

This supplement accompanies

Statistically-Consistent k-mer Methods for Phylogenetic Tree Reconstruction

by E.S. Allman, J.A. Rhodes, and S. Sullivant.

November 2015

Elizabeth S. Allman
Department of Mathematics and Statistics
University of Alaska Fairbanks
e.allman@alaska.edu

John A. Rhodes
Department of Mathematics and Statistics
University of Alaska Fairbanks
j.rhodes@alaska.edu

Seth Sullivant
Department of Mathematics
North Carolina State University, Raleigh, NC, 27695
smsulli2@ncsu.edu

Contents

1	Simulation results using d_{JC} + Neighbor Joining, 1000 bp	3
1.1	JC with no indel process.	3
1.2	JC with mutation rate $\mu = 0.01$ for the indel process.	3
1.3	JC with mutation rate $\mu = 0.05$ for the indel process.	4
1.4	JC with mutation rate $\mu = 0.10$ for the indel process.	4
2	Simulation results using $d_{JC}^{k,B}$ + Neighbor Joining, 1000 bp	5
2.1	JC with no indel process.	5
2.2	JC with indel rate $\mu = 0.01$. Lavalette parameters $a = 1.1, M = 100$	7
2.3	JC with indel rate $\mu = 0.01$. Lavalette parameters $a = 1.5, M = 100$	8
2.4	JC with indel rate $\mu = 0.01$. Lavalette parameters $a = 1.8, M = 100$	9
2.5	JC with indel rate $\mu = 0.05$. Lavalette parameters $a = 1.5, M = 100$	10
2.6	JC with indel rate $\mu = 0.05$. Lavalette parameters $a = 1.8, M = 100$	11
2.7	JC with indel rate $\mu = 0.10$. Lavalette parameters $a = 1.5, M = 100$	12
2.8	JC with indel rate $\mu = 0.10$. Lavalette parameters $a = 1.8, M = 100$	13
3	Simulation results using other k-mer distances + Neighbor Joining/UPGMA, 1000 bp	14
3.1	JC with no indel process.	14
3.2	JC with indel rate $\mu = 0.01$. Lavalette parameters $a = 1.1, M = 100$	16
3.3	JC with indel rate $\mu = 0.01$. Lavalette parameters $a = 1.5, M = 100$	18
3.4	JC with indel rate $\mu = 0.01$. Lavalette parameters $a = 1.8, M = 100$	20
3.5	JC with indel rate $\mu = 0.05$. Lavalette parameters $a = 1.5, M = 100$	22
3.6	JC with indel rate $\mu = 0.05$. Lavalette parameters $a = 1.8, M = 100$	24
3.7	JC with indel rate $\mu = 0.10$. Lavalette parameters $a = 1.5, M = 100$	26
3.8	JC with indel rate $\mu = 0.10$. Lavalette parameters $a = 1.8, M = 100$	28
4	Simulation results using $d_{JC}^{k,B}$ + Neighbor Joining, 10,000 bp	30
4.1	JC with no indel process.	30
4.2	JC with indel rate $\mu = 0.01$. Lavalette parameters $a = 1.5, M = 100$	32
4.3	JC with indel rate $\mu = 0.05$. Lavalette parameters $a = 1.8, M = 100$	34

November 2015

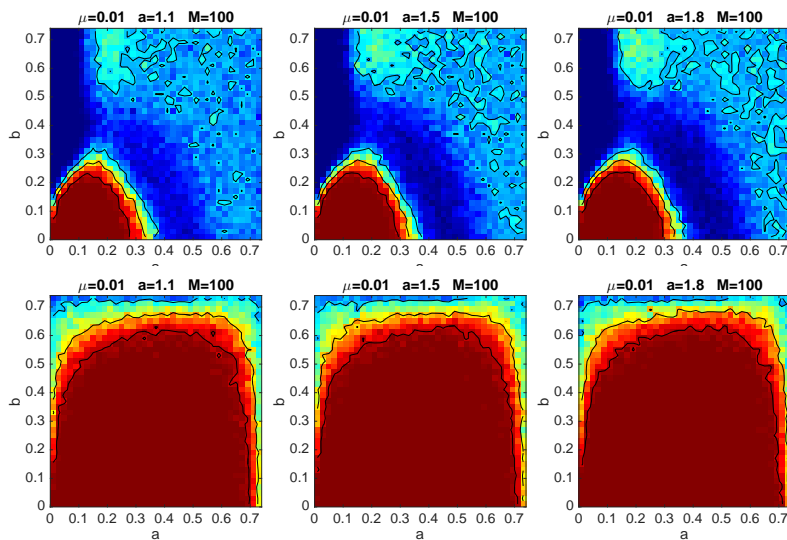
1 Simulation results using d_{JC} + Neighbor Joining, 1000 bp

Each pair of rows corresponds to 100 datasets simulated with a fixed rate of insertions/deletions. Columns correspond to varying Lavalette parameters. In each pair of rows, the top row shows alignment+NJ, and the bottom row shows true alignment+NJ. Sequences have length 1000 bp.

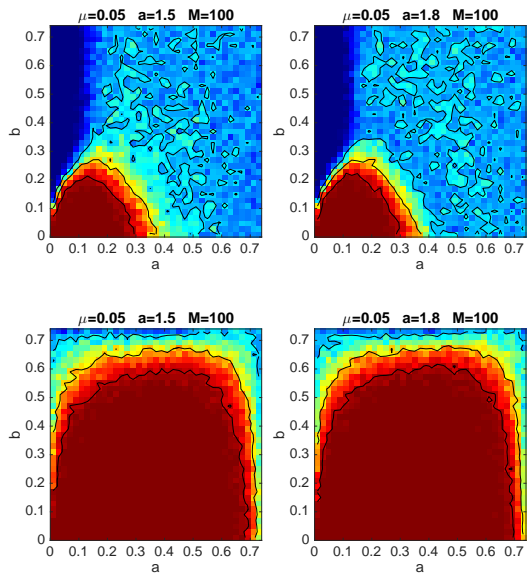
1.1 JC with no indel process.



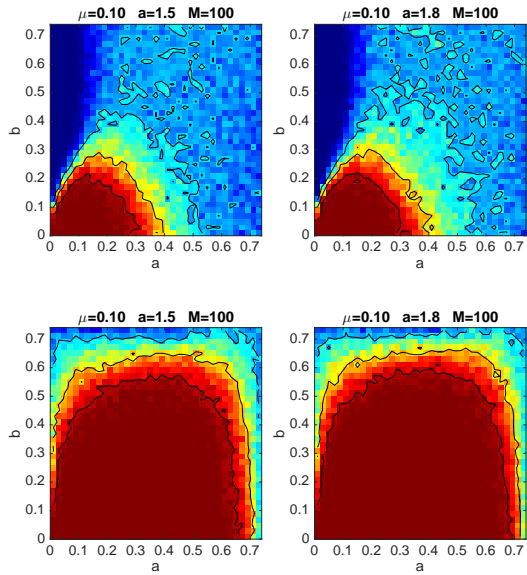
1.2 JC with mutation rate $\mu = 0.01$ for the indel process.



1.3 JC with mutation rate $\mu = 0.05$ for the indel process.



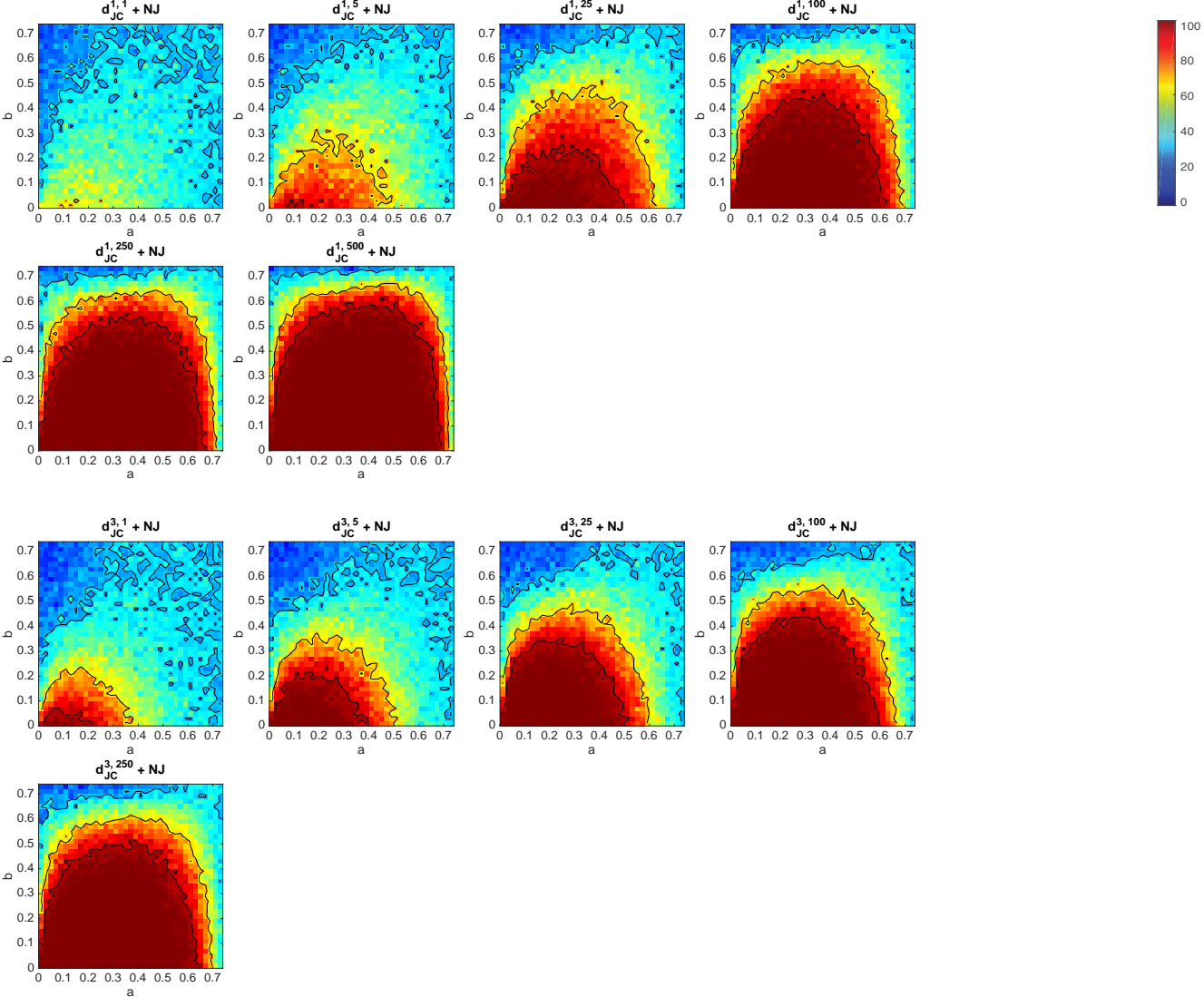
1.4 JC with mutation rate $\mu = 0.10$ for the indel process.



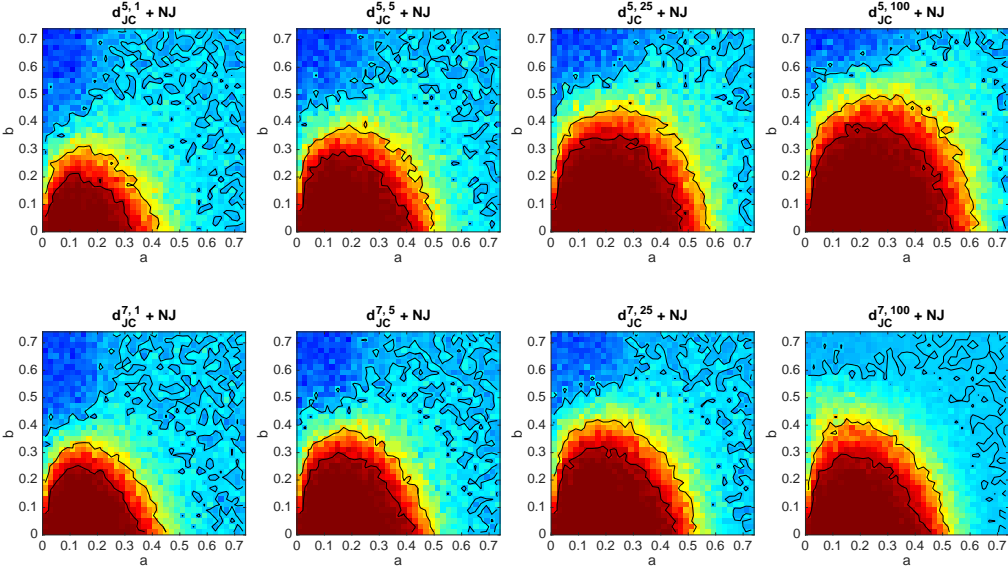
2 Simulation results using $d_{JC}^{k,B} + \text{Neighbor Joining}$, 1000 bp

Each subgroup of figures corresponds to a fixed setting for the model parameters used to generate sequences. Titles indicate the value of k and the number of blocks used for computing $d_{JC}^{k,B}$. Sequences have length 1000 bp.

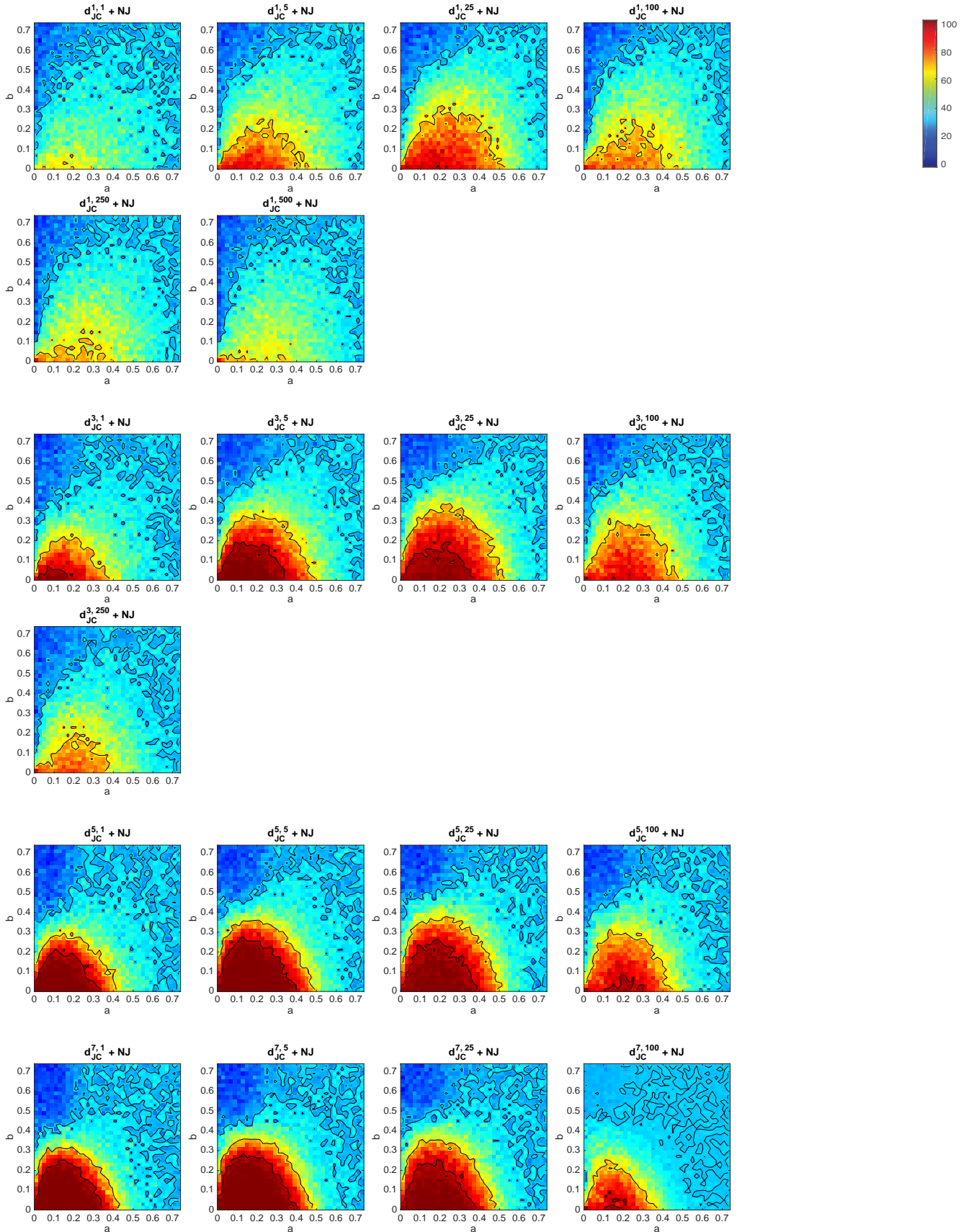
2.1 JC with no indel process.



