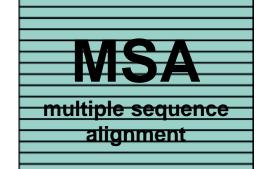
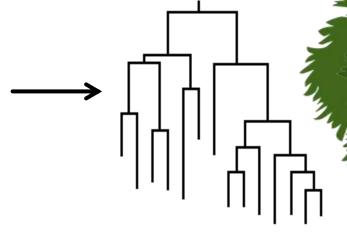
My research progress



A quick remainder of what I am doing

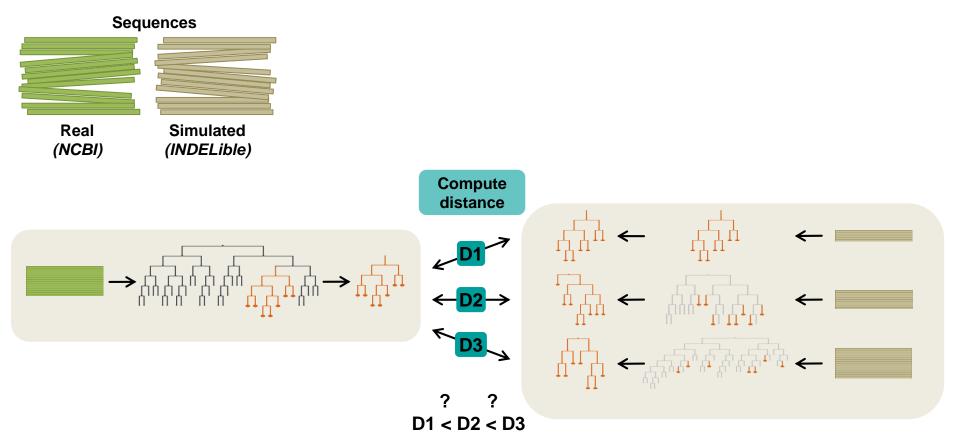








Examining the effect of sequence divergence



MSA size: 100, 800, 2000, 5000

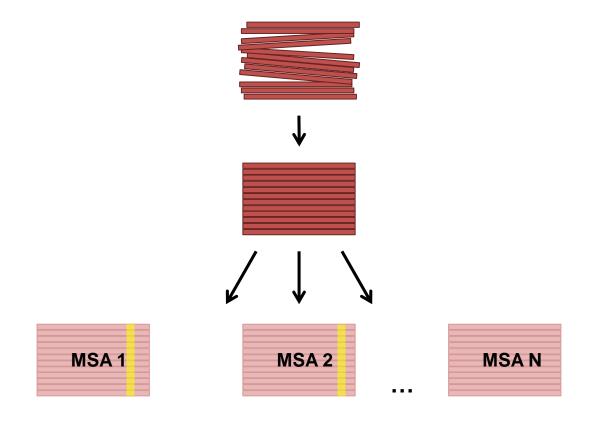
Distance: 0.38 < 0.43 < 0.441 < 0443

D1 < D2 < D3 < D4

Which sequences are simulated?

