

## **SUPPLEMENTAL MATERIAL**

### **TABLES**

**[Supplemental Table 1.](#) Primers used for DNA sequencing, PCR, and RT-PCR**

**[Supplemental Table 2.](#) Protein and taxon sampling used to generate HMMs**

**[Supplemental Table 3.](#) Sequence comparison of A-class and B-class chromosomes**

**[Supplemental Table 4.](#) Codon usage of *D. papillatum* mitochondrial genes**

### **FIGURE LEGENDS**

### **REFERENCES**

### **FIGURES**

**[Supplemental Figure 1.](#) Sequence repeats within chromosomes A3207 (A) and B3209 (B)**

**[Supplemental Figure 2.](#) GC-skew of chromosomes A3207 (A) and B3209 (B)**

**[Supplemental Figure 3.](#) Multiple protein alignment of Nad8**

**Supplemental Table 1. Primers used for DNA sequencing, PCR, and RT-PCR**

Name	Sequence	Used for
dp1	GCTGTTGATGCTCAGCAGAGCGGT	Primer walking, chromosome A and B
dp2	GCACTCGCTGTGGAGTGCGTGTAT	Primer walking, chromosome A and B
dp3	GATAGCATACCATCTACCGTCTACA	Primer walking, chromosome A
dp4	GCTCTGCTGCTCCGCTGCACAGTG	Primer walking, chromosome A
dp5	CCTCTACACGTAGTAGATGCATATC	Primer walking, chromosome A
dp6	GTAGTGCTACGAGGAGCGGTAGAT	Primer walking, chromosome A
dp7	GCTGTAGAGCTACGTGCTGCTACAT	Primer walking, chromosome B
dp8	CGTGCATCCATGCACAGCCGGTC	Primer walking, chromosome B
dp11	GGTAGATGGACGGTATGCCTA	Primer walking, chromosome B
dp15	GCATCTACATCCTACCCTCTACGT	Primer walking, chromosome B
dp17	CCGCTCTGCTGAGCATCAACAGC	Primer walking, chromosome A and B
dp18	ATACACGCACTCCACAGCGAGTGC	Primer walking, chromosome A
dp19	GCTACATGCTATCCTCTCCCGTAG	Primer walking, chromosome B
dp20	GCACCACAGGATGCTAGTACATGTACAGC	Primer walking, chromosome B
dp21	GCGGTGTATTGGTACTAGCAGTACGGCGT	Primer walking, chromosome B
dp22	GGGATATGCACAGGGAATGCA	Primer walking, chromosome A and B
dp40	CGTAGCAGCAGCACGTAGTGTATGCA	Primer walking, chromosome A
dp44	GGGTACAGATCCACTTGGGA	Primer walking, chromosome A
dp45	GCGTAGTAAGGTAATGGTAT	Primer walking, chromosome A
dp63	GCTAGTACCACCTCGTAGCATG	Primer walking, chromosome B
dp64	GGCATCAACGCACCTGTGCAT	Primer walking, chromosome A
dp26	GGTCTTCGGGCACCCAGAG	PCR of <i>cox1</i>
dp27	TACCCTACGTACACGTGCAC	PCR of <i>cox1</i>
dp38	GACTACCAGTATACCACAGG	PCR of <i>cox1</i>
dp39	AGCATCCATGCATCTGGAGG	PCR of <i>cox1</i>
dp47	CATGATGGGTGTTGCCTTCC	PCR of <i>cob</i>
dp50	GCTGTAACCCATCCACTAACGT	PCR of <i>cob</i>
dp59	CACGTTGTGGCCAACCAAGTG	PCR of <i>cox2</i>
dp62	CCGTACAGGGTGCCAGTAACG	PCR of <i>cox2</i>
dp76	GCCATCCACTAGCAGCCATGCTAC	RT-PCR of <i>cox1</i>
dp87	AAGAGCATGAGGATGCCATGC	RT-PCR of <i>cox1</i>
dp90	GCTGATGTGCAGCTCTCCCTTTGG A	PCR of <i>nad7</i>
dp91	GAGTGCCTCCTGCAGACCTAGGA	PCR of <i>nad7</i>
dp93	CTGGTGCTAGCTGAGACCTTAGG	PCR of <i>nad7</i>
dp94	CGCTCCGGAGTAGGACACCAGATG	PCR and RT-PCR of <i>nad7</i>
dp98	CCTGTACAGAGCTACTGCAGC	PCR of <i>nad7</i>
dp100	CCAGTGTAGTGCTCCGGAT	RT-PCR of <i>nad7</i>