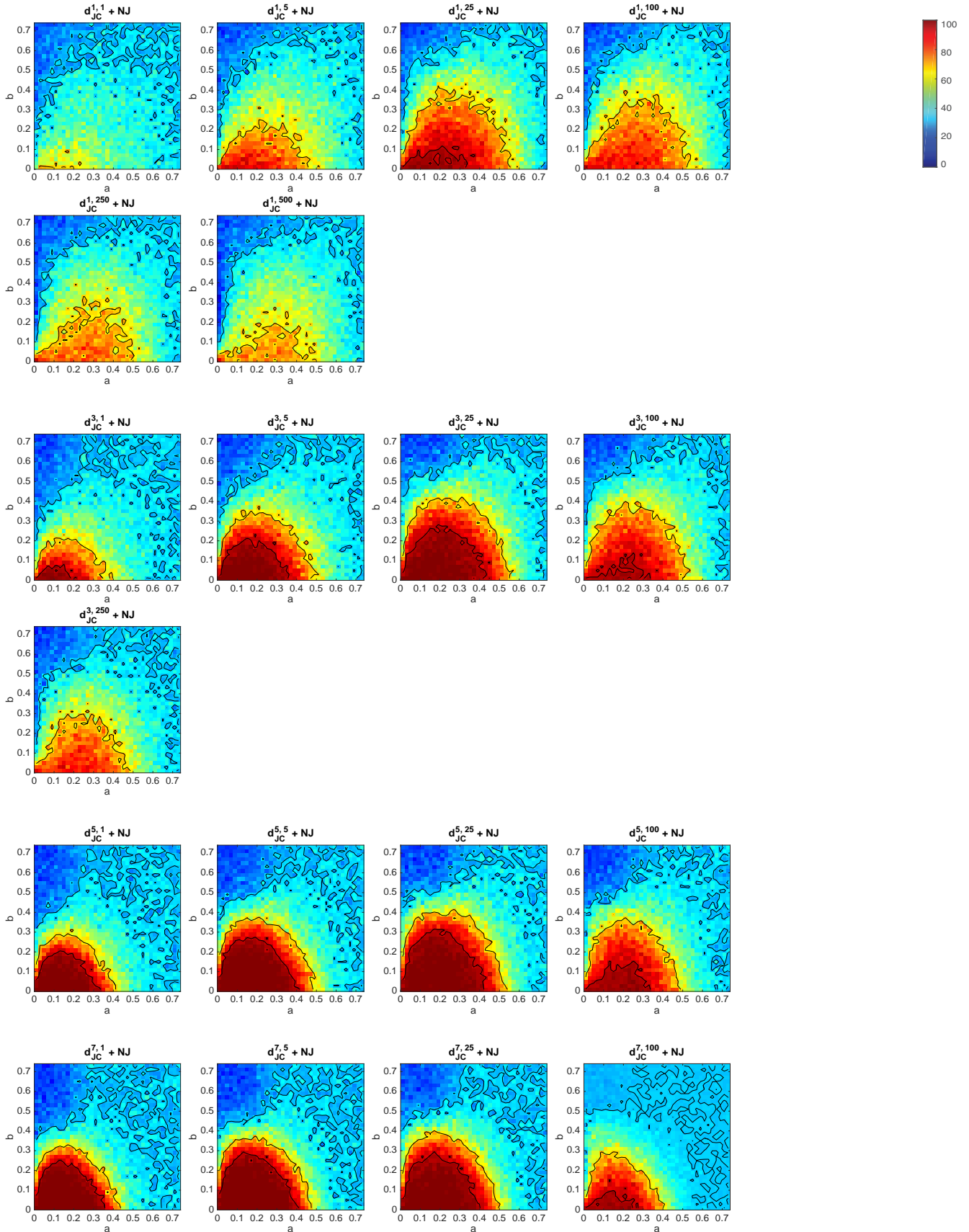
Simulation results using $d_{JC}^{k,B}$ + Neighbor Joining, 1000 bp



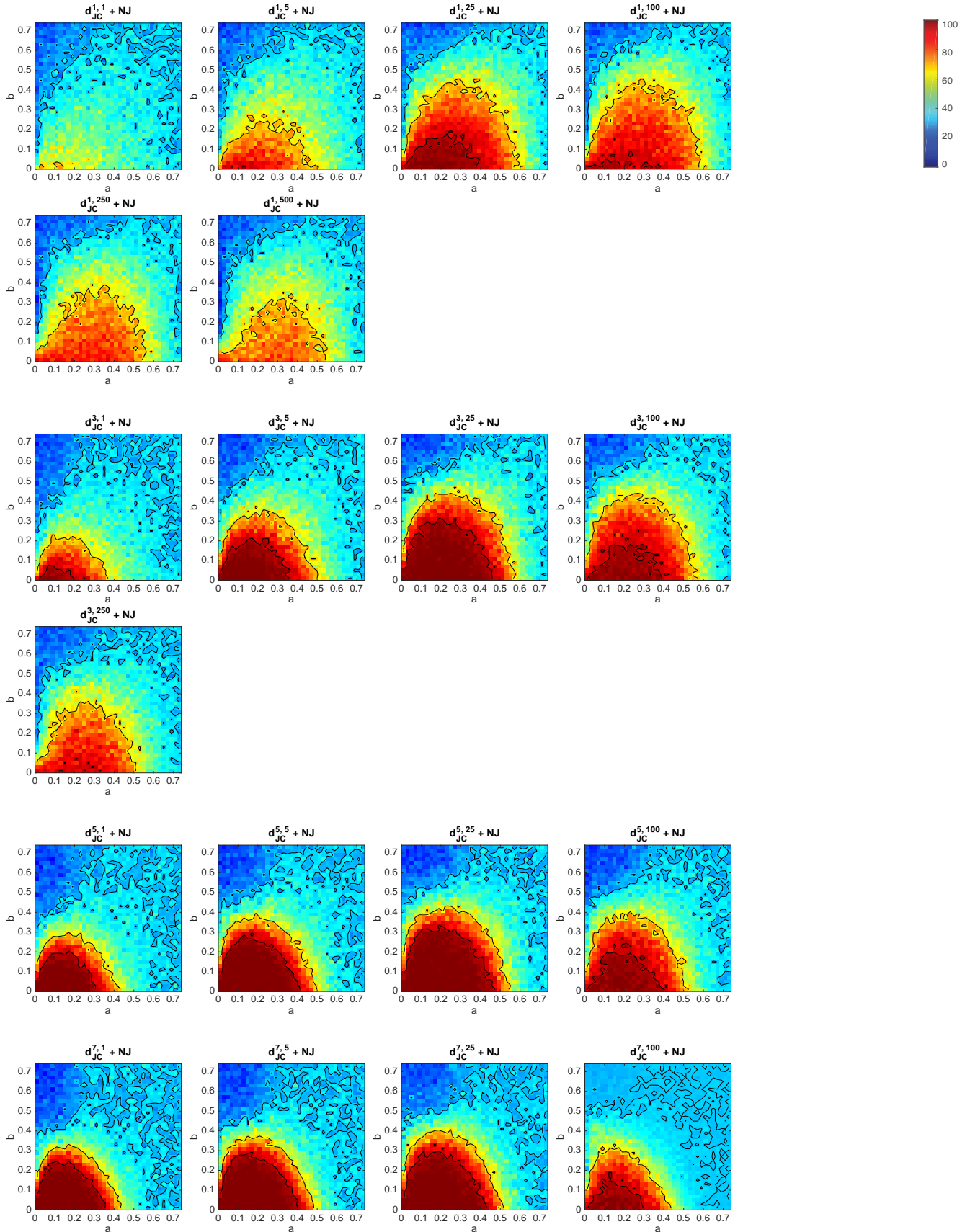
2.2 JC with indel rate $\mu = 0.01$. Lavalette parameters $a = 1.1$, $M = 100$.



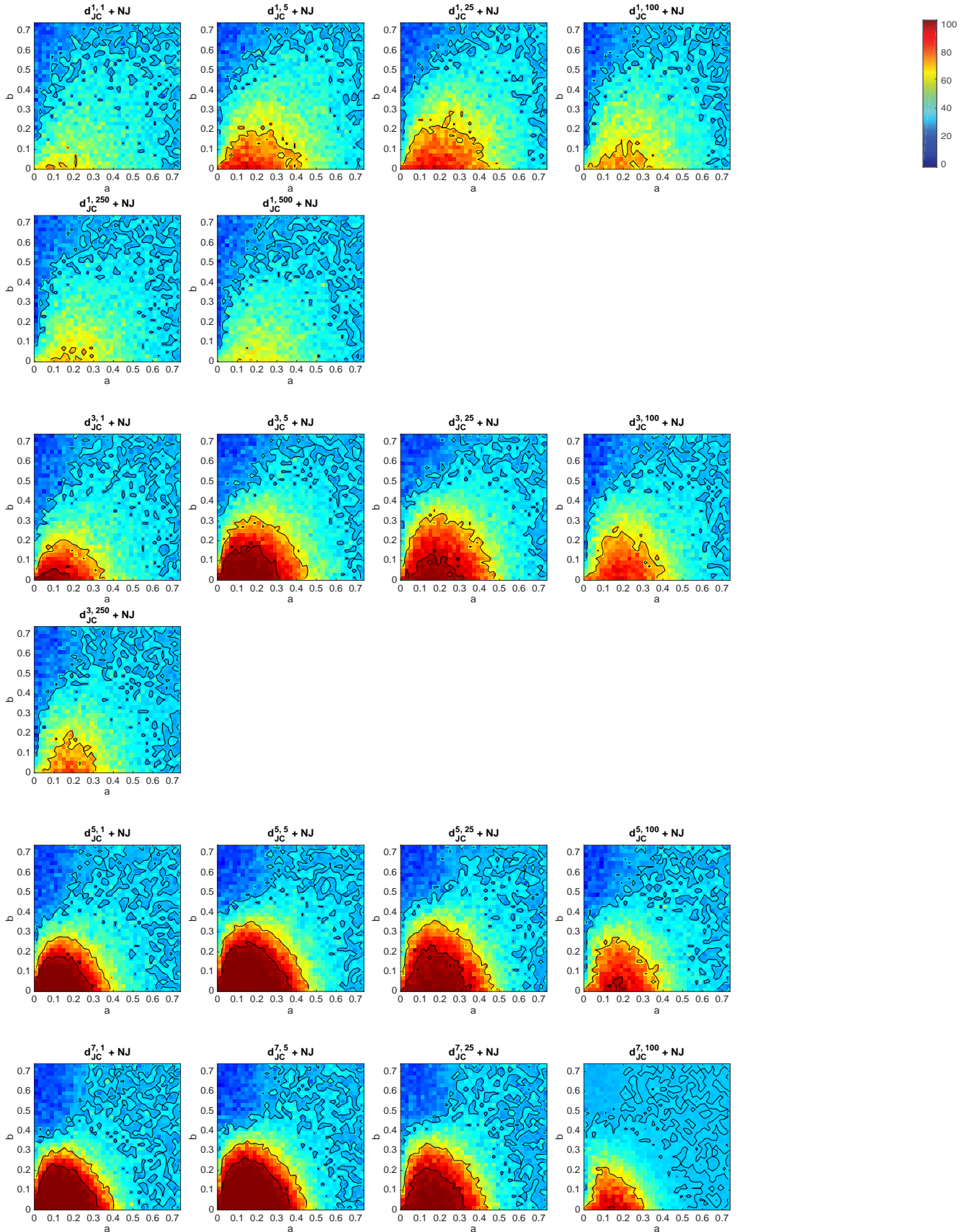
2.3 JC with indel rate $\mu = 0.01$. Lavalette parameters $a = 1.5$, $M = 100$.



