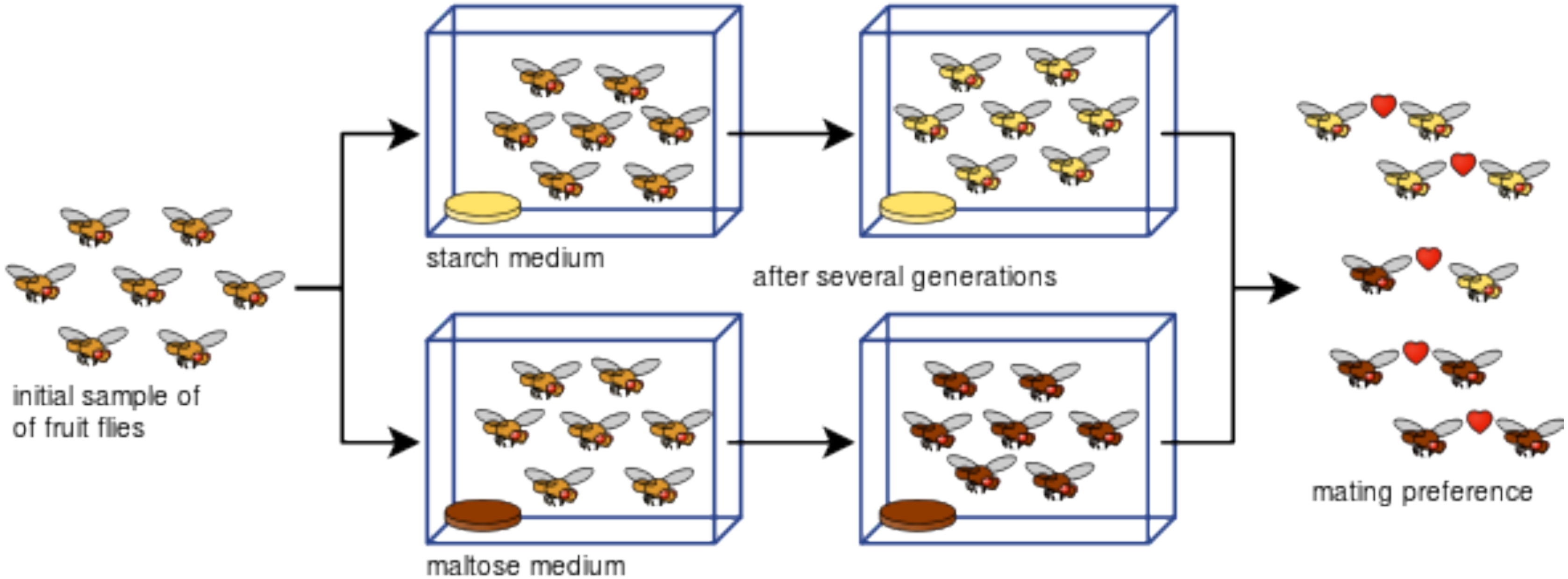
Speciation in spatially extended populations

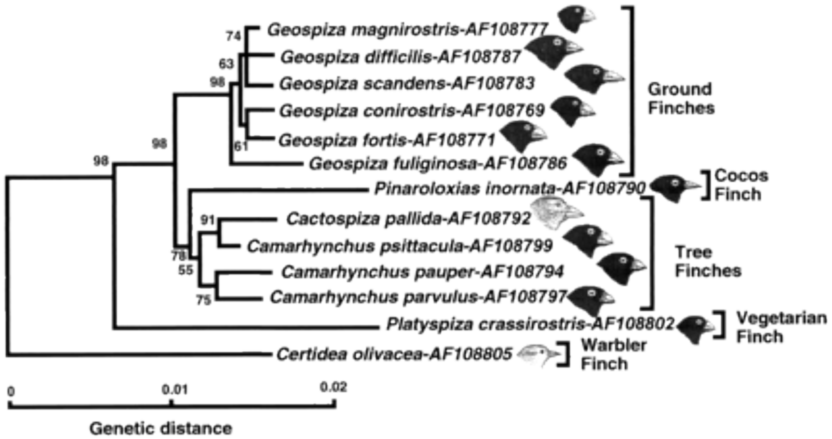
Allopatry: the basic mechanism of speciation:



Geographic isolation leads to Reproductive isolation.

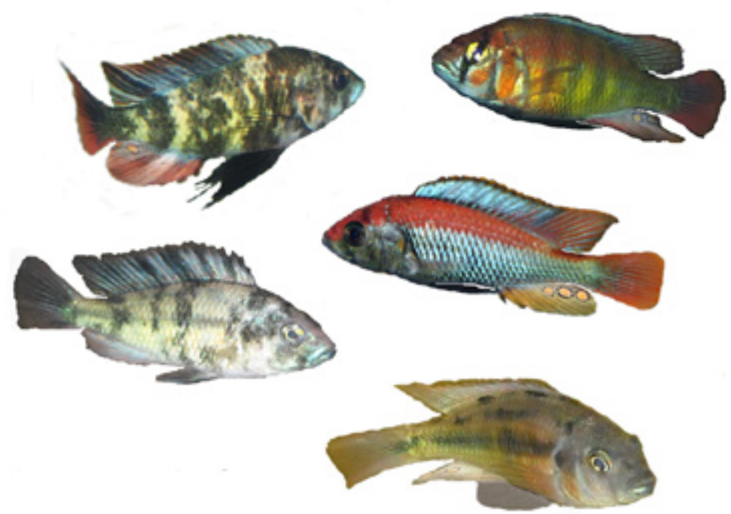
Classic example: Darwin finches on the Galápagos islands

T=ground, A=tree, V=vegetarian, C=singer



Sato et al, PNAS 1998

Another example: Cichlid fishes in lake Victoria (Tanzania, Africa)



There are approximately 400 species of cichlids with a common ancestral that lived only 14 thousand years ago.

Super fast speciation.
No geographic barriers.
No evidence of strong competition.