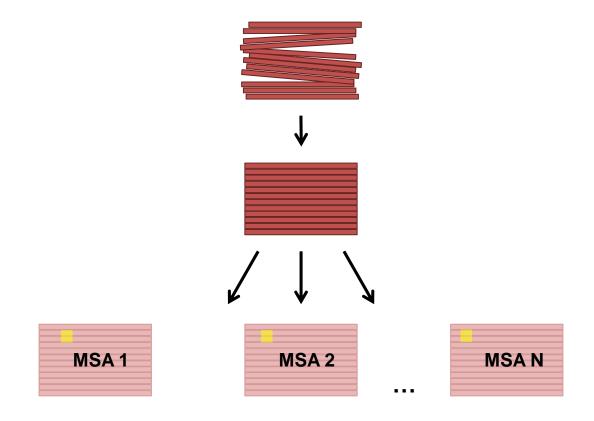
10051	_			-CGAA	CTCGTGACCCCCTTTT	mmmmcccccm» cccc	CAACCC_CMC	comcmmes cm/
13351	CCAATCCTG	ATCGACA			CTCGTGACCCCCTTTT TTATCATCACTCGTCGTC			
4424				GCGAACACG				
105750	-CGATGACC	CACTAGGCT			CTTGTGATATTGTTCC			
16752	TGATGCGTA	CCAAAACAAGAT			CATGTGACCCCTTGTT			
12953	TCGAGGACC	GAACGAGAGGGTA		GCGAA	CATAGTGACCCGCGCA			
22294		ACCCCTGCAAAGCAGCACGAC		GTGAA	TATGTCAATTGTCATC			
112846		GAAAATGCAGAGCAGCACGAC		GTGAA	CTTGTGAATTATCATT			
3429	T				CCTCTG-CCCCCGCTGTG			
13083	TCGCAAACC	AGAGC			PTCTGTCACTGAAGTTTG			
3426	TCGAGGCCC	-CTCCCTGCTAAAAGCAGAAC		GCGAA	GGAGTGAAAGCATTGC			
13419	TCGTTGCCT				CCTCTGCCCCCCGCCGTG			
13260	TGATGCCTA	CCAAAAACAAGA <mark>T</mark>	CGACCGAA	AGCGAA	CATGTGAAACCTTGCG	TTCTCCGCCCGCGCG	CGGGGCC-CC	TGCCCC
13559	TCGATACCT	A <mark>T</mark> AAAAG	GCGAGACCA	GCGAATTGT	PATCTGTTTTTTTCTTTG	TTGCATCGTATCGAG	AGATATA-TAC	CGGTGCCTCC
43891	TCGAAGCCT	GCAACGGCAGAAC	CGACCC	GCGAA	CTCGTCACCGATCGCA	CCCGCGGGGGGGGAC	CGTCCGG-CGC	CCCGGCGTC
2708	TTGAAACCT	GCC-CAGCAGAAC	CGACCC	GTGAA	CCAG-AGATATCACCG	GCGGCGGGGAGGGGGG	ATGCGTC-CGC	CAGCGGG
2711	TCGAAACCT	GCC-CAGCAGAAC	CGACCC	GCGAA	CCAGTTGATATCACCG	GCGGCGGGGAGGGGGG	ACGCGCC-CGC	CAGCGGG
43166	TCGACCTGC	CAGCAGA-C	CGACCC	GCGAA	CCAGTTGATATCACCG	GCGGCGGGGGGGGG	GCGCGTC-CGC	AGCGGG
105808	TCGAAACCT	GCCTAGCAGAAC	CGACCC	GCGAA	TACGTGAACGTGCTGA	GGGGGGACGACGGGC	GCTCGCG-CAC	CCTTTGCTC
85280	TCGAAACCT	GCC-TAGCAGAAC		GTGAA	TTTGTGAATGTTTTGG	GGGGCGAGGGCTGTG	CTTGTGC-AG	TGTCGT
13654	TCGAAACCT	GCCTAGCAGAAT	GACCA	AGCGAA	CTTGGTCGGATACGGA	TACGGCCATCGGCCT	PTCGTAG-TTC	GCCCTGTTT
3760	TCGAAACCT	GCC-TAGCAGAAC		GGAAC		GGGGGGGCGAGGGGT		
3486	TCGATGCCT	GCAACAGCAGAAC	CGACCC	GGAA	CACGTACCTTGGGTGG	GCGAGAGGAGCTTGC'	CCTTGG-ACC	CCCCCTCAC
3483	TCGAAACCT	GCAACAGCAGAAC		GGAA	CACGTG-CTTGGGCGG	GCGAGAGGAGCTTGC'	CCTTGG-ACC	CCCCCCCCAC
3460	TCGATGCCT				CGCGTGTTTTCAATTCTA			
3813	CCGC	AAAAAGAG		GAGAA	CCGGTTGGAACAAAAC			
3879		CATGCAGTCCAACAGGAATCA			CCTCTTGGTTCAGAGG			
63479	TCGAATCCT	ACAGAGCGAAC		GCGAA	CTCGTTTATCCACGGG		CCTGCCC-CCC	CCCGTC
55175								
		GACCGCGCAGGCCCACCCC	CMAG	CACGCAMA		GAGTCGAGTC		-CGCACGGATC
50993		GACCGCGCAGGCCCACCC	CTAG	-CACGCATA	CCCCCGCACAACCGAGA			-CGCACGGATC
50993 4400	GCCCTCCACC	GGCCGCGCAGGGCCTCTCC	CTAG	CACGCATA	CCCCCGCACAACCGAGA	GAGCCCAGCC	GTGG-	-AGCAAGGATC
50993 4400 50174	GCCCTCCACCC	GGCCGCGCAGGGCCTCTCC GTCCGCGCAACGCCGCTCC	CTAG	CACGCATA GACGCATA	CCCCCGCACAACCGAGA CCCCCGCGCCTTCGAGA CGCCCGCACACGCGAGA	GAGCCCAGCC GAGCCCAGGA	G <mark>T</mark> GG-	-AGCAAGGATC -CGCAAGGATC
50993 4400 50174 45172	GCCCTCCACCC	GGCCGCGCAGGGCCTCTCC GTCCGCGCAACGCCGCTCC GTCCGCGCAGGGCTACCAC	CTAG CTAG	CACGCATA GACGCATA CACGCATA	CCCCCGCACAACCGAGA CCCCCGCGCCTTCGAGA CGCCCGCACACGCGAGA	GAGCCCAGCC GAGCCCAGGA GAATCTAGTT	GTGG- GGGG- GCGG-	-AGCAAGGATC -CGCAAGGATC -AGCAAGGTTC
50993 4400 50174 45172 3592	GCCCTCCACCCGCCTTCTACTCGCCCTCCACCC	GGCCGCGCAGGGCCTCTCC GTCCGCGCAACGCCGCTCC GTCCGCGCAGGGCTACCAC GCCCGCGCAGGGCCAACTC	CTAG CTAG CTAG CTAG	CACGCATA GACGCATA CACGCATA CACGCATA	CCCCCGCACAACCGAGA CCCCCGCGCCTTCGAGA CCCCCGCACACGCGAC CCCCGGCACCATCGAGA	GAGCCCAGCC GAGCCCAGGA GAATCTAGTT GAGCCTAGCG		-AGCAAGGATC -CGCAAGGATC -AGCAAGGTTC -GGCATGGGTC
50993 4400 50174 45172 3592 3597	GCCCTCCACCCGCCTTCTACTCGCCTTCCACCCGCCTTCTACCC	GGCCGCGCAGGGCTCTCC GTCCGCGCAACGCCGCTCC GTCCGCGCAGGGCTACCAC GCCCGCGCAGGGCCAACTC GACCGCGCAAGGCCAACCC	CTAG CTAG CTAG CTAG CTAG	CACGCATA GACGCATA CACGCATA CACGCATA TACGCATA	CCCCCGCACAACCGAGACCCCCGCGCCCTTCGAGACGCCCGCACACGCGCACCGCACCGCACCGC	GAGCCCAGCC GAGCCCAGGA GAATCTAGTT GAGCCTAGCG GAGCCTAGCT		-AGCAAGGATC -CGCAAGGATC -AGCAAGGTTC -GGCATGGGTC -AGCAGGGTTC
50993 4400 50174 45172 3592 3597 4283	GCCCTCCACCCGCCTTCTACTCGCCTTCTACCCGCCTTCTACCC	GGCCGCGCAGGGCCTCTCC GTCCGCGCAACGCCGCTCC GTCCGCGCAGGGCTACCAC GCCCGCCAGGGCCAACTC GACCGCGCAAGGCCACCC GCCCGCGCATGGCCACCCC	CTAGCTAGCTAGCTAG	CACGCATAGACGCATACACGCATACACGCATATACGCATACACGCATA	CCCCCGCACAACCGAGACCCCCGCGCCCTTCGAGACGCCCGCACCACGGGAGACCCCGGCACCATCGAGACCCCCGCACTCCCGACACCCCCGCACATCCCAAA	GAGC		-AGCAAGGATC -CGCAAGGATC -AGCAAGGTTC -GGCATGGGTC -AGCAGGGTTC -GGCAAGGATC
50993 4400 50174 45172 3592 3597 4283 4292	GCCTTCACCCGCCTTCACCCGCCTTCTACCCGCCTTCACCCGCCTCCACCCCCCCCCCCCCCCCCCCCCCCCCCCC	GGCCGCGCAGGGCCTCTCC GTCCGCGCAACGCCTCC GTCCGCGCAGGGCTACCAC GCCCGCGCAGGGCCAACTC GACCGCGCAGGGCCACCC GCCCGCGCATGGCCACCC GGCCGCGCATGGCCACCCC GGCCGCGCAGGGCCACCCC	CTAG CTAG CTAG CTAG CTAG CTAG CTAG CTAG	CACGCATAGACGCATACACGCATATACGCATACACGCATACACGCATACACGCATA		GAGCCCAGCC- GAGCCCAGGA- GAATCTAGCT GAGCCTAGCG CAGCTCAGGT CAGCCAGGTC CAGCCCGAGTC	GTGG- GGGG- GCGG- GCGG- GCGG- GCGG- TCAG-	-AGCAAGGATC -CGCAAGGATC -AGCAAGGTTC -GGCATGGGTC -AGCAGGGTTC -GGCAAGGATC -GGCAAGGATC
50993 4400 50174 45172 3592 3597 4283	GCCTCCACC GCCTTCTACT GCCTTCTACCC GCCTTCTACCC GCCTTCTCCCCCC GCCTCCACCC GCCTCCACCC GCCTCCACCC	GGCCGCGAGGGCCTCC GTCCGGCAACGCCTCC GTCCGCGCAGGGCTACCAC GCCCGCGCAGGGCCAACTC GACCGCGCAAGGCCACCC GCCCGCGCATGCCACCC GCCCGCGCATGCCACCC GGCCGCGCAGGGCCACCC GGCCGCGCAGGGCCACCC GGCCGCGCAGGGCCACCC	CTAG CTAG CTAG CTAG CTAG CTAG ATAG CTAG C	CACGCATA GACGCATA CACGCATA TACGCATA CACGCATA CACGCATA CACGCATA CACGCATA CACGCATA		GAGC	GTGG GGGG GCGG GCGG GCGG GCGG TCAG GCGG	-AGCAAGGATC -CGCAAGGATC -AGCAAGGTTC -GGCATGGGTC -AGCAGGGTTC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC
50993 4400 50174 45172 3592 3597 4283 4292	GCCTCCACC GCCTTCTACT GCCTTCTACCC GCCTTCTACCC GCCTTCTACCC GCCTTCTCCACCC GCCTTCCACCC GCCTTCCACCC GCCTTCCACCC GCCTTCCACCC GCCTTCACCC	GGCCGCGAGGGCCTCC GTCCGGCAACGCCTCC GTCCGCGCAGGGCTACCAC GCCCGGCAGGGCCACTC GCCCGGCAAGGCCACCC GCCCGGGCATGGCCACCCC GCCCGGCATGGCCACCC GGCCGCAGGGCACCCC GGCCGCAGGGCACCCC GGCCGCAGGGCCACCC GACCGCGCAGGGCCACCCC GACCGCGCAGGGCACCCC GACCGCGCAGAGGCACCCC	CTAG CTAG CTAG CTAG CTAG CTAG CTAG CTAG	-CACGCATA -GACGCATA -CACGCATA -CACGCATA -CACGCATA -CACGCATA -CACGCATA -CACGCATA -CACGCATA -CACGCATA -CACGCATA		GAGC	GTGG GGGG GCGG GCGG GCGG GCGG TCAG GCGG GCG	-AGCAAGGATC -CGCAAGGATC -AGCAAGGTTC -AGCAAGGTTC -AGCAAGGTTC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC -GGCATGGATC -CGCACGGATC
50993 4400 50174 45172 3592 3597 4283 4292 16924	GCCTCCACC GCCTTCTACT GCCTTCTACCC GCCTTCTACCC GCCTTCTACCC GCCTTCCACCC GCCTTCCACCC GCCTCCACCC GCCTCCACCC GCCTCCACCC GCCTCCACCC GCCTTCACCC	GGCCGCGCAGGGCCTCC GTCCGGCGAACGCCGCTCC GTCCGCGCAGGGCTACCAC GCCCGCGCAGGGCCAACTC GACCGCGCAAGGCCACCC GCCCGCGCAGGGCCACCCC GCCCGCGCAGGGCCACCCC GACCGCCAGGGCCACCCC GACCGCCAGAGGCCACCC GACCGCCAGAGGCACCCC GACCGCCAGAGGCATCCC GACCGCCAGAGGCATCCC GACCGCGCAGAGGCATCCC	CTAG CTAG CTAG CTAG ATAG CTAG CTAG CTAG	CACGCATA GACGCATA CACGCATA CACGCATA TACGCATA CACGCATA CACGCATA CACGCATA CACGCATA CACGCATA CACGCATA CACGCATA CACGCATA		GAGC	GTGG- GGGG- GCGG- GCGG- GCGG- GCGG- GCGG- GCGG- GCGG- GCGG-	-AGCAAGGATC -CGCAAGGATC -AGCAAGGTTC -GGCATGGGTC -AGCAGGGTTC -GGCAAGGATC -GGCAAGGATC -GGCATGGATC -CGCAAGGATC -CGCAAGGATC
