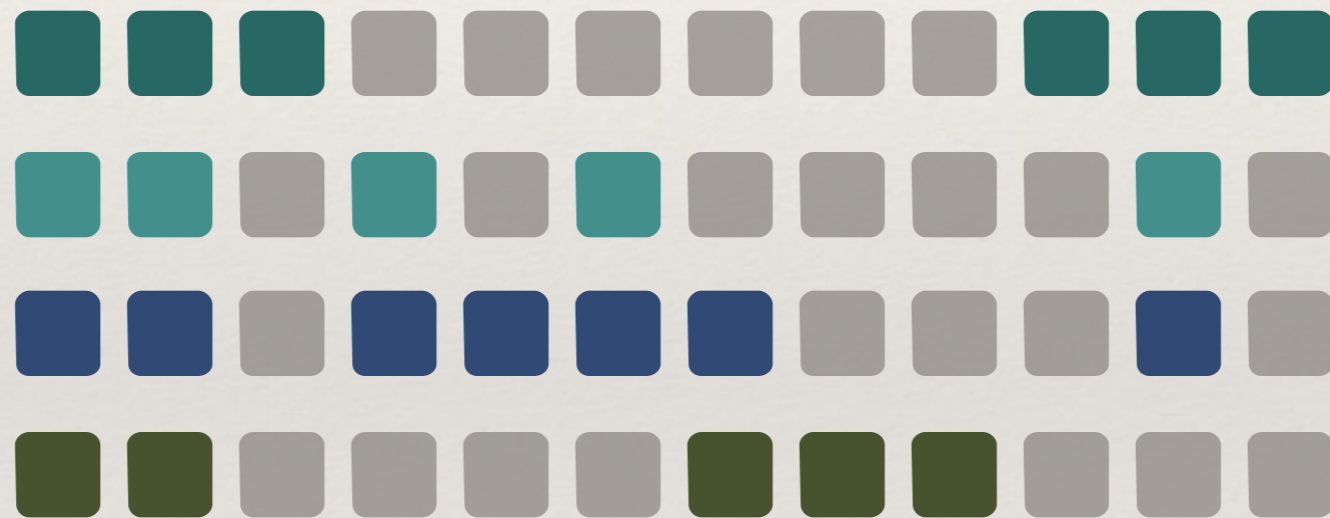


---

# Mutação

---

## Shift Gap Block Mutation

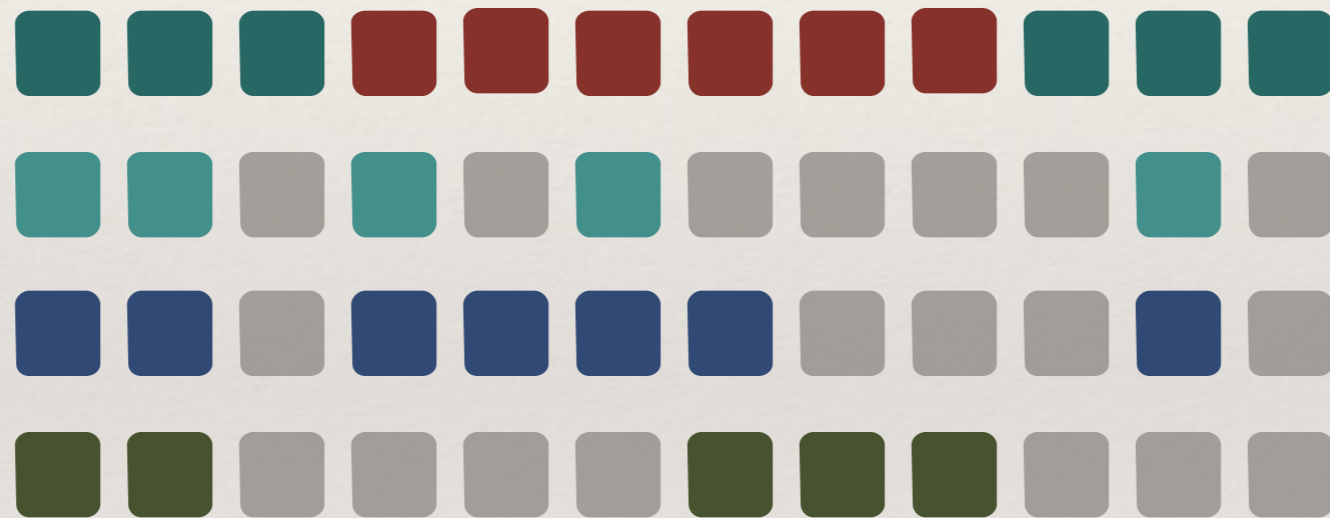


---

# Mutação

---

## Shift Gap Block Mutation

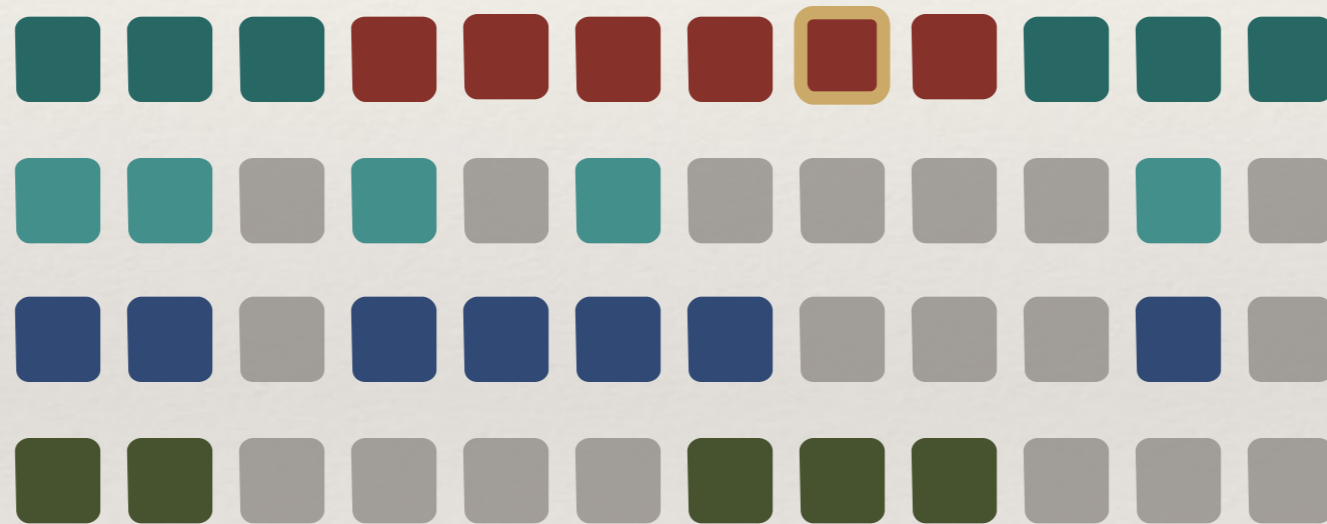


---

# Mutação

---

## Shift Gap Block Mutation



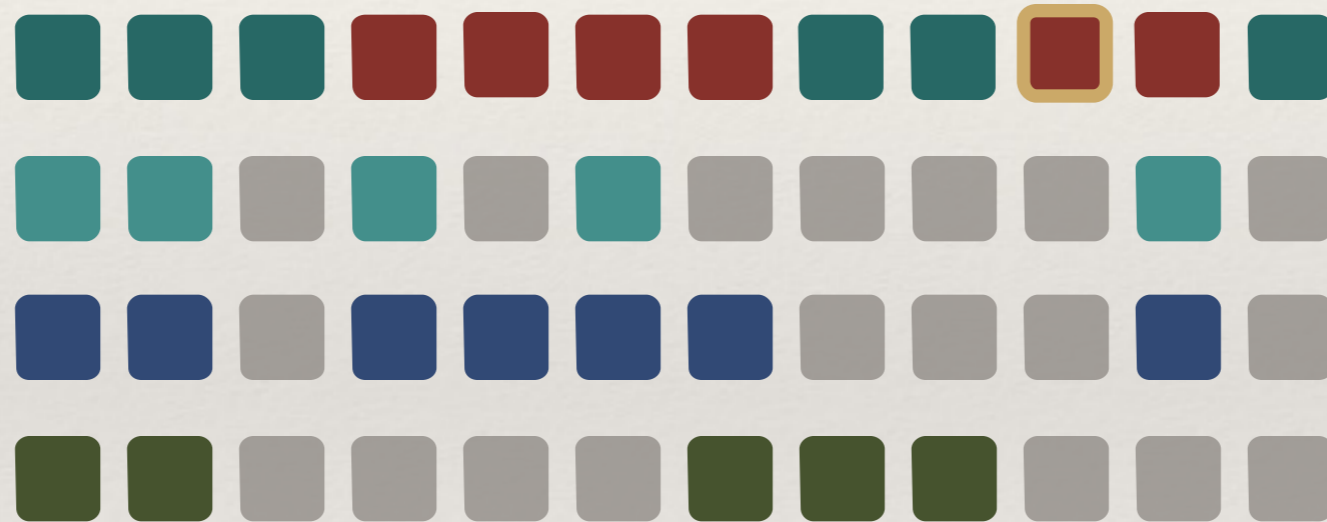


---

# Mutação

---

## Shift Gap Block Mutation

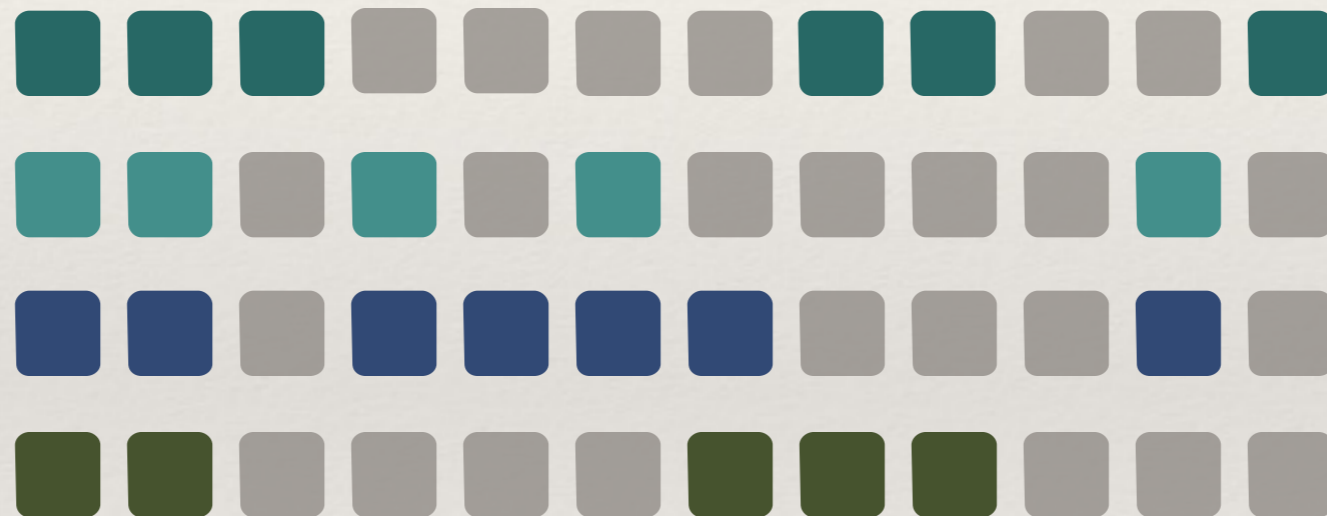


---

# Mutação

---

## Shift Gap Block Mutation

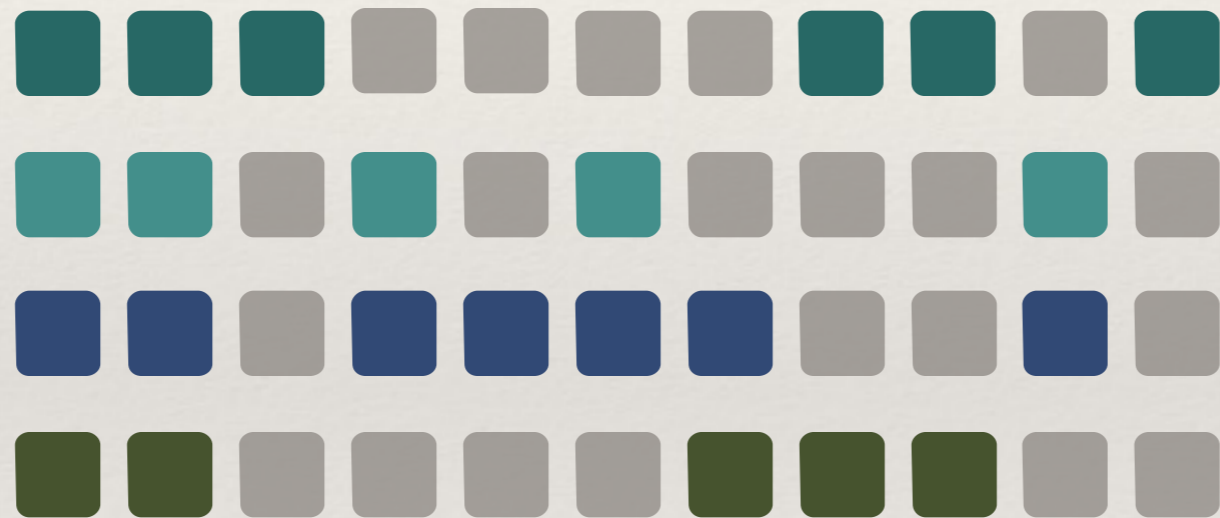


---

# Mutação

---

## Shift Gap Block Mutation

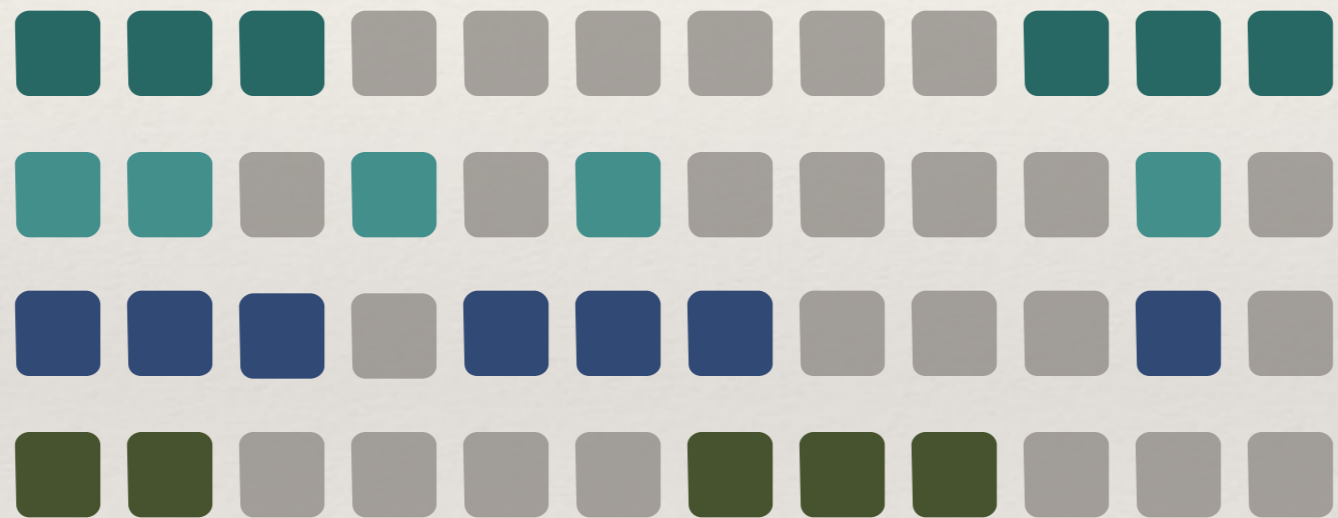




---

# Mutação

---

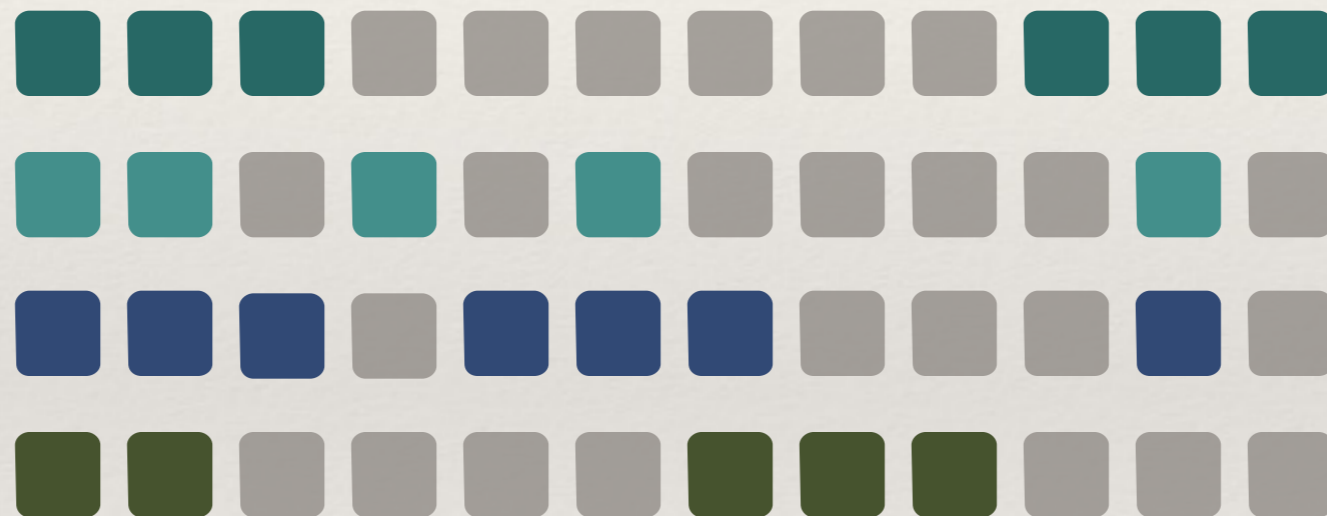


---

# Mutação

---

## Change Gap Block Mutation

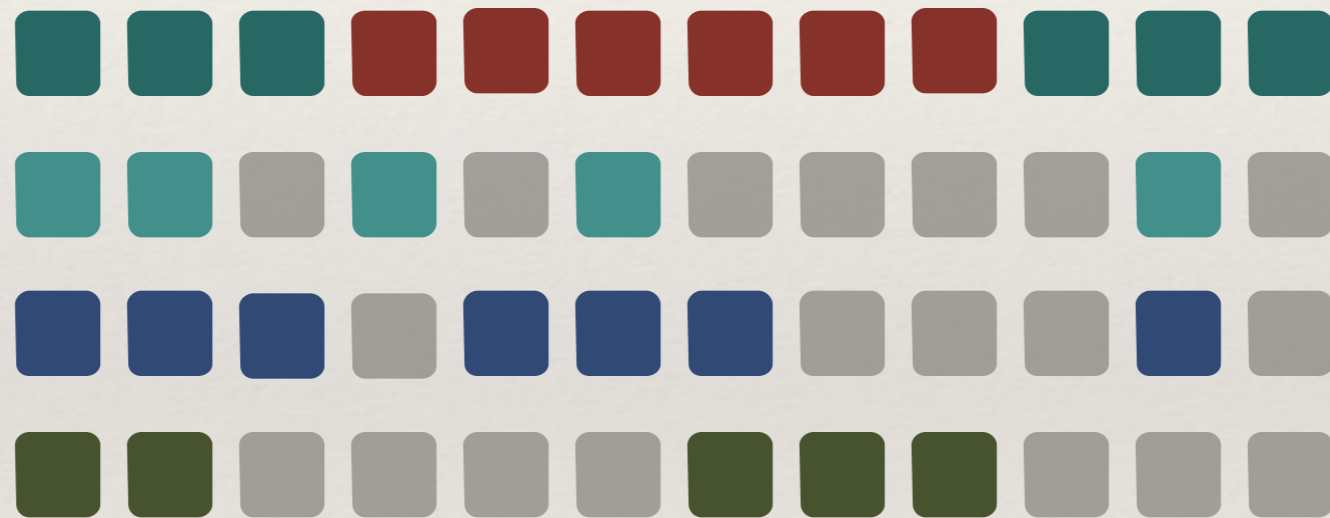


---

# Mutação

---

## Change Gap Block Mutation



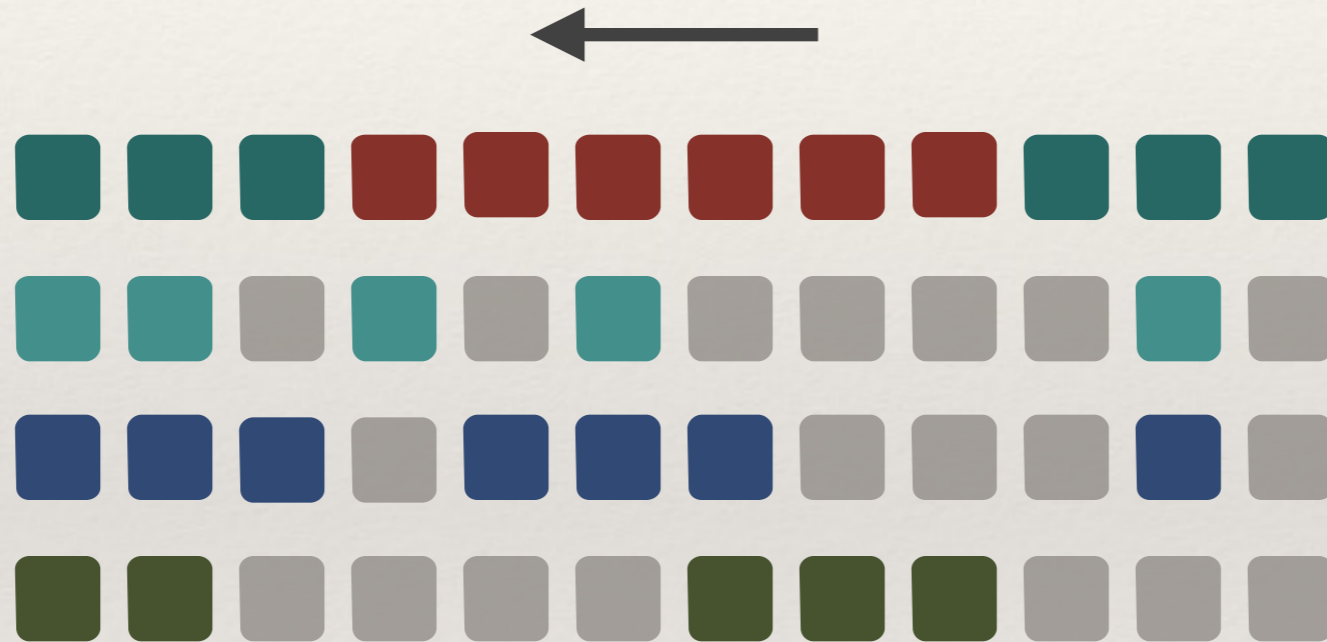


---

# Mutação

---

## Change Gap Block Mutation

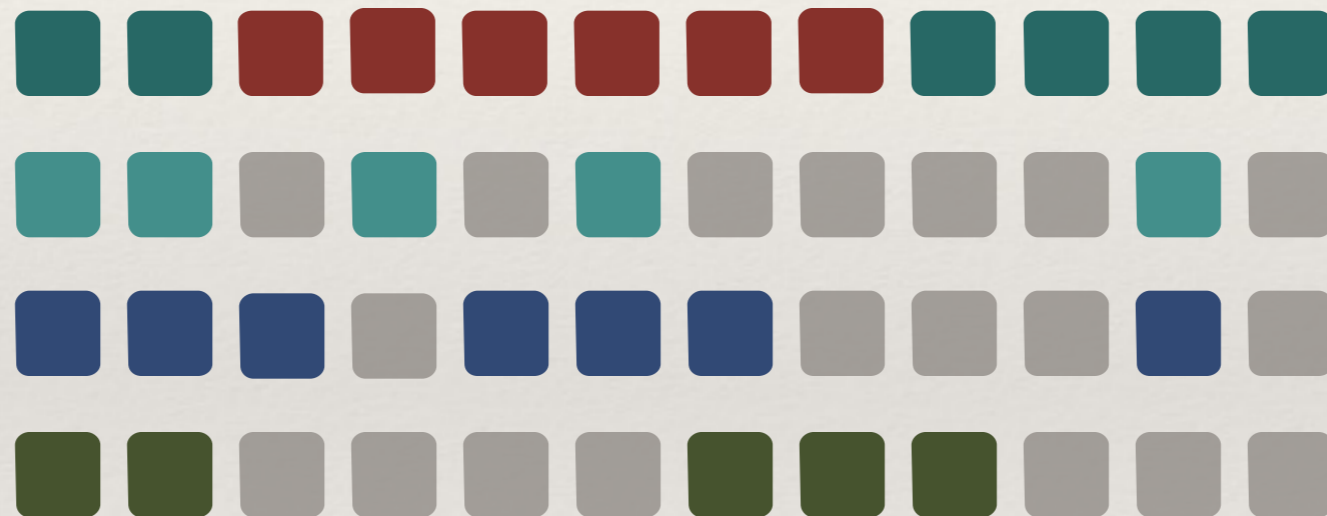


---

# Mutação

---

## Change Gap Block Mutation

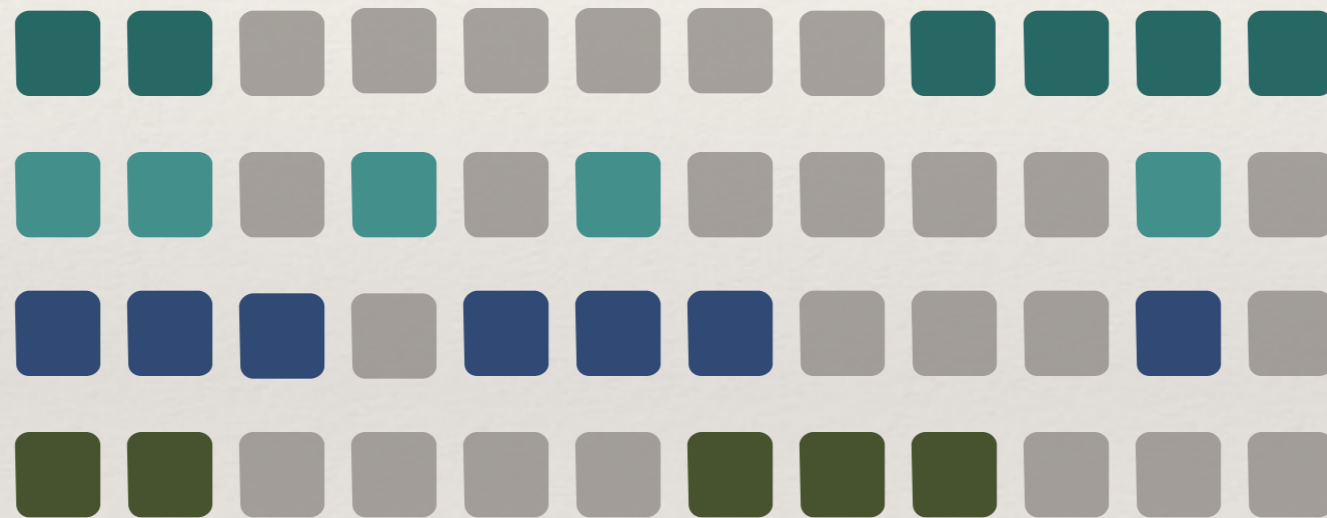


---

# Mutação

---

## Change Gap Block Mutation





---

# Crossover

---

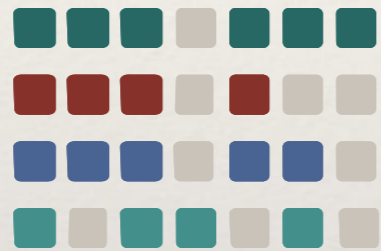
- ❖ Single Point Crossover