Geography and Genetics work together in Evolution

Dramatic examples of their intertwined roles are provides by ***RING SPECIES***,

where the population is restricted by partial geographic barriers

Greenish Warblers Birds in the Tibetan Plateau

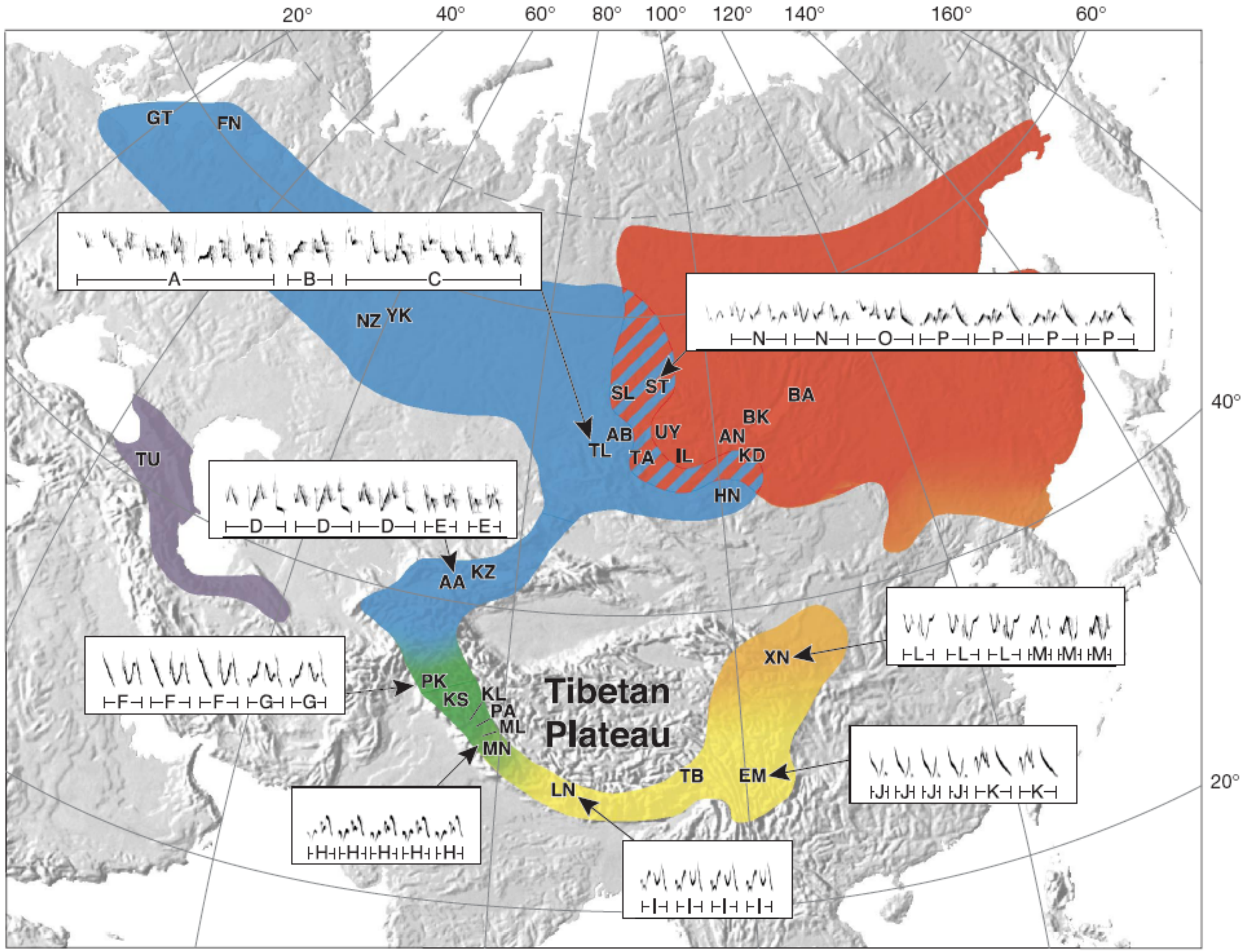
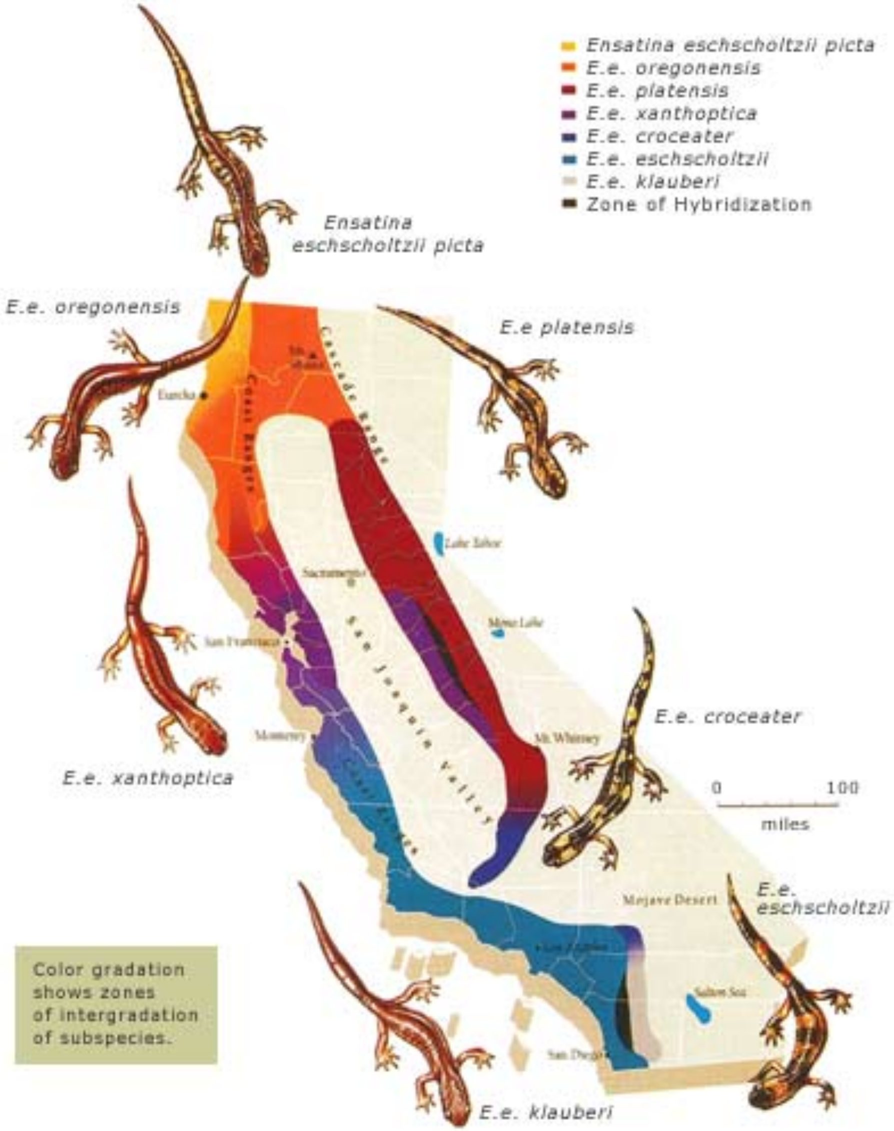
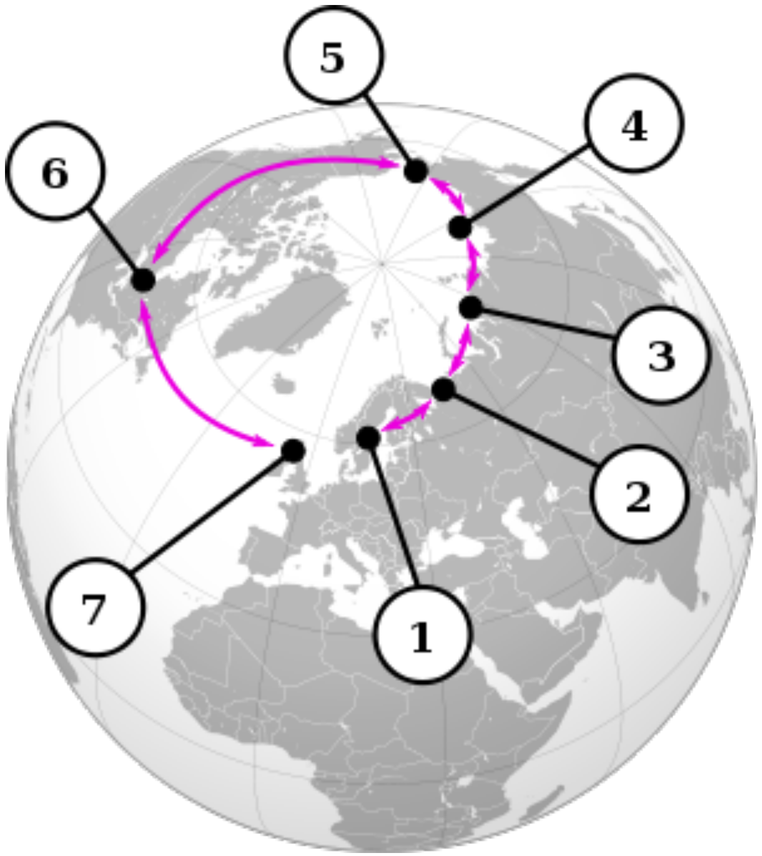


Figure 1 Geographic range of the greenish warbler species complex, along with research sites and representative song spectrograms. Different colours illustrate the ranges of six taxa commonly considered to be subspecies of *Phylloscopus trochiloides*[4]: purple, *nitidus*; blue, *viridanus*; green, *ludlowi*; yellow, *trochiloides*; orange, *obscuratus*; red, *plumbeitarsus*. Colours grade together in regions where Ticehurst[4] described gradual change between subspecies. Research sites are indicated by their two-letter designation. Also shown are representative song spectrograms (horizontal axis is time, vertical is frequency, darkness is amplitude) from eight locations[7]. Letters and brackets below the spectrograms indicate distinct song units. Song structure (for example, length of each unit, repetition of units, frequency range) differs between *viridanus* and *plumbeitarsus*, but there is a gradient in song around the southern side of the ring[7].