50993 4400 50174 45172 3592 3597 4283 4292 16924 3741	GCCTCCACC GCCTTCTACT GCCTTCTACCC GCCTTCTACCC GCCTTCTACCC GCCTTCCACCC GCCTTCCACCC GCCTCCACCC GCCTCCACCC GCCTCCACCC GCCTCCACCC GCCTTCACCC	GGCCGCGAGGGCCTCC GTCCGGCAACGCCTCC GTCCGCGCAGGGCTACCAC GCCCGGCAGGGCCACTC GCCCGGCAAGGCCACCC GCCCGGGCATGGCCACCCC GCCCGGCATGGCCACCC GGCCGCAGGGCACCCC GGCCGCAGGGCACCCC GGCCGCAGGGCCACCC GACCGCGCAGGGCCACCCC GACCGCGCAGGGCACCCC GACCGCGCAGAGGCACCCC	CTAG CTAG CTAG CTAG CTAG CTAG CTAG CTAG	-CACGCATA -GACGCATA -CACGCATA -CACGCATA -CACGCATA -CACGCATA -CACGCATA -CACGCATA -CACGCATA -CACGCATA -CACGCATA	CCCCCGCACAACCGAGACCCCCGCGCCTTCGAGACGCCCGCACACGCGCACATCGAGACGCCCGCACATCGAGA	GAGC	GTGG- GGGG- GCGG- GCGG- GCGG- GCGG- GCGG- GCGG- GCGG- GCGG-	-AGCAAGGATC -CGCAAGGATC -AGCAAGGTTC -AGCAAGGTTC -AGCAAGGTTC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC -GGCATGGATC -CGCACGGATC
50993 4400 50174 45172 3592 3597 4283 4292 16924 3741 4442	GCCTCCACC GCCTTCTACT GCCTTCTACC GCCTTCTACC GCCTTCTACC GCCTTCTCCCC GCCTCCACC GCCTCCACC GCCTCCACC GCCTCTCACC GCCTCCACC GCCTCCACC GCCATCTACC GCCATCTACC	GGCCGCGCAGGGCCTCC GTCCGGCGAACGCCGCTCC GTCCGCGCAGGGCTACCAC GCCCGCGCAGGGCCAACTC GACCGCGCAAGGCCACCC GCCCGCGCAGGGCCACCCC GCCCGCGCAGGGCCACCCC GACCGCCAGGGCCACCCC GACCGCCAGAGGCCACCC GACCGCCAGAGGCACCCC GACCGCCAGAGGCATCCC GACCGCCAGAGGCATCCC GACCGCGCAGAGGCATCCC	CTAG CTAG CTAG CTAG CTAG CTAG CTAG CTAG	CACGCATA GACGCATA CACGCATA CACGCATA TACGCATA CACGCATA CACGCATA CACGCATA CACGCATA CACGCATA CACGCATA CACGCATA CACGCATA		GAGC - CCAGCC GAGC - CCAGGA GAAT - CTAGTT GAGC - CTAGCT GAGC - CTAGCT CAGT - CAAGTG CAGC - CGAGTC CAGC - CGAGTG CAGT - CGAGTG	GTGG	-AGCAAGGATC -CGCAAGGATC -AGCAAGGTTC -GGCATGGGTC -AGCAGGGTTC -GGCAAGGATC -GGCAAGGATC -GGCATGGATC -CGCAAGGATC -CGCAAGGATC
50993 4400 50174 45172 3592 3597 4283 4292 16924 3741 4442 125047	GCCTCCACC GCCTTCTACT GCCTTCTACCC GCCTTCTACCC GCCTTCTACCC GCCTTCCACCC GCCTCCACCC GCCTCCACCC GCCATCTACCC GCCATCTACCC GCCATCTACCC	GGCCGCGCAGGGCCTCC GTCCGCGCAACGCCTCC GTCCGCGCAGGGCTACCAC GCCCGCGCAGGGCCAACTC GACCGCGCAAGGCCACCC GGCCGCGCAGGGCCACCCC GGCCGCGCAGGGCCACCCC GGCCGCGCAGGGCCACCCC GACCGCGCAGAGGCACCCC GACCGCCACAGAGCACCCC GCCCGCCACAGGGCATCCC GCCCGCCACAGGGCACCCC GCCCGCCACAGGGCACCCC GCCCGCCACGGGCACCCC	CTAG CTAG CTAG CTAG ATAG CTAG CTAG CTAG	CACGCATA GACGCATA CACGCATA CACGCATA TACGCATA CACGCATA		GAGC	GTGG	-AGCAAGGATC -CGCAAGGATC -AGCAAGGTTC -GGCATGGGTC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC -CGCAAGGATC -CGCAAGGATC -CGCAAGGATC -CGCAAGGATC -CGCAAGGATC
50993 4400 50174 45172 3592 3597 4283 4292 16924 3741 4442 125047 522403	GCCTCCACC GCCTTCTACT GCCTTCTACCC GCCTTCTACCC GCCTTCCCCCC GCCTCCCCCC GCCTCCACCC GCCTCCACCC GCCATCTACCC GCCATCTACCC GCCATCTACCC GCCATCTACCC GCCATCTACCC GCCCTCTACCC GCCCTCTACCC	GGCCGCGCAGGGCCTCC GTCCGGCAACGCCTCC GTCCGCGCAGGGCTACCAC GCCCGCGCAGGGCCAACTC GACCGCGCAGGGCCACCC GCCCGCGCATGGCCACCCC GGCCGCGCAGGGCCACCCC GACCGCGCAGGGCCACCCC GACCGCGCAGAGGCACCCC GACCGCGCAGAGGCATCCC GACCGCGCAGAGGCATCCC GACCGCGCAGAGGCACCCC GACCGCGCAGAGGCACCCC GACCGCGCAGGGGCACCC GACCGCGCAGGGGCACCC	CTAG CTAG CTAG CTAG CTAG CTAG CTAG CTAG	CACGCATA GACGCATA CACGCATA CACGCATA TACGCATA CACGCATA		GAGC	GTGG GGGG GCGG GCGG GCGG GCGG GCGG GCGG	-AGCAAGGATC -CGCAAGGATC -AGCAAGGTTC -GGCATGGGTC -GGCAAGGATC -GGCAAGGATC -GGCACGGATC -CGCACGGATC -CGCAAGGATC -CGCAAGGATC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC
50993 4400 50174 45172 3592 3597 4283 4292 16924 3741 4442 125047 522403 3772 4288	GCCTCCACC GCCTTCTACT GCCTTCTACCC GCCTTCTACCC GCCTTCTACCC GCCTCCACCC GCCTCCACCC GCCTCCACCC GCCTCCACCC GCCATCTACCC GCCATCTACCC GCCATCTACCC GCCATCTACCC GCCATCTACCC GCCATCTACCC GCCATCTACCC GCCATCTACCC GCCCTCCACCC GCCCTCCACCC	GGCCGCGCAGGGCCTCC GTCCGGCAACGCCTACCAC GTCCGGCAGGGCTACCAC GGCCGGCAGGGCCAACTC GGCCGGCAAGGCCACCC GGCCGGCAGGGCCACCCC GGCCGGCAGGGCCACCCC GGCCGGCAGGGCCACCCC GACCGCGCAGAGGCACCCC GACCGCGCAGAGGCACCCC GACCGCGCAGAGGCACCCC GACCGCGCAGAGGCACCCC GACCGCGCAGAGGCACCCC GACCGCCAGAGGCACCCC GACCGCCAGAGGCACCCC GACCGCCAGAGGCACCCC GACCGCCAGGGCACCCC GACCGCCAGGGCACCCC	CTAG CTAG CTAG CTAG ATAG CTAG CTAG CTAG	CACGCATA GACGCATA CACGCATA	CCCCCGCACAACCGAGA CCCCCGCACACCGAGA CCCCCGCACACGCGAGA CCCCCGCACACGCACA CCCCCGCACATCGAGA CCCCCGCACATCGAGA CCCCCGCACATCGAGA CCCCCGCACATCGAGA CCCCCGCACACTCGAGA CCCCCGCACACTCGAGA CCCCCGCACACTCGAGA CCCCCGCACATCGAGA CCCCCGCACATTCGAGA CCCCCGCACATTCGAGA CCCCCGCACATTCGAGA CCCCCGCACATTCGAGA CCCCCGCACATTCGAGA	GAGC - CCAGCC GAGC - CCAGGA GAAT - CTAGTT GAGC - CTAGCT GAGC - CTAGCT CAGT - CAAGTG CAGC - CGAGTG CAGC - CGAGTG CAGC - CGAGTG CACC - CGAGTG CACC - CGAGTG CACC - CAAGTG	GTGG	-AGCAAGGATC -CGCAAGGATC -AGCAAGGTTC -GGCATGGGTC -GGCAAGGATC -GGCAAGGATC -CGCAAGGATC -CGCACGGATC -CGCAAGGATC -CGCAAGGATC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC -TGCAAGGATC
50993 4400 50174 45172 3592 3597 4283 4292 16924 3741 4442 125047 522403 3772 4288 16901	GCCTCCACC GCCTTCTACT GCCTTCTACCC GCCTTCTACCC GCCTTCCACCC GCCTCCACCC GCCTCCACCC GCCACTCTACCC GCCATCTACCC GCCATCTACCC GCCATCTACCC GCCATCTACCC GCCATCTACCC GCCCTCTACCC GCCCTCTACCC GCCCTCTACCC	GGCCGCGAGGGCCTCCCCGCCCGCCAACGCCAACGCCAACGCCAACGCCAACGCCAACGCCCCCC	CTAG CTAG CTAG ATAG CTAG CTAG CTAG CTAG	CACGCATA GACGCATA CACGCATA CACGCATA TACGCATA CACGCATA		GAGC	GTGG	-AGCAAGGATC -CGCAAGGATC -GGCATGGGTC -GGCAAGGGTC -GGCAAGGATC -GGCAAGGATC -CGCAAGGATC -CGCAAGGATC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC -TGCAAGGATC -TGCAAGGATC
50993 4400 50174 45172 3592 3597 4283 4292 16924 3741 4442 125047 522403 3772 4288 16901 4146	GCCTCCACC GCCTTCTACT GCCTTCTACCC GCCTTCTACCC GCCTTCTACCC GCCTTCCACCC GCCTCCACCC GCCTCCACCC GCCATCTACCC GCCATCTACCC GCCATCTACCC GCCATCTACCC GCCATCTACCC GCCCTCTACCC GCCCTCTACCC GCCCTCTACCC GCCCTCTACCC GCCCTCTACCC GCCCTCTACCC	GGCCGCGAGGGCCTCC GTCCGGCAACGCCTCC GTCCGCGCAGGGCTACCAC GCCCGCGCAGGGCCAACTC GACCGCGCAAGGCCACCC GGCCGGCAGGGCACCC GGCCGGCAGGGCCACCC GGCCGGCAGGGCACCC GACCGCGCAGGGCACCC GACCGCGCAGGGCACCC GACCGCCAGGGCACCC GACCGCGCAGGGCACCC GACCGCGCAGGGCACCC GACCGCGCAGGGCACCC GACCGCGCAGGGCACCC GACCGCGCAGGGCACCC GACCGCGCAGGGCACCC GACCGCGAGGGCACCC GACCGCGAACGCCC	CTAG CTAG CTAG ATAG CTAG CTAG CTAG CTAG	CACGCATA GACGCATA CACGCATA		GAGC	GTGG	-AGCAAGGATC -CGCAAGGATC -AGCAAGGTTC -GGCATGGGTC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC -CGCAAGGATC -CGCAAGGATC -CGCAAGGATC -GGCAAGGATC -GGCAAGGATC -TGCAAGGATC -TGCAAGGATC -TGCAGCGATC
50993 4400 50174 45172 3592 3597 4283 4292 16924 3741 4442 125047 522403 3772 4288 16901 4146 3770	GCCTCCACC GCCTTCTACT GCCTTCTACT GCCTTCTACC GCCTTCTACC GCCTCCACC GCCTCCACC GCCTCCACC GCCTCCACC GCCTCCACC GCCATCTACC GCCATCTACC GCCATCTACC GCCATCTACC GCCTCTACC GCCTCTACC GCCTCCACC GCCTCCACC GCCTCCACC GCCTCCACC	GGCCGCGTAGGCCTCCCCCCCCCCCCCCCCCCCCCCCCC	CTAG CTAG CTAG CTAG ATAG CTAG CTAG CTAG	CACGCATA GACGCATA CACGCATA	CCCCGCACAACCGAGA CCCCCGCACACCGAGA CCCCCGCACACCGAGA CCCCCGCACACCGAGA CCCCCGCACACCGAGA CCCCCGCACACCGAGA CCCCCGCACACCGAGA CCCCCGCACACTCGAGA CCCCCGCACATCGAGA CCCCCGCACACTCGAGA	GAGC - CCAGCC GAGC - CCAGGA GAAT - CTAGTT GAGC - CTAGCT GAGC - CTAGCT CAGT - CAAGTG CAGC - CGAGTG CACC - CGAGTG CACC - CAAGTG CACC - CAAGTC CACC - CAAGTG CACC - CAAGTC	GTGG GGGG GCGG GCGG GCGG GCGG GCGG GCGG	-AGCAAGGATC -CGCAAGGATC -AGCAAGGTTC -GGCATGGGTC -GGCAAGGATC -GGCAAGGATC -CGCAAGGATC -CGCAAGGATC -CGCAAGGATC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC -TGCAAGGATC -TGCAAGGATC -TGCAAGGATC -TGCAAGGATC -TGCAAGGATC -TGCAAGGATC -TGCAAGGATC -CGCAACGGTC