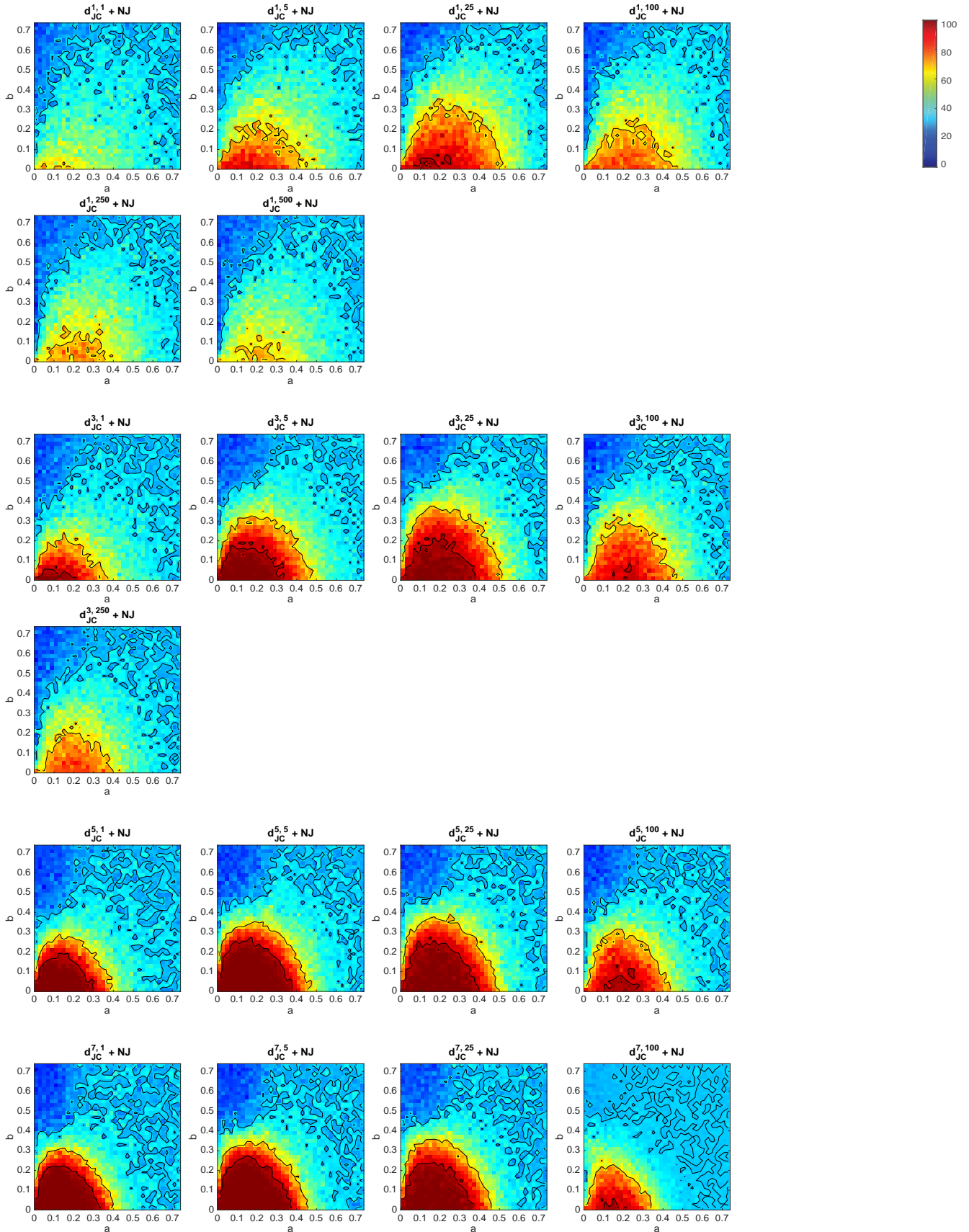
2.4 JC with indel rate $\mu = 0.01$. Lavalette parameters $a = 1.8$, $M = 100$.



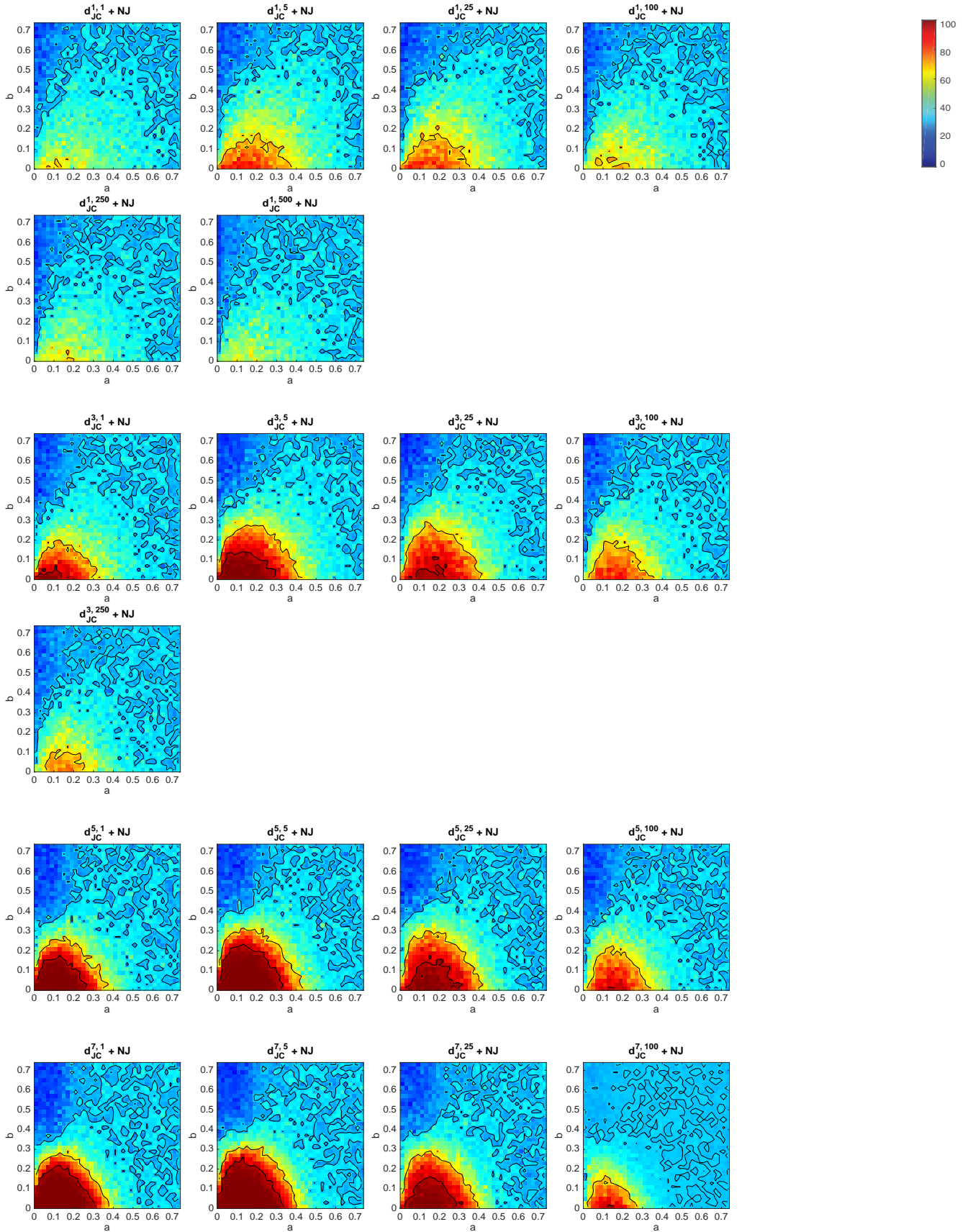
2.5 JC with indel rate $\mu = 0.05$. Lavalette parameters $a = 1.5$, $M = 100$.



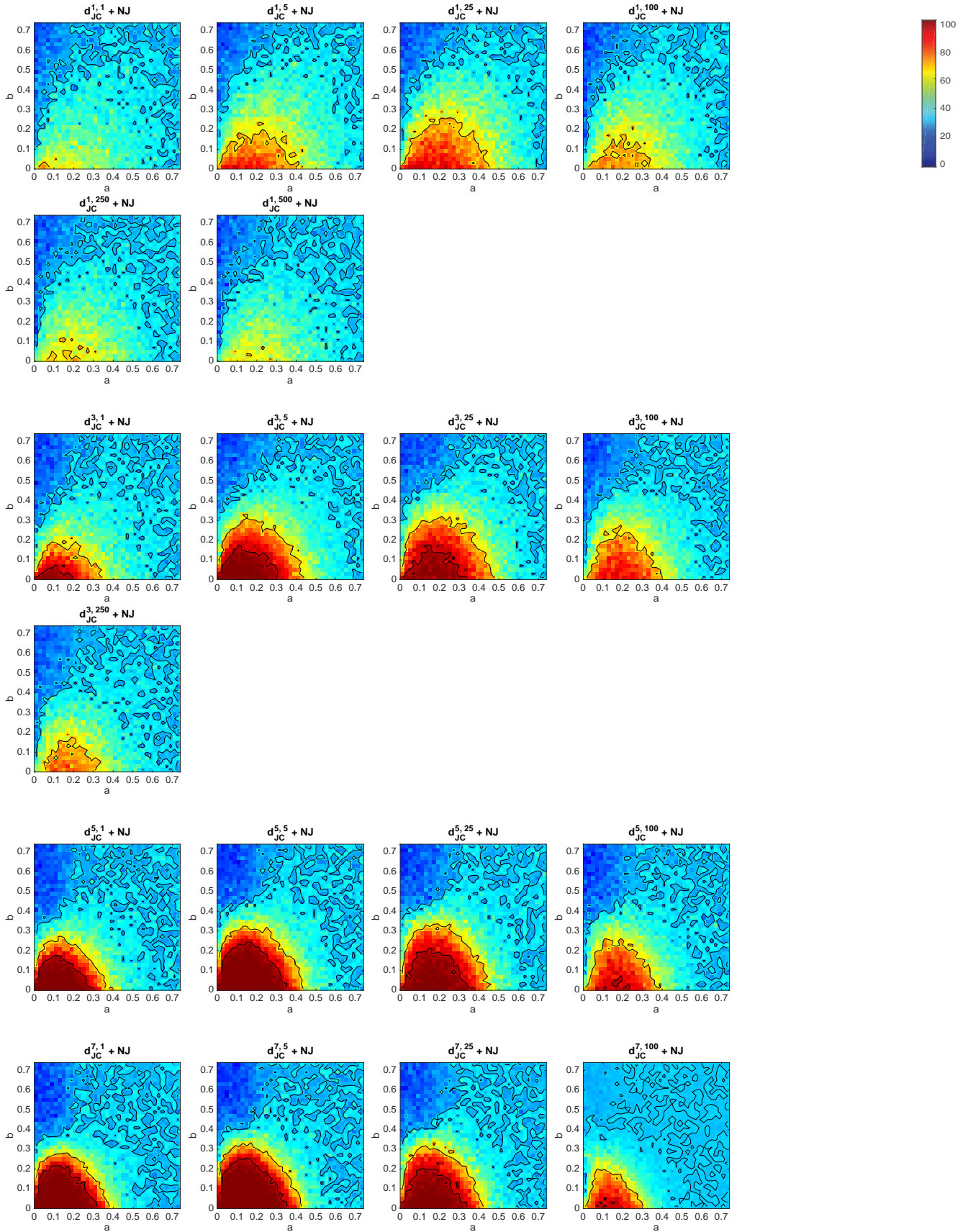
2.6 JC with indel rate $\mu = 0.05$. Lavalette parameters $a = 1.8$, $M = 100$.



2.7 JC with indel rate $\mu = 0.10$. Lavalette parameters $a = 1.5$, $M = 100$.



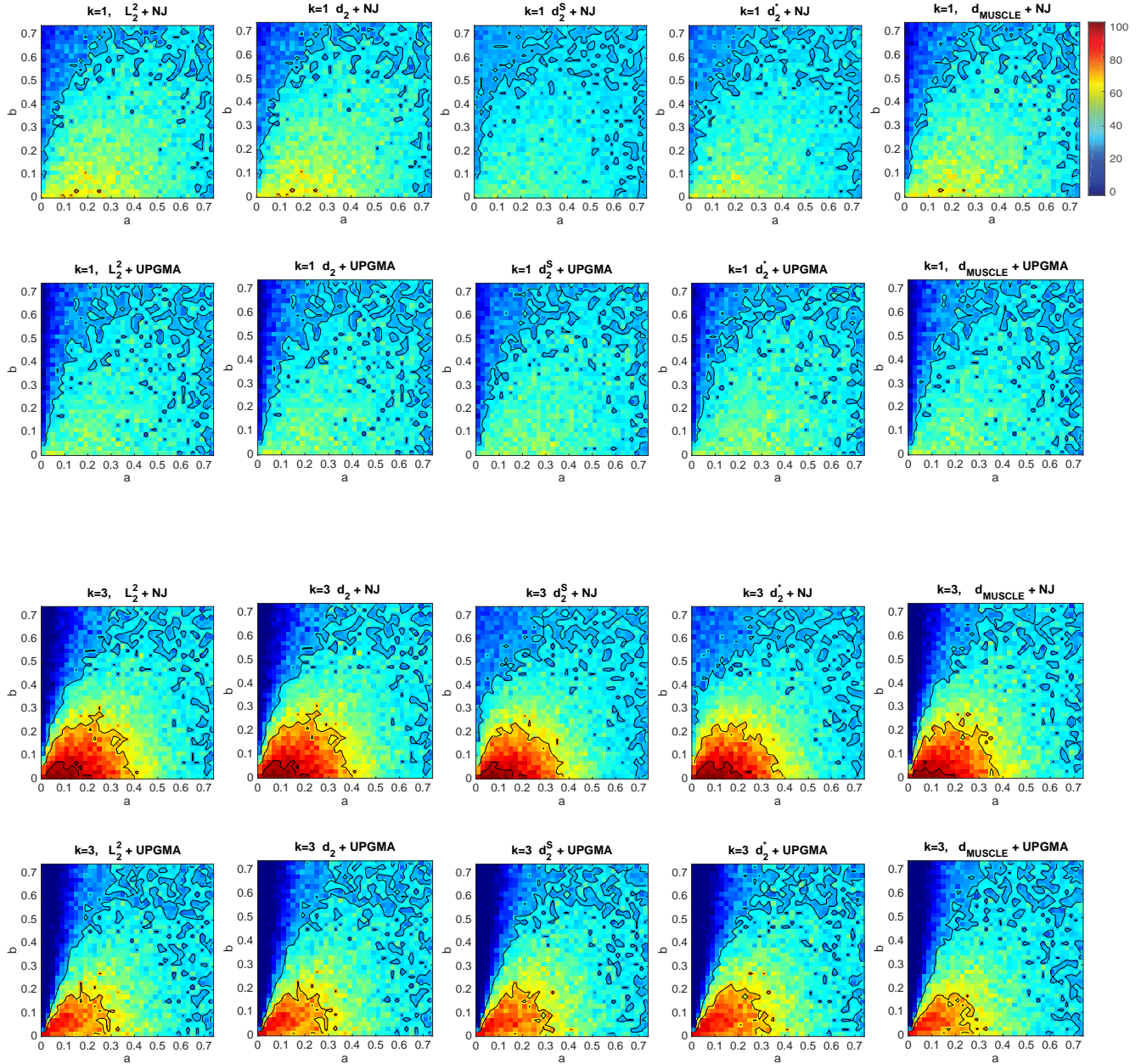
2.8 JC with indel rate $\mu = 0.10$. Lavalette parameters $a = 1.8$, $M = 100$.