- ❖ Best Partial Alignment Crossover
- ❖ Sequence Similarity Crossover

---

# Crossover

---

## Single Point Crossover

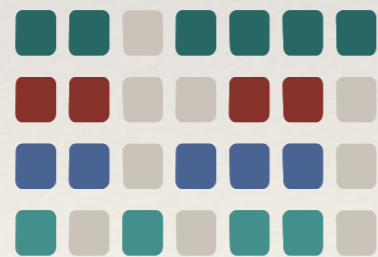
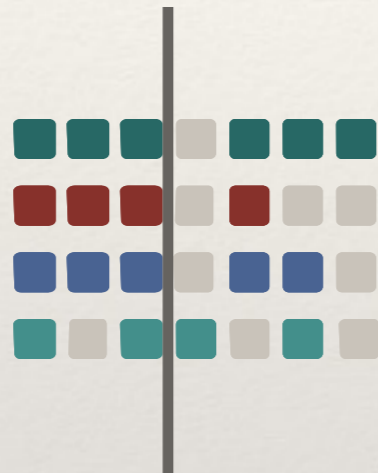


---

# Crossover

---

## Single Point Crossover



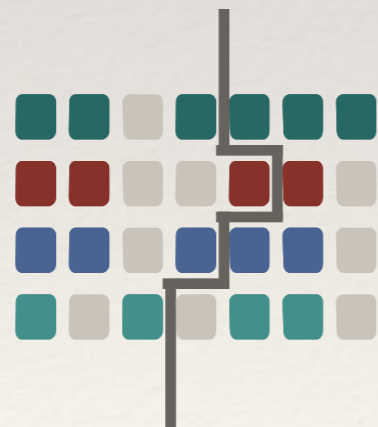
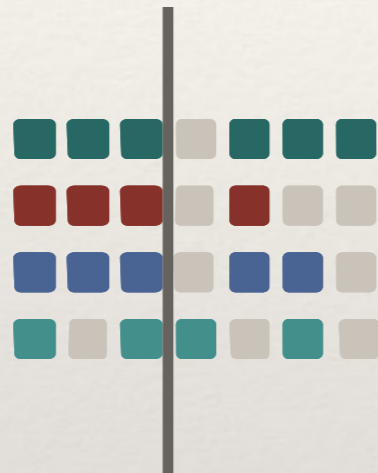


---

# Crossover

---

## Single Point Crossover

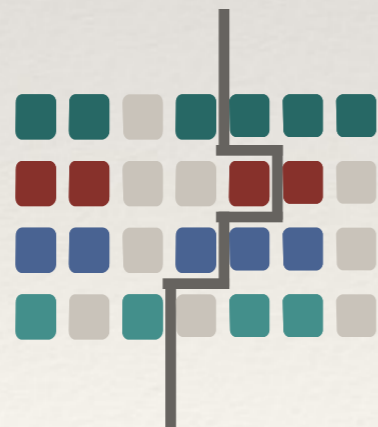
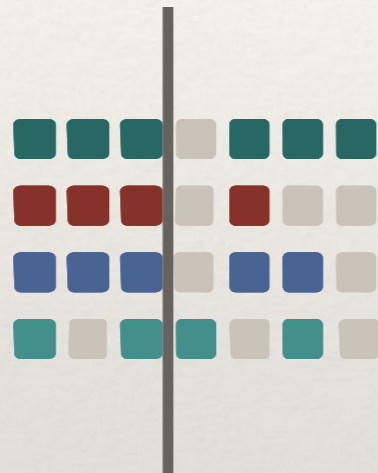


---

# Crossover

---

## Single Point Crossover

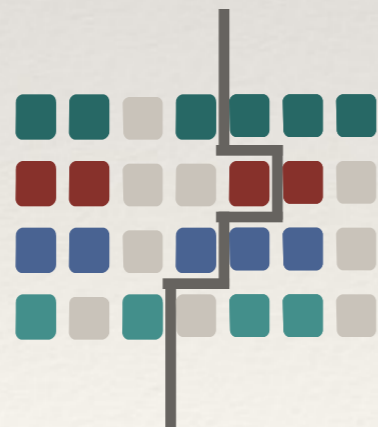
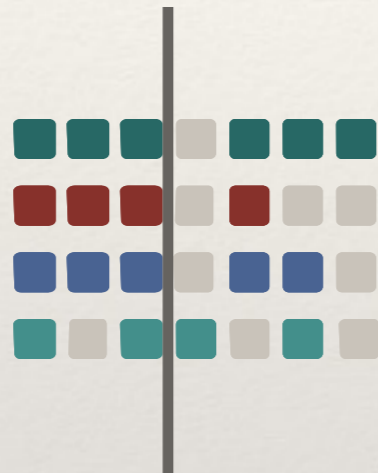


---

# Crossover

---

## Single Point Crossover



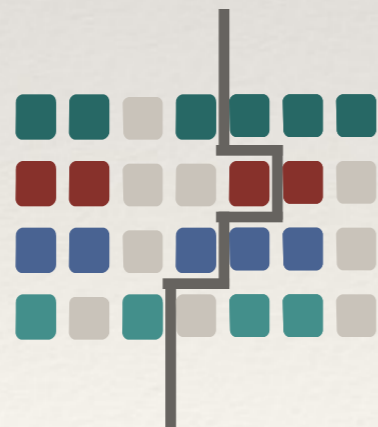
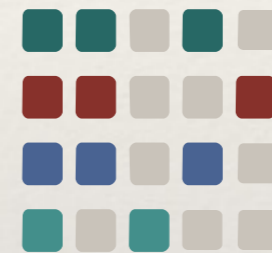
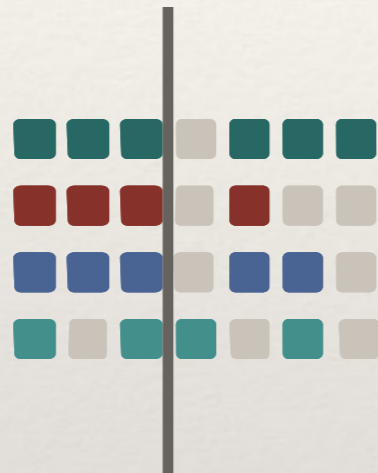


---

# Crossover

---

## Single Point Crossover

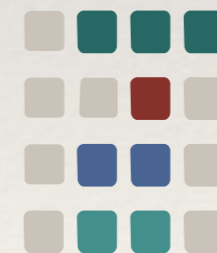
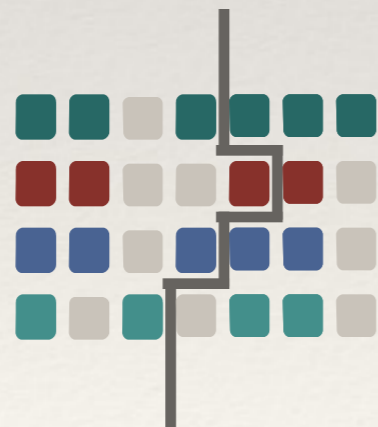
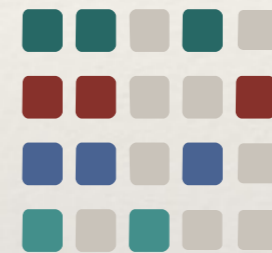
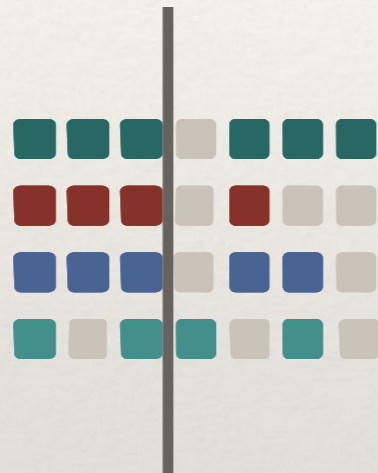


---

# Crossover

---

## Single Point Crossover

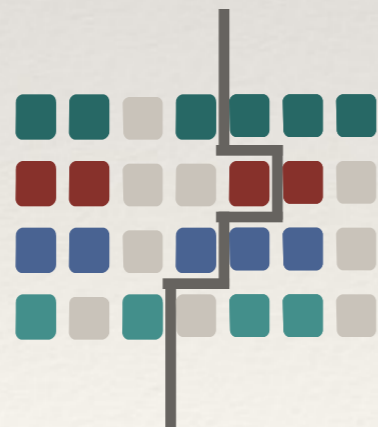
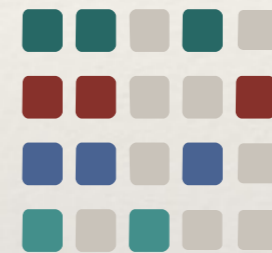
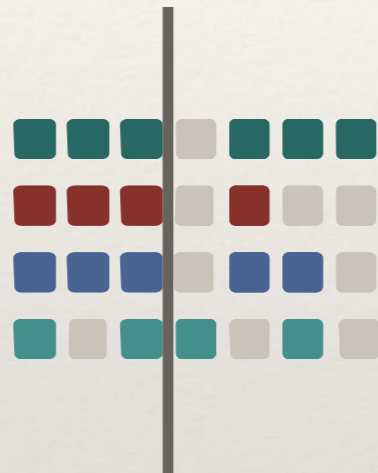