50993 4400 50174 45172 3592 3597 4283 4292 16924 3741 4442 125047 522403 3772 4288 16901 4146 3770 39347	GCCTCCACC GCCTTCTACT GCCTTCTACT GCCTTCTACCC GCCTTCTACCC GCCTTCTACCC GCCTCCACCC GCCTCCACCC GCCTCTACCC GCCTCTACCC GCCTCTACCC GCCTCTACCC GCCTCTACCC GCCTCTACCC GCCTCTACCC GCCTCTACCC GCCTCTACCC GCCCTCTACCC GCCTCTACCC GCCTCCACCC	GGCCGCGCAGGGCCTCCCCGTCCGCCGCAACGCCCCCGGCAGGCCAACTCCCCGGCAGGGCCACCCCGGCCAGGGCCACCCCCGGCCGCAGGGCCACCCCCGGCCGCAGGGCCACCCCCGGCCGCAGGGCCACCCCCGGCCGCAGGGCCACCCCCGGCCGCAGGGCCACCCCCGGCCGCAGGGCCACCCCCGACCGCAGGGCACCCCCGGCCGCAGGGCACCCCCGACCGCAGGGCACCCCCGACCGCAGGGCACCCCCGACCGCAGGGCACCCCCGGACGGCAGGGCACCCCGGCAGGGCACCCCGGCAGGGCACCCCGGCAGGGCACCCCGGCCAGGGCACCCCGGCCGCAGGGCACCCCCGGCCGCAGGCCACCCCCGCCCGCCAGCCCCCC	CTAG CTAG CTAG ATAG CTAG CTAG CTAG CTAG	CACGCATA GACGCATA CACGCATA TACGCATA CACGCATA TACGCATA	CCCCCGCACAACCGAGA CCCCCGCACACCGAGACCCGAGACCCGAGACCCGACACCGAGACCCGACACCGAGACCCGAGACCCGAGACCCCGACACCCGAGACCCCCGCACACCAC	GAGC - CCAGCC GAGC - CCAGGA GAGT - CTAGTT GAGC - CTAGCT GAGC - CTAGCT CAGT - CAAGTG CAGC - CGAGTG CAGC - CGAGTG CAGC - CGAGTG CACC - CAAGTG CAGC - CCAGTG CAGC - CCAGTG CAGG - CCAGTG CAGG - CCAGTG CAGG - CCAGTG CAGG - CCAGTG	GTGG	-AGCAAGGATC -CGCAAGGATC -GGCATGGGTC -GGCAAGGGTC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC -TGCAGGATC -TGCAGGATC -TGCAGGATC -TGCAGGATC -TGCAGGATC -TGCAGGATC -TGCAGGATC
50993 4400 50174 45172 3592 3597 4283 4292 16924 3741 4442 125047 522403 3772 4288 16901 4146 3770 39347 443381	GCCTCCACC GCCTTCTACT GCCTTCTACT GCCTTCTACCC GCCTTCTACCC GCCTTCACCC GCCTTCACCC GCCTTCACCC GCCTTCACCC GCCTTCACCC GCCATCTACCC GCCATCTACCC GCCATCTACCC GCCTTCACCC GCCTCTACCC GCCTCACCC GCCTCCACCC	GGCCGCGAGGGCCTCCC GTCCGGCAACGCCTCCC GTCCGCGCAGGGCTACCAC GCCCGCGCAGGGCCAACTC GACCGCGCAGGGCCACCCC GGCCGGCAGGGCCACCCC GGCCGGCAGGGCCACCCC GACCGCGCAGGGCCACCCC GACCGCGCAGGGCACCCC GACCGCGCAGGGCACCCC GACCGCGCAGGGCACCCC GACCGCGCAGGGCACCCC GACCGCGCAGGGCCACCCC GACCGCGTAGTGCCCC GCCCGCGTACTGCCCC GCCCGCGCAGGGCCACCCC GCCCGCGCAGGGCCACCCC GCCCGCGCAGGCCACCCC GCCCGCGCAGGTCCACCC GCCCGCGCAGGTCCACCC GCCCGCGCAGGTCCACCC GCCCGCGCAGGTCCACCC GCCCGCGCAGGTCCACCC GCCCGCGCAGGTCCACCC GCCCGCGCAGGTCCACCC GCCCGCCAGGTCCACCC GCCCGCCAGGTCCACCC GCCCGCCAGGTCCACCC GCCCGCCAGGTCCACCC GCCCGCCAGGTCCACCC GCCCGCCAGGTCCACCC GCCCGCCAGGTCCACCC	CTAG CTAG CTAG ATAG CTAG CTAG CTAG CTAG	CACGCATA GACGCATA CACGCATA		GAGC - CCAGCC GAGC - CCAGGA GAGT - CTAGTT GAGC - CTAGCT GAGC - CTAGCT CAGT - CAAGTG CAGC - CGAGTC CAGC - CGAGTG CAGC - CGAGTG CAGC - CGAGTG CAGC - CAAGTG CACC - CCAGTC CAGG - CGAGTG CAGG - CGAGTG	GTGG	-AGCAAGGATC -CGCAAGGATC -GGCATGGGTC -GGCAAGGGTC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC -CGCAAGGATC -GGCAAGGATC -GGCAAGGATC -TGCAAGGATC -TGCAAGGATC -TGCAAGGATC -TGCAAGGATC -TGCAGCGATC -TGCAGCGATC -CGCAACGGTC -CGCAACGGTC -CGCAACGGTC -CGCAACGGTC -AGCAGCGGTC -AGCAGCGGTC
50993 4400 50174 45172 3592 3597 4283 4292 16924 3741 4442 125047 522403 3772 4288 16901 4146 3770 39347 443381 39388	GCCTCCACC GCCTTCTACT GCCTTCTACCC GCCTTCTACCC GCCTTCCACCC GCCTCCACCC GCCTCCACCC GCCATCTACCC GCCATCTACCC GCCATCTACCC GCCATCTACCC GCCATCTACCC GCCATCTACCC GCCCTCCACCC GCCCTCCACCC GCCCTCCACCC GCCTCCACCC GCCTCCACCC GCCTCTACCC	GGCCGCGCAGGGCCTCCCCGCCGCAACGCCCCCGGCAGGCCACCCCCGGCAGGGCCACCCCCGGCCAGGGCCACCCCCGGCCAGGGCCACCCCCGGCCAGGGCCACCCCCGGCCAGGGCCACCCCCGGCCAGGGCAGGCCACCCCCGGCCAGGGCAGGCACCCCCGGCCAGGGCAGGCACCCCCGGCCAGGGCACCCCCGGCCAGGGCACCCCCGGCCAGGGCACCCCCGGCCAGGGCACCCCCGGCCAGGGCACCCCCGGCCAGGGCCACCCCCGGCCAGGGCCACCCCCGGCCAGGGCCACCCCCGGCCAGGGCCACCCCCC	CTAG CTAG CTAG CTAG ATAG CTAG CTAG CTAG	CACGCATA GACGCATA CACGCATA	CCCCCGCACAACCGAGA CCCCCGCACACCGAGACCCGAGACCCGAGACCCGACACCGAGACCCGAGACCCGAGACCCGAGACCCGAGACCCGAGACCCCGACAATCGAGACCCCCCGCACAATCGAGACCCCCCGCACAATCGAGACCCCCCGCACAATCGAGACCCCCGCACAATCGAGACCCCCGCACAATCGAGACCCCCGCACAATCGAGACCCCCGCACAATCGAGACCCCCGCACAATCGAGACCCCCGCACAATCGAGACCCCCGCACAATCGAGACCCCCGCACAATCGAGACCCCCGCACAATCGAGACCCCCGCACAATCGAGACCCCCCGCACAATCGAGACCCCCCGCACAATCGAGACCCCCCGCACAATCGAGACCCCCGCACAACCGAGACCGACAACCGAGACCGCCCGCACAACCGAGACCGACACCGCCCGCACAACCGAGACCGCCCGCACAACCGACACCGCCCGCACAACCGACACCGCCCGCACAACCGACACCGCCCGCACAACCGACACCGCCCCGCACAACCGACACACCGACACACCGACACCGACACCGACACCGACACCGACACCGACACCGACACACCGACACCAC	GAGC - CCAGCC GAGC - CCAGGA GAAT - CTAGTT GAGC - CTAGCT GAGC - CTAGCT CAGT - CAAGTG CAGC - CGAGTG CAGC - CGAGTG CAGC - CGAGTG CAGC - CGAGTG CACC - CGAGTG CACC - CAAGTG CACC - CAAGTC CACC - CAAGTG CACC - CCAGTG CAGC - CCAGTG CAGG - CCAGTG	GTGG GGGG GCGG GCGG GCGG GCGG GCGG GCGG	-AGCAAGGATC -CGCAAGGATC -AGCAAGGATC -GGCATGGGTC -GGCAAGGATC -GGCAAGGATC -CGCAAGGATC -CGCAAGGATC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC -TGCAAGGATC -TGCAAGGATC -TGCAAGGATC -TGCAAGGATC -CCCAACGGTC -CCCAACGGTC -CGCAACGGTC -AGCAGCGTC -CGCAACGGTC -CGCAACGGTC -CGCAACGGTC -CGCAACGGTC
50993 4400 50174 45172 3592 3597 4283 4292 16924 3741 4442 125047 522403 3772 4288 16901 4146 3770 39347 443381 39388 4151	GCCTCCACC GCCTTCTACT GCCTTCTACT GCCTTCTACCC GCCTTCTACCC GCCTTCTACCC GCCTCCACCC GCCTCCACCC GCCATCTACCC GCCATCTACCC GCCATCTACCC GCCATCTACCC GCCATCTACCC GCCATCTACCC GCCTCCACCC GCCTCGACCC GCCTCGACCC GCCCTCGACCC GCCCTCGACCC	GGCCGCGCAGGGCCTCCCCGTCCGCCGCAGGGCAGGCCACCCCCGGCAGGGCCACCCCCGGCCAGGGCCACCCCCGGCCGGCAGGCCACCCCCGGCCGGCAGGGCCACCCCCGGCCGGCAGGGCCACCCCCGGCCGGCAGGGCACCCCCGGCCGGCAGGGCACCCCCGGCCGGCAGGGCACCCCCGGCCGGGCAGGGCACCCCCGGCCGGAGGGCACCCCCGGCCGGAGGGCACCCCCGGCCGGAGGGCACCCCCGGCCGGAGGGCACCCCCGGCCGGAGGGCACCCCCGGCCGGAGGGCACCCCCGGCCGGCAGGGCACCCCCGGCCGGCAGGGCCACCCCCGGCCGGCAGGGCCACCCCCGGCCGGCAGGGCCACCCCCGCCG	CTAG CTAG CTAG CTAG CTAG CTAG CTAG CTAG	CACGCATA GACGCATA CACGCATA ACCCCATA	CCCCCGCACAACCGAGA CCCCCGCACAACCGAGA CCCCCGCACACGCGAGA CCCCCGCACACGCAGA CCCCCGCACACCCGAGA CCCCCGCACACCCGAGA CCCCCGCACACCCGAGA CCCCCGCACAATCGAGA CCCCCGCACAATCGAGA CCCCCGCACACTCGAGA CCCCCGCACACTCGAGA CCCCCGCACACTCGAGA CCCCCGCACATTCGAGA CCCCCGCACATTCGAGA CCCCCGCACATTCGAGA CCCCCGCACATTCGAGA CCCCCGCACATTCGAGA CCCCCGCACATTCGAGA CCCCCGCACATTCGAGA CCCCCGCACATTCGAGA CCCCCGCACAATCGAGA CCCCCGCACAATCGAGA CCCCCGCACAATCGAGA CCCCCGCACAACCGAGA CCCCCGCACAACCGAGA CCCCCGCACAACCGAGA CGCCCGCACAACCGACA CGCCCGCACACGGACA CCCCCGCACACCGACA	GAGC - CCAGCC GAGC - CCAGGA GAGT - CTAGTT GAGC - CTAGCT GAGC - CTAGCT CAGT - CAAGTG CAGC - CGAGTG CAGC - CGAGTG CAGC - CGAGTG CACC - CGAGTG CACC - CAAGTG CAGC - CAAGTG CAGC - CAAGTG CAGC - CCAGTG CAGG - CGAGTG CAGG - CGAGTG CAGG - CCAGTG	GTGG GGGG GCGG GCGG GCGG GCGG GCGG GCGG	-AGCAAGGATC -CGCAAGGATC -GGCATGGGTC -GGCATGGGTC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC -TGCAAGGATC -TGCAAGGATC -TGCAGCGATC -CGCAAGGATC -CGCAAGGATC -CGCAACGGTC -CGCAACGGTC -CGCAACGGTC -CGCAACGGTC -GGCAACGGTC -GGCAACGGTC -GGCAACGGTC -GGCAACGGTC