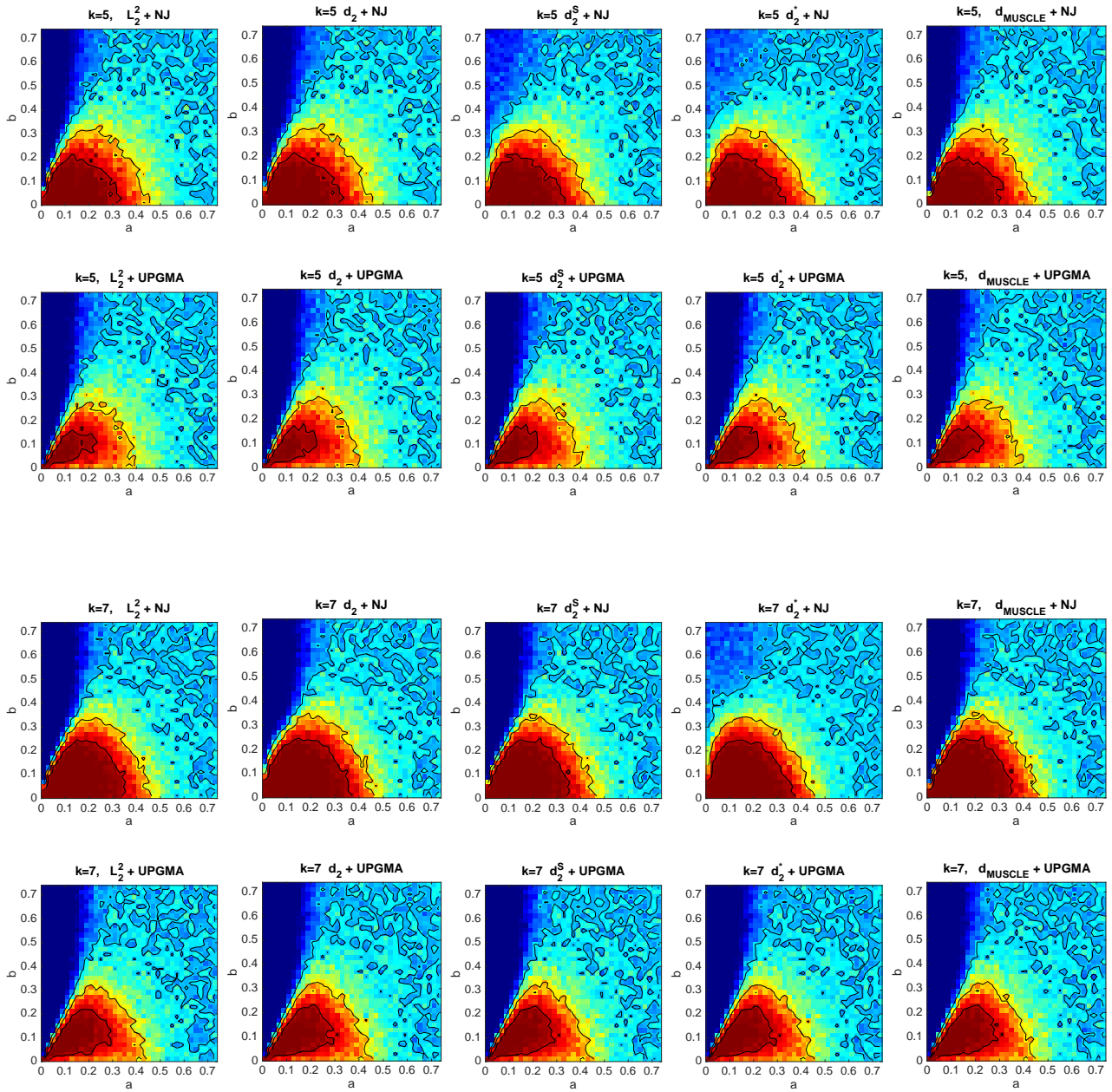
3 Simulation results using other k -mer distances + Neighbor Joining/UPGMA, 1000 bp

Each subgroup of figures corresponds to a fixed setting for the model parameters used to generate sequences. Titles indicate the value of k , the distance used, and whether tree construction was performed using NJ or UPGMA. Sequences have length 1000 bp.

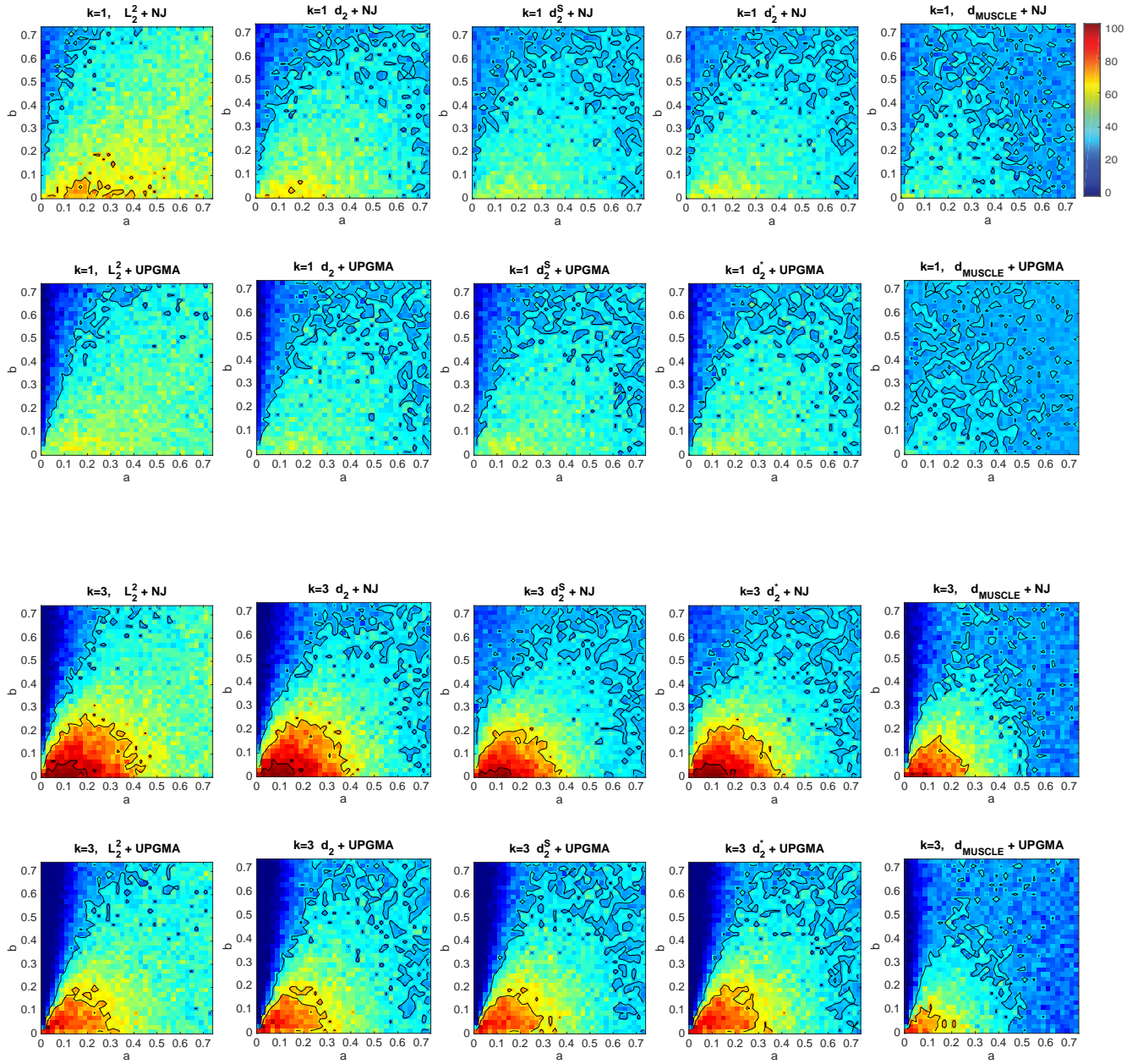
3.1 JC with no indel process.



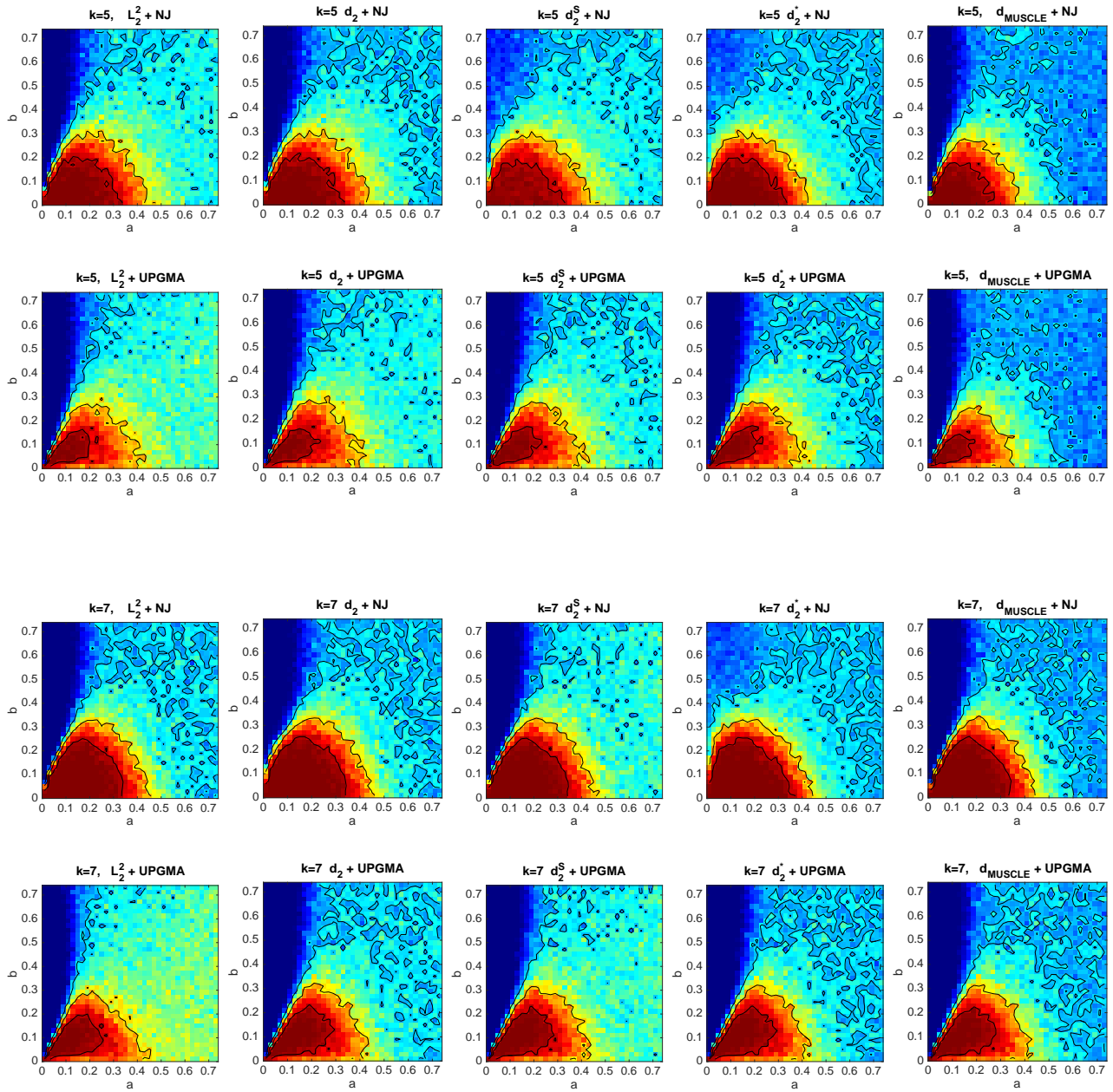
Simulation results using other k -mer distances, 1000 bp



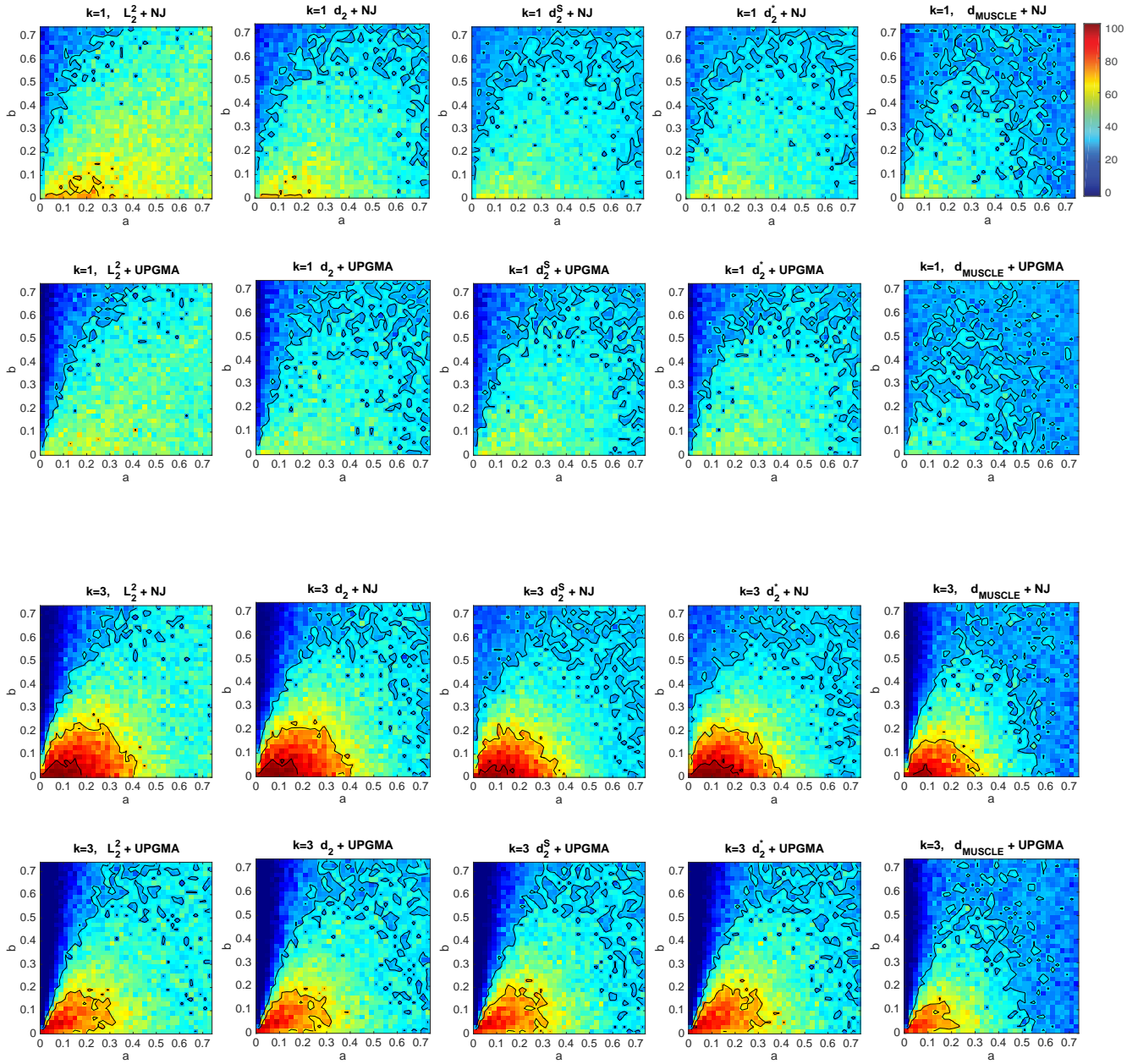
3.2 JC with indel rate $\mu = 0.01$. Lavalette parameters $a = 1.1$, $M = 100$.