---

# Crossover

---

## Single Point Crossover



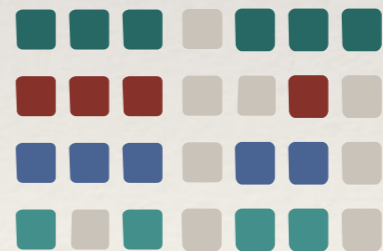
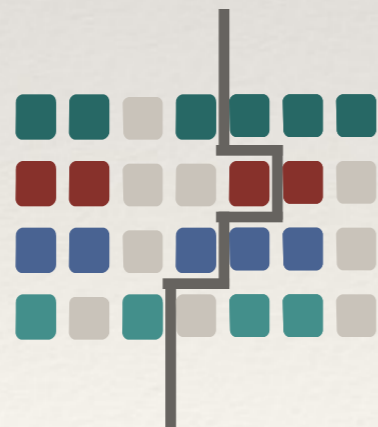
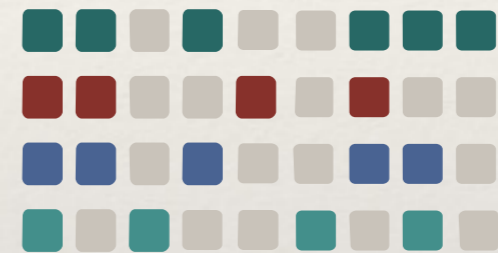
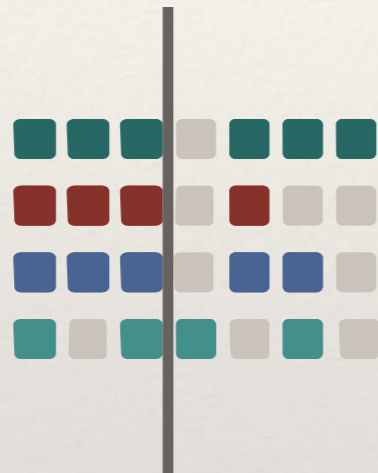


---

# Crossover

---

## Single Point Crossover

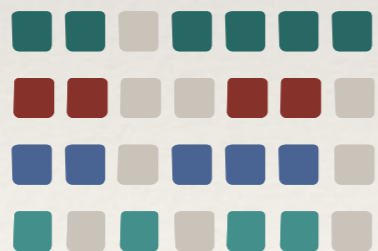
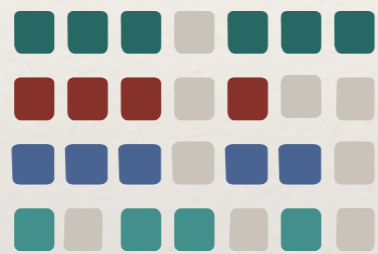


---

# Crossover

---

## Best Partial Alignment Crossover

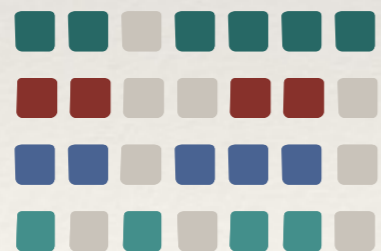
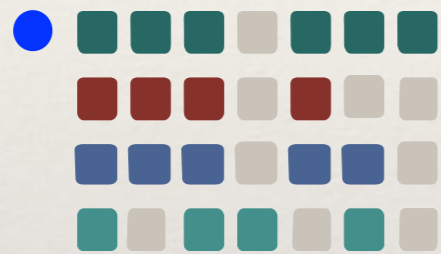