50993 4400 50174 45172 3592 3597 4283 4292 16924 3741 4442 125047 522403 3772 4288 16901 4146 3770 39347 443381 39388 4151 197815	GCCTCCACC GCCTTCTACT GCCTTCTACT GCCTTCTACCC GCCTTCTACCC GCCTTCTACCC GCCTCCACCC GCCTCCACCC GCCATCTACCC GCCATCTACCC GCCATCTACCC GCCATCTACCC GCCATCTACCC GCCTCTACCC GCCTCTACCC GCCTCTACCC GCCTCCACCC GCCCTCCACCC GCCTCCACCC GCCTCCACCC	GGCCGCGAGGGCCTCCC GTCCGGCAACGCCTCCC GTCCGCGCAGGGCTACCAC GCCCGCGCAGGGCCAACTC GACCGCGCAAGGCCAACTC GGCCGGCAAGGCCACCC GGCCGGCAGGGCCACCC GGCCGGCAGGGCCACCC GACCGCGCAGGGCACCCC GACCGCGCAGGGCACCCC GACCGCGCAGGGCACCCC GACCGCGCATGGTACGCC GACCGCGTACTGCCCC GCCCGCGTACTGCCCC GCCCGCGCAGGGCCACCCC GCCCGCGCAGGTCCCC GCCCGCGCAGGTCCCC GCCCGCGCAGGTCCACCC GCCCGCGCAGGTCCACCC GCCCGCCAGGTCCACCC GCCCGCCAGGTCCACCC GCCCGCCAGGTCCACCC GCCCGCCAGGTCCACCC GCCCGCCAGGTCCACCC GCCCGCGCAGGTCCACCC	CTAG CTAG CTAG CTAG CTAG CTAG CTAG CTAG	CACGCATA GACGCATA CACGCATA		GAGC - CCAGCC GAGC - CCAGGA GAGT - CTAGTT GAGC - CTAGCT GAGC - CTAGCT CAGT - CAGTG CAGC - CGAGTG CAGC - CGAGTG CAGC - CGAGTG CAGC - CAGTG CAGC - CAGTG CACC - CAGTG CAGC - CAGTG CAGG - CAGGTG CAGG - C	GTGG GGGG GCGG GCGG GCGG GCGG GCGG GCGG	-AGCAAGGATC -CGCAAGGATC -GGCATGGGTC -GGCAAGGGTC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC -TGCAAGGATC -TGCAAGGATC -TGCAGCGATC -CGCAACGGTC
50993 4400 50174 45172 3592 3597 4283 4292 16924 3741 4442 125047 522403 3772 4288 16901 4146 3770 39347 443381 39388 4151 19388 4151 201512	GCCTCCACC GCCTTCTACT GCCTTCTACT GCCTTCTACCC GCCTTCTACCC GCCTTCCACCC GCCTTCCACCC GCCTTCACCC GCCTTCACCC GCCTTCACCC GCCTTCACCC GCCTTCACCC GCCTTCACCC GCCTTCACCC GCCTTCACCC GCCTCTACCC GCCTCTACCC GCCTTCACCC	GGCCGCGAGGGCCTCCC GTCCGGCAACGCCTCCC GTCCGGCAACGCCTACCAC GCCCGCGCAGGGCCAACTC GACCGCGCAGGGCCAACTC GACCGCGCAGGGCCACCCC GGCCGCGCAGGGCCACCCC GACCGCGCAGGGCCACCCC GACCGCGCAGGGCACCCC GACCGCGCAGGGCACCCC GACCGCGCAGGGCACCCC GACCGCGCAGGGCCACCCC GACCGCGCAGGGCCACCCC GACCGCGTAGTGCCACCCC GCCCGCGTAGTGCCACCCC GCCCGCGCAGGGCCACCCC GCCCGCGCAGGTCCACCCC GCCCGCGCAGGGCCCC GCCCGCGCAGGGCCCC GCCCGCGCAGGGCCCC GCCCGCCAGGGCCCC	CTAG CTAG CTAG ATAG CTAG CTAG CTAG CTAG	CACGCATA GACGCATA CACGCATA	CCCCGCACAACCGAGA CCCCGGCACAACCGAGA CCCCGGCACACGGAGA CCCCGGCACACGGAGA CCCCGGCACATCGAGA CCCCGGCACATCGAGA CCCCGGCACATCGAGA CCCCGCACATCGAGA CCCCGCACATCGAGA CCCCGCACATTCGAGA CCCCGCACATTCGAGA CCCCGCACATTCGAGA CCCCGCACATTCGAGA CCCCCGCACATTCGAGA CCCCCGCACATTCGAGA CCCCGGCACATTCGAGA CCCCGGCACATTCGAGA CCCCGGCACATTCGAGA CCCCGGCACATTCGAGA CCCCGGCACATTCGAGA CGCCGGCACATTCGAGA CGCCGCACATTCGAGA CGCCGCACATTCGAGA CGCCGCACATTCGAGA CGCCGCACATTCGACA CGCCGCACATTCGAGA CGCCGCACATCGGACA CGCCGCACATCGACA CGCCGCACATCGACA CCCCGCACATACCGACA CCCCGCACATACGACA CCCCGCACATAACGACA CCCCGCACATAACGACA CCCCGCACATACGACA CCCCGCACATACGACA CCCCGCACATACGACA CCCCGCACATACGACA CCCCCGCACATACGACA CCCCCGCACATACGACA CCCCCGCACATACGACA CCCCCGCACATACGACA CCCCCCCCACATACGACA CCCCCCCCACATACGACA CCCCCCCCACATACGACA CCCCCCCCCACATACGACA CCCCCCCCCACATACGACA CCCCCCCCCC	GAGC - CCAGCC GAGC - CCAGGA GAGT - CTAGTT GAGC - CTAGCT GAGC - CTAGCT CAGT - CAAGTG CAGC - CGAGTC CAGC - CGAGTG CAGC - CGAGTG CAGC - CGAGTG CACC - CAAGTG CAGC - CGAGTG CAGG - CGAGTG CAGG - CGAGTG CAGG - CCAGTG CAGG - CAAGCT - CAAGCT - CAGG - CAAGCT - CAAGCT - CAAGCT - CAAGCT - CAAGCT - CAAGCT - CAAG	GTGG GGGG GCGG GCGG GCGG GCGG GCGG GCGG	-AGCAAGGATC -CGCAAGGATC -GGCATGGGTC -AGCAAGGATC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC -TGCAAGGATC -TGCAAGGATC -TGCAGCGATC -CCCAACGGTC -CCCACCGGTC -CCCACCGGTC -CGCAACGGTC
50993 4400 50174 45172 3592 3597 4283 4292 16924 3741 4442 125047 522403 3772 4288 16901 4146 3770 39347 443381 39388 4151 197815 201512 39353	GCCTCCACC GCCTTCTACT GCCTTCTACCC GCCTTCTACCC GCCTTCTACCC GCCTTCTACCC GCCTCCACCC GCCTCCACCC GCCTCCACCC GCCATCTACCC GCCATCTACCC GCCATCTACCC GCCATCTACCC GCCATCTACCC GCCTCCACCC GCCTCCACCC GCCTCCACCC GCCTCCACCC GCCTCCACCC GCCTCCACCC GCCTCCACCC GCCTCCACCC GCCTCGACCC GCCTCGACCC GCCTCGACCC GCCTCGACCC GCCTCGACCC GCCTCGACCC GCCTCGACCC GCCTCGACCC GCCTCGACCC	GGCCGCGAGGGCCTCCCCGCCGCAGGGCAGGCACCCCCGGCAGGTCACCCCGGCAGGCCACCCCGGCCAGGGCCACCCCGGCCAGGCCAGGCCACCCCCGGCCGCAGGCCACCCCCGGCCGCAGGCCACCCCCGGCCGCAGGCCACCCCCGACCCCCGGCCAGGCCACCCCCGACCCCCGGCCAGGCAGGCACCCCCGACCCCCGCCAGGCAGGCACCCCCGACCGCAGGGCACCCCCGACCGCAGGGCACCCCCGCCAGGCCACCCCCGCCAGGCCACCCCCGCCAGGCCACCCCCGCCAGGCCACCCCCGCCAGGCCACCCCCGCCAGGCCACCCCCGCCAGGCCACCCCCGCCAGGCCAGGCCACCCCCGCCAGGCCACCCCCC	CTAG CTAG CTAG CTAG CTAG CTAG CTAG CTAG	CACGCATA GACGCATA CACGCATA	CCCCCGCACAACCGAGA CCCCCGCACAACCGAGA CCCCCGCACACCGAGAC CCCCCGCACACCGAGAC CCCCCGCACACCCGAGA CCCCCGCACACCCGAGA CCCCCGCACACCCGAGA CCCCCGCACACCGAGA CCCCCGCACACTCGAGA CCCCCGCACACTCGAGA CCCCCGCACACTCGAGA CCCCCGCACACTCGAGA CCCCCGCACATTCGAGA CCCCCGCACATTCGAGA CCCCCGCACATTCGAGA CCCCCGCACATTCGAGA CCCCCGCACATTCGAGA CCCCCGCACATTCGAGA CCCCCGCACATTCGAGA CCCCCGCACATTCGAGA CCCCCGCACATTCGAGA CCCCCGCACACTCGAGA CCCCCGCACACTCGAGA CCCCCGCACACTCGAGA CCCCCGCACAACCGAGA CCCCCGCACAACCGAGA CCCCCGCACAACCGACA CGCCCGCACAACCGACA ATCCCGCACTAGGACA ATCCCGCATTAGACGACA CCCCCGCACTTAGACGACA CCCCCGCACTTAGACGACA CCCCCGCACTTAGACGACA CCCCCGCACTTAGACGACA CCCCCGCACTTAGACGACA CCCCCGCACTTAGACGACA CCCCCGCACTTAGACGACA CCCCCGCACTTAGACGACA CCCCCGCACTTAGACGACA CCCCCCGCACTTAGACGACA CCCCCGCACTTAGACGACA CCCCCCGCACTTAGACGAGA CCCCCCGCACTTAGACGAGA CCCCCCGCACTAGACCGAGA	GAGC - CCAGCC GAGC - CCAGGA GAGT - CTAGTT GAGC - CTAGCT GAGC - CTAGCT CAGT - CAAGTG CAGC - CGAGTC CAGC - CGAGTG CACC - CGAGTG CACC - CGAGTG CACC - CAAGTG CAGC - CCAGTG CAGG - CCAGTG CAGC - CCAGTG CA	GTGG GGGG GCGG GCGG GCGG GCGG GCGG GCGG	-AGCAAGGATC -CGCAAGGATC -GGCATGGGTC -GGCAAGGGTC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC -TGCAGCGATC -TGCAGCGATC -TGCAGCGATC -CGCAACGGTC -CGCAACGGTC -CGCAACGGTC -GGCAACGGTC -GGCAACGGTC -GGCAACGGTC -GGCAACGGTC -GGCAACGGTC -GGCAACGGTC -GGCAACGGTC -GGCAACGGTC -GGCAACGGTC -CGCAACGGTC
50993 4400 50174 45172 3592 3597 4283 4292 16924 3741 4442 125047 522403 3772 4288 16901 4146 3770 39347 443381 39388 4151 19388 4151 201512	GCCTCCACC GCCTTCTACT GCCTTCTACT GCCTTCTACC GCCTTCTACC GCCTTCTACC GCCTTCTACC GCCTTCACC GCCTCCACC GCCTCTACC GCCTCCACC	GGCCGCGAGGGCCTCCC GTCCGGCAACGCCTCCC GTCCGGCAACGCCTACCAC GCCCGCGCAGGGCCAACTC GACCGCGCAGGGCCAACTC GACCGCGCAGGGCCACCCC GGCCGCGCAGGGCCACCCC GACCGCGCAGGGCCACCCC GACCGCGCAGGGCACCCC GACCGCGCAGGGCACCCC GACCGCGCAGGGCACCCC GACCGCGCAGGGCCACCCC GACCGCGCAGGGCCACCCC GACCGCGTAGTGCCACCCC GCCCGCGTAGTGCCACCCC GCCCGCGCAGGGCCACCCC GCCCGCGCAGGTCCACCCC GCCCGCGCAGGGCCCC GCCCGCGCAGGGCCCC GCCCGCGCAGGGCCCC GCCCGCCAGGGCCCC	CTAG	CACGCATA GACGCATA CACGCATA	CCCCGCACAACCGAGA CCCCGGCACAACCGAGA CCCCGGCACACGGAGA CCCCGGCACACGGAGA CCCCGGCACATCGAGA CCCCGGCACATCGAGA CCCCGGCACATCGAGA CCCCGCACATCGAGA CCCCGCACATCGAGA CCCCGCACATTCGAGA CCCCGCACATTCGAGA CCCCGCACATTCGAGA CCCCGCACATTCGAGA CCCCCGCACATTCGAGA CCCCCGCACATTCGAGA CCCCGGCACATTCGAGA CCCCGGCACATTCGAGA CCCCGGCACATTCGAGA CCCCGGCACATTCGAGA CCCCGGCACATTCGAGA CGCCGGCACATTCGAGA CGCCGCACATTCGAGA CGCCGCACATTCGAGA CGCCGCACATTCGAGA CGCCGCACATTCGACA CGCCGCACATTCGAGA CGCCGCACATCGGACA CGCCGCACATCGACA CGCCGCACATCGACA CCCCGCACATACCGACA CCCCGCACATACGACA CCCCGCACATAACGACA CCCCGCACATAACGACA CCCCGCACATACGACA CCCCGCACATACGACA CCCCGCACATACGACA CCCCGCACATACGACA CCCCCGCACATACGACA CCCCCGCACATACGACA CCCCCGCACATACGACA CCCCCGCACATACGACA CCCCCCCCACATACGACA CCCCCCCCACATACGACA CCCCCCCCACATACGACA CCCCCCCCCACATACGACA CCCCCCCCCACATACGACA CCCCCCCCCC	GAGC - CCAGCC GAGC - CCAGGA GAGT - CTAGTT GAGC - CTAGCT GAGC - CTAGCT CAGT - CAAGTG CAGC - CGAGTG CAGC - CGAGTG CAGC - CGAGTG CACC - CAAGTG CAGC - CCAGTG CAGG - CCAGTG CAGC - CCAGTG	GTGG GGGG GCGG GCGG GCGG GCGG GCGG GCGG	-AGCAAGGATC -CGCAAGGATC -GGCATGGGTC -AGCAAGGATC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC -GGCAAGGATC -TGCAAGGATC -TGCAAGGATC -TGCAGCGATC -CCCAACGGTC -CCCAACGGTC -CCCAACGGTC -CGCAACGGTC