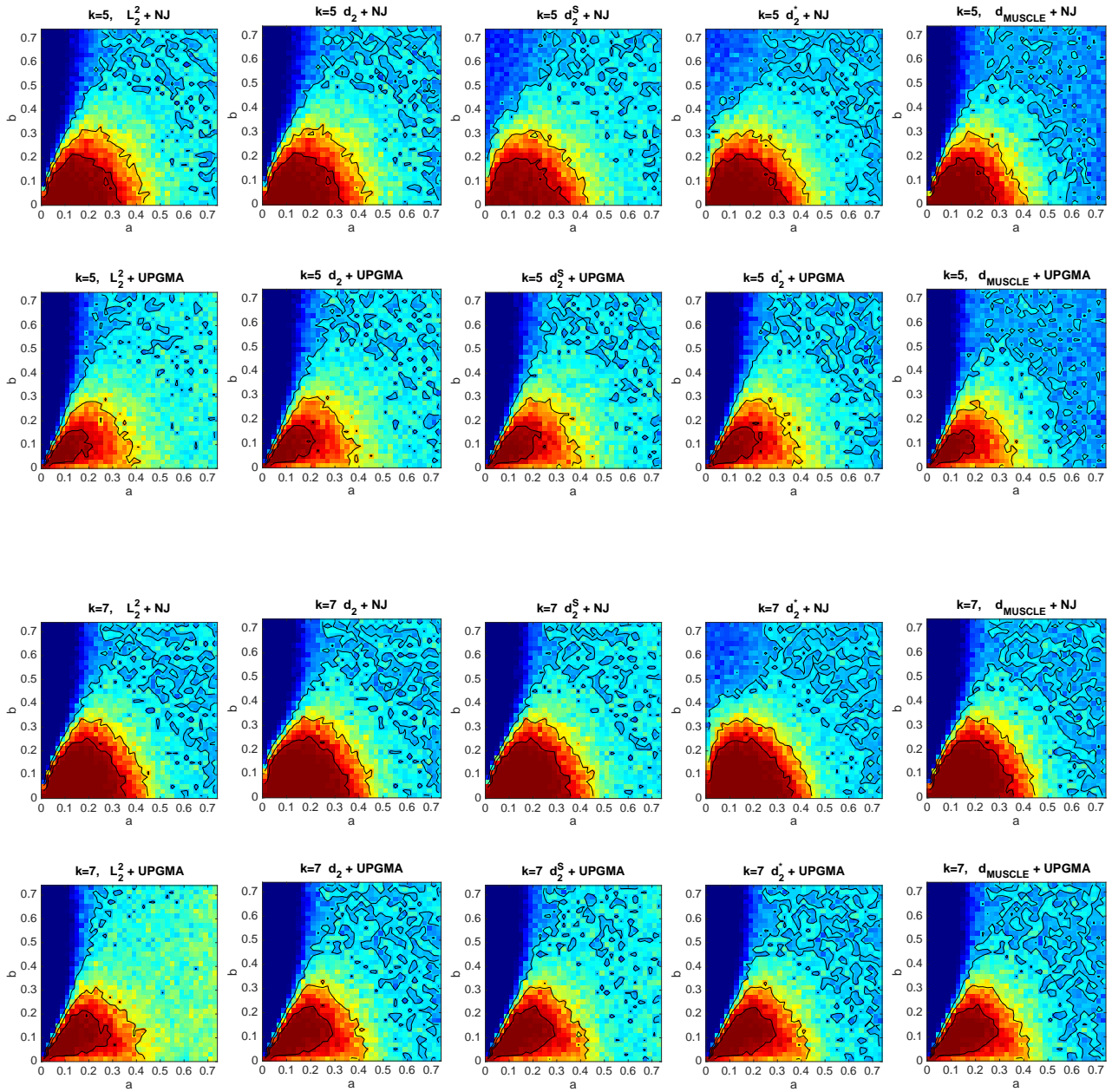
Simulation results using other k -mer distances, 1000 bp



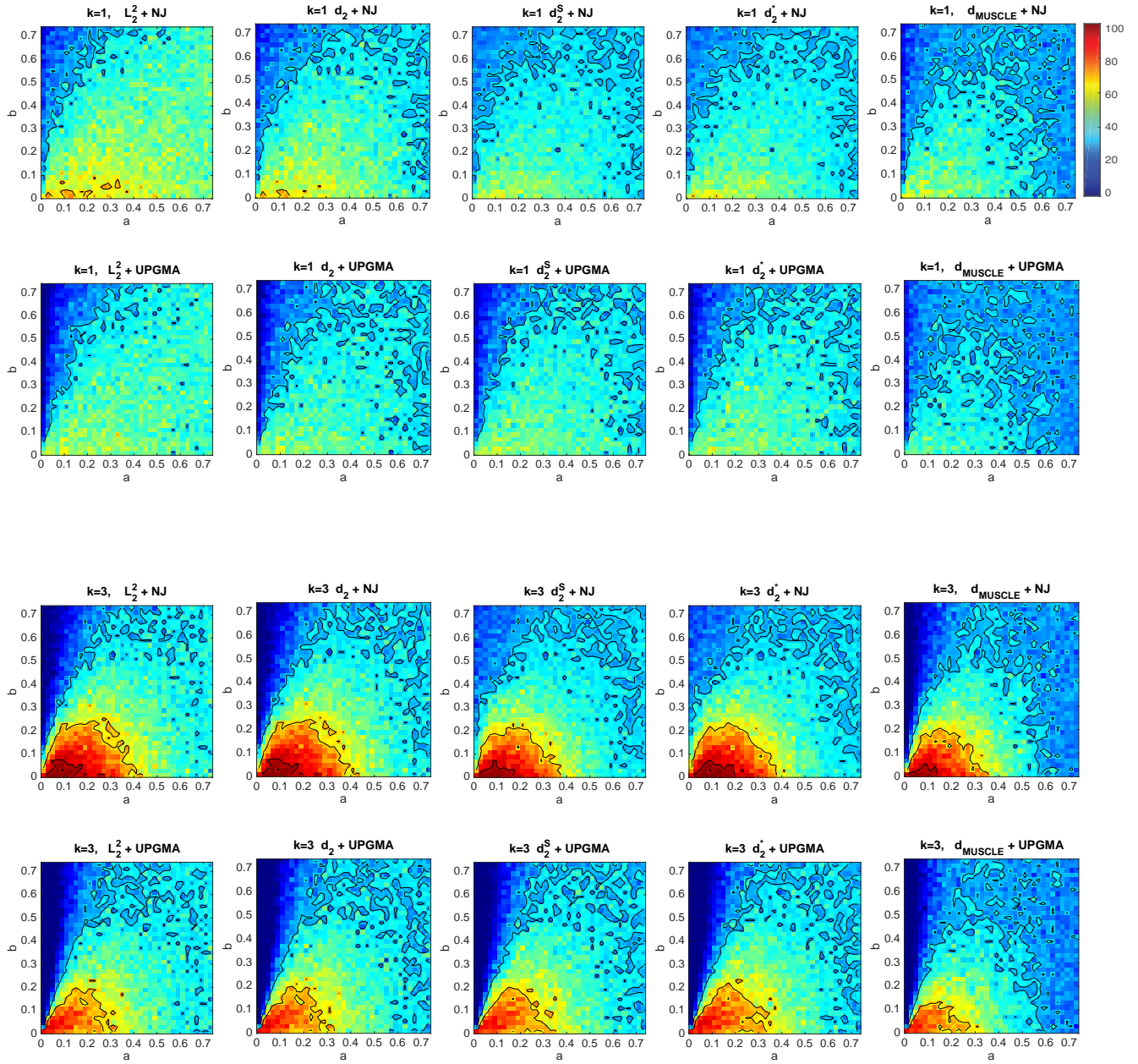
3.3 JC with indel rate $\mu = 0.01$. Lavalette parameters $a = 1.5$, $M = 100$.



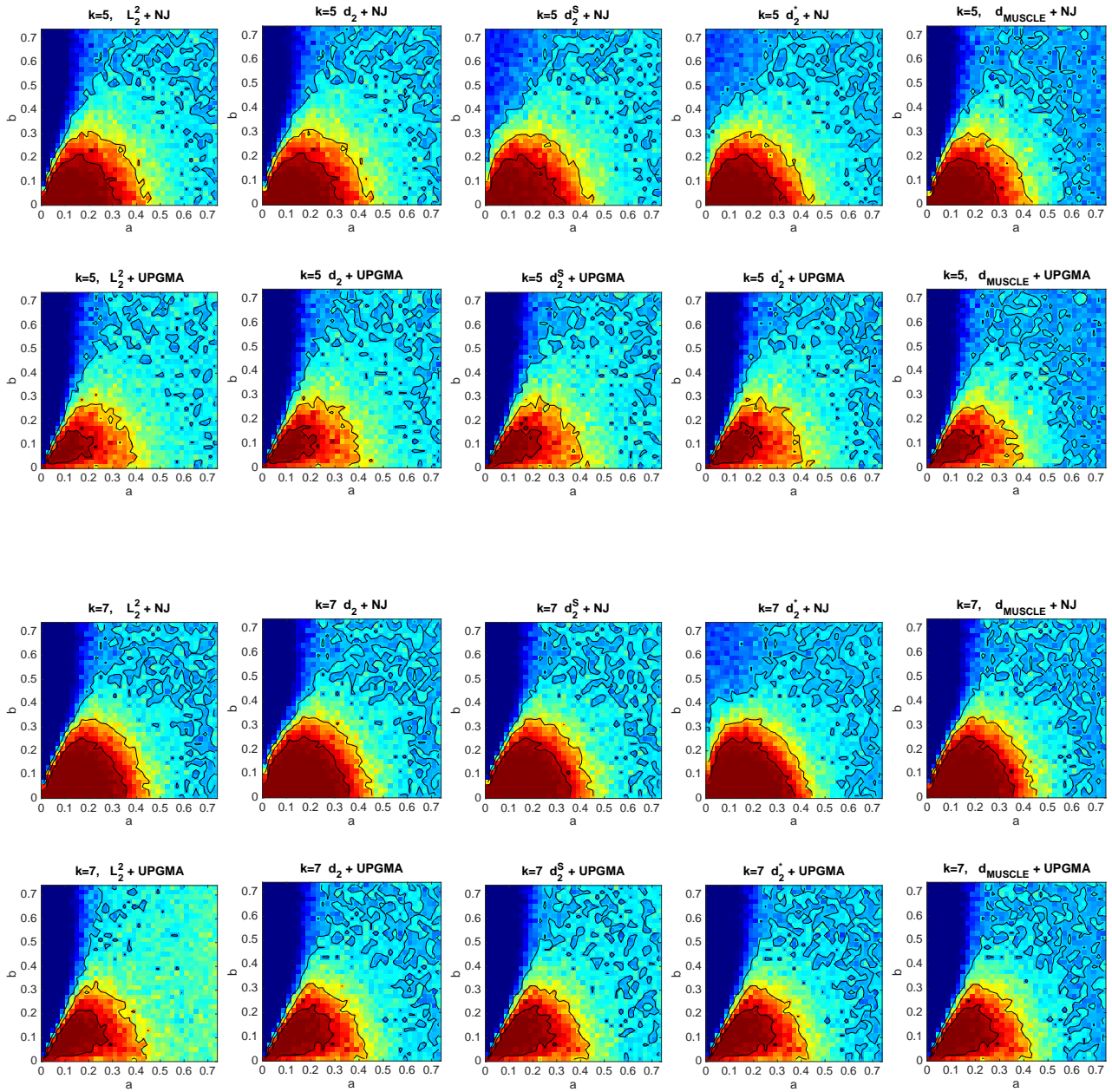
Simulation results using other k -mer distances, 1000 bp



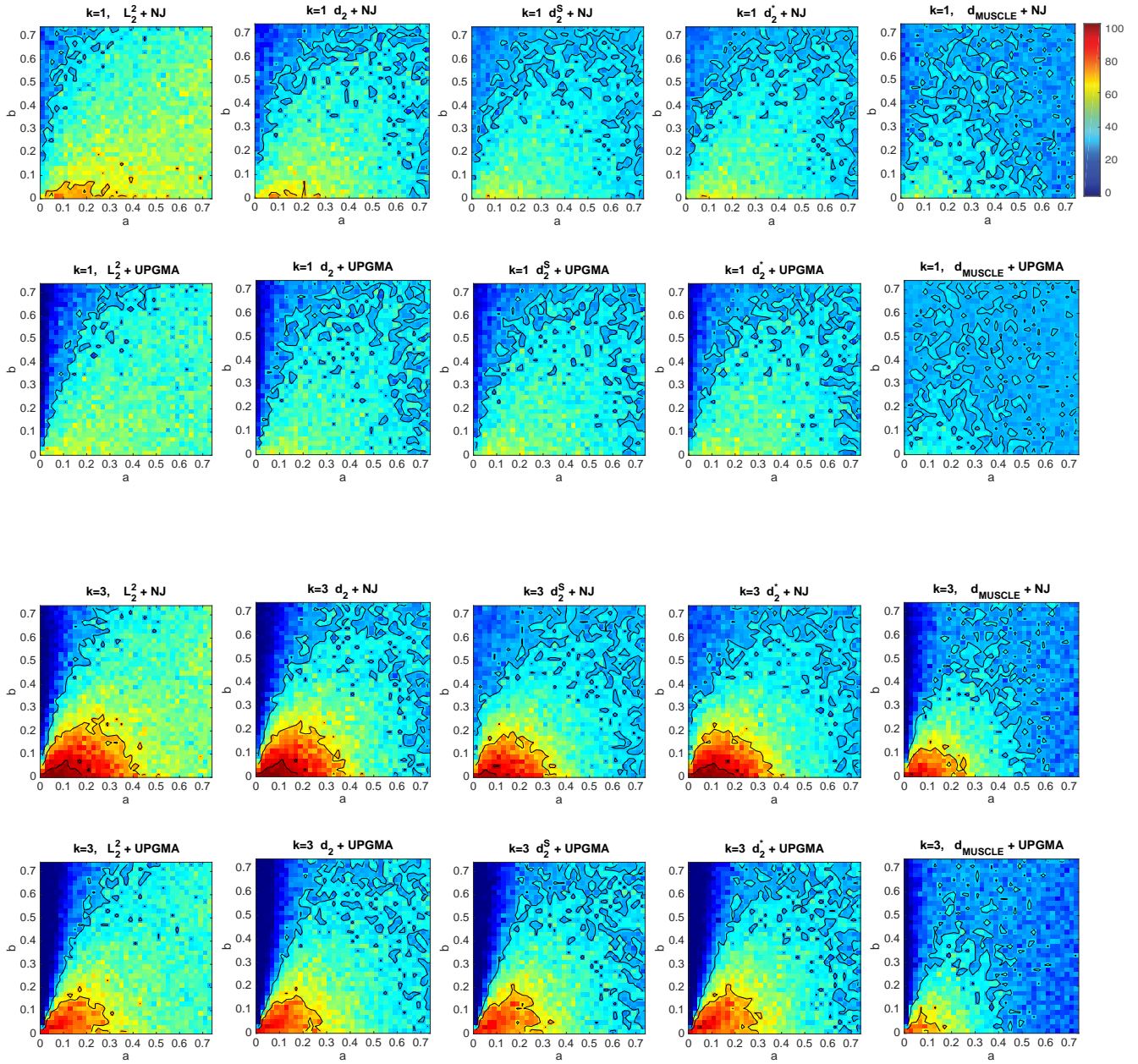
3.4 JC with indel rate $\mu = 0.01$. Lavalette parameters $a = 1.8$, $M = 100$.