---

# Crossover

---

## Best Partial Alignment Crossover



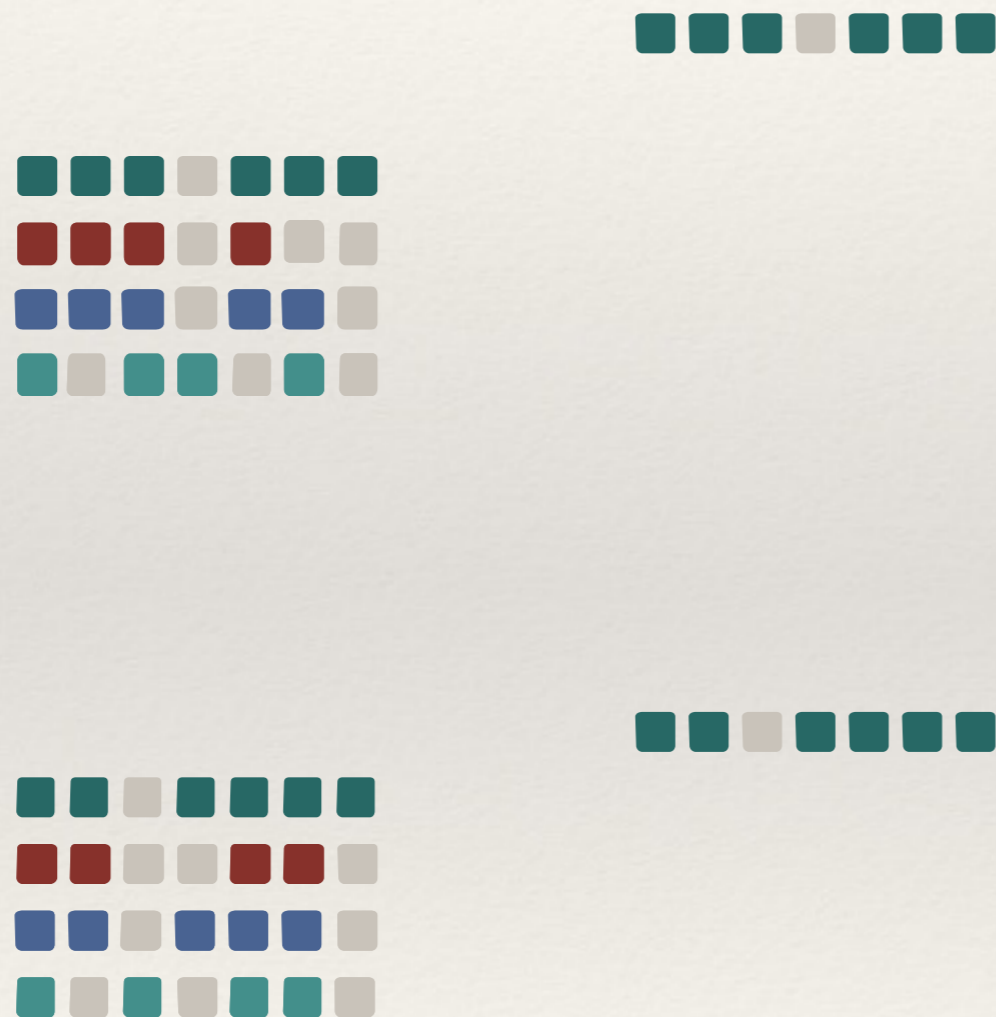


---

# Crossover

---

## Best Partial Alignment Crossover

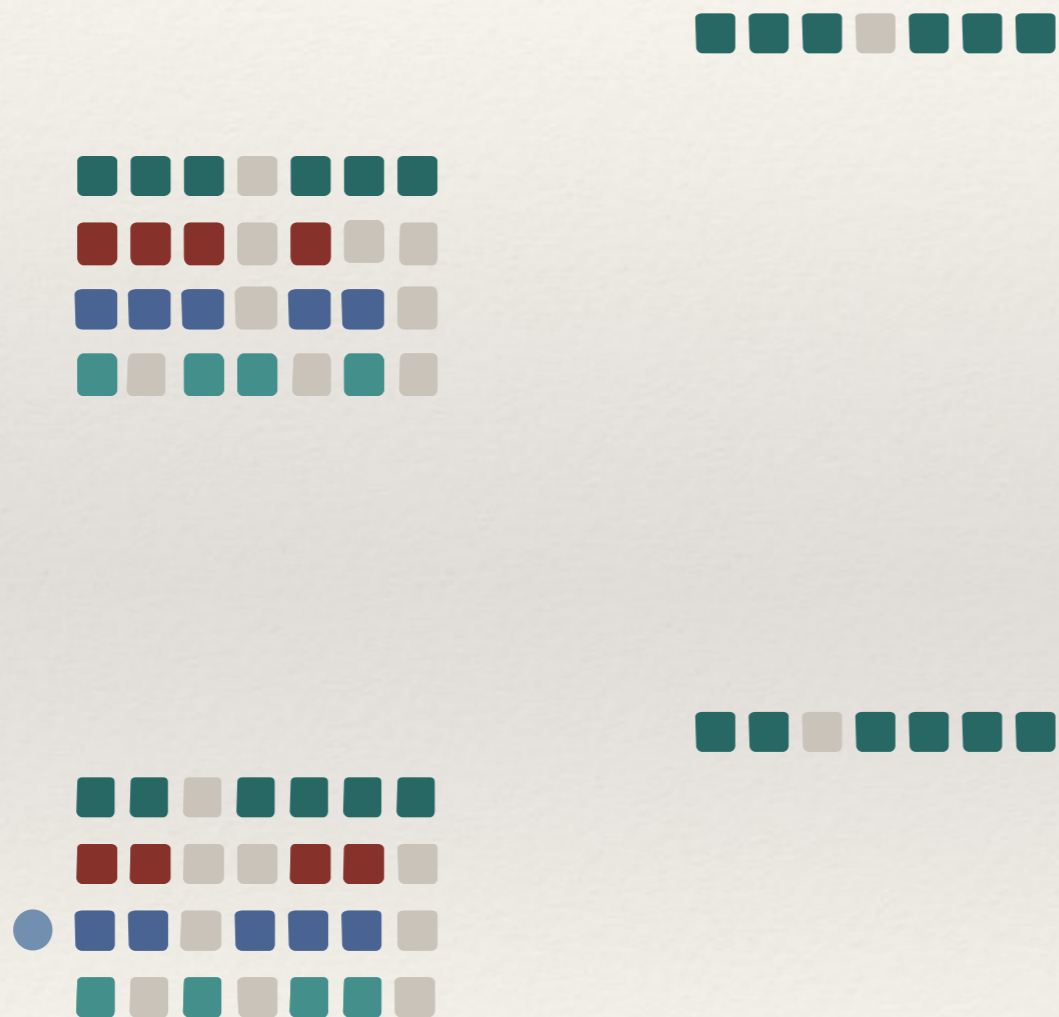


---

# Crossover

---

## Best Partial Alignment Crossover

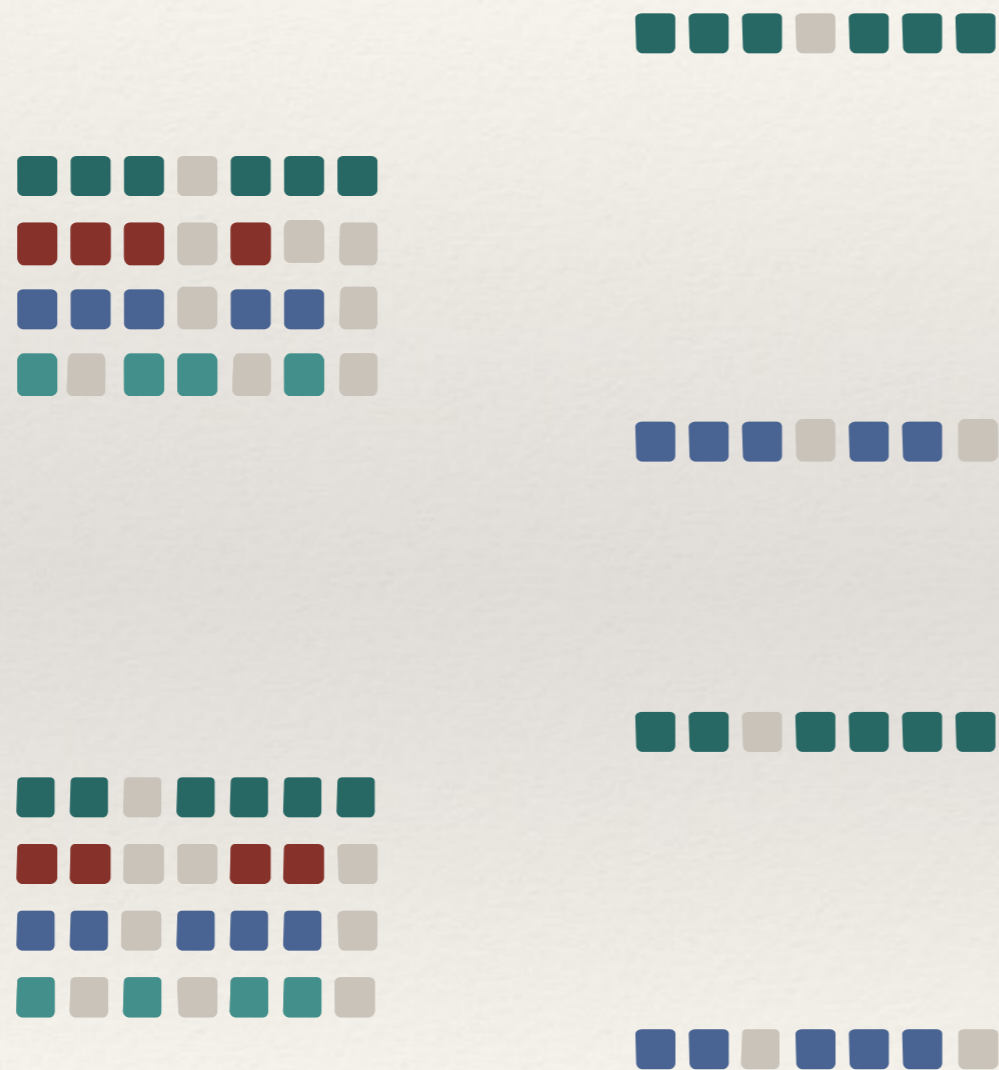


---

# Crossover

---

## Best Partial Alignment Crossover



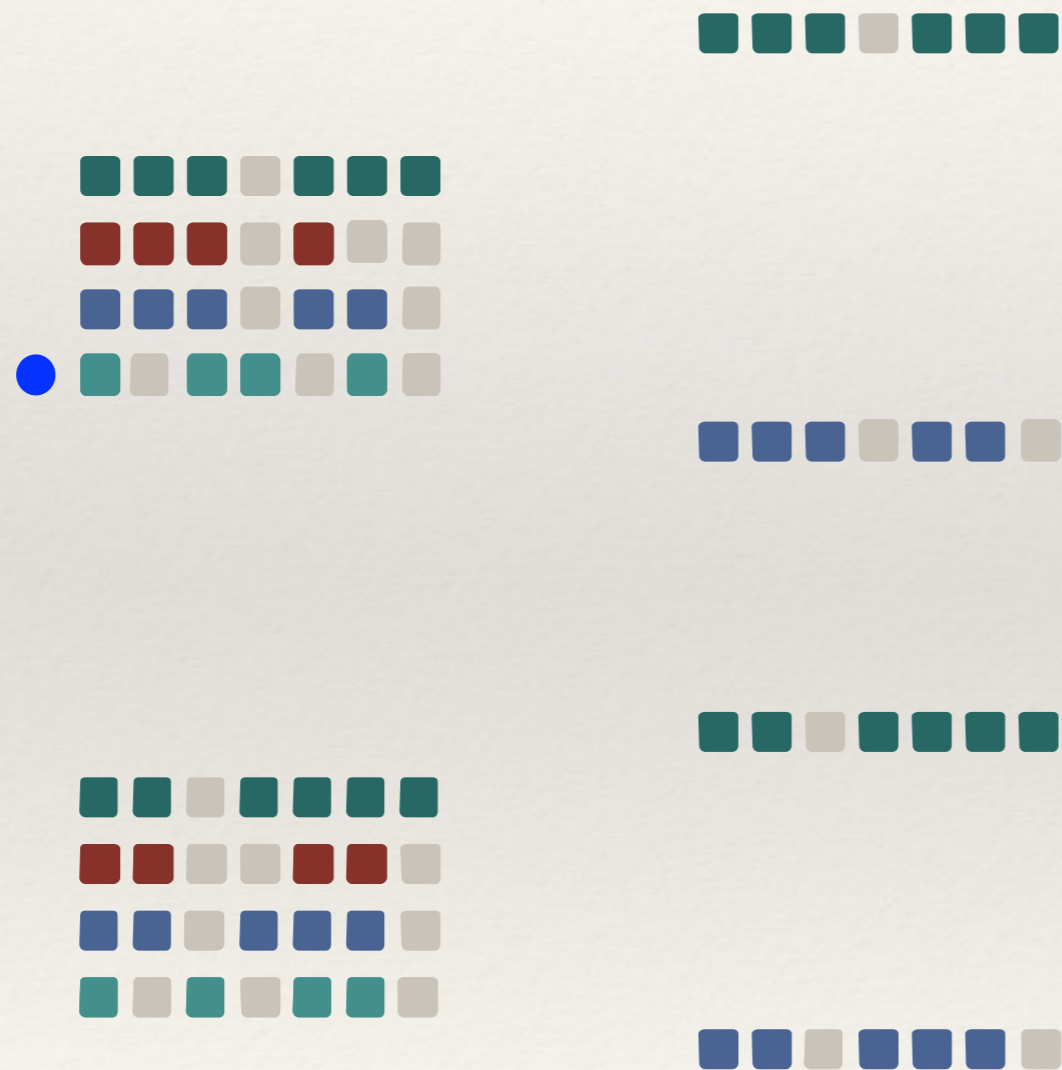


---

# Crossover

---

## Best Partial Alignment Crossover

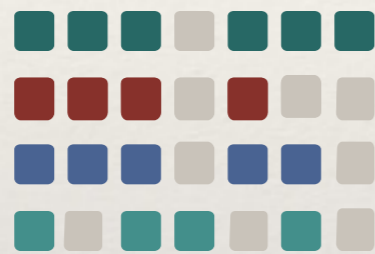


---

# Crossover

---

## Best Partial Alignment Crossover

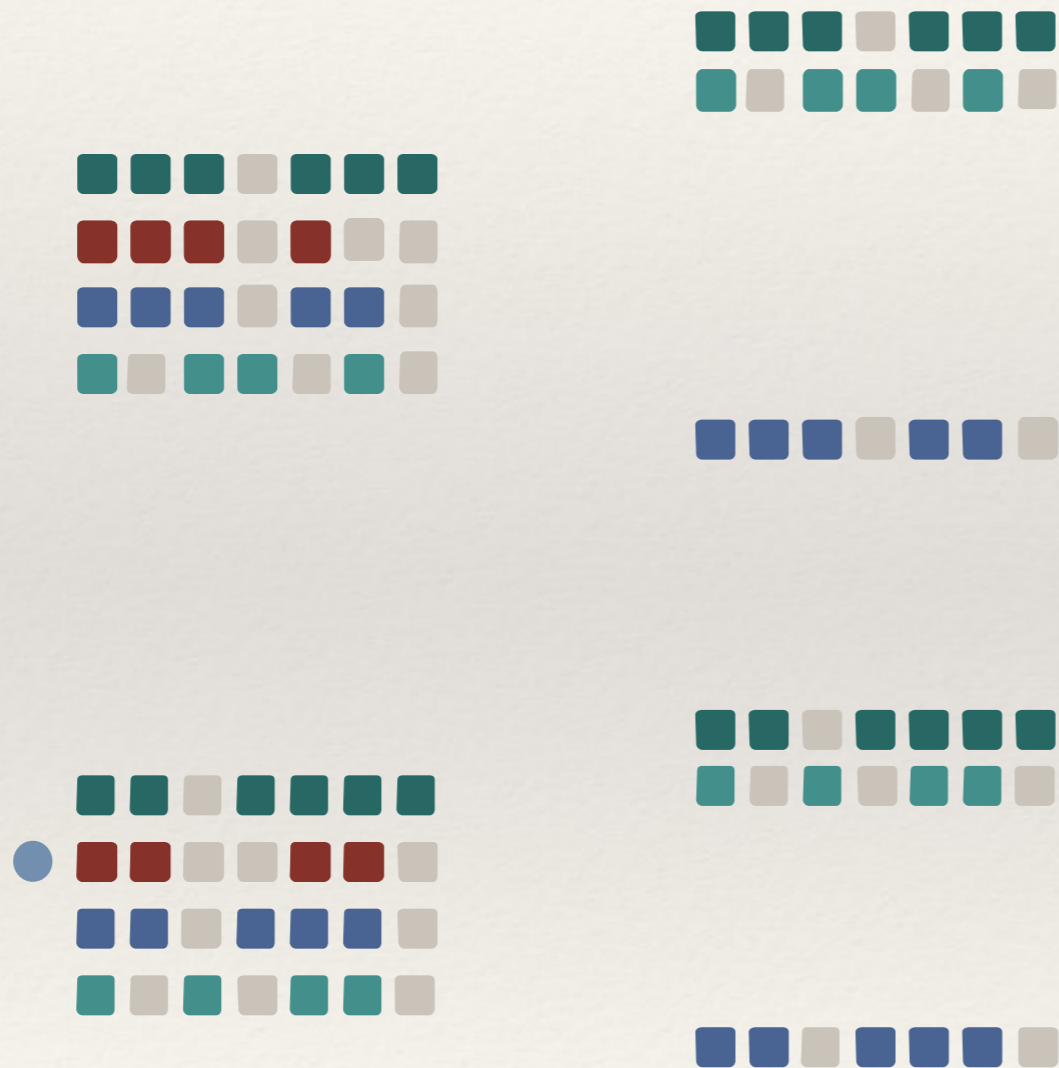


---

# Crossover

---

## Best Partial Alignment Crossover



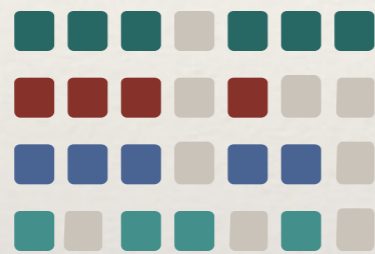


---

# Crossover

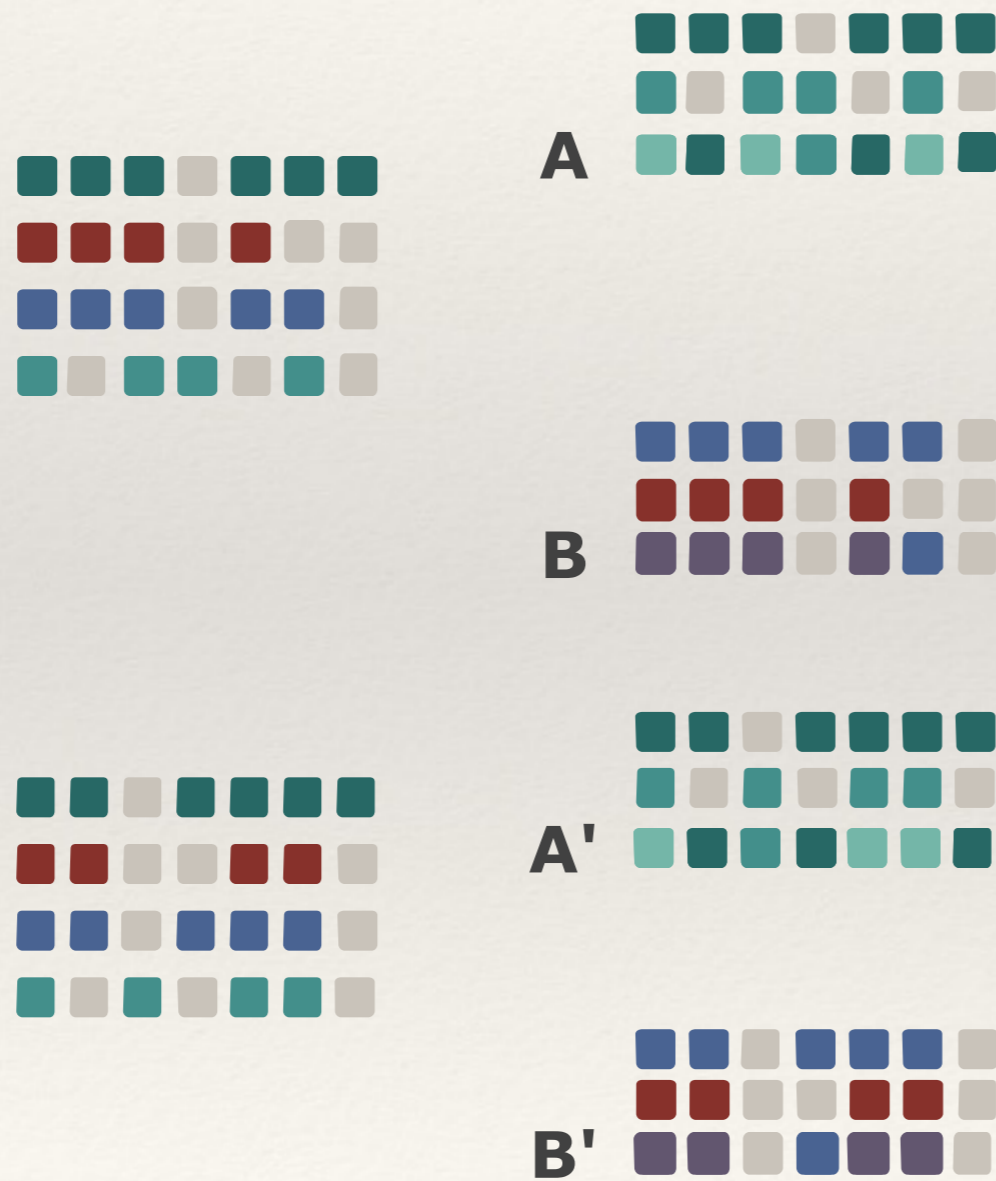
---

## Best Partial Alignment Crossover



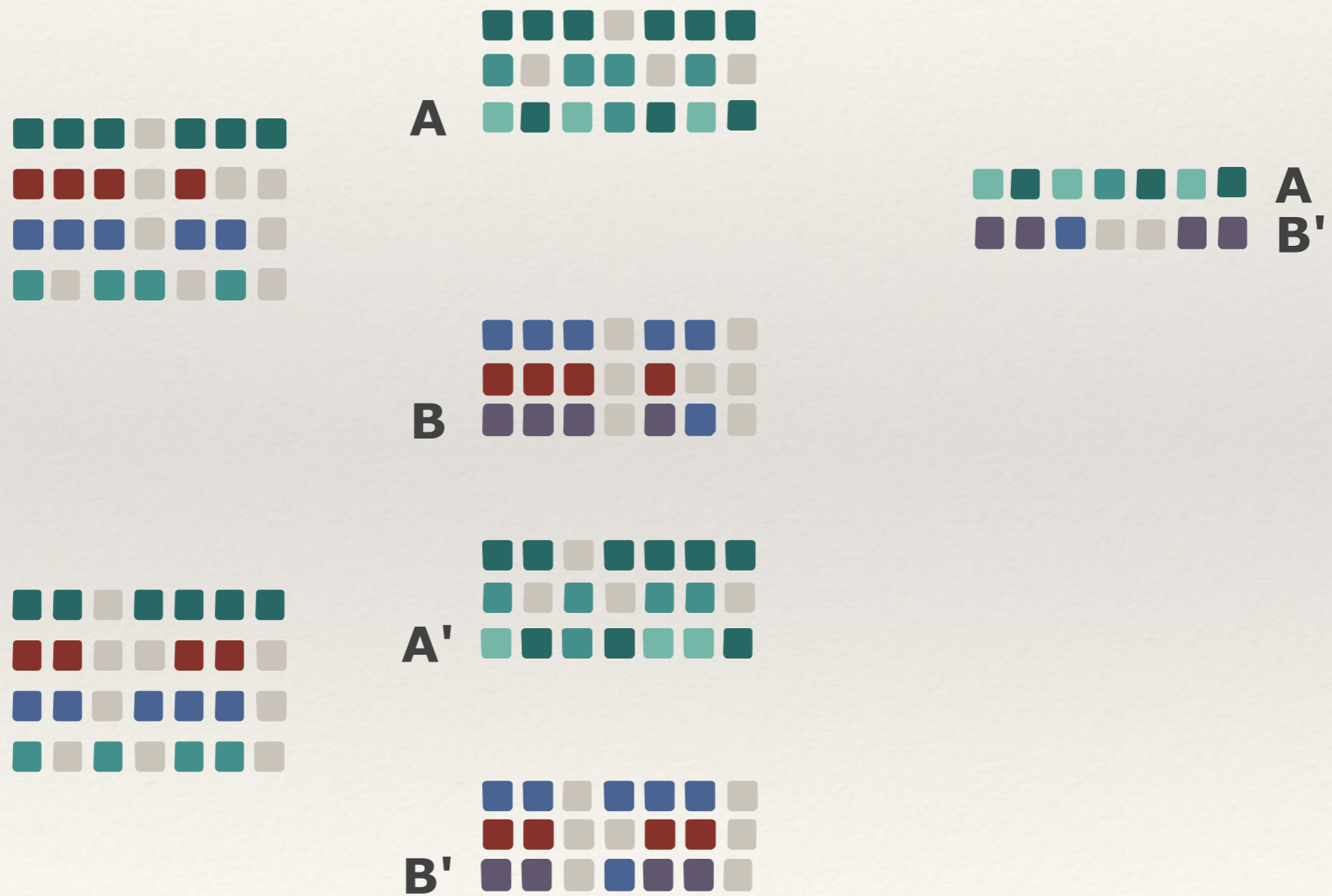
# Crossover

## Best Partial Alignment Crossover



# Crossover

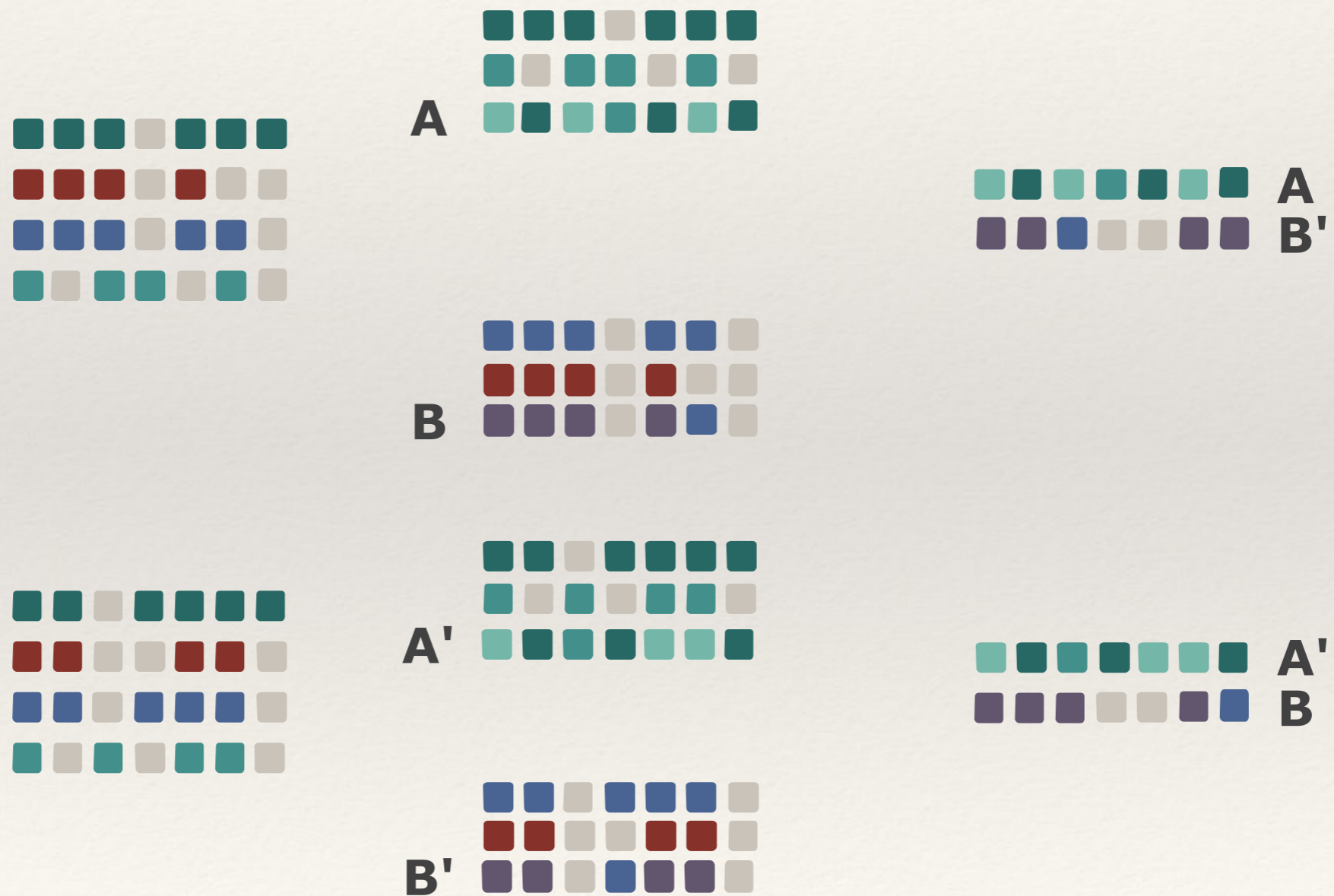
## Best Partial Alignment Crossover





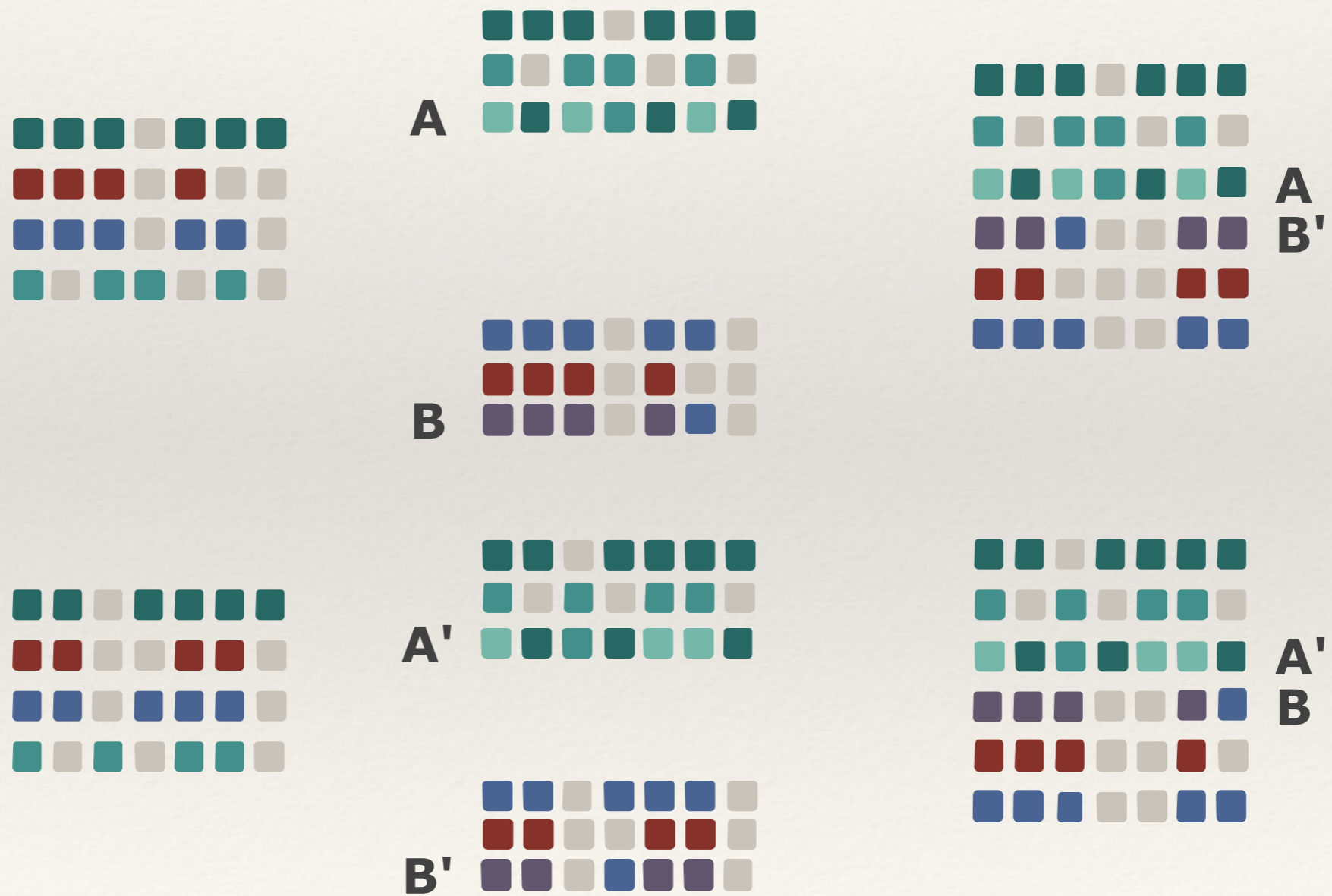
# Crossover

## Best Partial Alignment Crossover



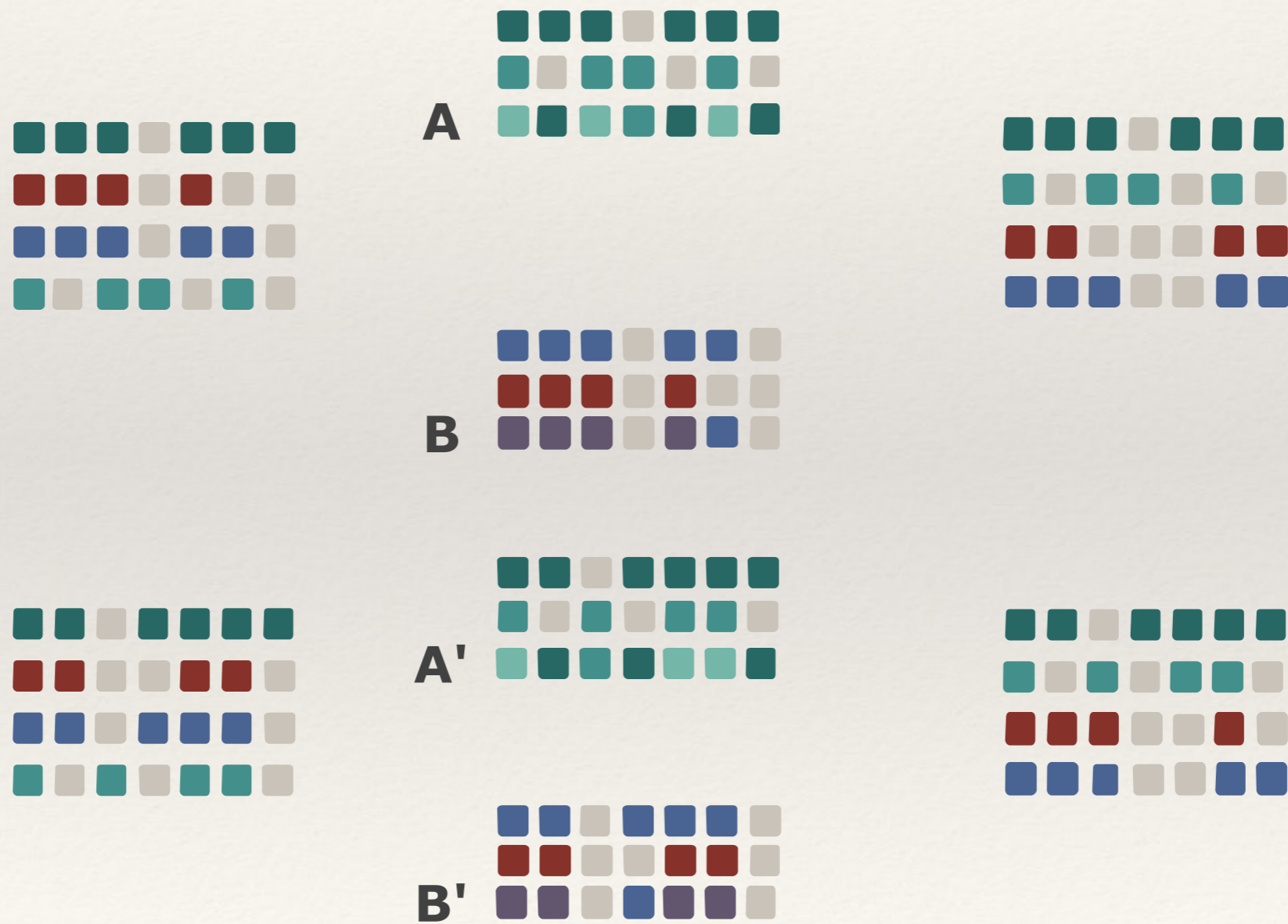
# Crossover

## Best Partial Alignment Crossover



# Crossover

## Best Partial Alignment Crossover



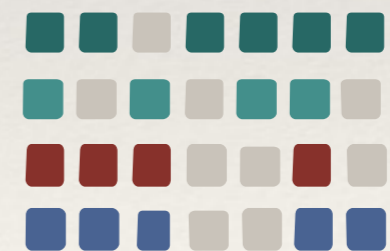
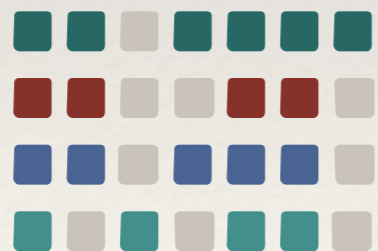
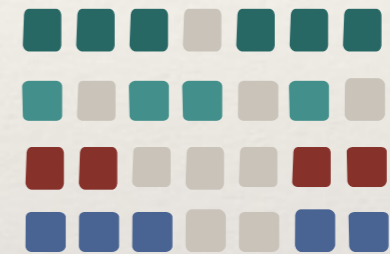
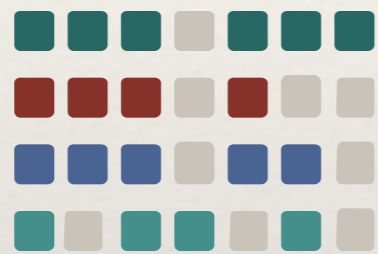


---

# Crossover

---

## Best Partial Alignment Crossover

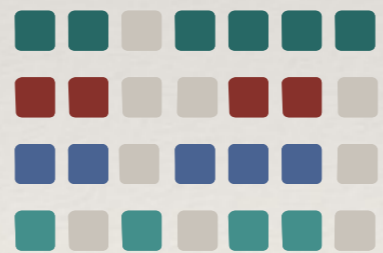
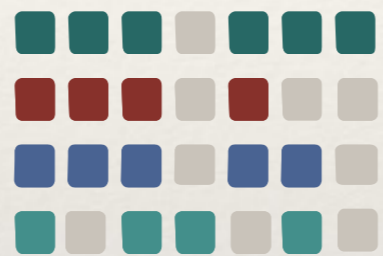


---

# Crossover

---

## Sequence Similarity Crossover

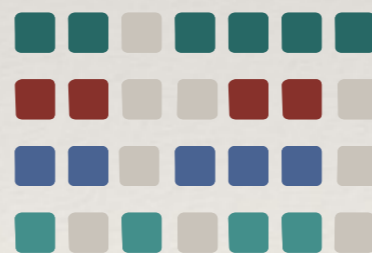