GUIDANCE as a measure of aligning difficulty



Column score: 2/3

GUIDANCE as a measure of aligning difficulty



Residue score: 3/3

GUIDANCE as a measure of aligning difficulty

Real sequences

CGAA CTCGTGACCCCTTTTTTTGGGGGTAGGGCGAAGGG CTGCTCTCAGT -ATCGACACAC -CCGCGAACACGTTATCATCCACTCTGGTGGGTTGCCCACCCCGGGTGGG CGTCCCCCACC ACTAGGCTGAC -CGGCGAA CTTGGTATTGTTCCTTGAGGGTAGGGGTCATTCCTTTCAGTATACGGAA CGATGACC CACTAGGCTGAC CGGCGAA TGATGCGTA CCAAAACAAGATCGACCGAAGCGAA TCGAGGACC GAACGAGGGTAGAC CCGCGAA TGGAGAACCGAACCCCTGCAAAGCAGCACGACTG GTGTGAA TATGTCAATTGTCATCTGGGGGTGGGGACGCTAGCGCC TAG CTTGTGAATTATCATTTGGGGGCGGGGACACTAGTGCC TAG 112846 PCGCDBACC 3426 TCGTTGCCT GCAACGGCAGACGAC CCGCGAA GCC CAGCAGAACGAC CCGTGAA TCGAAACCT -CCAGTTGATATCACCGGCGGGGGGGGGGGGGGGGGCCC-CGC -CCAGTTGATATCACCGGCGGCGGGAGGGGGGGCGCTC-CGC -TACGTGAACGTGCTGAGGGGGGGACGACGGCGCTCCGCG-CAC GCC TAGCAGAACGAC CTGTGAA-GCCTAGCAGAATGAC CAGCGAA TTTGTGAATGTTTTGGGGGGCGAGGGCTGTGCTTGTGC-AGC CTTGGTCGGATACGGATACGGCCATCGGCCTTTCGTAG-TTG TCGAAACCT -GCC-TAGCAGAACGAC---CCGGAAC-----TAGCGGGGGGGGGGGGGGTCCTGCGG-CTC - CACGTACCTTGGGTGGGCGAGAGGAGCTTGCTCCTTGG - ACCCGCCCTCAC - CACGTG - CTTGGGCGGGCGAGAGGAGCTTGCTCCTTGG - ACCCGCCCGCAC TCGAAACCT GCAACAGCAGAACGAC GCAACAGCAGAACGAC TCGATGCCT -GCATAGCCCCGCGGGTTTTTCAATTCTACGGGAGCCTCCCGCGGGAGAGC-TGTCCCCTCGC CCCAGAA CCGGTTGGAACAAAACTTTGGCGGGGGAGGCCGAGGCC-CCACCCCCG TCGGCTTA CCTCTTGGTTCAGAGGAAGACGACGAAGTGC CCGC AAAAAGAGCGAC CCGAGAA TCGATGCCTTACATGCAGTCCAACAGGAATCAACACCTCGGCTTA TCGAATCCT ACA GAGCGAACGAC CCGCGAA -CTCGTTTATCCACGGGCGTCGGGCGCGGGGTCCTGCCC-CCGCCCGTC-

TGCCCCCGT	
CGACGCCCC	
TGACGCCCC	
C	
CCAGCCTGA	
CCGCCCCAG	
.C	
TGCCCC	
GGTGCCTCC	
CCCGGCGTC	
AGCGGG	
AGCGGG	
AGCGGG	
CCTTTGCTC	
TGTCGT	
GCCCTGTTT	
CTTGTC-	

Simulated sequences

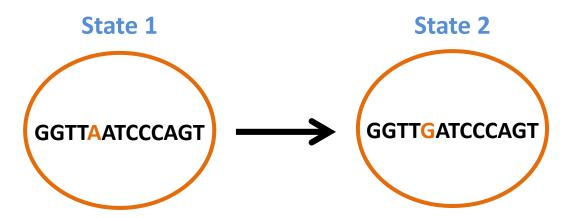
50993	GCCCTTCACCGACCGCGCAGGCCCACCCC	CTAG CACGCATA	CCCCCGCACAACCGAGAGAGT	CGAGTC	GGGG CGCACGGATC
4400	GCCCTCCACCGGCCGCGCAGGGCCTCTCC	CTAG CACGCATA	CCCCCGCGCCTTCGAGAGAGC	CCAGCC	GTGG AGCAAGGATC
50174	GCCTTCTACTGTCCGCGCAACGCCGCTCC	CTAG GACGCATA		CCAGGA	GGGG CGCAAGGATC
45172	GCCCTCCACCGTCCGCGCAGGGCTACCAC	CTAG CACGCATA	CCCCGGCACCATCGAGAGAAT	CTAGTT	GCGG AGCAAGGTTC
3592	GCCTTCTACCGCCCGCGCAGGGCCAACTC	CTAG CACGCATA	CGCCCGCACAGTCGAGAGAGC	CTAGCG	GCGG GGCATGGGTC
3597	GCCTTCTACCGACCGCGCAAGGCCACCCC	ATAG TACGCATA		CTAGCT	GCGG AGCAGGGTTC
4283	GCCCTCCCCCGCCCGCGCATGGCCACCCC	CTAG CACGCATA	CGCCCGCACAATCGAAACAGT	CAAGTG	GCGG GGCAAGGATC
4292	GCCCTCCACCGGCCGCGCAGGGCCACCCC	CTAG CACGCATA		CGAGTC	TCAG GGCAAGGCTC
16924	GCCCTCCGCCGGCCGCGCGCGGGCCACCCC	CTAG CACGCATA	CCCCCGCACAATCGAGACAGC	CGAGTG	GCGG GGCATGGATC
3741	GCCATCTACCGACCGCGCAGAGGCACCCC	CTAG CACGCATA		CGAGTG	GGGG CGCACGGATC
4442	GCCATCTACCGACCGCGCAGAGGCATCCC	CTAG CACGCATA	CCCCCGCACACTCGAGACAGT	CGAGTG	GTGGCGCAAGGATC
125047	GCCATCTACCGCCCGCGCAGGGGCACCCC	CTAG CACGCATA	CCCCCGCACAATCGAGACACC	CGAGTG	GTGG GGCAAGGATC
522403	GCCCTCTACCGACCGCGCATGGGTACGCC	CTAG CATGCATA	CCCCCGCACAGTCGAGACACC	CAAGTC	GCGG GGCAAGGATC
3772	GCCCTCTGCCGACCGCGCAGGGCCACCCC	CTAG CACGCATA	CTCCTGCACAGTCGAGGCAGC	CAAGTA	GAGG TGCAAGGATC
4288	GCCCTCCACCGACCGCGTAGTGCCACCCC	CTAG CACGCATA	CCCCCGCACAATCGAGACACC	CAAGCG	GCGG TGCAGCGATC
16901	GCCCTCCACCGCCCGCGTAGTGCCTCCCC	CTAG CACGCATA		CAAGGG	GCGG TGCAGCGATC
4146	GCCCTCTACCGCCCGCGCATGTCCACCAC	CTAG CACGCATA	CCCCCGCACAACCGAGACAGC	CGAGTG	GAGG-CGCAACGGTC
3770	GCCCTCGACCGCCCGCGCAGGGCCACCCC	CTAG TACGCATA		CCAGTC	GGGG CCCACCGGTC
39347	GCCCTCGACTGTCCGCGCAGGTCCACCCC	CTAG CATGCATA	CGCCCGCACATTCGACCCAGG	CGAGTG	GTGA AGCAGCGGTC
443381	GCCCTCGACCGCCCGCGCAGGTCCATCCC	CTAG CACGCATA	CGCCCGCACAACCGACACAGG	CGAGTG	GAGA CGCAACGGTC
39388	GCCCTCGACCGCCCGCGCAGGTCCACCCC	CTAG CACGCATA	CGCCCGCACAAGCGACACAGA	CGAGTG	GTGACGCAACGGTC
4151	GCCCTCGCCGGGCCGCGCAAGTCCACCCC	CTAG AACGCATA	ATCCCGCATAAACGATACAGG	CCAGTG	GAGA GGCAACGGTC
197815	GCCCTCGGCCGACCGCGGAGGTCGGCCCC	CTAG CACGCATA	TGCCCGCATGGCCGAGACAGG	CAAGCT	GGGACGCAACGATC
201512	GCCCTCGACCGCCCGCGCAGGGCCACCCC	CTAG CACGCATA	CTCCCGCACATACGAGACAGC	CGAGTC	GTGG CGCAACGTTC
39353	GCCCTCGACCGCCCGCGCAGGTCCACACC	CTAG CACGCATA		CCAGTG	GAGG CGCAACGCTC
28536	GCCCTCGACCGTCCGCGCAGGACCACCCC	CTAG CACGCATA		CCAGTG	GCGA TGCAGCGGTC
49121	GCCCTCGACCGACCGCGCAGGACCTCCCC	CTAG CACGCATA		CGAGTG	GGGG CGCAGCGGTC