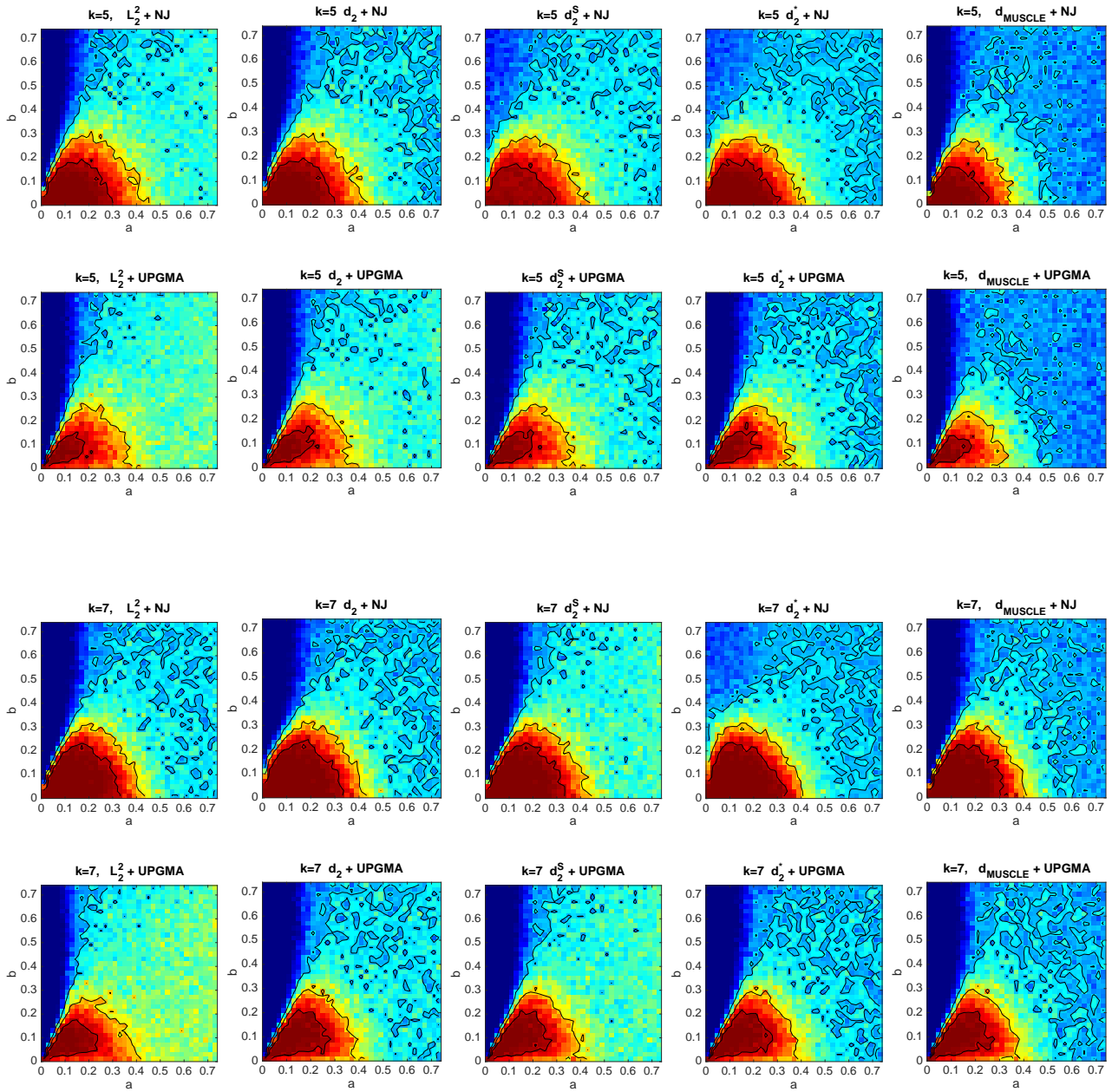
Simulation results using other k -mer distances, 1000 bp



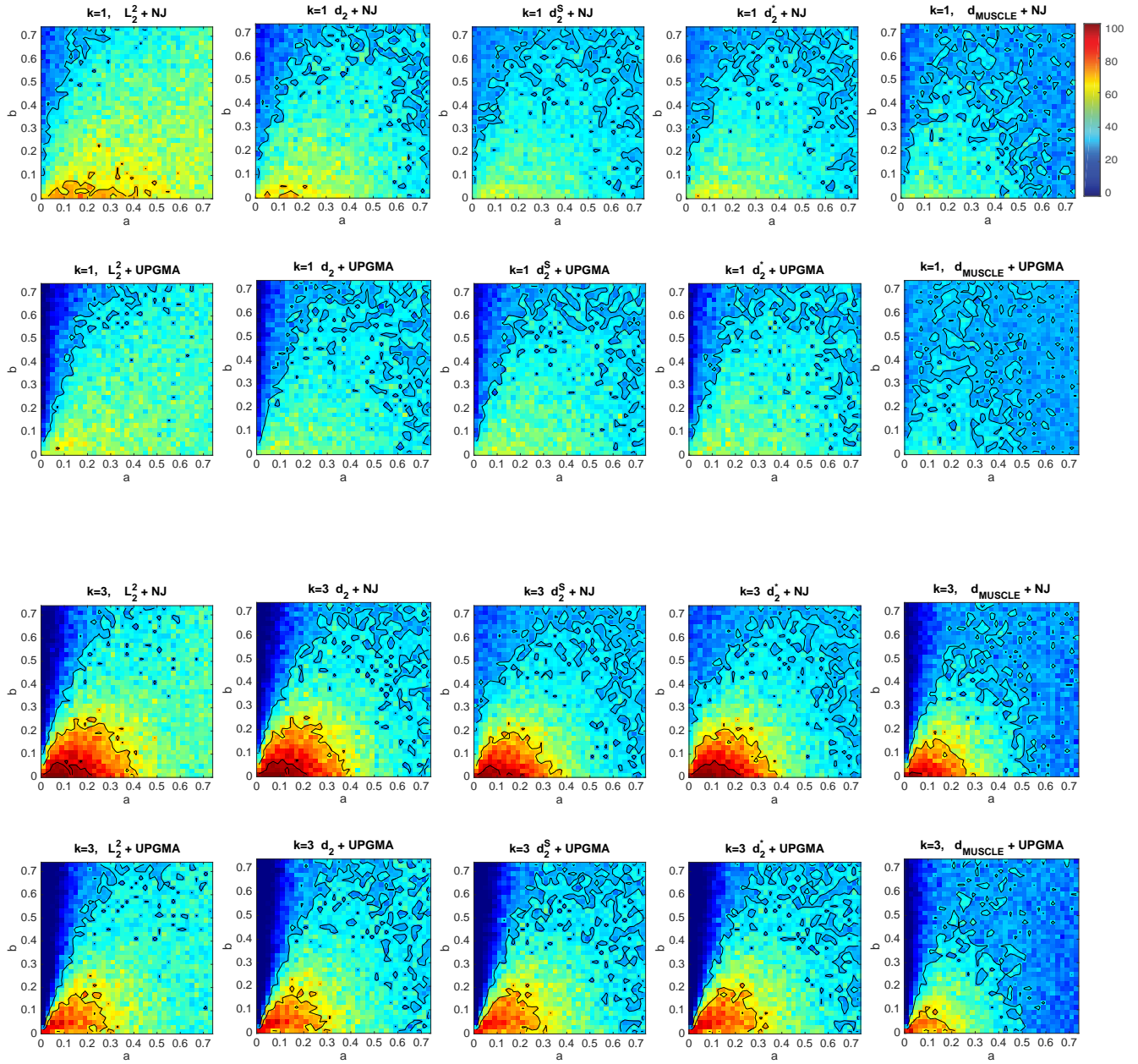
3.5 JC with indel rate $\mu = 0.05$. Lavalette parameters $a = 1.5$, $M = 100$.



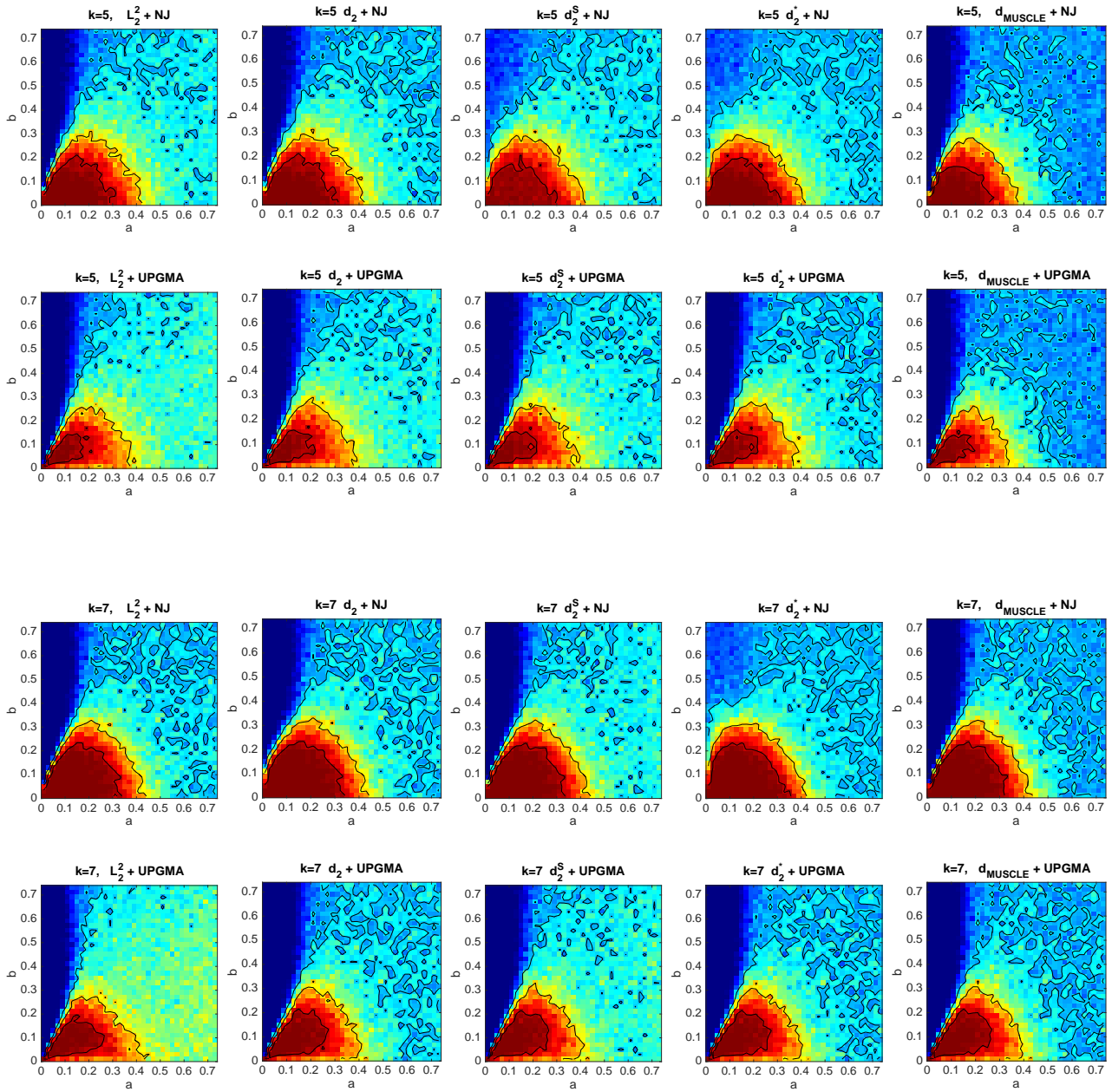
Simulation results using other k -mer distances, 1000 bp



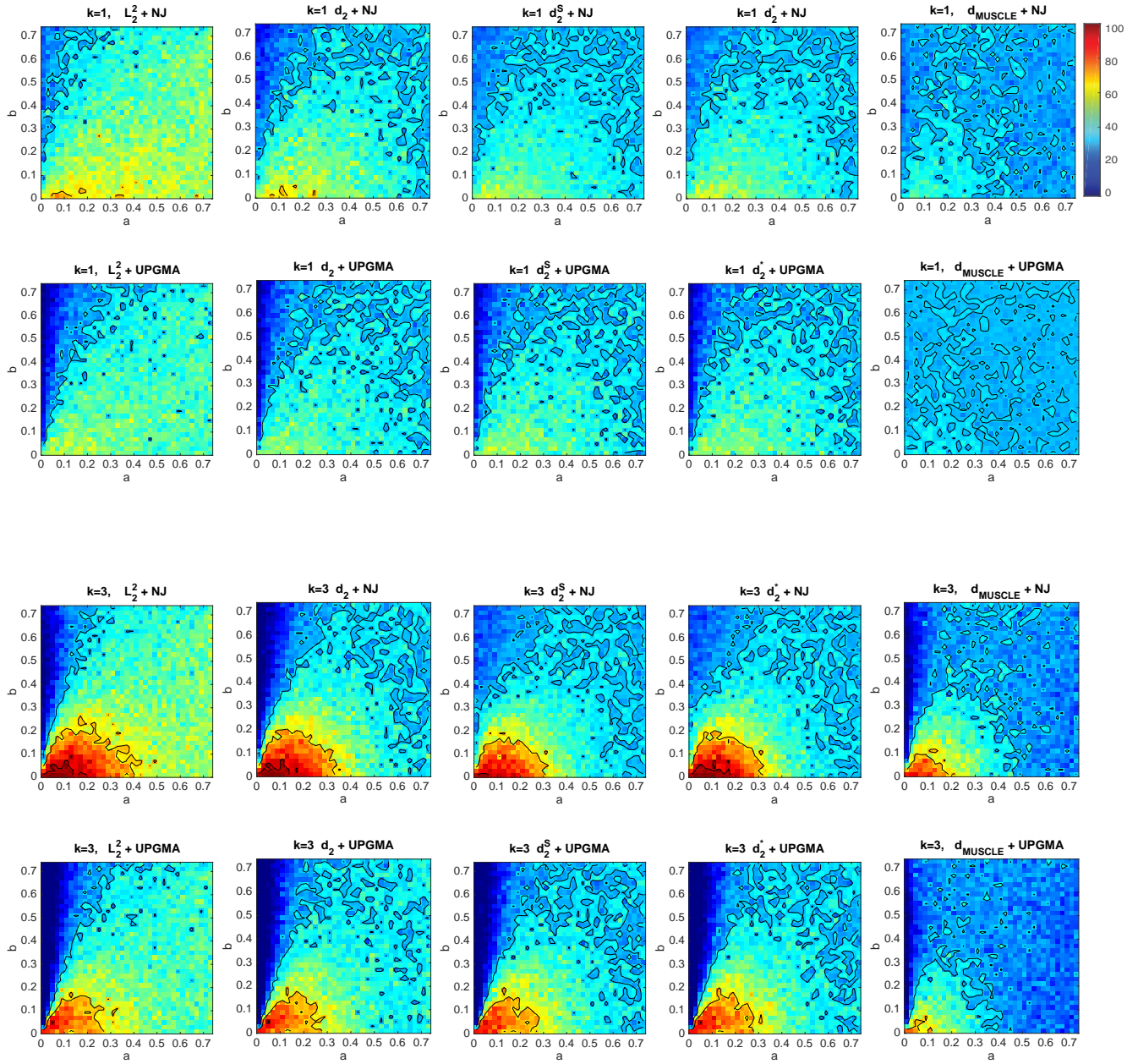
3.6 JC with indel rate $\mu = 0.05$. Lavalette parameters $a = 1.8$, $M = 100$.