Salamander, Ensatina (California)



Hering Gull (polar circle)



Euphrobia Tithymaloides (plant)

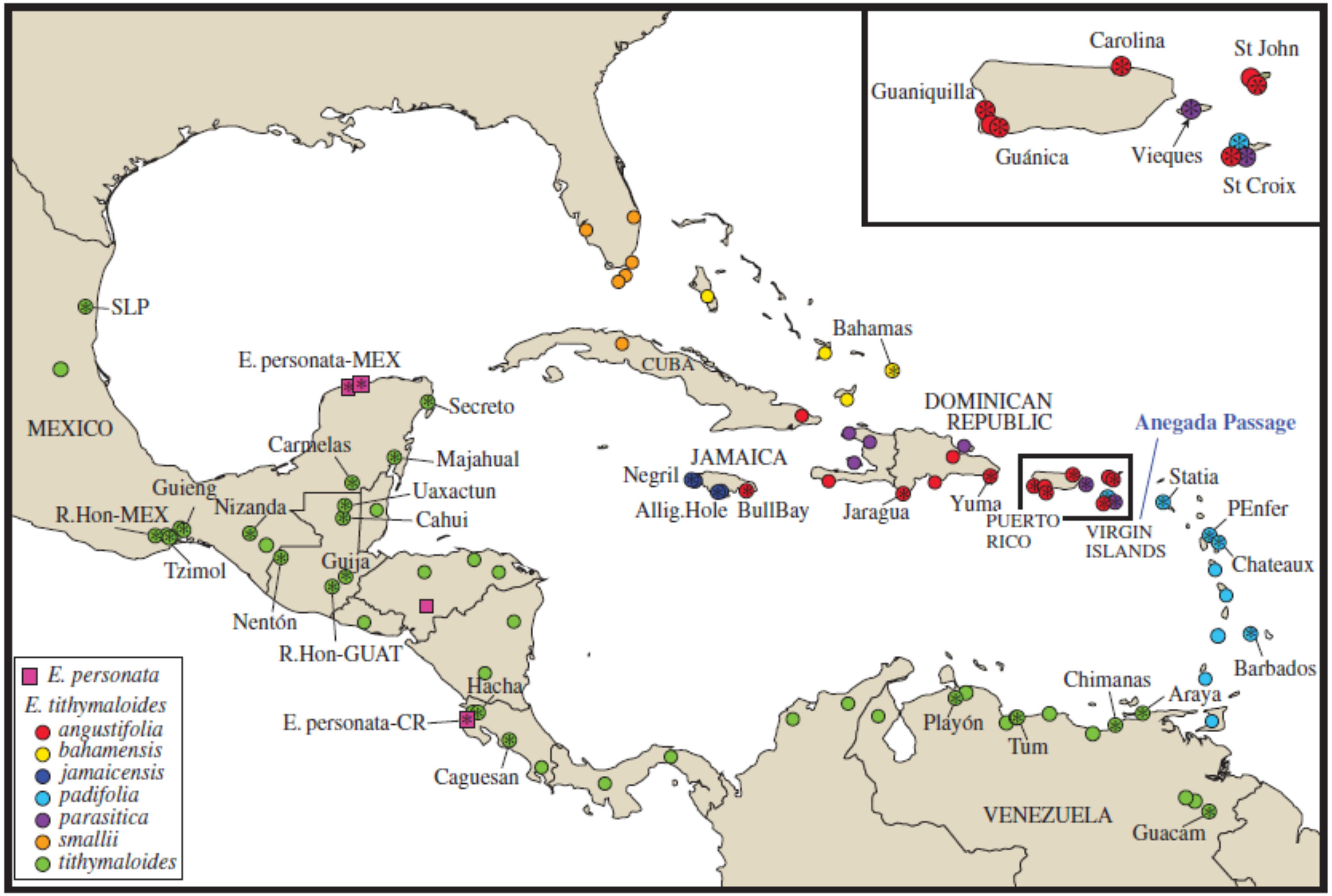


Figure 2. Documented populations of *E. tithymaloides* in the Caribbean are represented in full circles, colour-coded by subspecies. The 42 populations included in this study are marked with an asterisk and have been assigned names representative of their localities.