Gene	Residues score	Column score
18S	0.99955	0.854151
26S	0.999351	0.742893
atpB	0.999953	0.947041
matK	0.999471	0.959492
ITS	0.893573	0.050596

Gene	Residues score	Column score
185	0.987769	0.001852
26S	0.862116	0.000057
atpB	0.995953	0.350238
matK	0.997872	0.075541
ITS	0.492671	0.00093

Markov chain

State of the chain = whole sequence



Simulating Markov chains using Gillespie's algorithm



Given:

GGTTGATCCTGCCAGTAGTCATCT

sequence

branch length (t)

Denote:
$$\lambda = I + D + S$$

Insertion rate

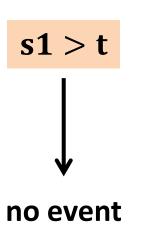
Deletion rate

Substitution rate

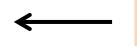
Waiting time:

s1

 $s \sim \exp(\lambda)$, mean=1/ λ



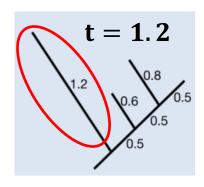




s1 < t

$$\frac{I}{\lambda'}$$
 $\frac{D}{\lambda'}$ $\frac{S}{\lambda}$

draw a location



 $\lambda = I + D + S$





uniformly proportional to (L+1 positions) the substitution rate at each site



insert, delete or substitute

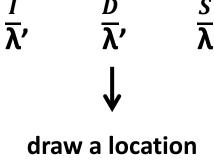
update sequence length L, time = t-s1

draw an event with probabilities $\frac{I}{D} = \frac{S}{D}$



s1 < t

repeated until s1 + s2 + ... > t







uniformly

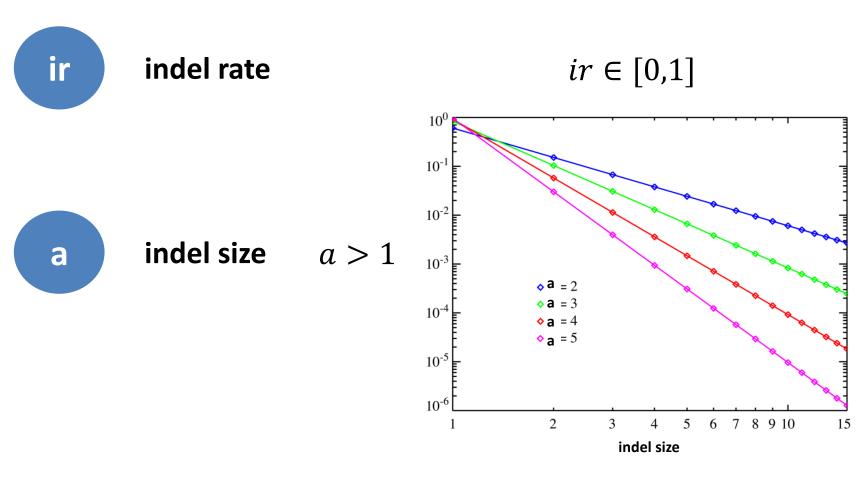
proportional to the substitution rate at each site