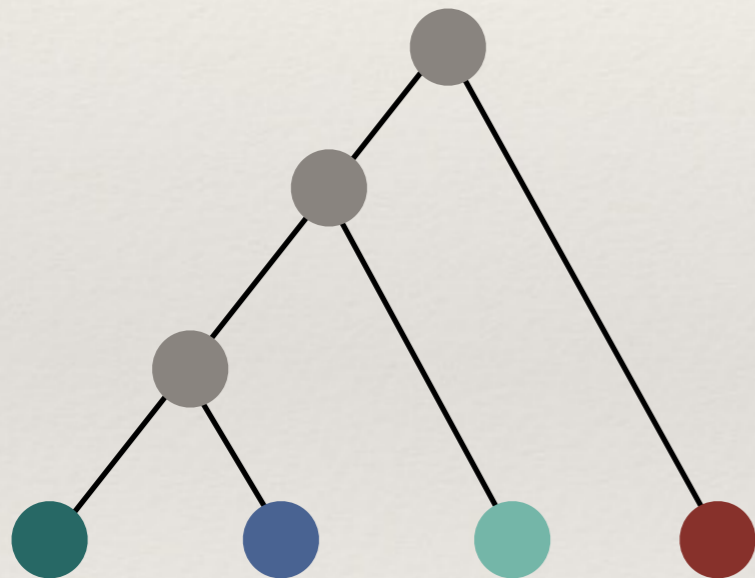


---

# Crossover

---

## Sequence Similarity Crossover



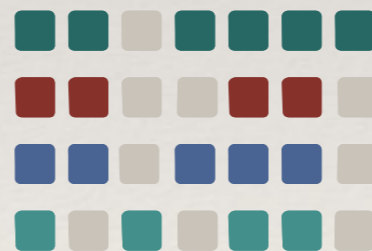
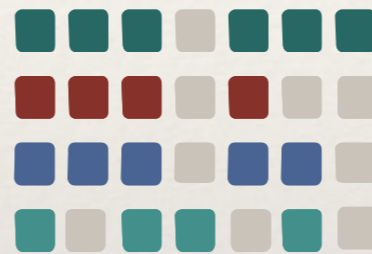
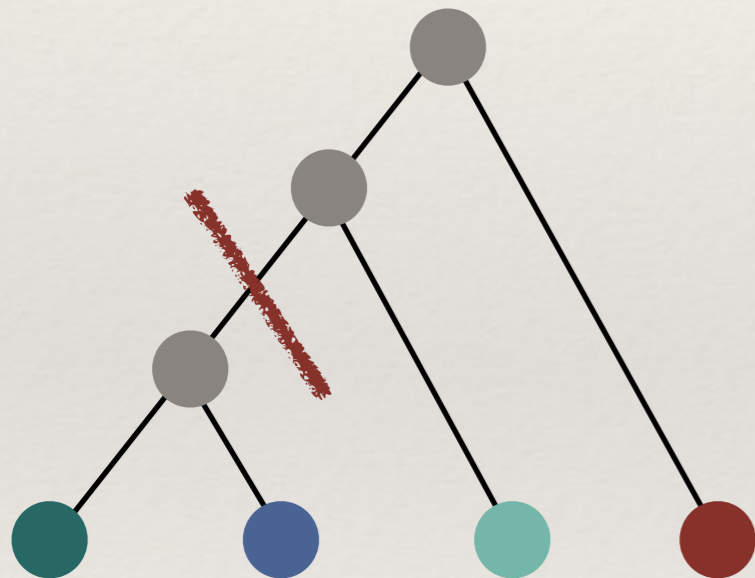


---

# Crossover

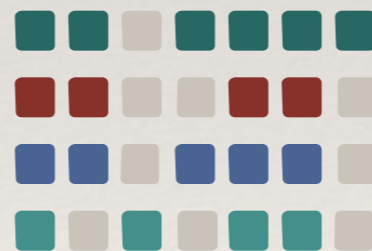
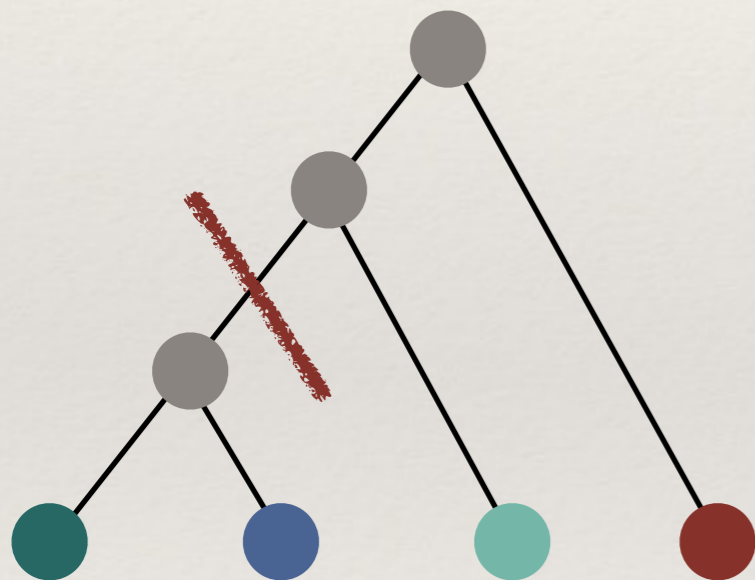
---

## Sequence Similarity Crossover



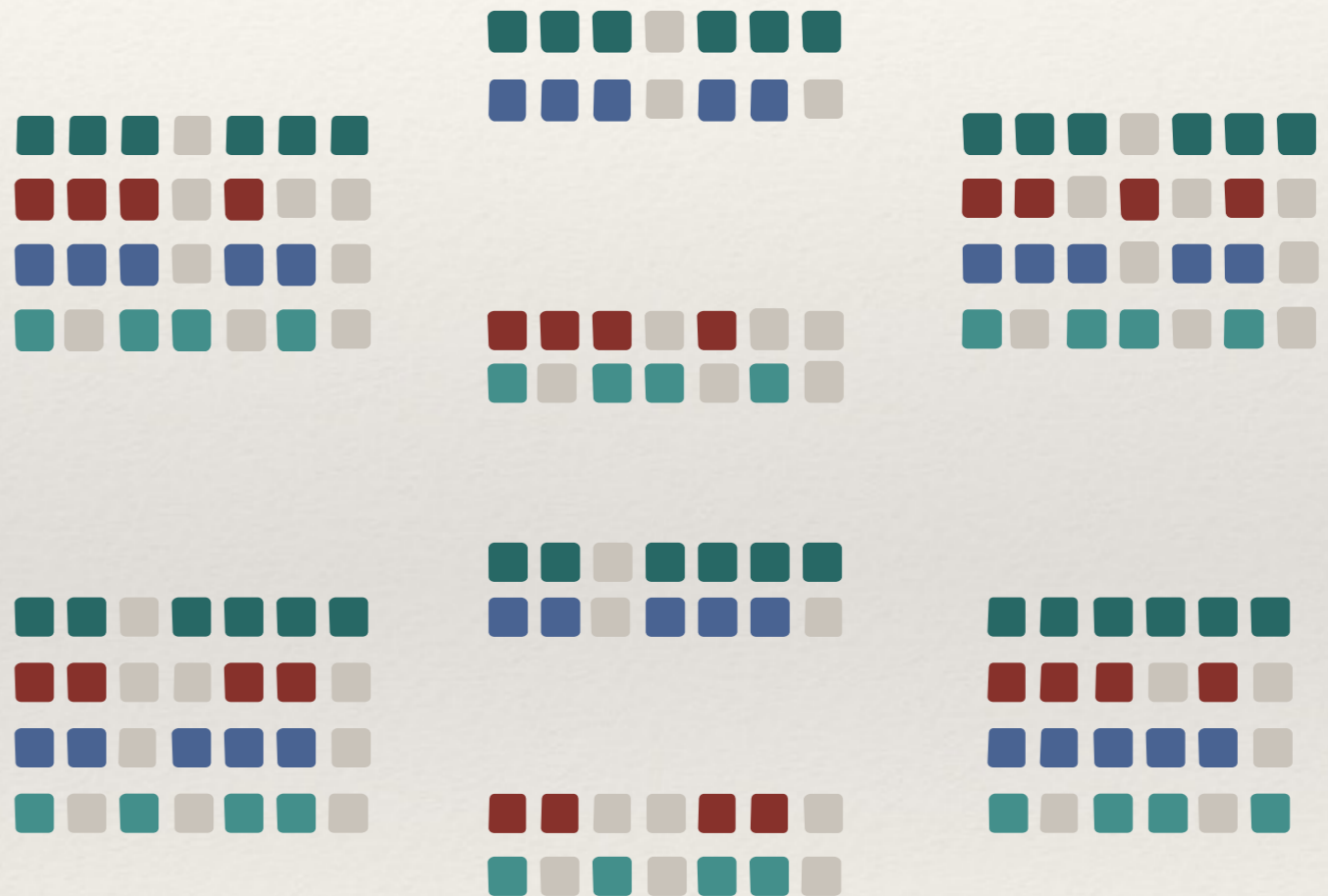
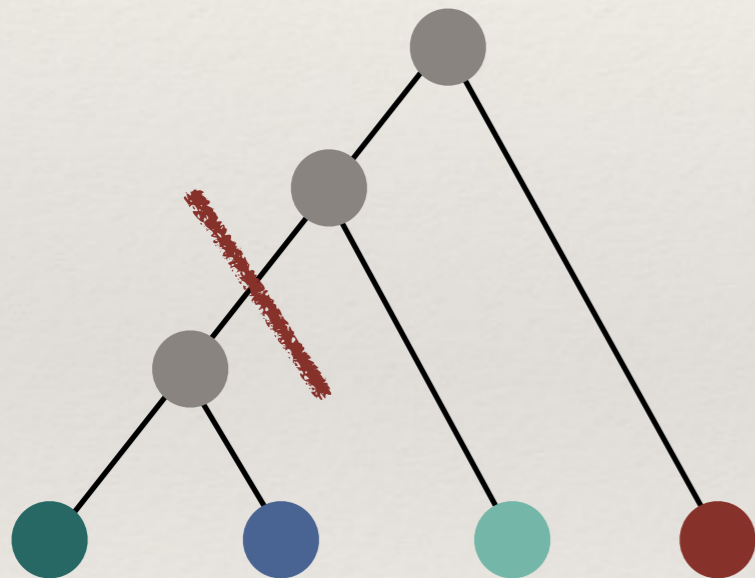
# Crossover

## Sequence Similarity Crossover



# Crossover

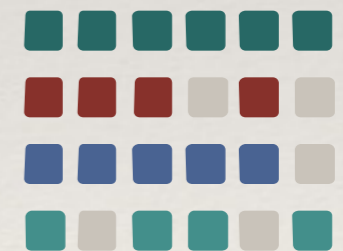
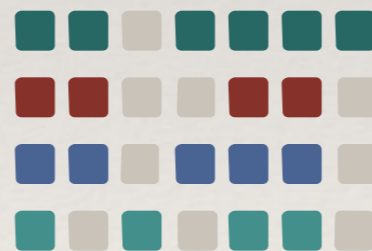
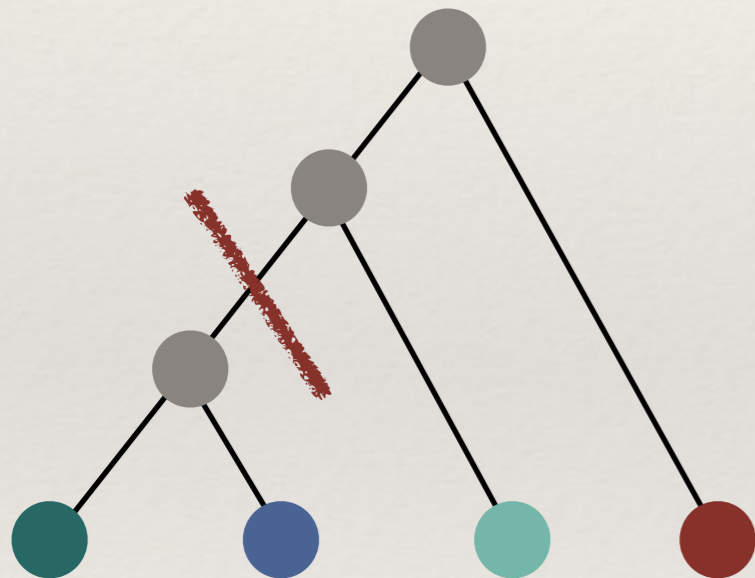
## Sequence Similarity Crossover





# Crossover

## Sequence Similarity Crossover



---

# Função Objetivo

---

- ❖ Sum of Pair Matrix Based Score
- ❖ Gap Affinity Sum of Pairs Matrix Based Score

---

# Função Objetivo

---

## Sum Of Pairs Matrix Based Score

$$fitness = \sum_{i=1}^m \sum_{j=1}^{n-1} \sum_{k=j+1}^n \sigma(seq_j^i, seq_k^i)$$



---

# Função Objetivo

---

## Gap Affinity Sum Of Pairs Matrix Based Score

$$score_{gap} = gop + n \times gep$$

$$\sigma(c_1, c_2) = \begin{cases} gop + gep & \text{se } c_1 \text{ ou } c_2 \text{ inicia um bloco de } gaps, \\ gep & \text{se } c_1 \text{ ou } c_2 \text{ estende um bloco de } gaps. \end{cases}$$

---

# Seleção

---

- ❖ Roulette Selection
- ❖ Most Fitted Selection

---

# Seleção

---

## Roulette Selection



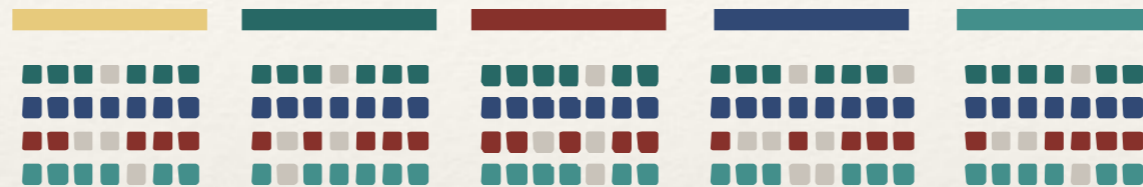


---

# Seleção

---

## Roulette Selection



---

# Seleção

---

## Roulette Selection

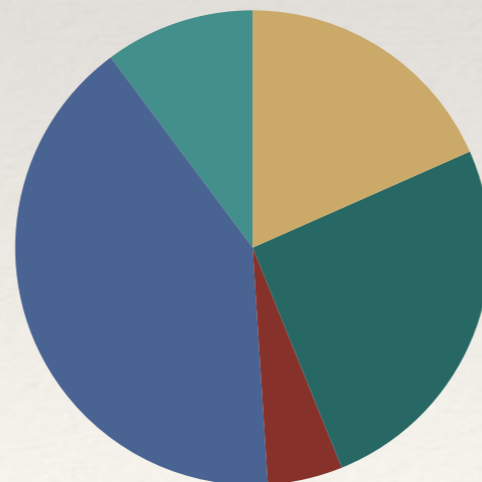
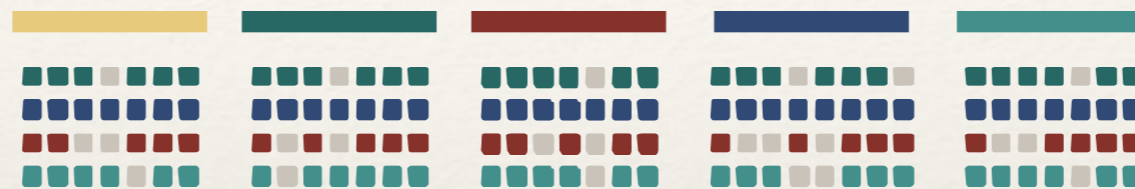