Trumpetfish around the world

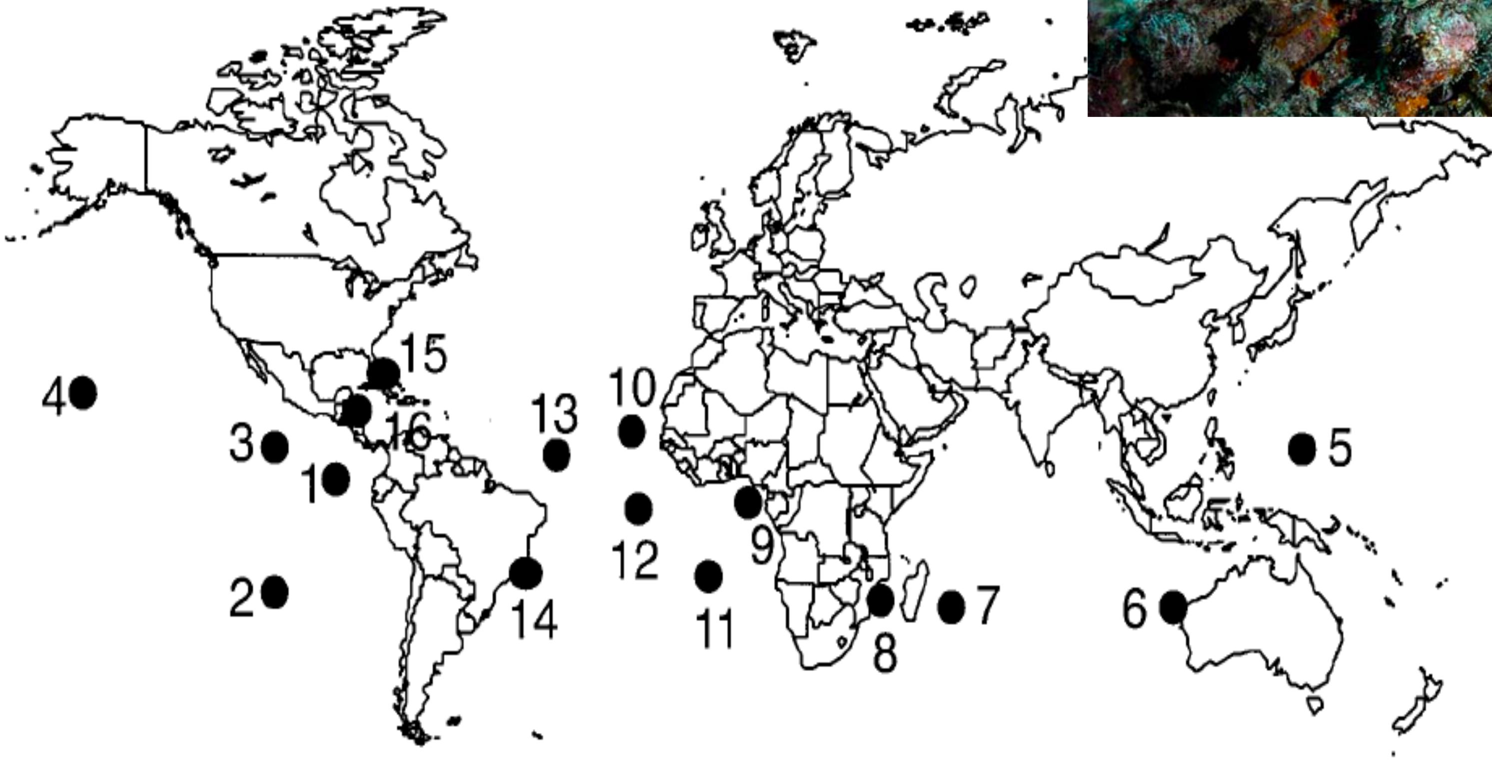


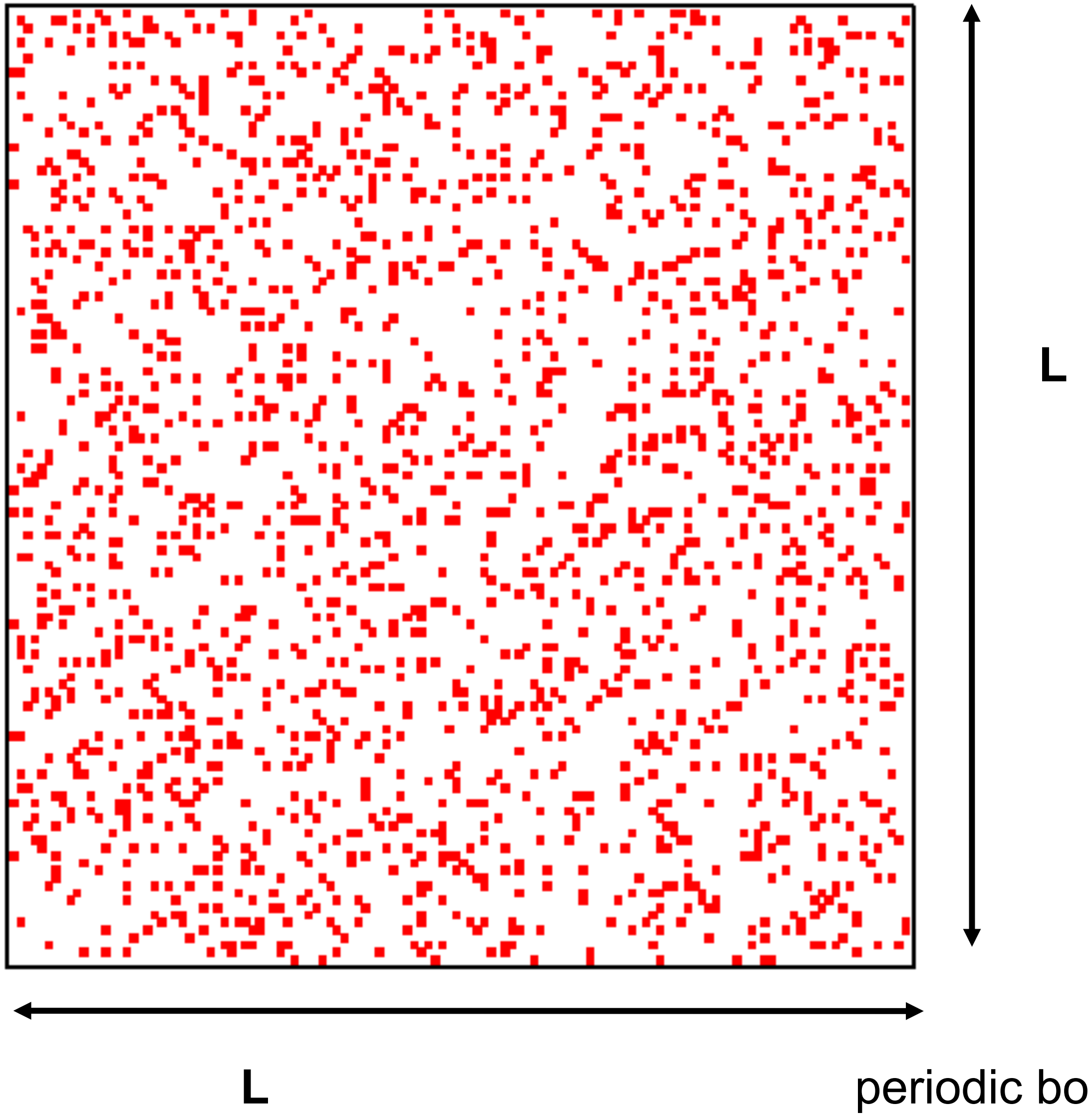
FIG. 1. The 16 sample locations for *Aulostomus* species. By conventional distributions, *A. maculatus* inhabits the West Atlantic, *A. strigosus* inhabits the East Atlantic, and *A. chinensis* inhabits the Indian and Pacific Oceans. *Aulostomus chinensis* was collected from: (1) Isla del Coco (Pacific Costa Rica, $n = 9$); (2) Easter Island ($n = 5$); (3) Clipperton Island ($n = 5$); (4) Hawaii ($n = 15$); (5) Guam ($n = 18$); (6) Ningaloo Reef (West Australia, $n = 7$); (7) Reunion Island ($n = 7$); and (8) Punta del Orro (Mozambique, $n = 2$). *Aulostomus strigosus* was collected from: (9) São Tome (Gulf of Guinea, West Africa, $n = 16$); (10) Cape Verde ($n = 16$); (11) St. Helena ($n = 19$); (12) Ascension Island ($n = 19$); and, based on a revision of species distributions (13) St. Paul's Rocks (Brazil, $n = 14$) and (14) Espirito Santo (Brazil, $n = 22$). *Aulostomus maculatus* was collected from: (15) Florida Keys (U.S.A., $n = 15$); and (16) Belize Barrier Reef ($n = 7$).

Summary

- Speciation without barriers
- Partial barriers and Ring species
- Future directions