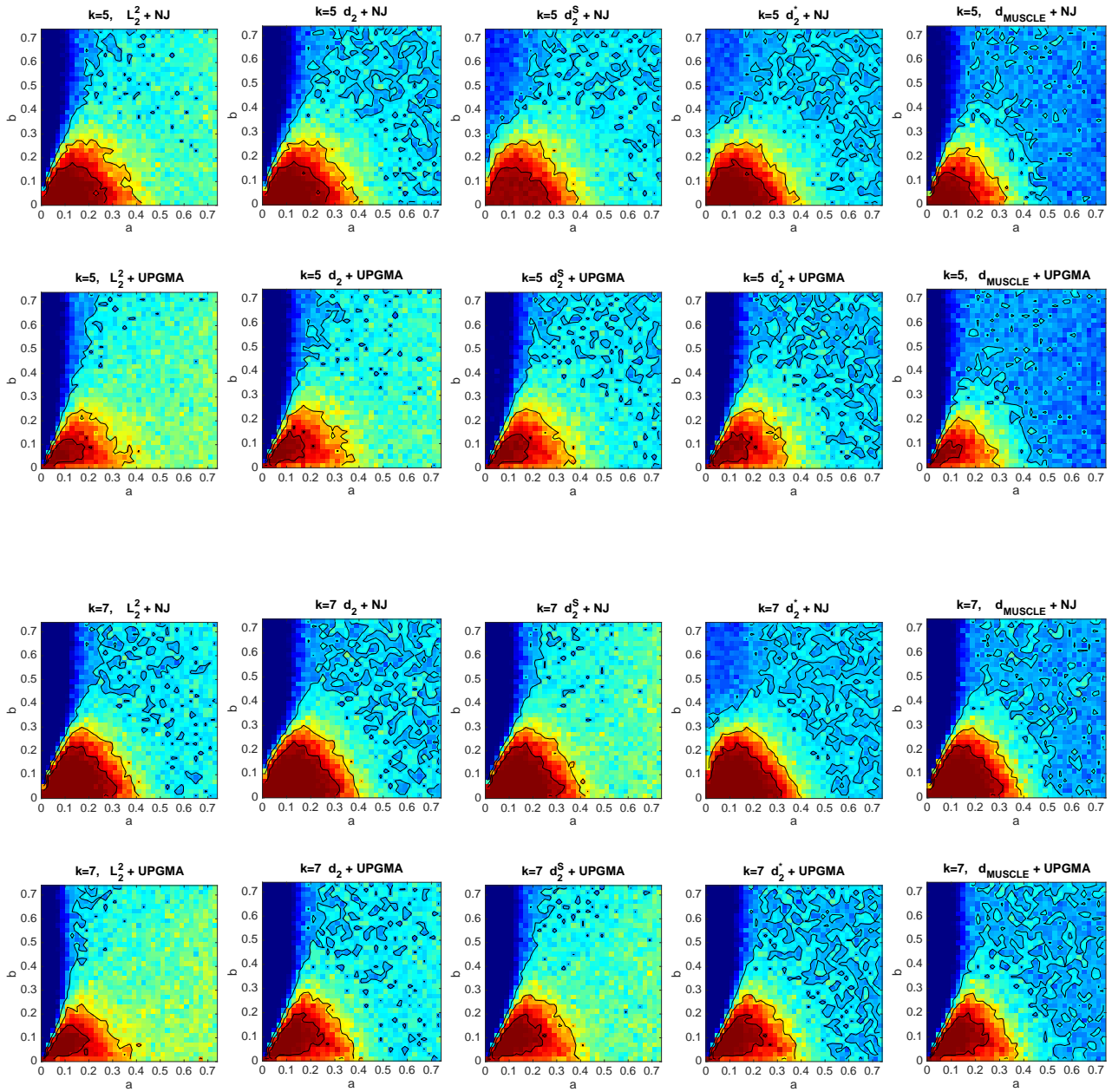
Simulation results using other k -mer distances, 1000 bp



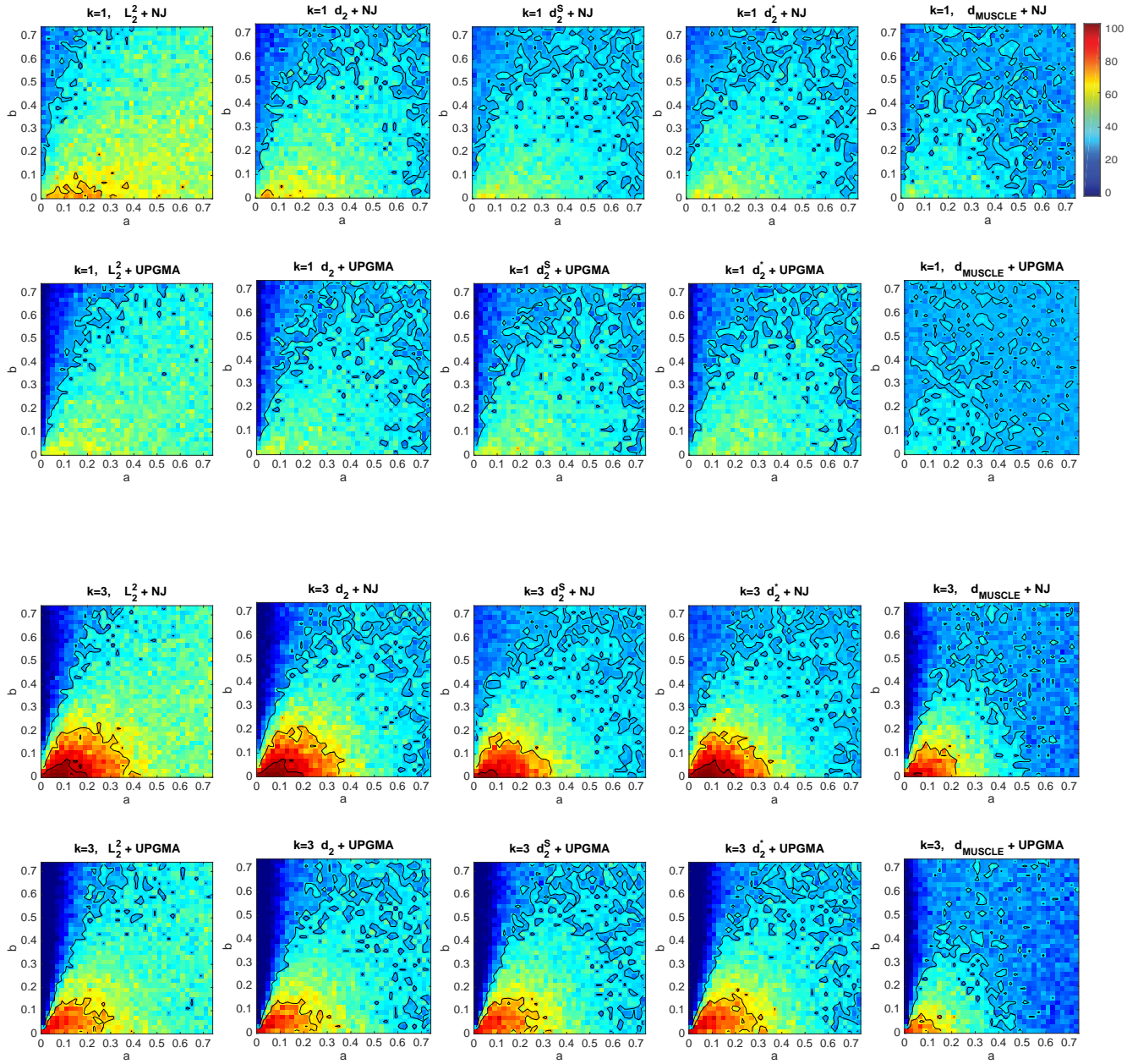
3.7 JC with indel rate $\mu = 0.10$. Lavalette parameters $a = 1.5$, $M = 100$.



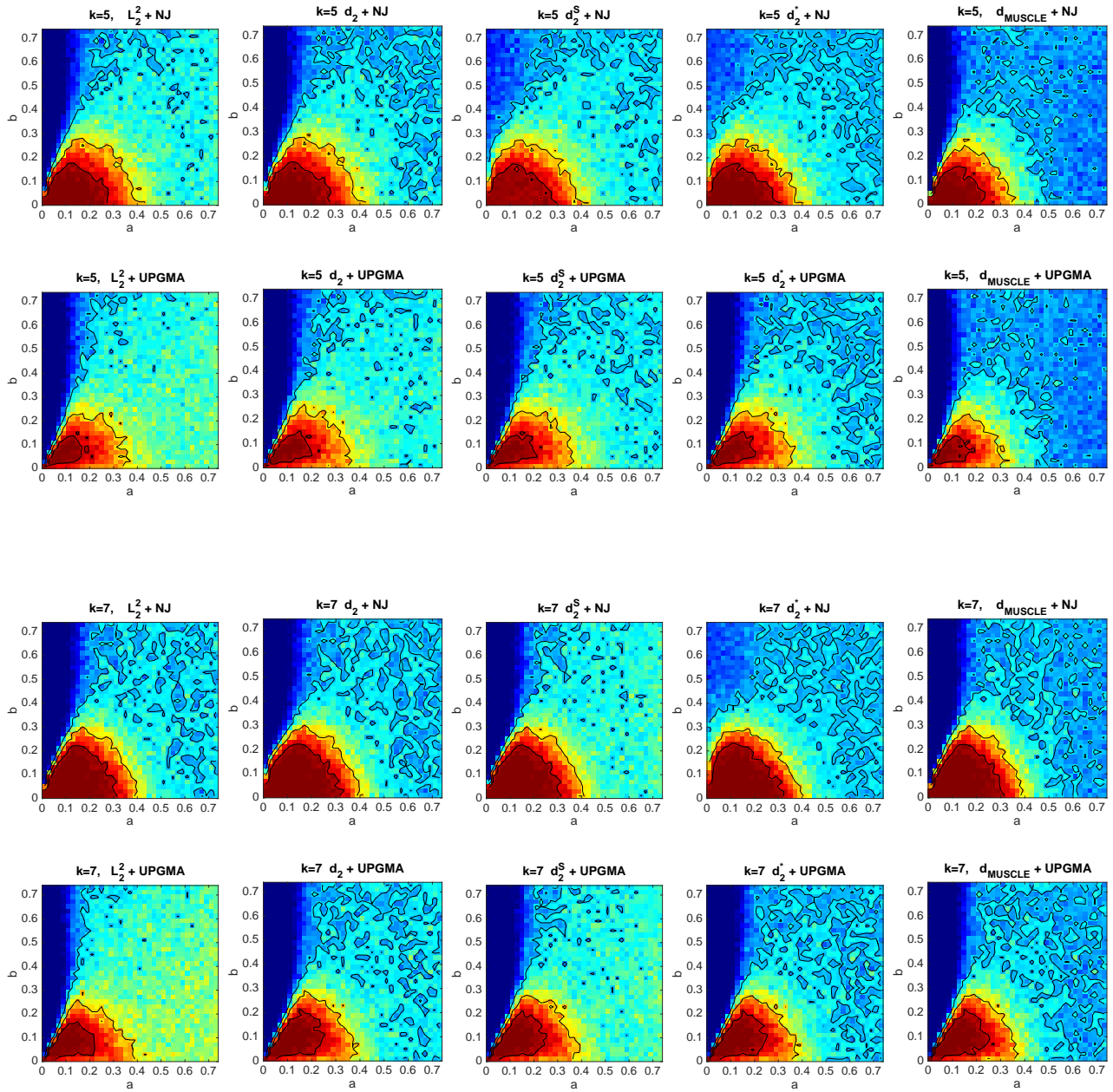
Simulation results using other k -mer distances, 1000 bp



3.8 JC with indel rate $\mu = 0.10$. Lavalette parameters $a = 1.8$, $M = 100$.



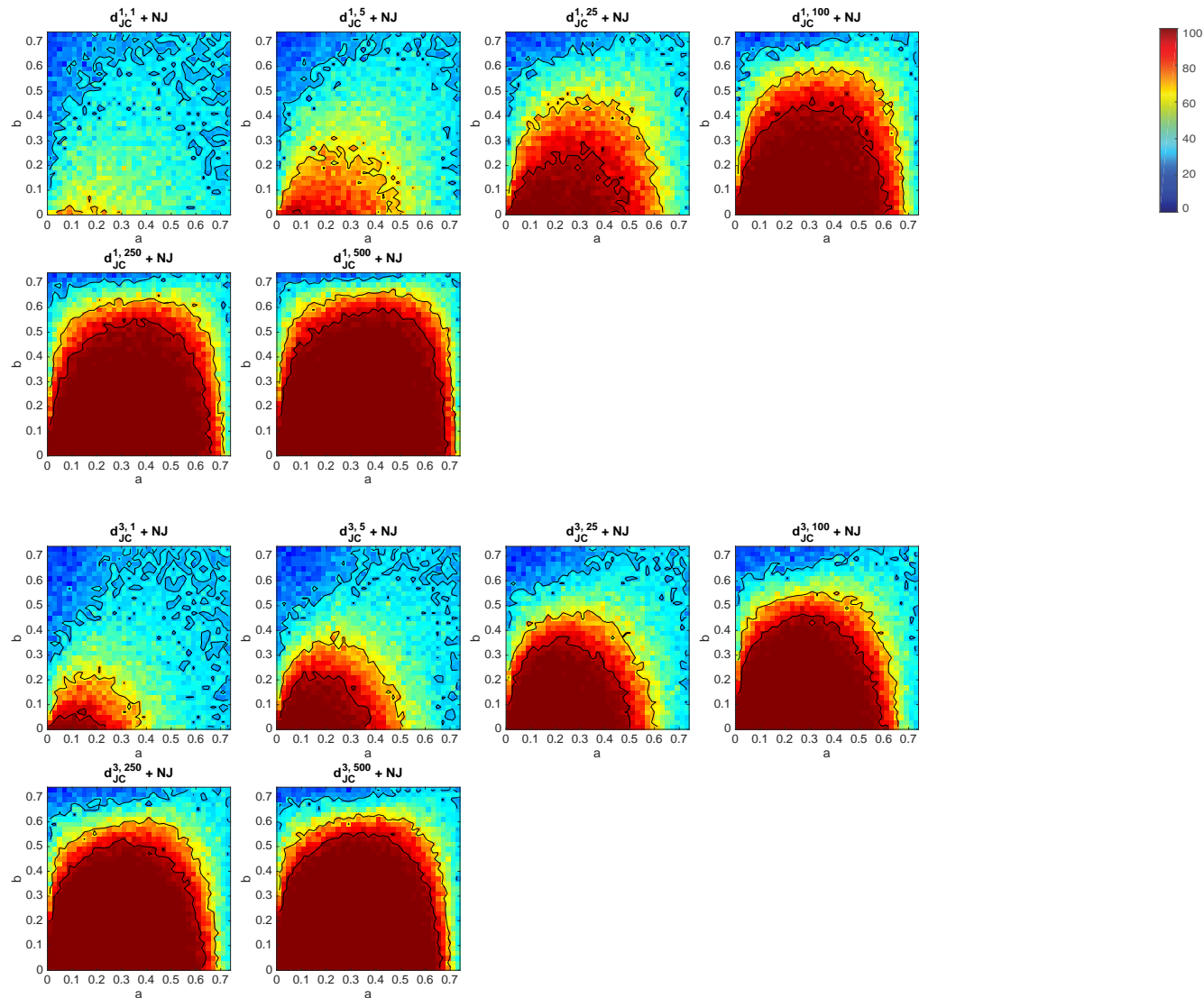
Simulation results using other k -mer distances, 1000 bp