---

# Seleção

---

## Roulette Selection



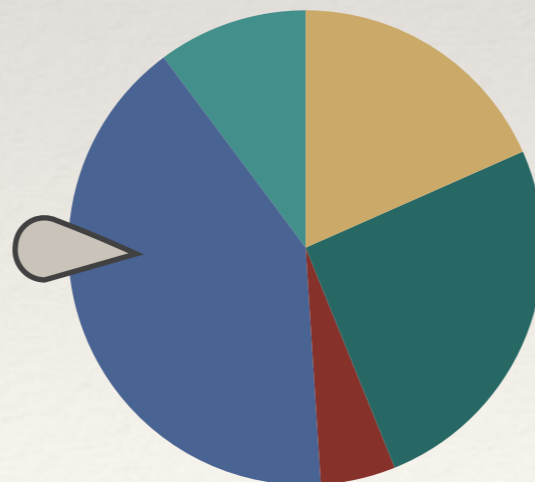
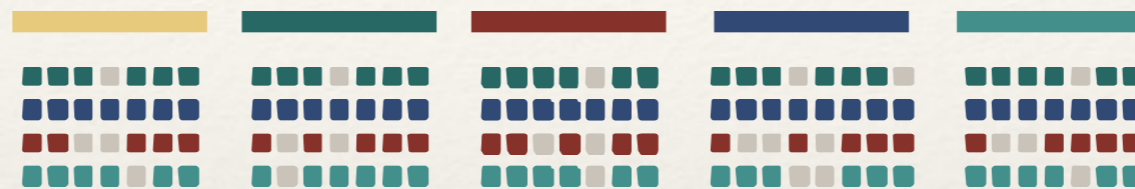


---

# Seleção

---

## Roulette Selection

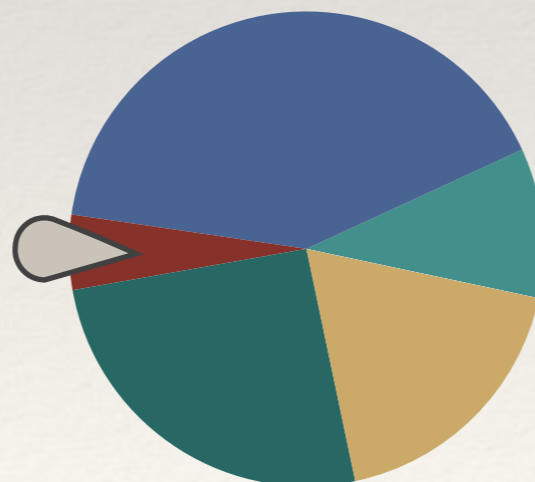
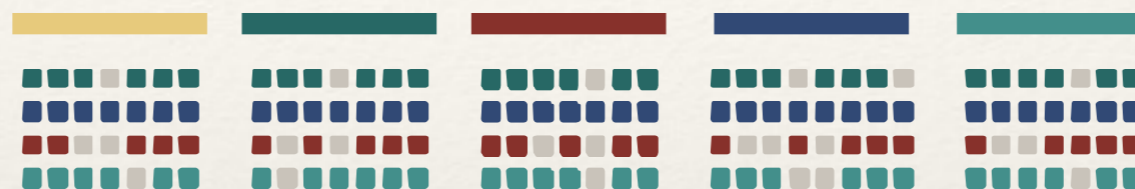


---

# Seleção

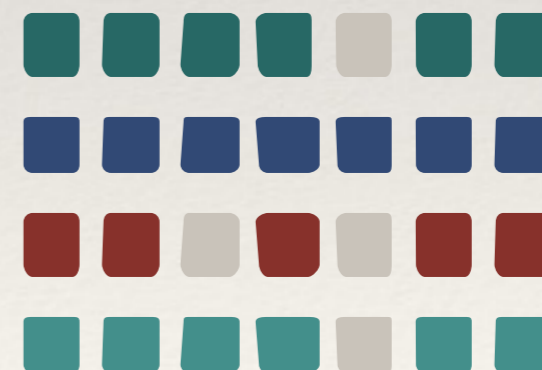
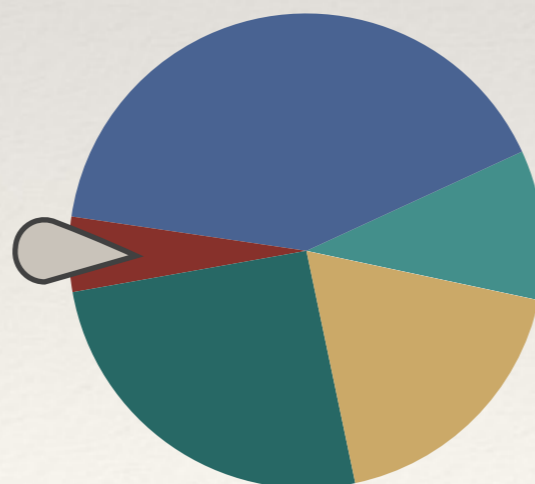
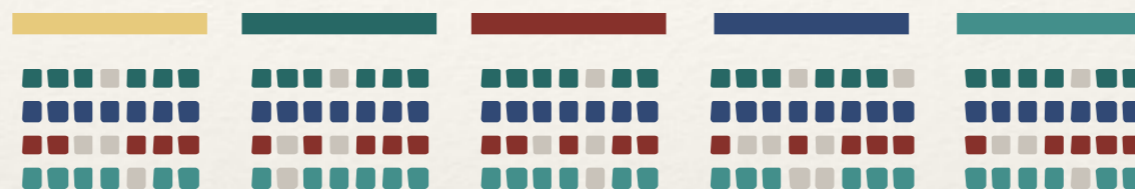
---

## Roulette Selection



# Seleção

## Roulette Selection





---

# Seleção

---

## Most Fitted Selection

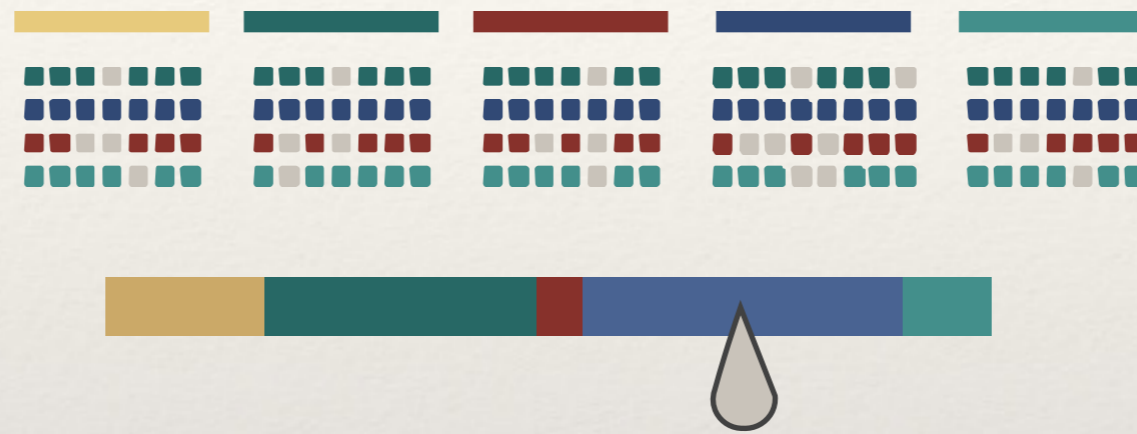


---

# Seleção

---

## Most Fitted Selection

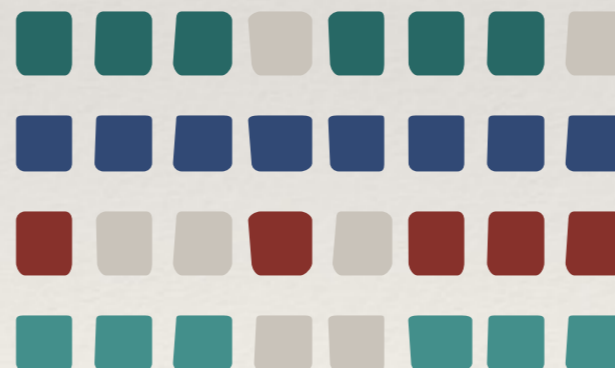
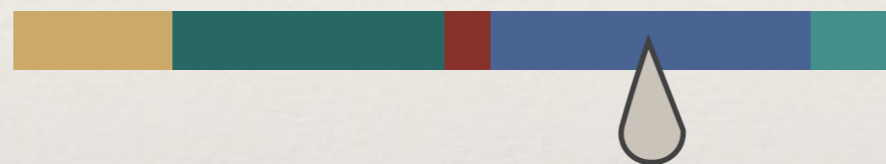
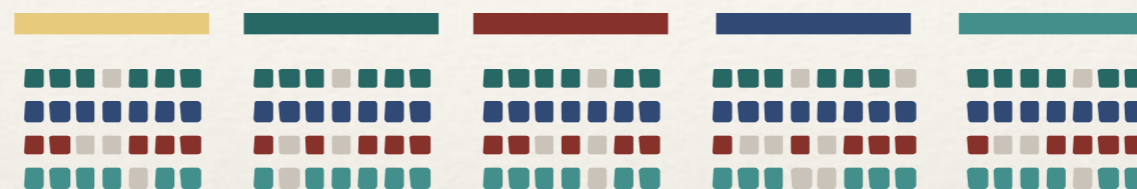


---

# Seleção

---

## Most Fitted Selection





---

# Primeiros Resultados

---

- ❖ GAADT: Santos, 1998
  - ❖ Mesmo escopo
    - ❖ Mestrado
    - ❖ Algoritmos Genéticos
  - ❖ Código fonte não disponível
    - ❖ Poucos casos de teste para comparação

# Primeiros Resultados

	1aab	1fjlA	1hpi	1csy	1tgxA	média
GAADT	88,1%	81,4%	70,8%	70,3%	69,2%	76,0%
ALGAe (média)	88,4%	93,8%	88,3%	76,2%	68,5%	83,0%
ALGAe (máximo)	89,6%	100,0%	96,2%	80,7%	77,2%	88,7%

fonte: Santos, 1998

# Primeiros Resultados

BAliBASE	T-COFFEE	Muscle	ClustalW	ALGAe
RV11 (SP)	47,1%	47,9%	43,4%	38,1%
RV12 (SP)	85,5%	84,7%	81,7%	76,1%

fonte: Hang, 2008



# Melhorias

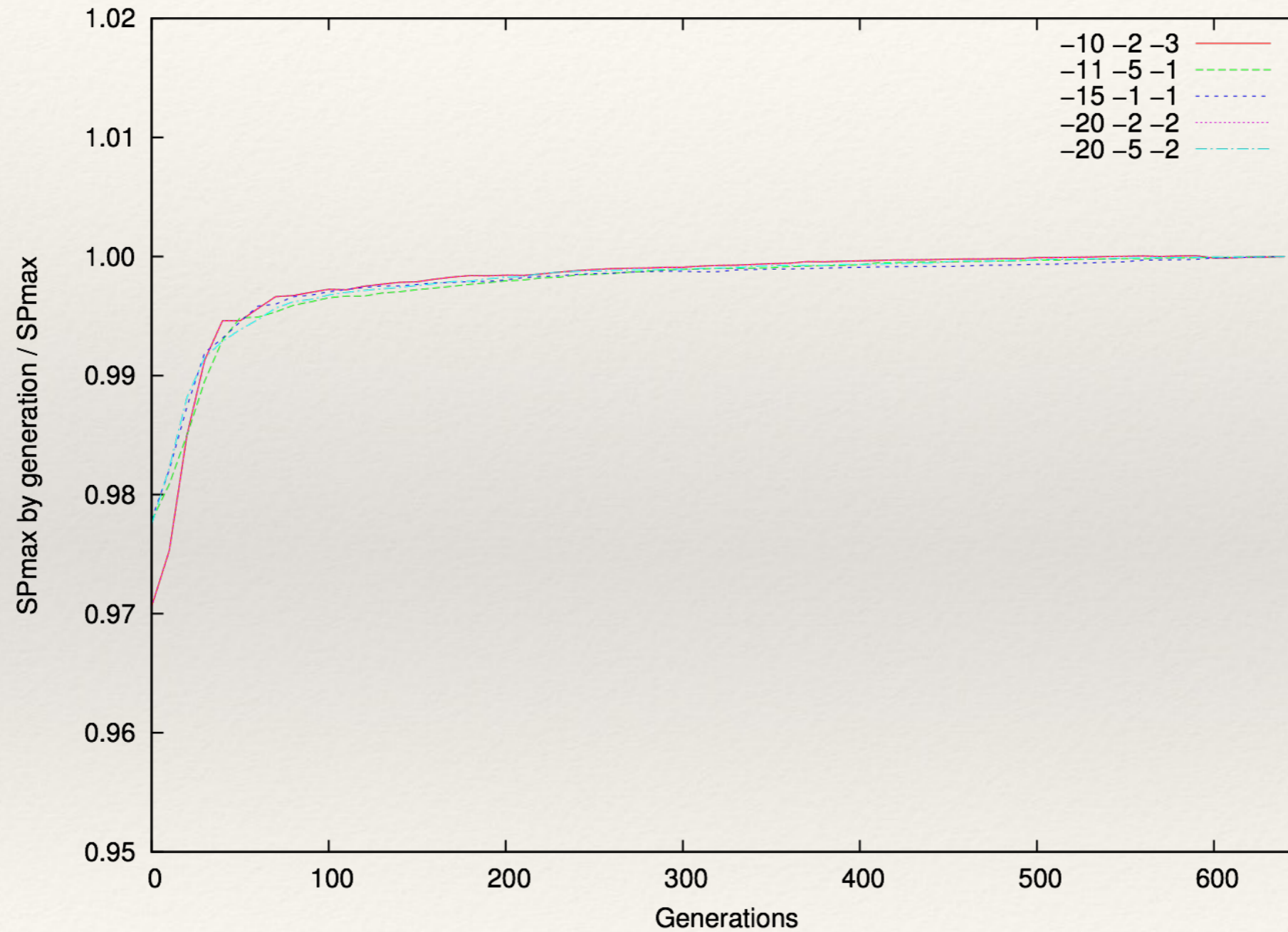
---

# Análise Inicial

---

$$m(gen) = \frac{SP_{max}(gen)}{SP_{max}(n)}$$

# Análise Inicial





---

# Penalização de Gaps

---

- ❖ Trabalhos relacionados
  - ❖ Agrawal e colegas (2008) e Reese e Pearson (2002)
  - ❖ Buscas em bancos de sequências

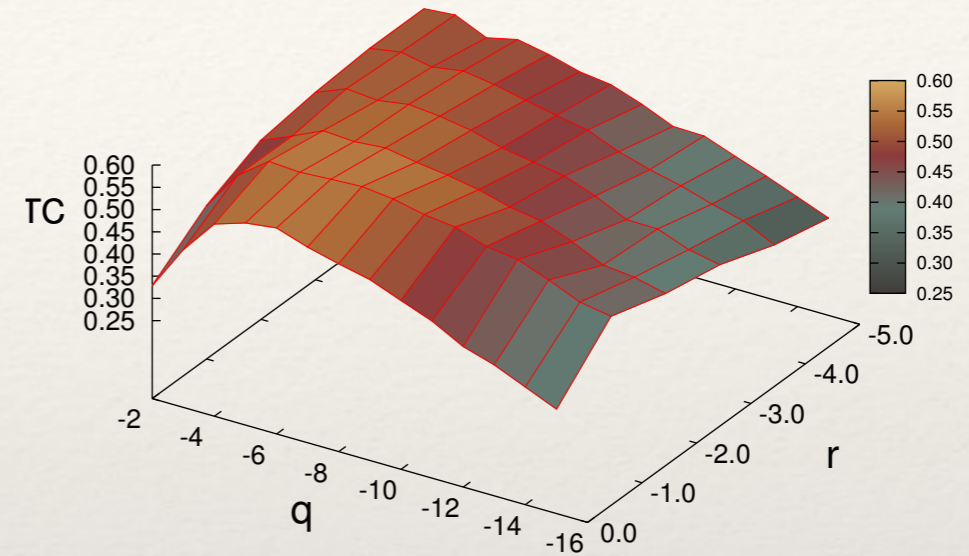
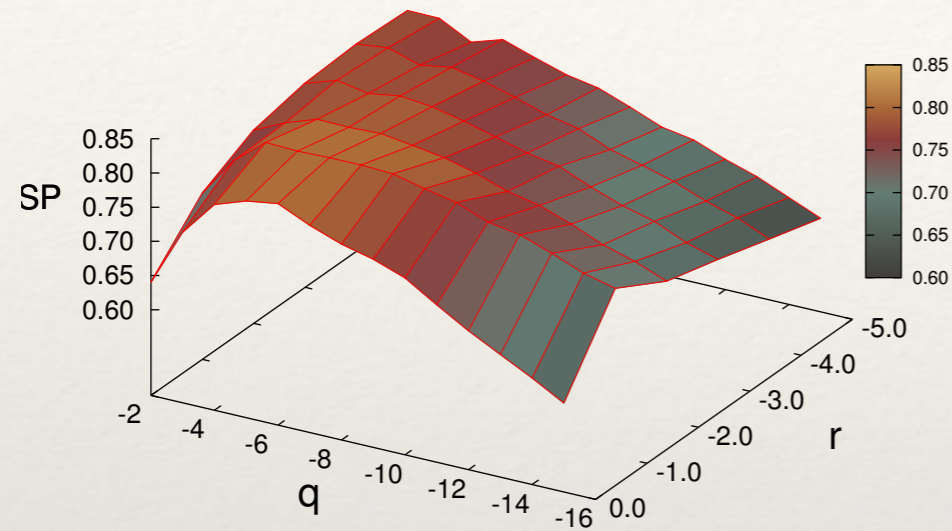
---