**Supplemental Table 2. Protein and taxon sampling used to generate HMMs**

Proteins	<i>atp1, atp6, atp8, atp9, cob, cox1-3, nad1-4, nad4L, nad5-11, rps12, sdh1, sdh2</i>
Taxa	<i>Adriamonas vulgaris, Allomyces macrogynus, Amastigamonas bermudensis, Amastigamonas sp., Andalucia godoyi, Bigelowiella natans, Capsaspora owcarzaki, Paracercomonas marina, Chaetosphaeridium globosum, Chara vulgaris, Chondrus crispus, Chrysodidymus synuroideus, Compsopogon caeruleus, Cyanidioschyzon merolae, Cyanophora paradoxa, Desmarestia viridis, Emiliana huxleyi, Geodia neptuni, Glaucocystis nostochinearum, Guillardia theta, Histiona ariudes, Jakoba bahamensis, Jakoba libera, Klebsormidium flaccidum, Malawimonas californiana, Malawi jakobiformis, Marchantia polymorpha, Mesostigma viride, Monosiga brevicollis, Naegleria gruberi, Nephroselmis ovilacea, Nuclearia sp., Ochromonas danica, Ostreococcus lucimarinus, Pavlova lutheri, Physcomitrella patens, Phytophthora infestans, Porphyra purpurea, Prototheca, Reclinomonas americana NZ, Rhodomonas, Rhizopus, Saprolegnia, Seculamonas ecuadoriensis, Thalassiosira pseudonana</i>

**Supplemental Table 3. Sequence comparison of A-class and B-class chromosomes**

Chromosomes (size)	Gene module contained in chromosomes	Length of		substitutions	Number of	
		cassette	constant region (shared length)		indels (length)	overall nt differences
A3207 (5,852 bp) vs A3208 (5,802 bp)	<i>cox1</i> -m9	311 bp	5,541 bp	63	5 (1-48 bp)	2.0 %
A3207 (5,852 bp) vs A4001 (5,794 bp)	<i>cox1</i> -m9 <i>nad7</i> -m6	311 bp 295 bp	5,541 bp 5,794 bp (100%)	71 <sup>a</sup>	28 (1-48 bp) <sup>a</sup>	3.3 % <sup>a</sup>
B3209-population (7,182 bp) <sup>b</sup> vs A3207	<i>cox1</i> -m4 <i>cox1</i> -m9	310 bp /	6,872 bp (41%; 2.57 kbp)	54 44 <sup>c</sup>	2 (1-50 bp) 13 (1-12 bp) <sup>c</sup>	1.8 % 3.2 % <sup>c</sup>
B3209-population	<i>cox1</i> -m4					

<sup>a</sup> The cassette regions were not included in the comparison

<sup>b</sup> Average length. The population was sequenced collectively, see Methods.

<sup>c</sup> Numbers refer to shared regions only.

**Supplemental Table 4.** Codon frequency of *D. papillatum* mitochondrial genes<sup>a</sup>

F	TTT	5.3	S	TCT	5.8	Y	TAT	16.3	C	TGT	37.5
F	TTC	94.7	S	TCC	27.7	Y	TAC	83.8	C	TGC	62.5
L	TTA	2.3	S	TCA	9.5	*	TAA	66.7	W	TGA	31.1
L	TTG	4.9	S	TCG	3.6	*	TAG	33.3	W	TGG	68.9
L	CTT	2.3	P	CCT	32.7	H	CAT	38.8	R	CGT	26.0
L	CTC	24.9	P	CCC	16.3	H	CAC	61.3	R	CGC	10.0
L	CTA	30.2	P	CCA	34.7	Q	CAA	2.8	R	CGA	4.0
L	CTG	35.5	P	CCG	16.3	Q	CAG	97.2	R	CGG	20.0
I	ATT	0.0	T	ACT	15.9	N	AAT	10.7	S	AGT	11.7
I	ATC	50.9	T	ACC	31.9	N	AAC	89.3	S	AGC	41.6
I	ATA	49.1	T	ACA	34.5	K	AAA	8.3	R	AGA	8.0
M	ATG	100	T	ACG	17.7	K	AAG	91.7	R	AGG	32.0
V	GTT	4.3	A	GCT	30.7	D	GAT	30.8	G	GGT	37.8
V	GTC	25.5	A	GCC	21.3	D	GAC	69.2	G	GGC	21.0
V	GTA	26.8	A	GCA	38.0	E	GAA	11.4	G	GGA	25.9
V	GTG	43.3	A	GCG	10.0	E	GAG	88.6	G	GGG	15.4

<sup>a</sup>Codon frequencies in % from *cob*, *cox1*, *cox2*, *cox3*, *nad7*. Shaded codons are underrepresented.