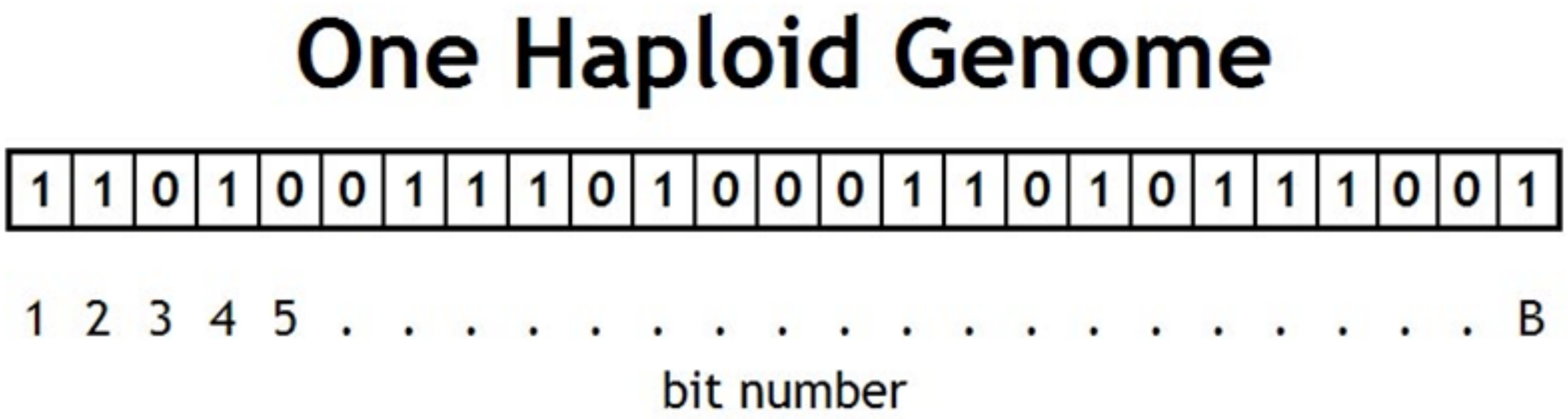
1. Speciation without barriers

a) finite population with initially identical individuals randomly distributed in space



periodic boundary conditions

b) haploid and hemafroditic individuals with genome represented by a sequence of B loci with two alleles



(c) sexual reproduction occurs only if:

- individuals are not too different genetically (**G**)
- individuals are not too distant from each other (**S**)

Genetic distance:

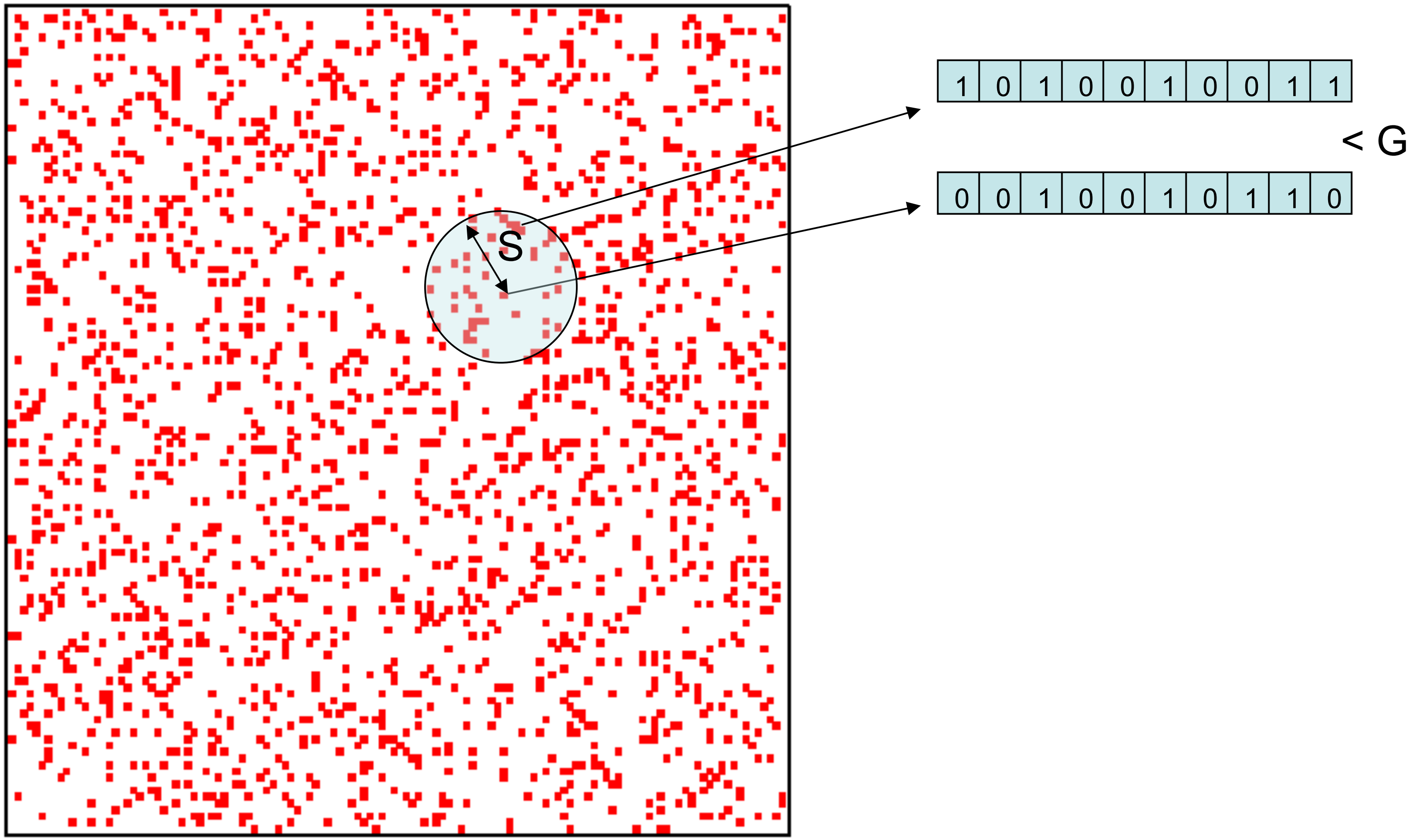
	1	1	0	1	0	0	1	1	1	0	1	0	0	0	1	1	0	1	0	1	1	1	0	0	1
-	1	0	0	0	1	0	1	0	0	0	1	0	1	1	1	1	0	0	0	0	1	0	0	1	0

…

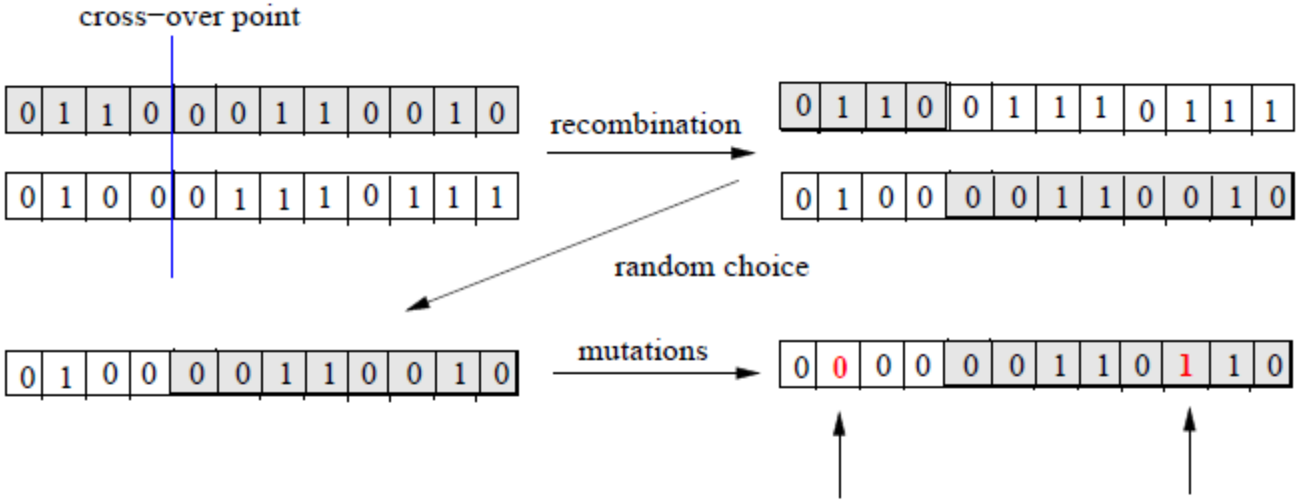
= 0 + 1 + 0 + ... + 0 + 1 + 1 = 12

(c) sexual reproduction with recombination and mutation occurs if:

- individuals are not too different genetically (**G**)
- individuals are not too distant from each other (**S**)



Recombination and Mutation
(diploid phase)



S = spatial restriction on mating

G = genetic restriction on mating

μ= mutation rate

(d) The total number of individuals in the population is kept constant.