# Penalização de Gaps

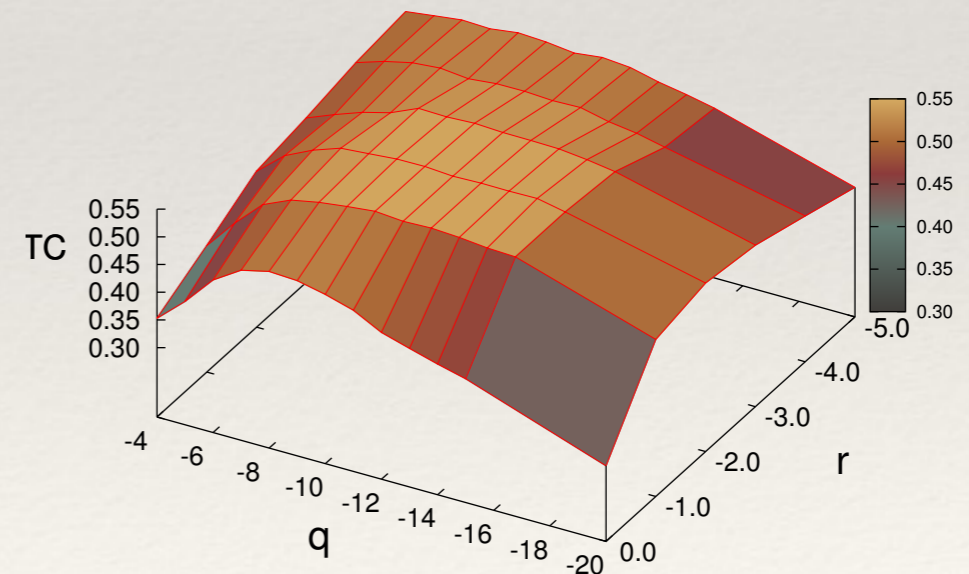
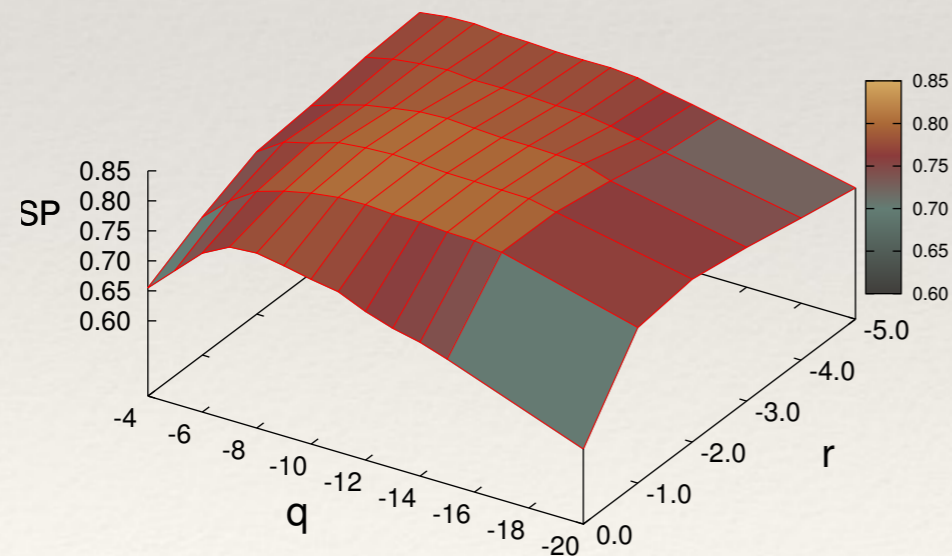
---

- ❖ Penalizações de Gaps
  - ❖ Abertura ( $q$ )
  - ❖ Extensão ( $r$ )
- ❖ BLOSUM62 e BLOSUM80
- ❖ BAliBASE 3 - RV12

# Penalização de Gaps



BLOSUM62



BLOSUM80



# Penalização de Gaps

	Melhor resultado (q,r)	1 - Agrawal e colegas (q,r)	2 - Reese e Pearson (q,r)	Melhoria em relação a 1 (%)	Melhoria em relação a 2 (%)
BLOSUM62	(5,1)	(11,2)	(9,5)	4,0	14,0
BLOSUM80	(9,2)	(13,2)	(13,5)	0,7	6,0

---

# Resultados

---

- ❖ Experimentos inconclusivos
  - ❖ Seleção de matriz par a par
  - ❖ Função objetivo baseada em estrutura

---

# Resultados

---

- ❖ Possíveis melhorias
  - ❖ Ferramental estatístico
  - ❖ Comparação visual contra *benchmark*



---

# Resultados

---

- ❖ Problema
  - ❖ Visualizadores existentes
    - ❖ Poucas ferramentas (comparação de alinhamentos)
    - ❖ Não permitiam as visualizações desejadas
    - ❖ Difícil configuração / customização

# Anubis

---

# Anubis

---





---

# Anubis

---

- ❖ Comparador visual
  - ❖ *Benchmark* x alinhamento alvo
  - ❖ Java
  - ❖ Extensível / flexível

---

# Anubis

---

- ❖ *Plug-ins*
  - ❖ Baixo acomplamento
  - ❖ Mensagens
- ❖ Visualizadores
  - ❖ Camadas

# Anubis

The screenshot displays the Anubis software interface. The main window shows two sequence alignment panels. The top panel aligns sequences 1BBT3, 1AYM3, 1BBT2, 1AYM1, and 1BBT1 from position 120 to 140. The bottom panel aligns the same sequences from position 130 to 150. A central sequence viewer shows the sequence: Q L G T N P H A F W D P G L N P T T R F S I P Y I S. The right sidebar contains settings for alignment, including 'Default', 'Compressed Alphabet', 'Column Matches', 'Secondary Structure', and 'Success %'. The bottom status bar shows 'Sum of Pairs with Gap Affinity' with values 2059.0 and 2069.0.

File View Plugins Help

120 130 140

1BBT3 A H C I - - H A E W D T G L N S K F T F S I P Y L S A  
1AYM3 M L G T - - H V V W D V G L Q S T V S L V V P W I S A  
1BBT2 T L F P - - H Q F I N P R T N M T A H I T V P F V G V  
1AYM1 Q S G T N A S V F W Q H G - Q P F P R F S L P F L S I  
1BBT1 N T - T N P T A Y H K - - - A P L T R L A L P Y T A P

Q L G T N P H A F W D P G L N P T T R F S I P Y I S

130 140 150

1BBT3 - - C I H A - E - W D T G L N S K F T F S I P Y L S -  
1AYM3 G - - T H V - V - W D V G L Q S T V S L V V P W I S -  
1BBT2 - - F P H Q - F - I N P R T N M T A H I T V P F V G -  
1AYM1 G - - T N A S V F W Q H G - Q P F P R F S L P F L S I  
1BBT1 Y T A P H R - V - L A T V Y N G E C R Y - - - - - S -

G T C P H A S V F W D T G L N S T C R F S I P F I S

message Sum of Pairs with Gap Affinity 2059.0 2069.0



# Anubis

The screenshot displays the Anubis software interface. The main window shows a sequence alignment of five sequences: 1BBT3, 1AYM3, 1BBT2, 1AYM1, and 1BBT1. The alignment is presented in a grid format with columns color-coded by amino acid type. A red oval highlights the top portion of the alignment, specifically the first 140 columns. Below the main alignment, a secondary alignment is visible, showing a different set of sequences. The interface includes a menu bar (File, View, Plugins, Help), a toolbar (Left, Bottom, Right, Top), and a settings panel on the right. The settings panel includes options for 'Default', 'Compressed Alphabet', 'Column Matches', 'Secondary Structure', and 'Success %'. The bottom status bar shows 'Sum of Pairs with Gap Affinity' with values 2059.0 and 2069.0.

Sequence	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	
1BBT3	A	H	C	I	-	-	H	A	E	W	D	T	G	L	N	S	K	F	T	F	S	I	P	Y	L	S	A
1AYM3	M	L	G	T	-	-	H	V	V	W	D	V	G	L	Q	S	T	V	S	L	V	V	P	W	I	S	A
1BBT2	T	L	F	P	-	-	H	Q	F	I	N	P	R	T	N	M	T	A	H	I	T	V	P	F	V	G	V
1AYM1	Q	S	G	T	N	A	S	V	F	W	Q	H	G	-	Q	P	F	P	R	F	S	L	P	F	L	S	I
1BBT1	N	T	-	T	N	P	T	A	Y	H	K	-	-	-	A	P	L	T	R	L	A	L	P	Y	T	A	P

Q L G T N P H A F W D P G L N P T T R F S I P Y I S

Sequence	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150						
1BBT3	-	-	C	I	H	A	-	E	-	W	D	T	G	L	N	S	K	F	T	F	S	I	P	Y	L	S	-
1AYM3	G	-	-	T	H	V	-	V	-	W	D	V	G	L	Q	S	T	V	S	L	V	V	P	W	I	S	-
1BBT2	-	-	F	P	H	Q	-	F	-	I	N	P	R	T	N	M	T	A	H	I	T	V	P	F	V	G	-
1AYM1	G	-	-	T	N	A	S	V	F	W	Q	H	G	-	Q	P	F	P	R	F	S	L	P	F	L	S	I
1BBT1	Y	T	A	P	H	R	-	V	-	L	A	T	V	Y	N	G	E	C	R	Y	-	-	-	-	-	-	S

G T C P H A S V F W D T G L N S T C R F S I P F I S

message Sum of Pairs with Gap Affinity 2059.0 2069.0

# Anubis

The screenshot shows the Anubis software interface. The main window displays a sequence alignment for five sequences: 1BBT3, 1AYM3, 1BBT2, 1AYM1, and 1BBT1. The alignment is shown in a grid format with columns numbered 120, 130, 140, and 150. The sequences are color-coded by column. A red oval highlights a region of the alignment, specifically the area between columns 120 and 150 for sequences 1BBT3, 1AYM3, 1BBT2, 1AYM1, and 1BBT1. The interface includes a menu bar (File, View, Plugins, Help), a toolbar (Left, Bottom, Right, Top), and a status bar at the bottom showing 'Sum of Pairs with Gap Affinity' and values 2059.0 and 2069.0.

Sequence	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150
1BBT3	A	H	C	I	-	-	H	A	E	W	D	T	G	L	N	S	K	F	T	F	S	I	P	Y	L	S	A	-	-	-	-
1AYM3	M	L	G	T	-	-	H	V	V	W	D	V	G	L	Q	S	T	V	S	L	V	V	P	W	I	S	A	-	-	-	-
1BBT2	T	L	F	P	-	-	H	Q	F	I	N	P	R	T	N	M	T	A	H	I	T	V	P	F	V	G	V	-	-	-	-
1AYM1	Q	S	G	T	N	A	S	V	F	W	Q	H	G	-	Q	P	F	P	R	F	S	L	P	F	L	S	I	-	-	-	-
1BBT1	N	T	-	T	N	P	T	A	Y	H	K	-	-	-	A	P	L	T	R	L	A	L	P	Y	T	A	P	-	-	-	-

Q L G T N P H A F W D P G L N P T T R F S I P Y I S

Sequence	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150
1BBT3	-	-	C	I	H	A	-	E	-	W	D	T	G	L	N	S	K	F	T	F	S	I	P	Y	L	S	-	-	-	-	-
1AYM3	G	-	-	T	H	V	-	V	-	W	D	V	G	L	Q	S	T	V	S	L	V	V	P	W	I	S	-	-	-	-	-
1BBT2	-	-	F	P	H	Q	-	F	-	I	N	P	R	T	N	M	T	A	H	I	T	V	P	F	V	G	-	-	-	-	-
1AYM1	G	-	-	T	N	A	S	V	F	W	Q	H	G	-	Q	P	F	P	R	F	S	L	P	F	L	S	I	-	-	-	-
1BBT1	Y	T	A	P	H	R	-	V	-	L	A	T	V	Y	N	G	E	C	R	Y	-	-	-	-	-	S	-	-	-	-	-

U T C P H A S V F W D T G L N S T C R F S I P E L S

message Sum of Pairs with Gap Affinity 2059.0 2069.0





# Anubis

The screenshot displays the Anubis software interface. The main window shows a sequence alignment of five sequences: 1BBT3, 1AYM3, 1BBT2, 1AYM1, and 1BBT1. The alignment is presented in a grid format with colored cells representing matches and gaps. The interface includes a menu bar (File, View, Plugins, Help), a toolbar (Left, Bottom, Right, Top), and a status bar at the bottom. The status bar shows 'message' and 'Sum of Pairs with Gap Affinity' with values 2059.0 and 2069.0.

	120	130	140	
1BBT3	A H C I - -	H A E W D T G L N S K F T F S I P Y L S A		
1AYM3	M L G T - -	H V V W D V G L Q S T V S L V V P W I S A		
1BBT2	T L F P - -	H Q F I N P R T N M T A H I T V P F V G V		
1AYM1	Q S G T N A S V F W Q H G -	Q P F P R F S L P F L S I		
1BBT1	N T - T N P T A Y H K - - -	A P L T R L A L P Y T A P		

Q L G T N P H A F W D P G L N P T T R F S I P Y I S

	130	140	150	
1BBT3	- - C I H A - E -	W D T G L N S K F T F S I P Y L S -		
1AYM3	G - - T H V - V -	W D V G L Q S T V S L V V P W I S -		
1BBT2	- - F P H Q - F -	I N P R T N M T A H I T V P F V G -		
1AYM1	G - - T N A S V F W Q H G -	Q P F P R F S L P F L S I		
1BBT1	Y T A P H R - V -	L A T V Y N G E C R Y - - - - - S -		

G T C P H A S V F W Q H G - Q P F P R F S L P F L S I

message Sum of Pairs with Gap Affinity 2059.0 2069.0

# Anubis

The screenshot shows the Anubis software interface. The main window displays two sequence alignment views. The top view shows sequences 1BBT3, 1AYM3, 1BBT2, 1AYM1, and 1BBT1 aligned from position 120 to 140. The bottom view shows the same sequences aligned from position 130 to 150. A control panel on the right side of the window is titled 'Alignment' and includes the following options:

- Default
- Compressed Alphabet
- Column Matches
- Secondary Structure
- Success %

The status bar at the bottom of the window displays the following information:

message Sum of Pairs with Gap Affinity 2059.0 2069.0

# Anubis

The screenshot displays the Anubis software interface. The main window shows a sequence alignment of five sequences: 1BBT3, 1AYM3, 1BBT2, 1AYM1, and 1BBT1. The alignment is color-coded by column, with columns grouped into regions labeled 120, 130, and 140. Below the alignment, a sequence logo is visible, showing the conservation of residues across the alignment. The right-hand side of the interface contains a settings panel with tabs for 'Left', 'Bottom', 'Right', and 'Top'. The 'Alignment' tab is selected, showing options for 'Default', 'Compressed Alphabet', 'Column Matches', 'Secondary Structure', and 'Success %'. The 'Column Matches' and 'Success %' options are checked. The bottom status bar shows the 'Sum of Pairs with Gap Affinity' metric, with values 2059.0 and 2069.0.

Sequence	120	130	140
1BBT3	A H C I - - H A E W D T G L N S K F T F S I P Y L S A		
1AYM3	M L G T - - H V V W D V G L Q S T V S L V V P W I S A		
1BBT2	T L F P - - H Q F I N P R T N M T A H I T V P F V G V		
1AYM1	Q S G T N A S V F W Q H G - Q P F P R F S L P F L S I		
1BBT1	N T - T N P T A Y H K - - - A P L T R L A L P Y T A P		

Sequence Logo: Q L G T N P H A F W D P G L N P T T R F S I P Y I S

Sequence	130	140	150
1BBT3	- - C I H A - E - W D T G L N S K F T F S I P Y L S -		
1AYM3	G - - T H V - V - W D V G L Q S T V S L V V P W I S -		
1BBT2	- - F P H Q - F - I N P R T N M T A H I T V P F V G -		
1AYM1	G - - T N A S V F W Q H G - Q P F P R F S L P F L S I		
1BBT1	Y T A P H R - V - L A T V Y N G E C R Y - - - - - S -		

Sequence Logo: G T C P H A S V F W D T G L N S T C R F S I P F I S

message Sum of Pairs with Gap Affinity 2059.0 2069.0



# Anubis

The screenshot shows the Anubis software interface. The main window displays two sequence alignment views. The top view shows sequences 1BBT3, 1AYM3, 1BBT2, 1AYM1, and 1BBT1 aligned from position 120 to 140. The bottom view shows the same sequences aligned from position 130 to 150. A right-hand panel contains settings for alignment, including 'Default', 'Compressed Alphabet', 'Column Matches', 'Secondary Structure', and 'Success %'. The bottom status bar shows 'Sum of Pairs with Gap Affinity' with values 2059.0 and 2069.0.

Sequence	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140						
1BBT3	A	H	C	I	-	-	H	A	E	W	D	T	G	L	N	S	K	F	T	F	S	I	P	Y	L	S	A
1AYM3	M	L	G	T	-	-	H	V	V	W	D	V	G	L	Q	S	T	V	S	L	V	V	P	W	I	S	A
1BBT2	T	L	F	P	-	-	H	Q	F	I	N	P	R	T	N	M	T	A	H	I	T	V	P	F	V	G	V
1AYM1	Q	S	G	T	N	A	S	V	F	W	Q	H	G	-	Q	P	F	P	R	F	S	L	P	F	L	S	I
1BBT1	N	T	-	T	N	P	T	A	Y	H	K	-	-	-	A	P	L	T	R	L	A	L	P	Y	T	A	P

Sequence	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150						
1BBT3	-	-	C	I	H	A	-	E	-	W	D	T	G	L	N	S	K	F	T	F	S	I	P	Y	L	S	-
1AYM3	G	-	-	T	H	V	-	V	-	W	D	V	G	L	Q	S	T	V	S	L	V	V	P	W	I	S	-
1BBT2	-	-	F	P	H	Q	-	F	-	I	N	P	R	T	N	M	T	A	H	I	T	V	P	F	V	G	-
1AYM1	G	-	-	T	N	A	S	V	F	W	Q	H	G	-	Q	P	F	P	R	F	S	L	P	F	L	S	I
1BBT1	Y	T	A	P	H	R	-	V	-	L	A	T	V	Y	N	G	E	C	R	Y	-	-	-	-	-	-	S

# Anubis

The screenshot displays the Anubis software interface. The main window shows a sequence alignment of four sequences: 1BBT3, 1AYM3, 1BBT2, and 1AYM1. The alignment is color-coded by column, with columns grouped into blocks labeled 120, 130, and 140. A secondary alignment is shown below, with columns labeled 130, 140, and 150. The interface includes a menu bar (File, View, Plugins, Help) and a settings panel on the right. The settings panel has tabs for Left, Bottom, Right, and Top, with the Alignment tab selected. The settings include: Default (selected), Compressed Alphabet, Column Matches (checked), Secondary Structure, and Success % (checked). The status bar at the bottom shows 'message' and 'Sum of Pairs with Gap Affinity' with values 2059.0 and 2069.0.