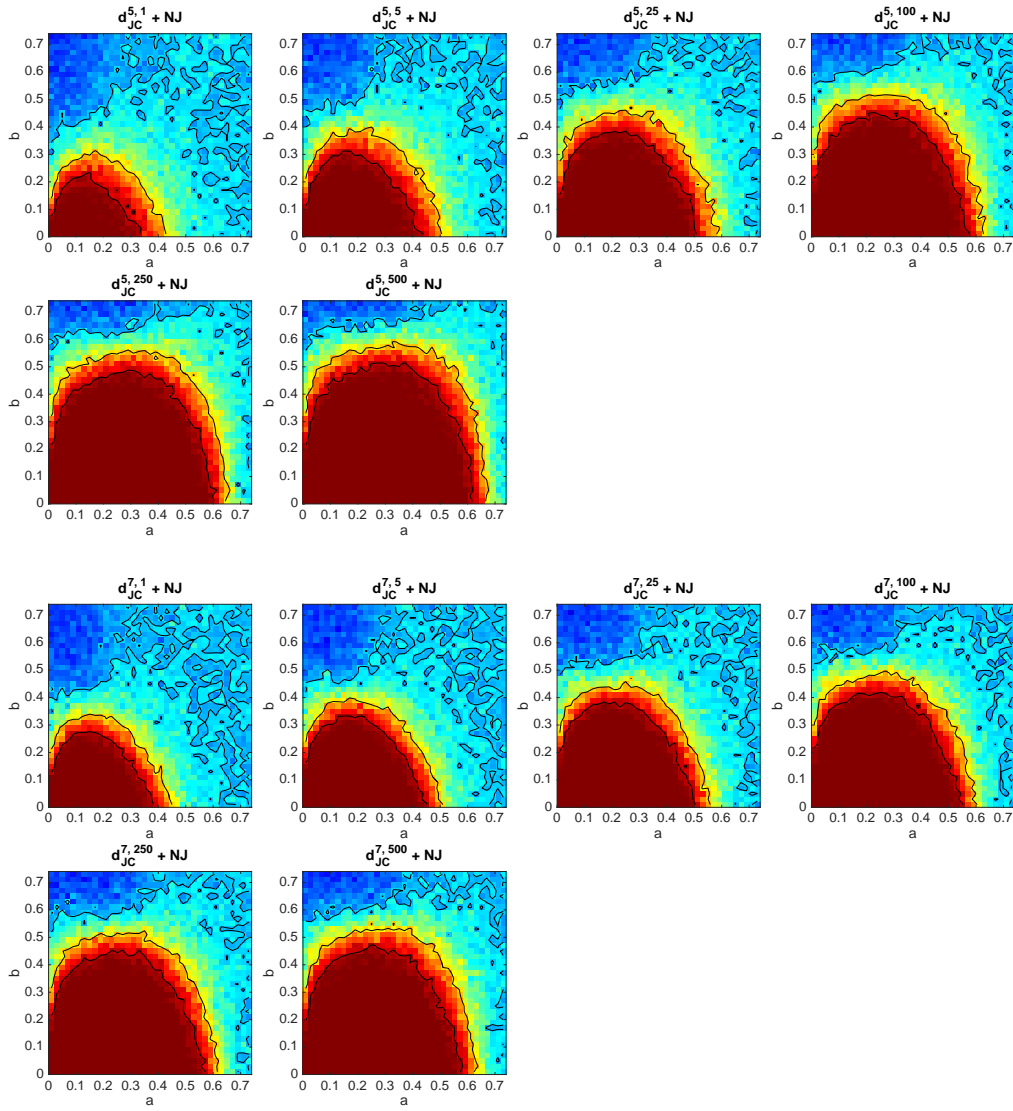
4 Simulation results using $d_{JC}^{k,B} + \text{Neighbor Joining}$, 10,000 bp

Each subgroup of figures corresponds to a fixed setting for the model parameters used to generate sequences. Titles indicate the value of k and the number of blocks used for computing $d_{JC}^{k,B}$. Sequences have length 10,000 bp.

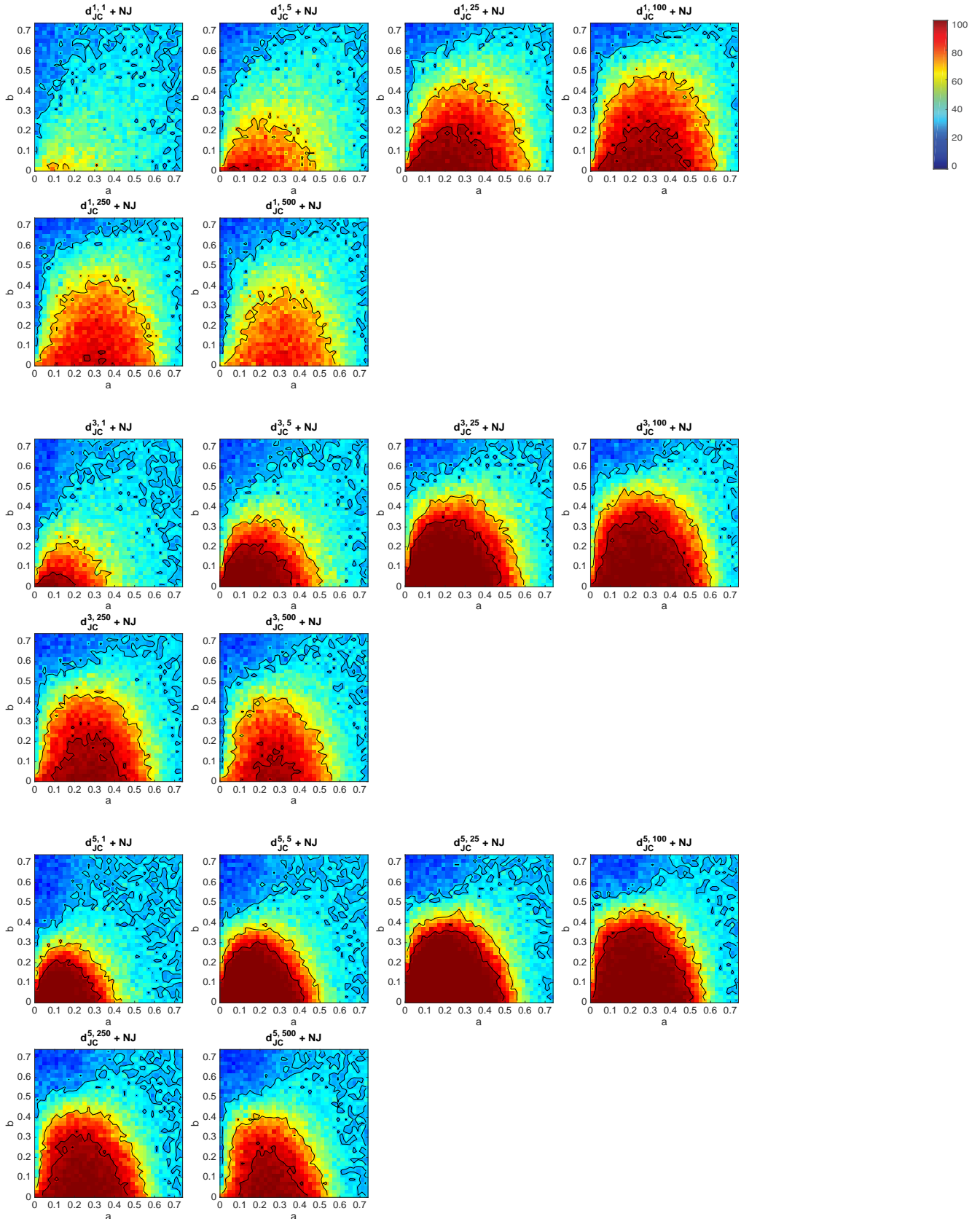
4.1 JC with no indel process.



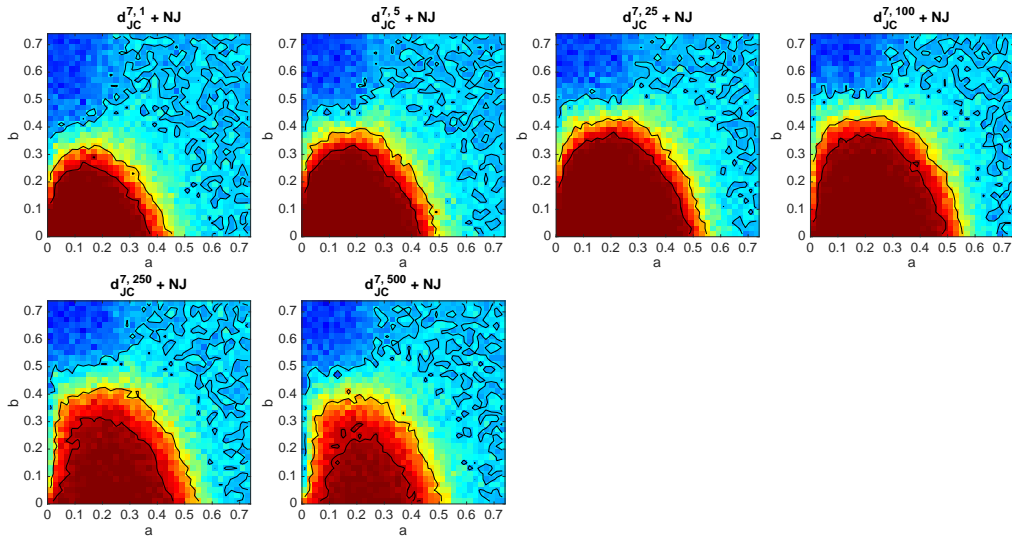
Simulation results using $d_{JC}^{k,B} + \text{Neighbor Joining}$, 10,000 bp



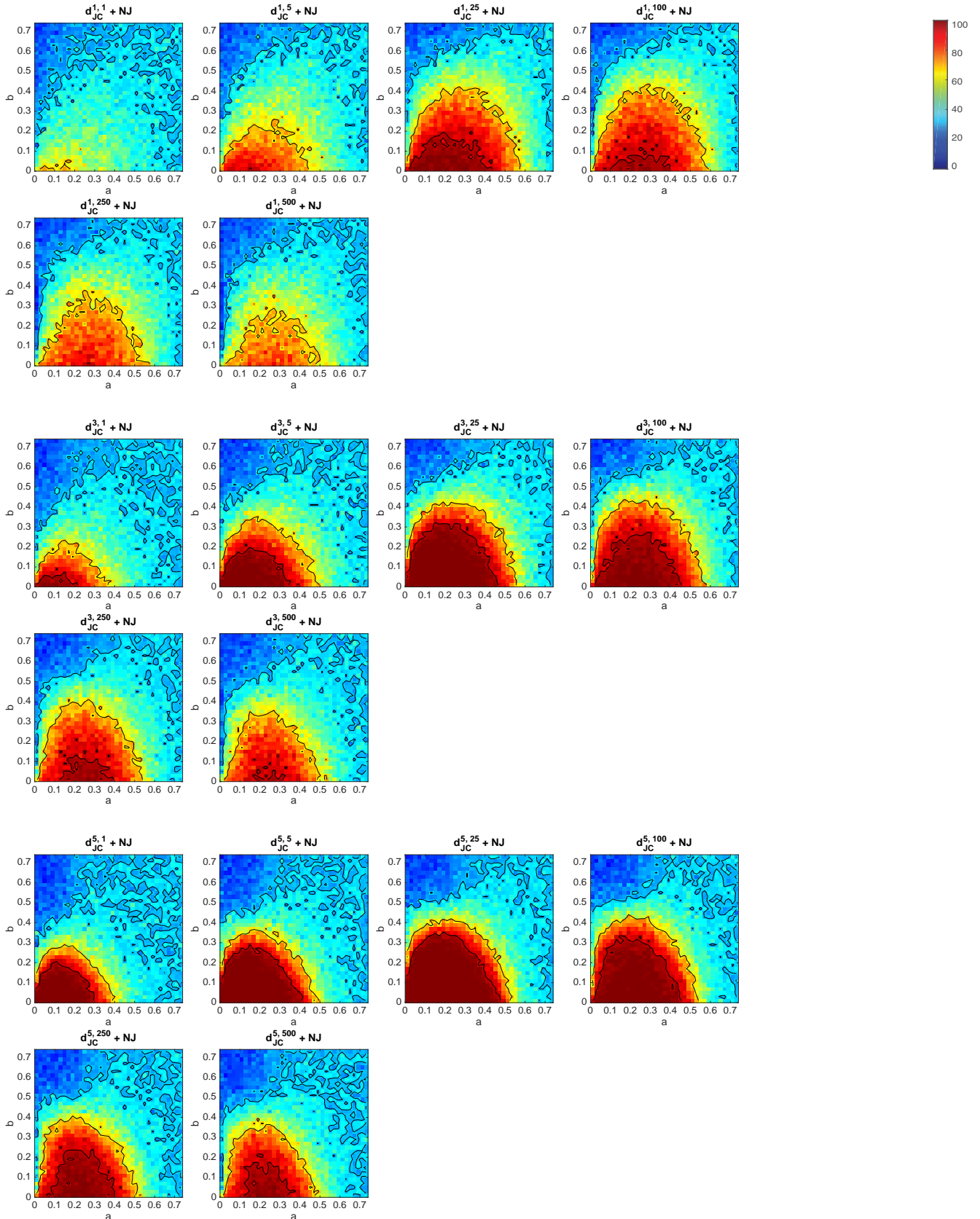
4.2 JC with indel rate $\mu = 0.01$. Lavalette parameters $a = 1.5$, $M = 100$.



Simulation results using $d_{JC}^{k,B} + \text{Neighbor Joining}$, 10,000 bp



4.3 JC with indel rate $\mu = 0.05$. Lavalette parameters $a = 1.8$, $M = 100$.



Simulation results using $d_{JC}^{k,B}$ + Neighbor Joining, 10,000 bp