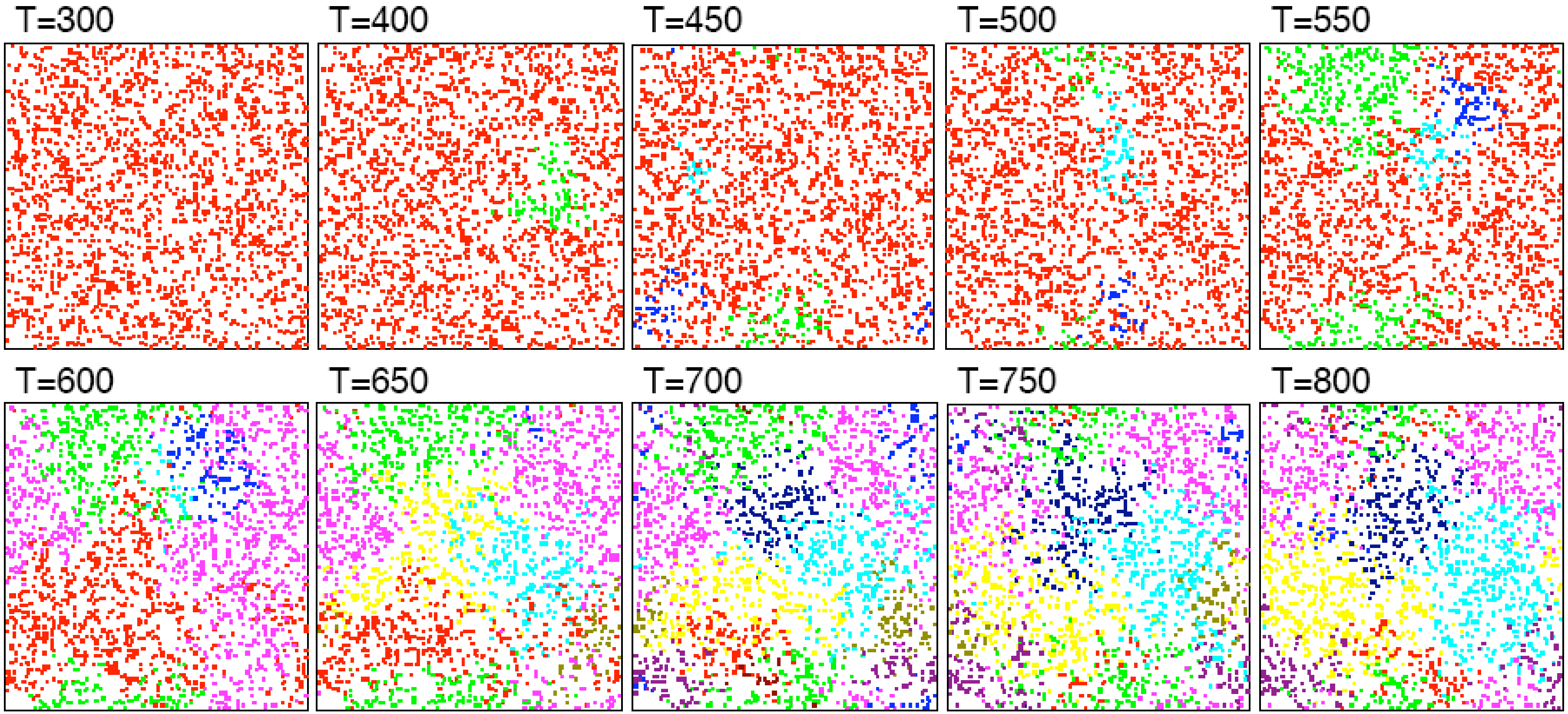
(e) Initial condition: population of genetically identical individuals

0	0	0	0	0	0	0	0	0	0

(f) at each generation every individual tries to reproduce, but there is a probability **Q** that it will not make it. In this case a neighbor is chosen to reproduce in its place. The total number of individuals in the population is kept constant.

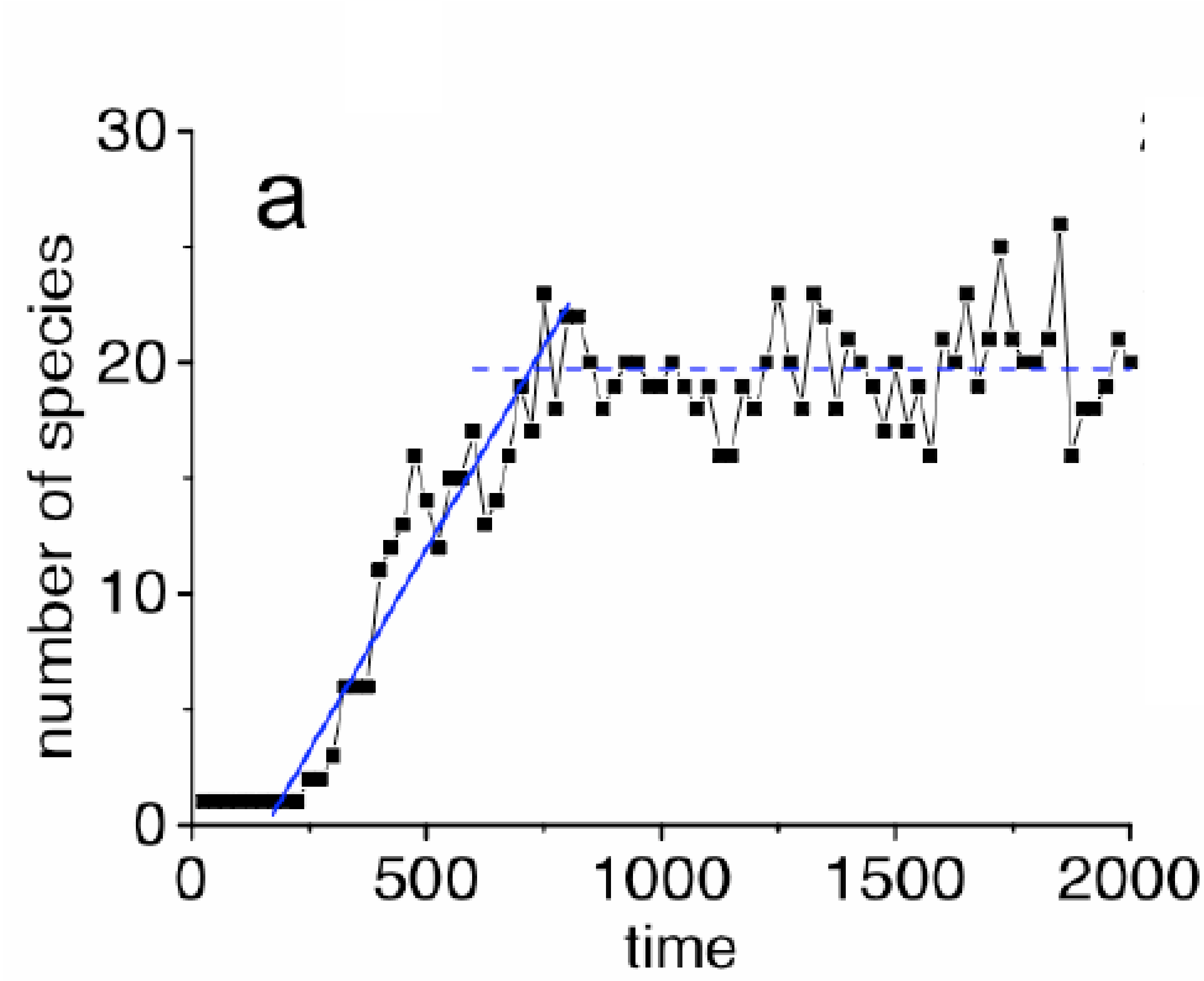
(g) Offspring can be placed exactly at position of the dying individual with probability **(1-D)** or somewhere nearby with probability **D**.