## FIGURE LEGENDS

**Supplemental Figure 1.** Sequence repeats within chromosomes A3207 (A) and B3209 (B). Positive diagonals indicate direct repeats. The dotplot was generated with Dotter v3.1 (<http://www.cgb.ki.se/cgb/groups/sonnhammer/Dotter.html>) (1), using default parameters.

**Supplemental Figure 2.** GC-skew of chromosomes A3207 (A) and B3209 (B). The cumulative value of (G-C)/(G+C) is plotted along the sequence. The graph was generated using the University of Pittsburgh bioinformatics software and web tools collection at the URL <http://bioinformatics2.pitt.edu/index.html>. Window size was 300. The origin of replication is generally associated with the global minimum of the cumulative GC skew (2).

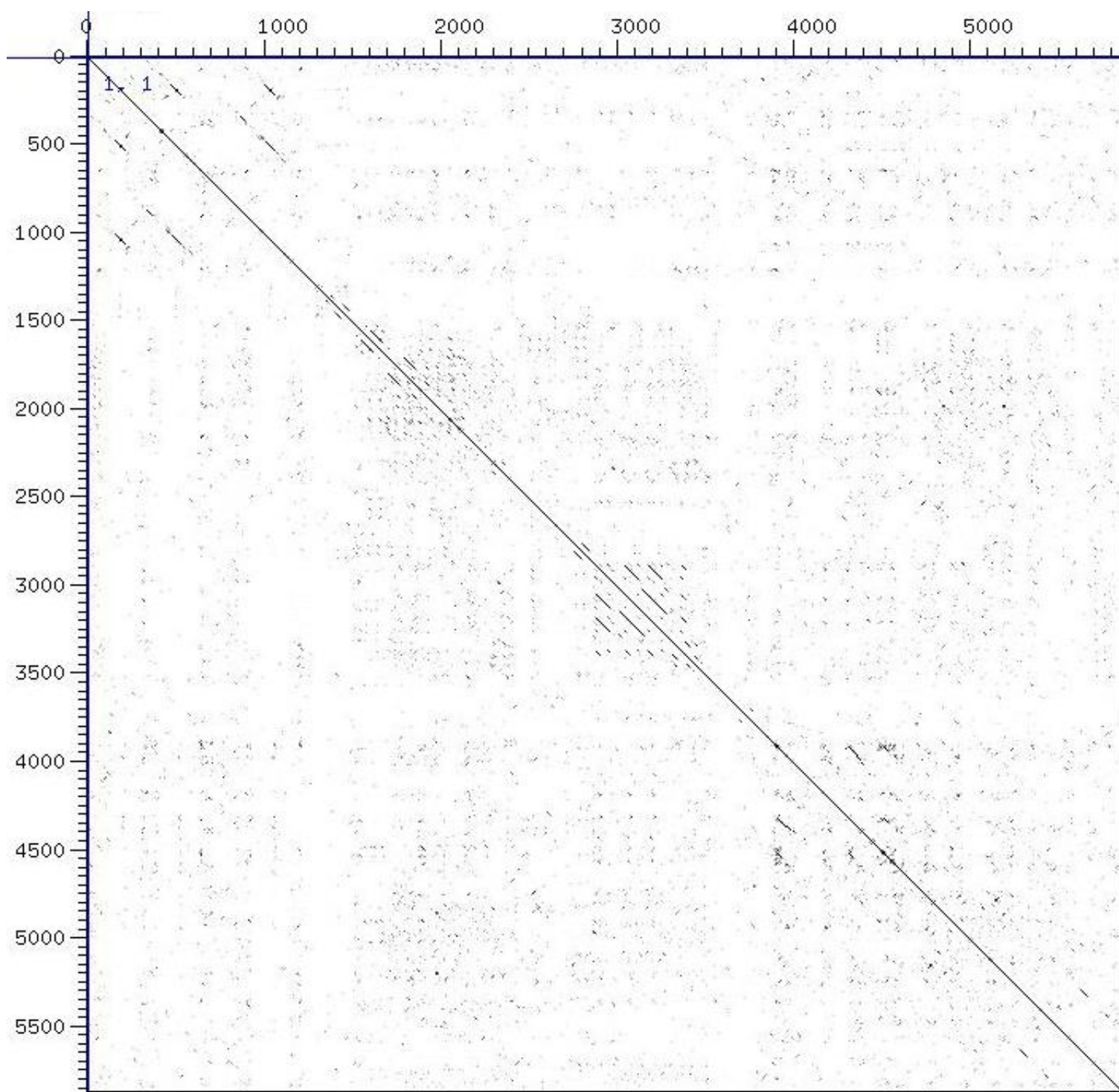
**Supplemental Figure 3.** Multiple protein alignment of Nad8. The protein regions best conserved across reference proteins are boxed. The predicted partial sequence from *D. papillatum* (labeled *Dipl.papi.*) has been constructed from the conceptual translation of two genomic cassettes, cassetteA67 and cassetteB44. Modules could not be precisely mapped within the cassettes, because the corresponding cDNA sequence is missing from our dataset. A third - the terminal - module indicated by x remains to be identified. Lowercase letters, protein sequence likely not corresponding to modules, but rather to module-flanking regions. Mitochondrion-encoded Nad8 protein sequences and GenBank acc. nos. (protein) are: *Rhod.sali.*, *Rhodomonas salina* (10444172); *Recli.amer*, *Reclinomonas americana* (11466496); *Tryp.bruc.*, *Trypanosoma brucei* (552291).