Sequence	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150
1BBT3	A	H	C	I	-	-	H	A	E	W	D	T	G	L	N	S	K	F	T	F	S	I	P	Y	L	S	A	-	-	-	-
1AYM3	M	L	G	T	-	-	H	V	V	W	D	V	G	L	Q	S	T	V	S	L	V	V	P	W	I	S	A	-	-	-	-
1BBT2	T	L	F	P	-	-	H	Q	F	I	N	P	R	T	N	M	T	A	H	I	T	V	P	F	V	G	V	-	-	-	-
1AYM1	Q	S	G	T	N	A	S	V	F	W	Q	H	G	-	Q	P	F	P	R	F	S	L	P	F	L	S	I	-	-	-	-
1BBT1	N	T	-	T	N	P	T	A	Y	H	K	-	-	-	A	P	L	T	R	L	A	L	P	Y	T	A	P	-	-	-	-

Sequence	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150									
1BBT3	-	-	C	I	H	A	-	E	-	W	D	T	G	L	N	S	K	F	T	F	S	I	P	Y	L	S	-	-	-	-
1AYM3	G	-	-	T	H	V	-	V	-	W	D	V	G	L	Q	S	T	V	S	L	V	V	P	W	I	S	-	-	-	-
1BBT2	-	-	F	P	H	Q	-	F	-	I	N	P	R	T	N	M	T	A	H	I	T	V	P	F	V	G	-	-	-	-
1AYM1	G	-	-	T	N	A	S	V	F	W	Q	H	G	-	Q	P	F	P	R	F	S	L	P	F	L	S	I	-	-	-
1BBT1	Y	T	A	P	H	R	-	V	-	L	A	T	V	Y	N	G	E	C	R	Y	-	-	-	-	-	-	S	-	-	-

# Anubis

The screenshot displays the Anubis software interface. The main window shows a sequence alignment of four sequences: 1BBT3, 1AYM3, 1BBT2, and 1AYM1. The alignment is color-coded by column, with columns 120-140 visible. Below the alignment, a sequence is shown: Q L G T N P H A F W D P G L N P T T R F S I P Y I S. The interface includes a menu bar (File, View, Plugins, Help) and a settings panel on the right. The settings panel has tabs for Left, Bottom, Right, and Top, with the Alignment tab selected. The Alignment tab contains the following options:

- Default
- Compressed Alphabet
- Column Matches (with a slider)
- Secondary Structure (with a slider)
- Success % (with a slider)

At the bottom of the window, a message box displays the following information:

message Sum of Pairs with Gap Affinity 2059.0 2069.0



# Anubis

The screenshot displays the Anubis software interface. The main window shows a sequence alignment of five protein sequences: 1BBT3, 1AYM3, 1BBT2, 1AYM1, and 1BBT1. The alignment is color-coded by amino acid type. A settings panel on the right is open, showing the 'Alignment' tab with options for 'Default', 'Compressed Alphabet', 'Column Matches', 'Secondary Structure', and 'Success %'. The 'Sum of Pairs with Gap Affinity' is displayed as 2059.0 and 2069.0.

File View Plugins Help

120 130 140

1BBT3 A H C I - - H A E Y D T G L N S K E T F S I P Y L S A  
1AYM3 M L G T - - H V V W D V G L Q S T V S L V V P W I S A  
1BBT2 T L F P - - H D P I N P R I N N T A H I T V P F V G  
1AYM1 Q S G T N A S V F W Q H G - Q P F P R F S L P F L S I  
1BBT1 N T - T I N P T A Y H X - - - A P L T R L A L P Y T A P

Q L G T N P H A F W D P G L N P T T R F S I P Y I S

130 140 150

1BBT3 - - C I H A - E - W D T G L N S K F T F S I P Y L S -  
1AYM3 G - - T H V - V - W D V G L Q S T V S L V V P W I S -  
1BBT2 - - F P H Q - F - I N P R T N M T A H I T V P F V G -  
1AYM1 G - - T N A S V F W Q H G - Q P F P R F S L P F L S I  
1BBT1 Y T A P H R - V - L A T V Y N G E C R Y - - - - - S -

G T C P H A S V F W D T G L N S T C R F S I P F I S

message Sum of Pairs with Gap Affinity 2059.0 2069.0

# Correspondência de Colunas

The screenshot displays the Anubis software interface for sequence alignment. The main window shows two panels of sequence alignment. The top panel shows sequences 1BBT3, 1AYM3, 1BBT2, 1AYM1, and 1BBT1 aligned from column 120 to 140. The bottom panel shows the same sequences aligned from column 130 to 150. Columns are color-coded to highlight matches between sequences. For example, in the top panel, column 120 has 'A' in 1BBT3, 'M' in 1AYM3, 'T' in 1BBT2, 'Q' in 1AYM1, and 'N' in 1BBT1. In the bottom panel, column 130 has 'C' in 1BBT3, 'G' in 1AYM3, 'F' in 1BBT2, 'G' in 1AYM1, and 'Y' in 1BBT1. The right panel shows alignment options: 'Left', 'Bottom', 'Right', and 'Top' tabs; 'Alignment' sub-tab; 'Default' (selected), 'Compressed Alphabet', 'Column Matches' (checked), 'Secondary Structure' (unchecked), and 'Success %' (checked). The bottom status bar shows 'message' and 'Sum of Pairs with Gap Affinity' with values 2059.0 and 2069.0.

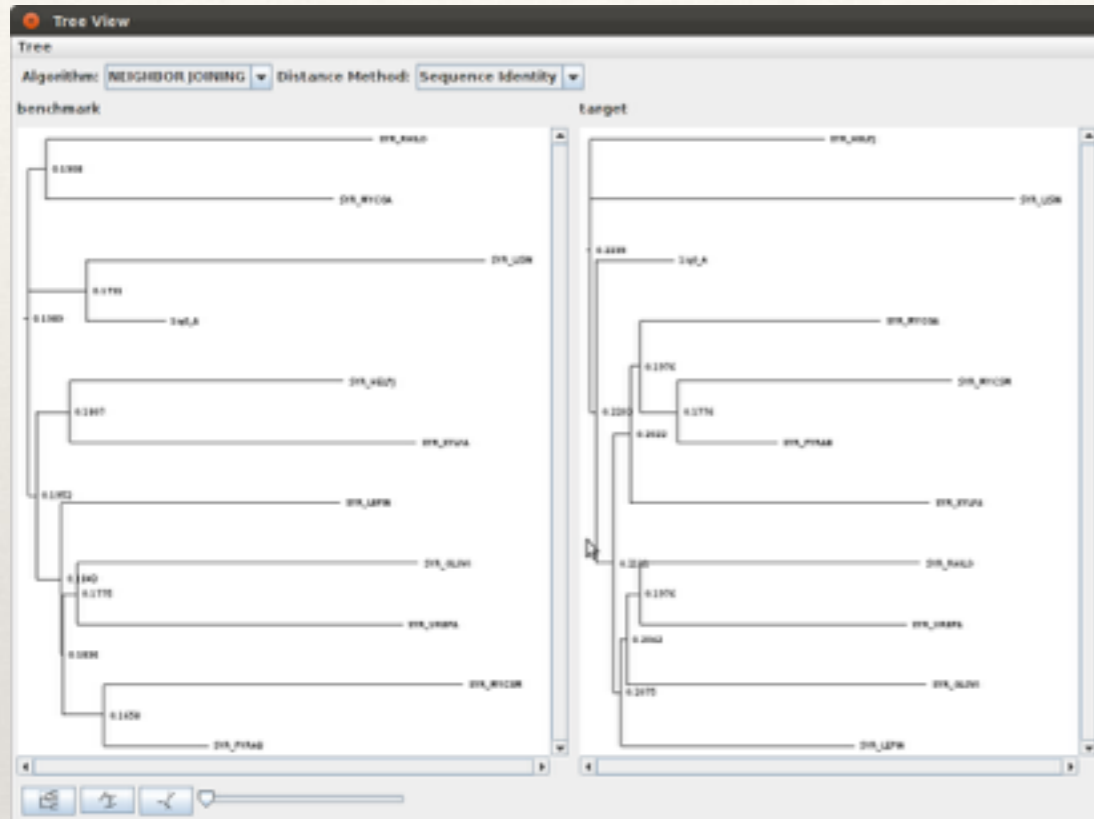
---

# Árvores Filogenéticas

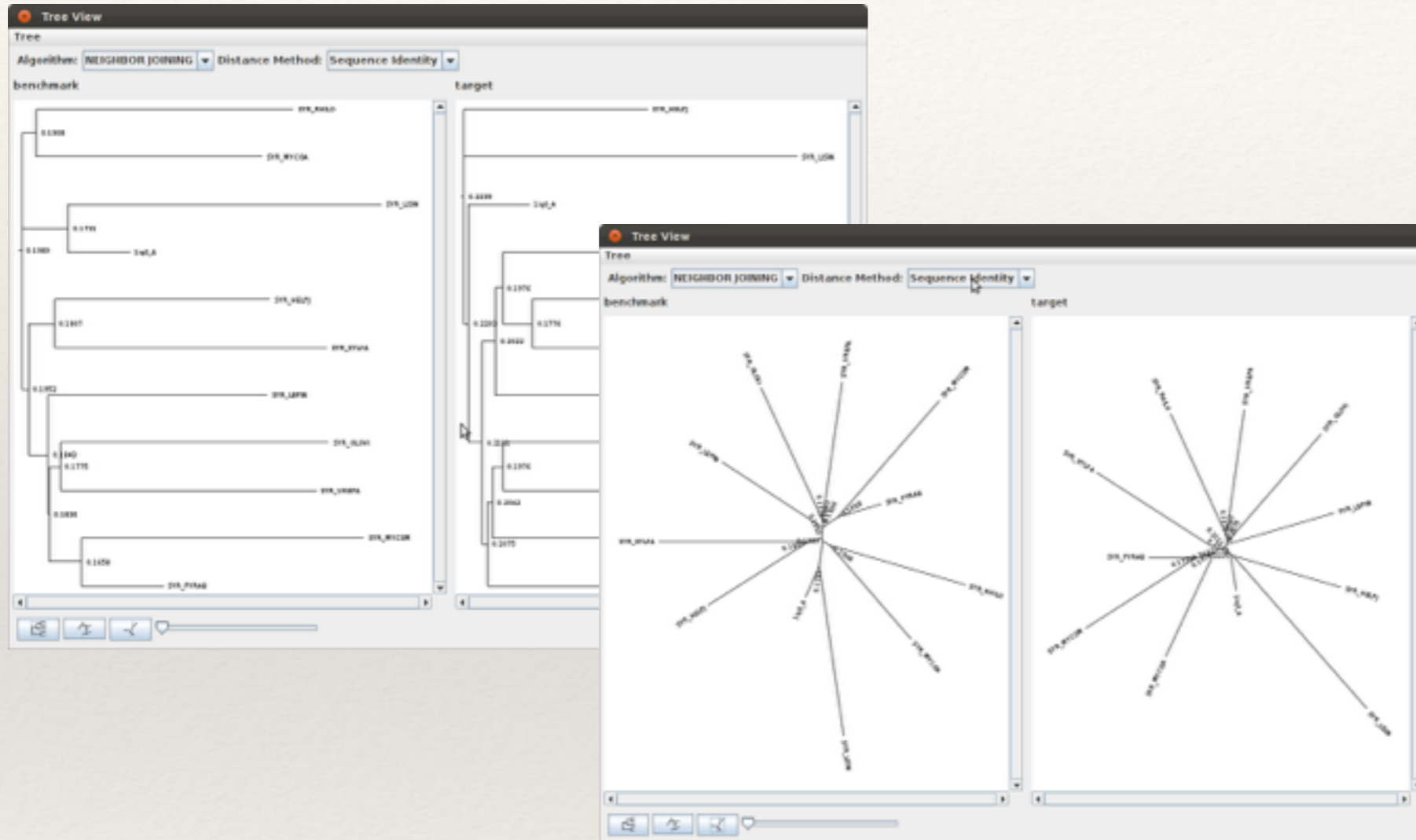
---



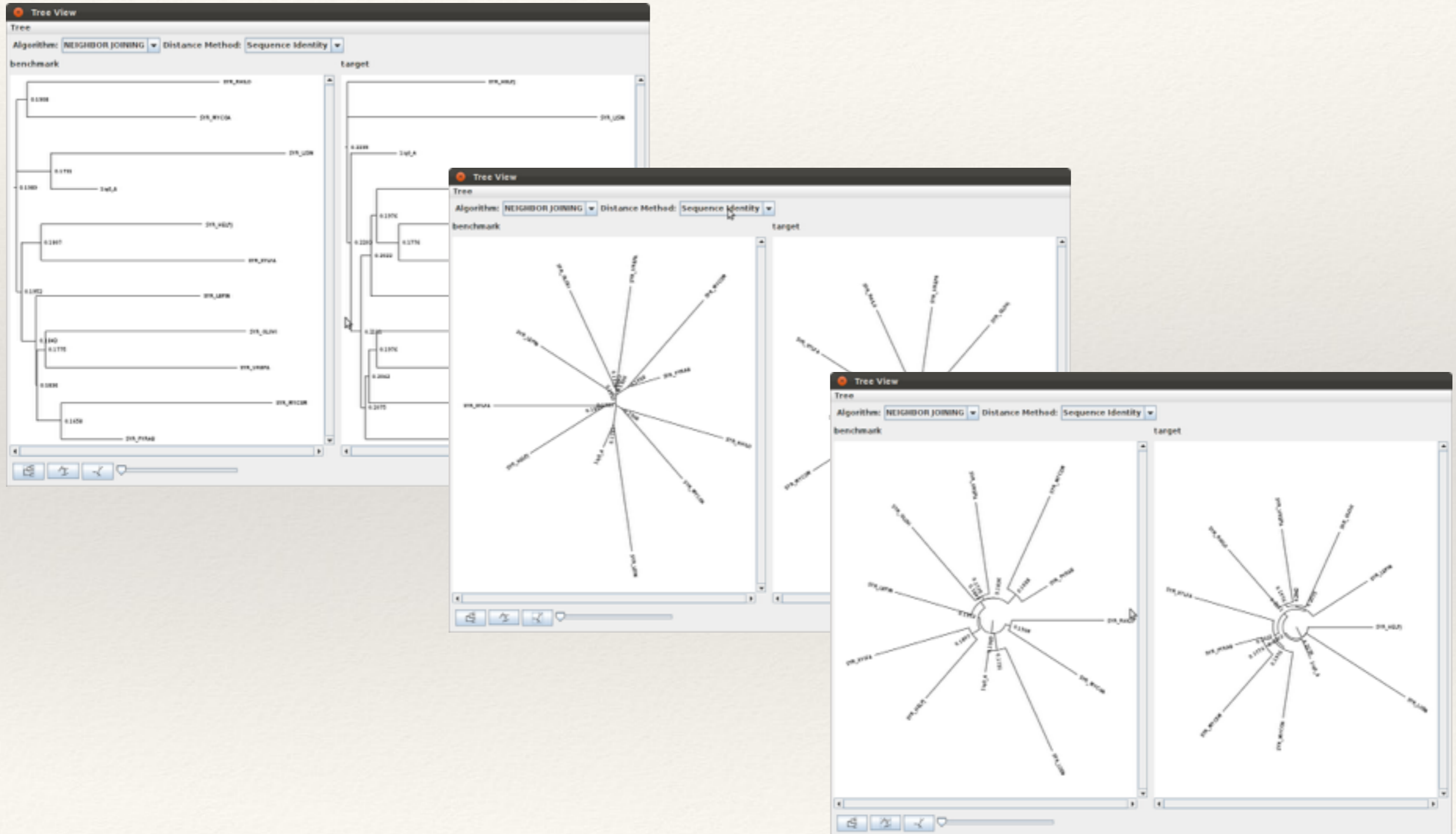
# Árvores Filogenéticas



# Árvores Filogenéticas



# Árvores Filogenéticas





# Estruturas Secundárias

The screenshot displays the Anubis software interface for protein sequence alignment. The main window shows two alignments of protein sequences. The top alignment is between sequences 16BT3, 1AYM3, 16BT2, 1AYM1, and 16BT1. The bottom alignment is between sequences 16BT3, 1AYM3, 16BT2, 1AYM1, and 16BT1. Both alignments show secondary structure elements (SSEs) highlighted in green and yellow. The SSEs are represented by arrows indicating their orientation. The alignment is scored using the Sum of Pairs with Gap Affinity method, with a score of 2059.0 for the top alignment and 2069.0 for the bottom alignment. The interface includes a menu bar (File, View, Plugins, Help), a toolbar (Left, Bottom, Right, Top), and a settings panel on the right with options for Alignment, Column Matches, Secondary Structure, and Success %.

16BT3 M S L A A K H M S N T F L A G L A Q Y Y T Q Y S G T I  
1AYM3 V D I A S H P L A T T L I G E I A S Y F T H W T G S L  
16BT2 L P T D H K - - - G V Y G S L T D S Y A Y M R N G W  
1AYM1 I N L Q E M - - - A Q I R R K F E M F T Y A R F D S  
16BT1 L D L M Q V P - S H T L V G G L L R A S T Y Y F S D L

L D L A Q K P M S N T L I G G L A E Y Y T Y Y R G D

16BT3 - A A K H - M - S N T F L A - G L A Q Y Y T Q Y S G T  
1AYM3 D I A S H P L - A T T L I G - E I A S Y F T H W T G S  
16BT2 H L L E L P T D H K G V Y G - S L T D S Y A Y M R N G  
1AYM1 N I N L Q E M - A Q I R R K F E M F T - Y A R F D S E  
16BT1 D L M Q V P S - - H T L V G - G L L R A S T Y Y F S D

D I A Q H P M D A N T L I G F G L A Q Y Y T Y Y S G

message Sum of Pairs with Gap Affinity 2059.0 2069.0

---

# Outras Funcionalidades

---

- ❖ Porcentagem de sucesso
- ❖ Similaridade em relação ao *benchmark*
- ❖ Colunas e sequências
- ❖ *Gaps* (opcional)

---

# Outras Funcionalidades

---

- ❖ Sequências consenso
- ❖ Alfabetos comprimidos
- ❖ Pontuação



---

# Outras Funcionalidades

---

- ❖ Visualização
  - ❖ Ocultar colunas / sequências
  - ❖ Cores
  - ❖ Sincronizar o *scrolling* dos alinhamentos

# Considerações Finais

---

# Contribuições - Ferramentas

---

- ❖ ALGAe
- ❖ Anubis
  - ❖ *Open-source*
  - ❖ GitHub



---

# Contribuições

---

- ❖ *ALGAe: A testbench environment for a Genetic Algorithm-based Multiples Sequence Aligner*
  - ❖ *S. Ordine, A. Grilo, A. Almeida e Z. Dias*
  - ❖ *Brazilian Symposium on Bioinformatics - BSB'2011*
    - ❖ *Brasília*

---

# Contribuições

---

- ❖ *An empirical study for gap penalty score using a multiple sequence alignment genetic algorithm*
  - ❖ *S. Ordine, A. Almeida e Z. Dias*
- ❖ *Brazilian Symposium on Bioinformatics - BSB'2012*
  - ❖ *Campo Grande*

---

# Trabalhos Futuros

---

- ❖ ALGAe
  - ❖ Função objetivo baseada em estrutura
- ❖ Anubis
  - ❖ Mesclar alinhamentos
- ❖ Ferramental estatístico



Obrigado