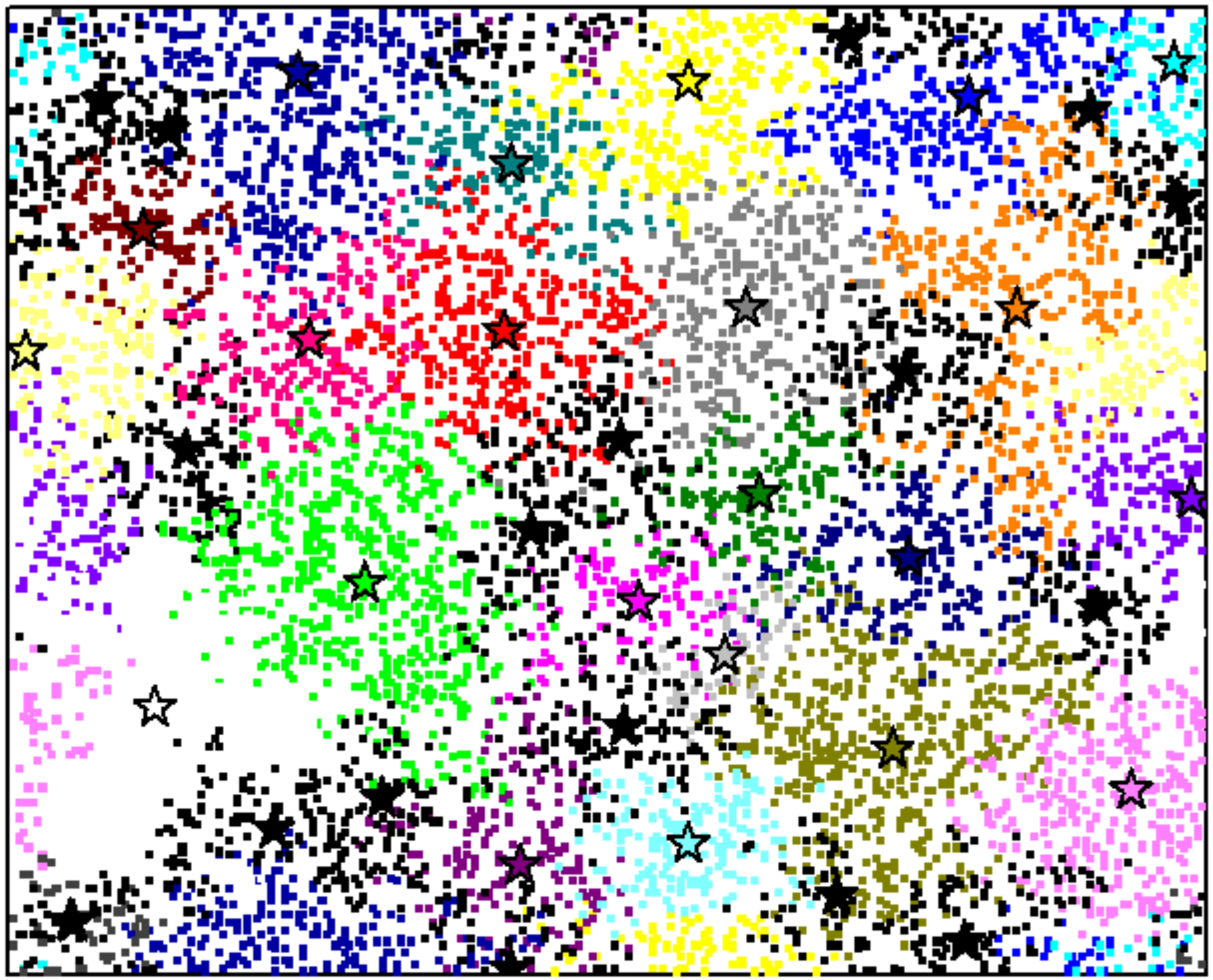
Time Evolution Measured in Number of Generations



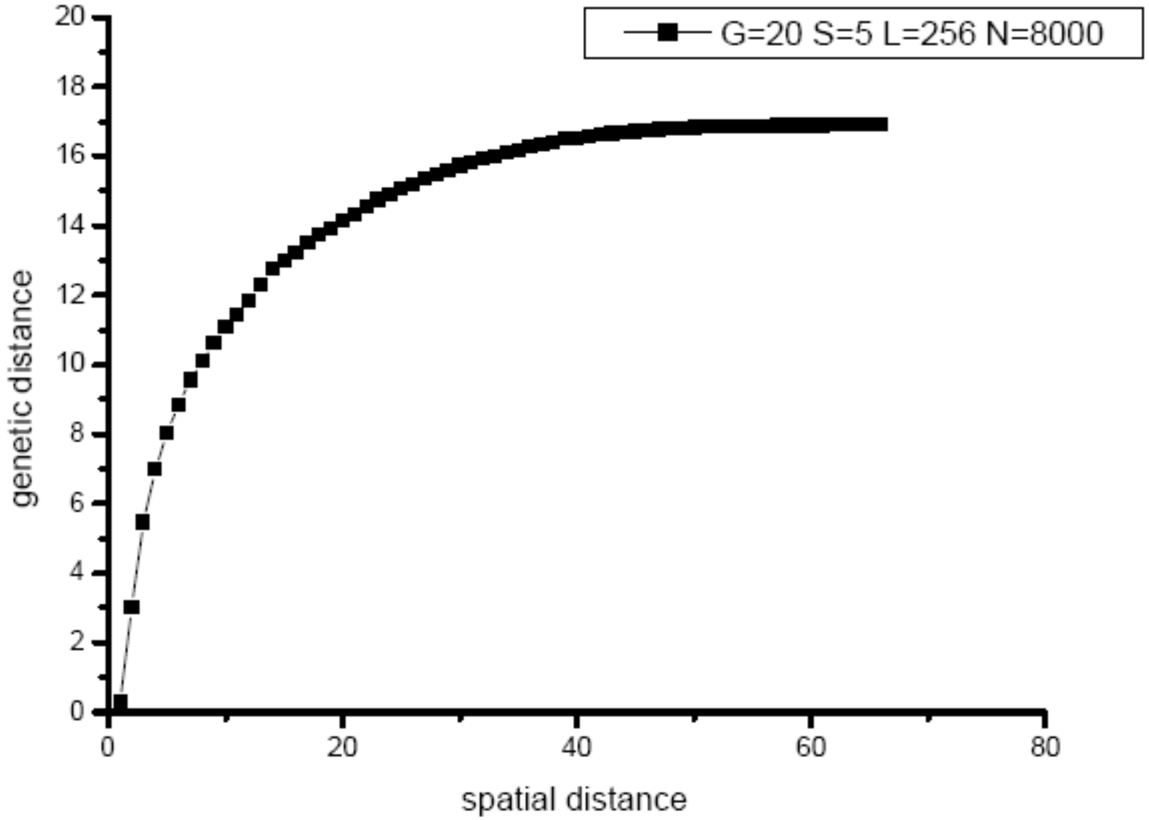
S=6 G=20 L=128 N=2000
Q=0.3 μ=0.001 D=0.01