## REFERENCES

1. Sonnhammer, E.L. and Durbin, R. (1995) A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene*, **167**, GC1-10.
2. Lobry, J.R. (1996) Asymmetric substitution patterns in the two DNA strands of bacteria. *Mol Biol Evol*, **13**, 660-665.

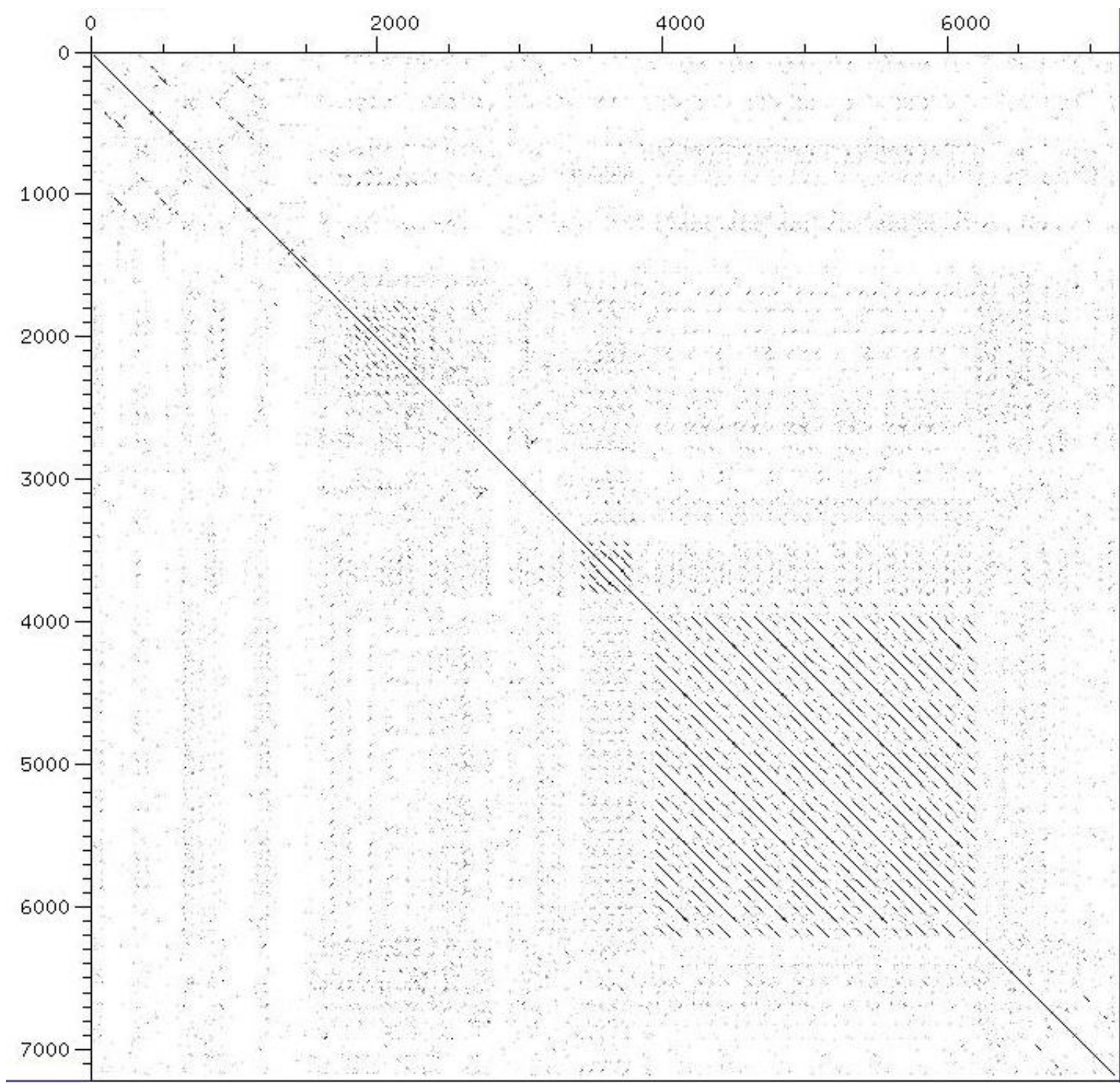
# Supplemental Figure 1A

A3207 vs A3207



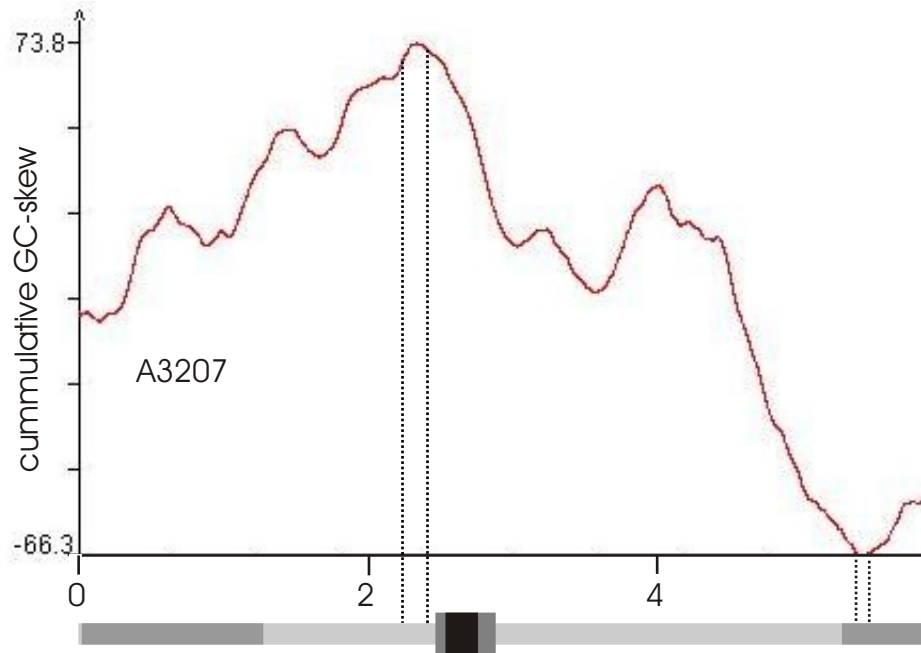
# Supplemental Figure 1B

B3209 vs B3209

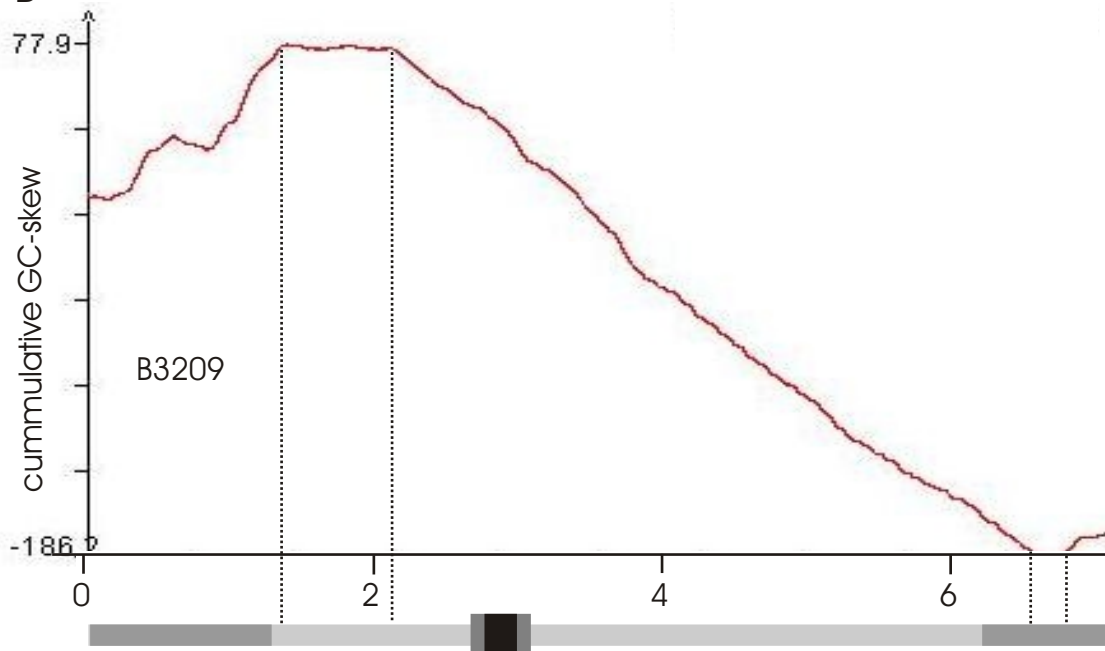


# Supplemental Figure 2

**A**



**B**





# Supplemental Figure 3

	1	11	21	31	41	51	60
Rhod.sali.							
	MSYFNFNVRILFLSELFRGLWMVLQYFFRRKVTLNYPFEKAVLSPRFR					RGEHALRRYFSG	