insert, delete or substitute

update sequence length L, time = t-s1

Insertion-deletion model



Zeta distribution

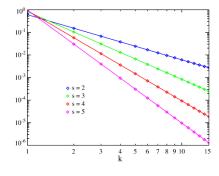
Insertion-deletion model

ir indel rate

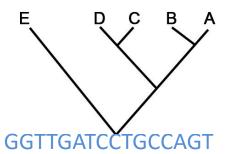
$$ir \in [0,1]$$

a indel size

a > 1



rl root length



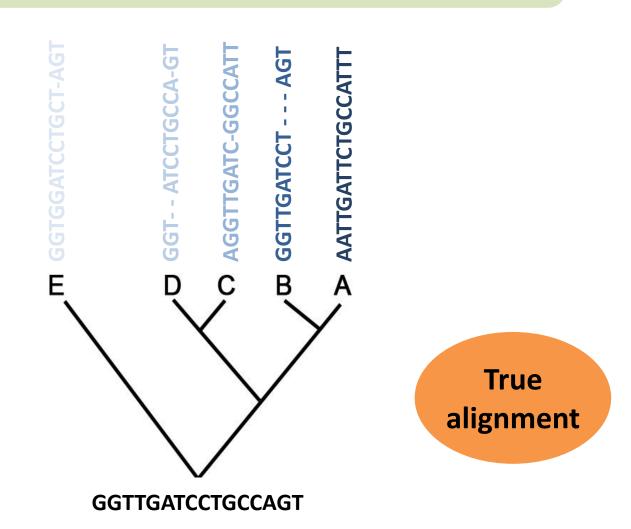
Insertion-deletion model

ir indel rate

a indel size

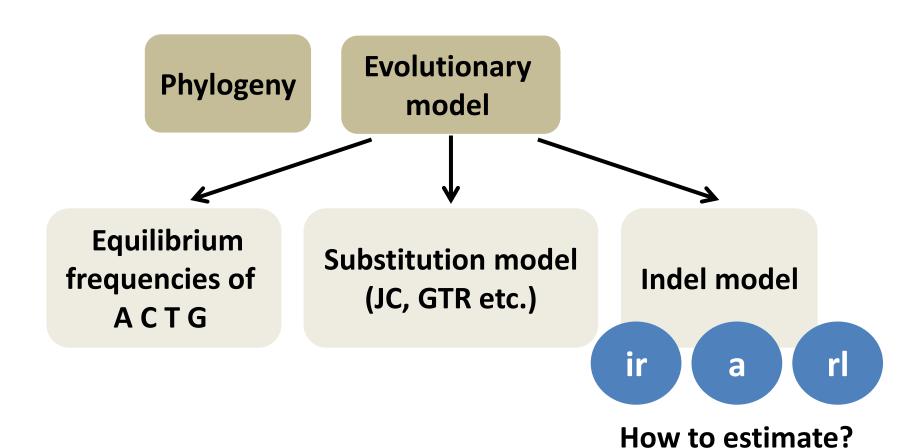
rl root length



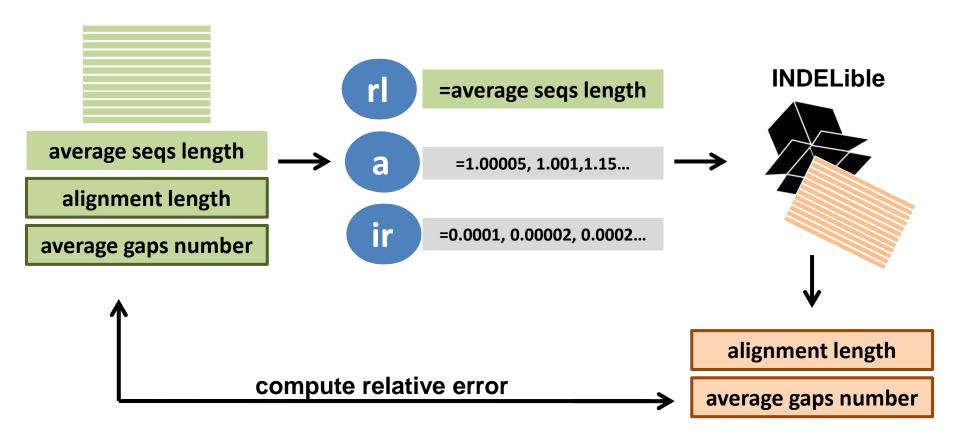


Gillespie algorithm

INDELible – estimating parameters



INDELible – estimating parameters



Choose the parameters a, ir which gave the lowest error

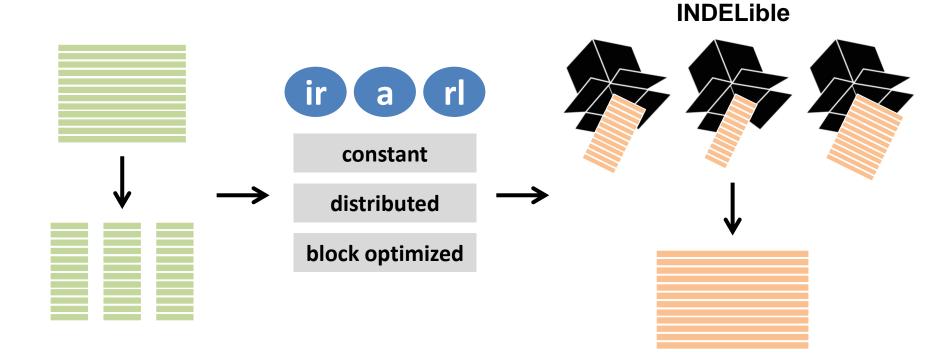
Getting more realistic simulations

Real sequences

Simulated sequences

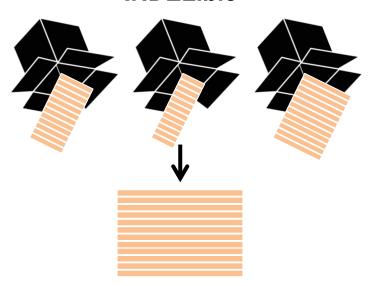
13351		CGAA CTCGTGACCCCCTTTTTTTTTGGGGGTAGGGGCGAAGGG	CIGCICITCAGE
1424	CCAATCCTG	ATCGACACAC	-CGTCCCCCGACC
105750	CGATGACC	CACTAGGCTGAC CGGCGAA CTTGTGATATTGTTCCTTGAGGGATGACGGTCTATCCT	TGCGATACCGAA
6752	TGATGCGTA	CCAAAACAAGATCGACCGAAGCGAA CATGTGACCCCTTGTTCTTTGCTTGCGTGCGAGGATTG	CCGTCTCGG
12953 12294	TCGAGGACC	GAACGAGAGGGTAGAC CCGCGAA CATAGTGACCCGCGCAAATCGGGAGGGGGGGGGG	-GGCTGCCCCGT
2254	TGGAGAACC	MACCCCTGCAAAGCAGCACGACTG GTGTGAA TATGTCAATTGTCATCTGGGGGTGGGGACGCTAGCGCC	TAGCGACGCCCC
12846 1429 3083 1426	TIGAGAACT	GAAAATGCAGAGCACGACTA GTGTGAA CTTGTGAATTATCATTTGGGGGGGGACACTAGTGCC	TAGTGACGCCCC
429	TT	CGACCCGCGAACCGACATCAGTTTGACCTGCCTCTG CCCCCGCTGTGCCCCCGGGTGCGGTCGGGTCGGGTGGG	AGAC
3083	PCGCAAACC	AGAGCCAC CGGTGAACCCGTTCTGTCACTGAAGTTTGCGGGTTGTGCATCTGAGGCCTC	GAACCAGCCTGA
426	TCGAGGCCC	CTCCCTGCTAAAAGCAGAACGAC CCGCGAA GGAGTGAAAGCATTGCTCTCGGTGGCGGAACGACTCTG	TCACCSCCCCAS
3419	TOSTIGOCT	CGACCGGCSAACCGATATCAGTCTGACCTGCCTCTGCCCCCCGGCGGTGCCTCGCGGTGCGGTCGGGTGGG	AGAC
3419 3260	TGATGCCTA	CCARANACAAGATCGACCGAAGCGAA CATGTGAAACCTTGCGTTCTCCGCCCGGGGGGGGGG	-CCGTGCCCC
3559	TOGATACOT	ATAAAAGCGAGACCAGCGAATTGTTATCTGTTTTTCTTTGTTGCATCGTATCGAGAGATATA	TACGGTGCCTCC
3891	TOGARGOOT	GCARCGGCAGAACGAC CCSCGAA CFCGTCACCGATCGCACCCSCGGGGGGGGGCGACCGTCCGG	-caccccsscsrc
3559 3891 708 711 3166	TEGRARACCE	GCC CAGCAGAACGAC COSTGAA CCAG AGATATCACOGGCGGCGGGGGGGGGGGGGGGGATGCGTC	-CGCAGCGGG
711	TOGARACOT	GCC CAGCAGAACGAC CCGCGAA CCAGTTGATATCACCGGCGGCGGGGGGGGGG	CGCAGDGGG
3166	TOGACCTGC	CAGCAGA CGAC COGCGAA CCAGTYGATA7CACOGGCGGCGGGGGGGGGGGGGGGCGCC	CGCAGCGGG
	regalacer	GCCTAGCAGAACGAC CCGCGAA TACGTGAACGTGCTGAGGGGGGACGACGGCGCTCGCG	CACCCPPTGCPC
5280 3654 760 486 483 460 813	TOGARACOT	GCC TAGCAGAACGAC CTGTGAA TTTGTGAATGTTTTGGGGGGGGGG	AGCTGTCGT
3654	TOGARACOT	GCCTAGCAGAATGAC CAGCGAA CTTGGTCGGATACGGATACGGCCATCGGCCTTTCGTAG	TIGGCCCIGTER
760	TOGARACOT	GCC TAGCAGAACGAC CCGGAAC TAGCGGGGGGGGGGG	CTCCTTGTC
49.6	TOGATGCCT	GCARCAGCAGAACGAC CCGGAA CACGTACCTTGGGTGGGCGAGAGGAGCTTGCTCCTTGG	ACCCGCCCTCAC
493	TOGARACOT	SCARCAGCAGACGAC COSGAA CACGTG CTTGGGCGGGGGGGGGGGCTTGCTCCTTGG	ACCCGCCCGCAC
460	TOGATGCCT	GCATAGCCCCGCGCGCGTGTTTTCAATTCTACGGGAGCCTCCCGCGGGAGAGC	TGTCCCCTCGCG
813	ccgc	AAAAAGAGCGAC CCGAGAA CCGGTTGGAACAAAACTTTGGCGGGGGAGGCCGAGGCC	CCACCCCCG
879		ACATGCAGTCCAACAGGAATCAACACCTCGGCTTA CCTCTTGGTTCAGAGGAAGACGACGAAGTGC	
3479	POGRAFICOR	aca gagogaacgac cogogaa crogograficacgagggggggggggggggggggggggggggggggg	- COGCCCCGTC

50993	GCCCTTCACCGACCGCCAGGCCCACCCC	CZAG CACGCAZA	DOCCOGGACANCOGNGNGNGNGT	CGAGTC	GGGG - CGCACGGATC
4400	GOTOTOCACOPRIOSOSCAGRACOTOTOC	CTAG CACGCATA		003600	GTGG - ASCRAGGATC
50174	GOCTTCTACTGTCCGCGCAACGCCGCTCC	CTAG SACGCATA	- CGCCCGCACACGCGAGAGAGC	CCAGGA	GGGGCGCAAGGA7C
15172	GCCCTCCACCGTCCGCGCAGGGCTACCAC	CTAG CACGCATA	CCCCGGCACCATCGAGAGAAT	CTAGTT	GOOG - ASCAAGGTTC
3592	GOCTTOTACOSCOSOSCAGSSCCAACTO	CTAG CACGCATA	OGCCCGTACASTCGAGAGAGAGC	CTAGOS	GCGG GSCATGGGTC
3597	GCCTTCTACCGACGGCGCAAGGCCACCCC	ATAG TACGCATA	- CCCCCGGCACTCCCGAGAGAGAGC	CTAGCT	GCGG - ASCAGGGTTC
1283	GOZOTODODOGOZOGOGOANGGOCA COCO	CTAG CACGCATA	- CGCCCGCACAATCGAAACAGT	CAAGIG	GCGG GGCAAGGA9C
292	GOZOTOCACOSGIOSCIAGOGICACCOC	CTAG CADGCATA	COCCOGGACCACCGAGACAGC	CGAGTC	TCAG - GSCAAGGCTC
6924	GOZOTODGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	CZAG CADGCAZA	DODDOOGRACAATOGRACAGO	CSAGTG	GCGG GSCATGGA7C
741	GOZATOTACOSACOSOSCAGAGSCACCCC	CTAG CACGCATA	DODDOGGACACTOGAGACAGT	CGAGTG	GGGG - CGCACGGATC
1442	GOZATONACOSACOSOSCAGAGSCATCOC-	CTAG CACGCATA		CSAGTG	GTGG - CSCRAGGRTC
25047	GOZATOVACOGOZOGOGOGGGGGGGCACCCC	CTAG CACGCATA	- COCCOGIACAATOGAGACACC	CHAGTG	GTGG - GSCNAGGATC
322403	GOZOTOWACOSACOSOSCATIGGGTACGOD	CTAG CATGCATA		CALGEC	GCGG GSCAAGGATC
772	GOZOTOTGOOSACOSOSCAGGGCCACCCC	CTAG CACGCATA	CTCCTGCACAGTCGAGGCAGC	CAAGTA	GAGG TECHASGATO
288	GOZOTOCACOGACOGOGTAGTGCCACCCC	CTAG CACGCATA		CAAGOG	GOGG TGCLGCGL7C
6901	GOZOTOCACOGOZOGOTAGTGCCTCCCC	CTAG CACGCATA	- COCCOGGAGAGOGAGAGAGA	CAAGGG	GOGG TGCAGOGATC
1146	GOZOTOWACOSCIOSOSCATGTCCACCAC	CTAG CACGCATA		CGAGTG	GAGG - CSCAACGGTC
3770	GOZCECGACOGOZOZOGCAGGGCCACCCC	CTAG TACGCATA		CCAGTC	GGGG - CCCACCGGTC
9347	GOZCTOSACTGTCCGCGCAGGTCCACCCC	CTAG CATGCATA	DGCCCGCACATTCGACCCAGG	CGAGTG	GTGA - AGCAGCGGTC
143381	GOZOTOGRACOGOZOGOGGGGTCCATCCC	CTAG CACGCATA	- CGCCCGCACAACCGACACAGG	CGAGTG	GAGA - CSCAACGGTC
9388	GCCCTCSACOSCCCGCGGGTCCACCCC	CTAG CACGCATA	CGCCCGCACAGGCGACACAGA	CGAGTG	GTGA - CSCAADGGTC
1151	GCCCTCGCCGGGGCCGCAAGTCCACCCC	CTAG AACGCATA	ATCCCGTATAAACGATACAGG	CCAGTG	GAGA GSCAACGGTC
197815	GOZOTOSGOOSACOSOSGAGSTOSGOCOC	CTAG CACGCATA	TGCCCGTATGGCCGAGACAGG	CAAGCT	GGGA - CGCAACGA7C
01512	GOZOTOGACOGOZOGOGOAGGGCCACCCC	CTAG CADGCATA	CPCCCGTACATACGAGACAGC	CGAGTC	GTGG - CSCAACGTTC
19353	GOZOTOGACOSCZOSCSCAGGTCCACACC	CTAG CACGCATA	- COCCOBEAGAACOGAGACAGC	CCAGTG	GAGG CECAACGCTC
18536	GODO TOGRADOS TOGROS CAGGACCACCCC	CZAG CACGCAZA		CCAGTG	GOGA TECHGOGOTO
49121	GOZOTOSACOSACOSOSCAGRACOTOCOC	CTAG CACGCATA	- COCCOCRIACACCOCRAGACACC	CGAGTG	9999 - OSCAGO9970



Getting more realistic simulations

INDELible



മ	
	Ш
္ပ	Ü
	Ž
S	
\subseteq	⋖
\subseteq	
=	=
2	
0	U
ŭ	

constant	distributed	block optimized
0.07004606	0.07046526	0.07672056
0.97901696 0.91552168	0.97916526 0.94023076	0.97672856 0.91669352
0.9883448	0.98878618	0.98792382
0.99055344	0.99287678	0.98774598
0.94787124	0.9599763	0.9362862
0.76490788	0.79192566	0.74172564
0.51005486	0.52628762	0.60550184

Changing route?