Recl.amer.	MTIINKTAQTLFLTELVKGMSLTLDYFFRKKVTLNYPFEKGPLSPRFR					RGEHALRRYQTG	
Naeg.grub.	MFSSFQYLYLSELMRGLFVALKNLFSKKVTINYPFEKGAI SPRFR					RGEHALRRYSNG	
Tryp.bruc.	MF	FFDFL	FFFFVCFY	MC	FVCCVTICLP	IELTIVSLLV	RGNHFLRFYWCG
Dipl.papi.	VQYSWWCTGLYMEYVVTLSCVLVYHASRSVLLTE						-VPASHWHRGEHILACHADG
cassetteA67	rssccvVQYSWWCTGLYMEYVVTLSCVLVYHASRSVLLTE						-VPASHWHRGEHILACHADG
cassetteB44							qhptsyqqlvpprsmwqyawcctachvv

	61	71	81	91	101	111	120
Rhod.sali.							
	EERICIACKLCEAVCPAQAITIE		ETESRADN	SRKTSRYD	IDMTKCIFCGLCQE	EACP	VDAIVE
Recl.amer.	EERICIACKLCEAICPAQAITIE		ESEPRIDG	SRRTTRYD	IDMTKCIYCGFCQE	EACP	VDAIVE
Naeg.grub.	EERICIACKLCEVVPALAITID		SAQQLN	SRQTTRYD	IDMTKCIYCGLCQE	EACP	VDAIVE
Tryp.bruc.	LERCIACRLCDLICPSLALDVRVGSF		GGHRFADWF	TLSYRRCI	YCGFCMHVCPTDAITH		
Dipl.papi.	DPRCIACRLCSVACPAYAITVLAGIA		STGSRTPAEYV	LCTTRCI	YCGWCDAVCP	THATIH	
cassetteA67	DPRCIAaehhl*						
cassetteB44	rpsgstCRLCSVACPAYAITVLAGIA		STGSRTPAEYV	LCTTRCI	YCGWCDAVCP	THATIH	

	121	131	141	151	163
Rhod.sali.					
	GPNYEYSVFKRQNLINKEKLLKNGDQWELEIRENRLSESMYR*				
Recl.amer.	GPNFEFATETHEELLYDKEKLLQNGDRWETEIAANLANEALYR*				
Naeg.grub.	SFNFEYSTFTHDELLYDKKLLLANGDKFEIEIYLNKKEILYR*				
Tryp.bruc.	SLFVMCFCLAMYLL-APKFLFLFGCCFMLDFYLCFV*				
Dipl.papi.	SVVLYCTGSTASTTLYxxxxxxxxxxxxxxxxxxxxxxxxxxxx*				
cassetteB44	SVVLYCTGSTASTTLYlptraryti*				