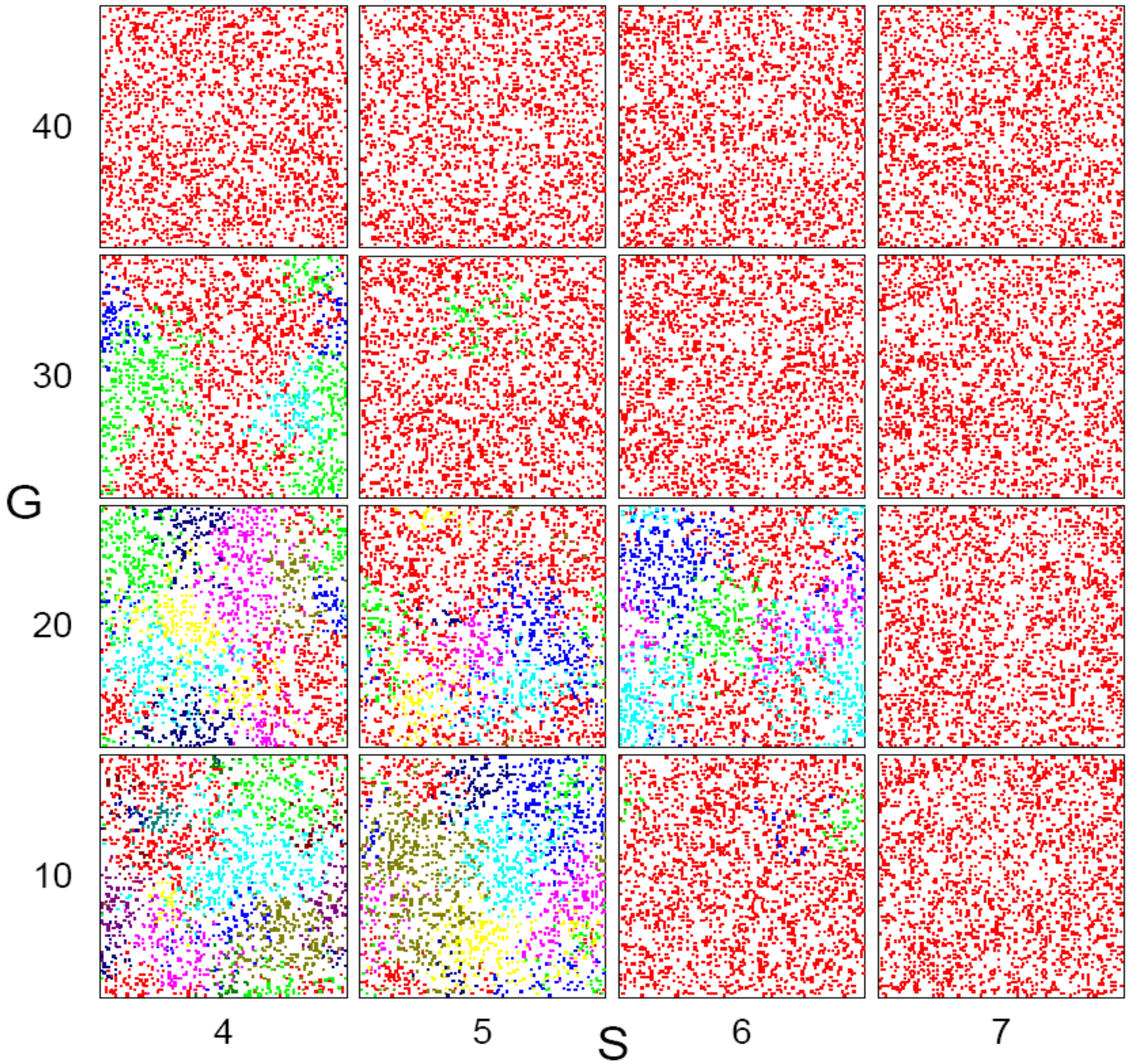
Number of Species as a function of time





Genetic distance from individuals of the same species
as measured from the center of population

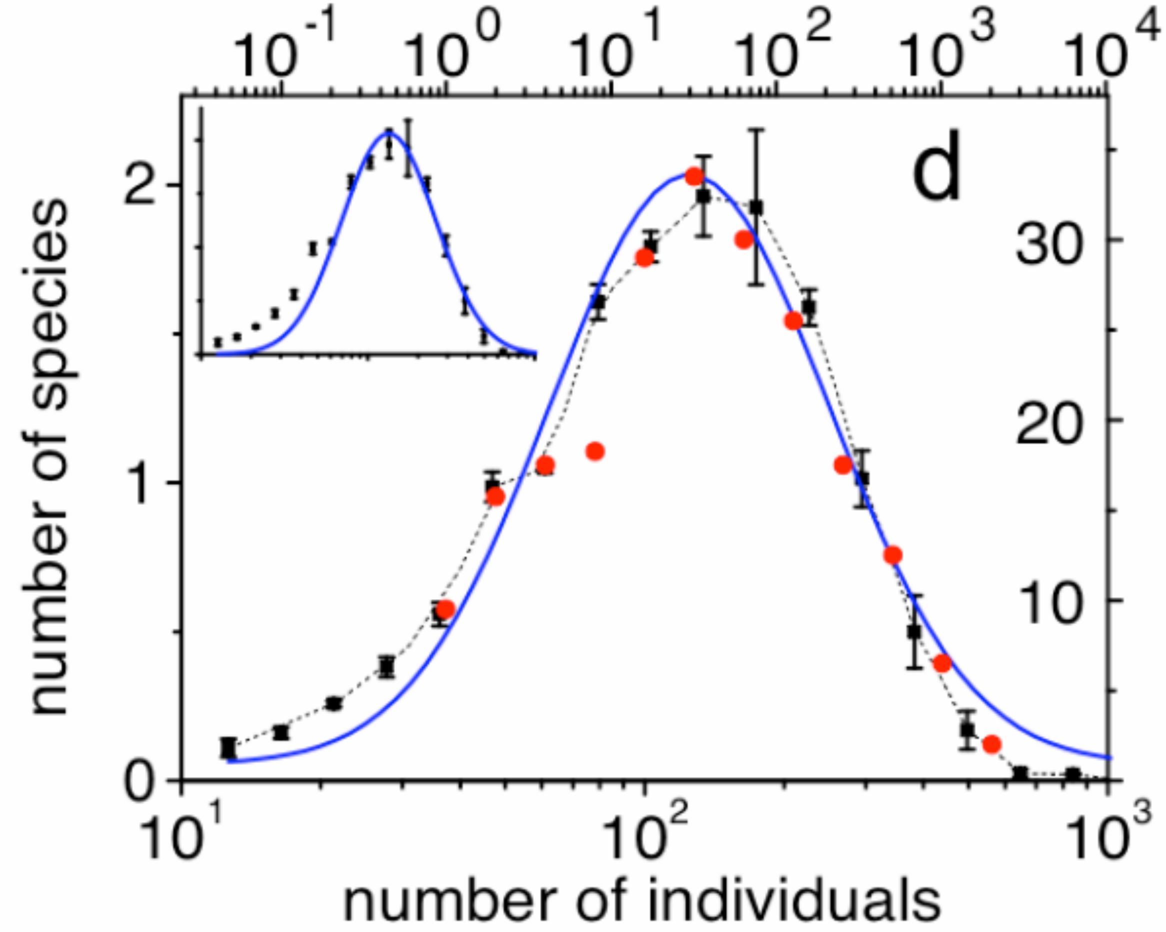


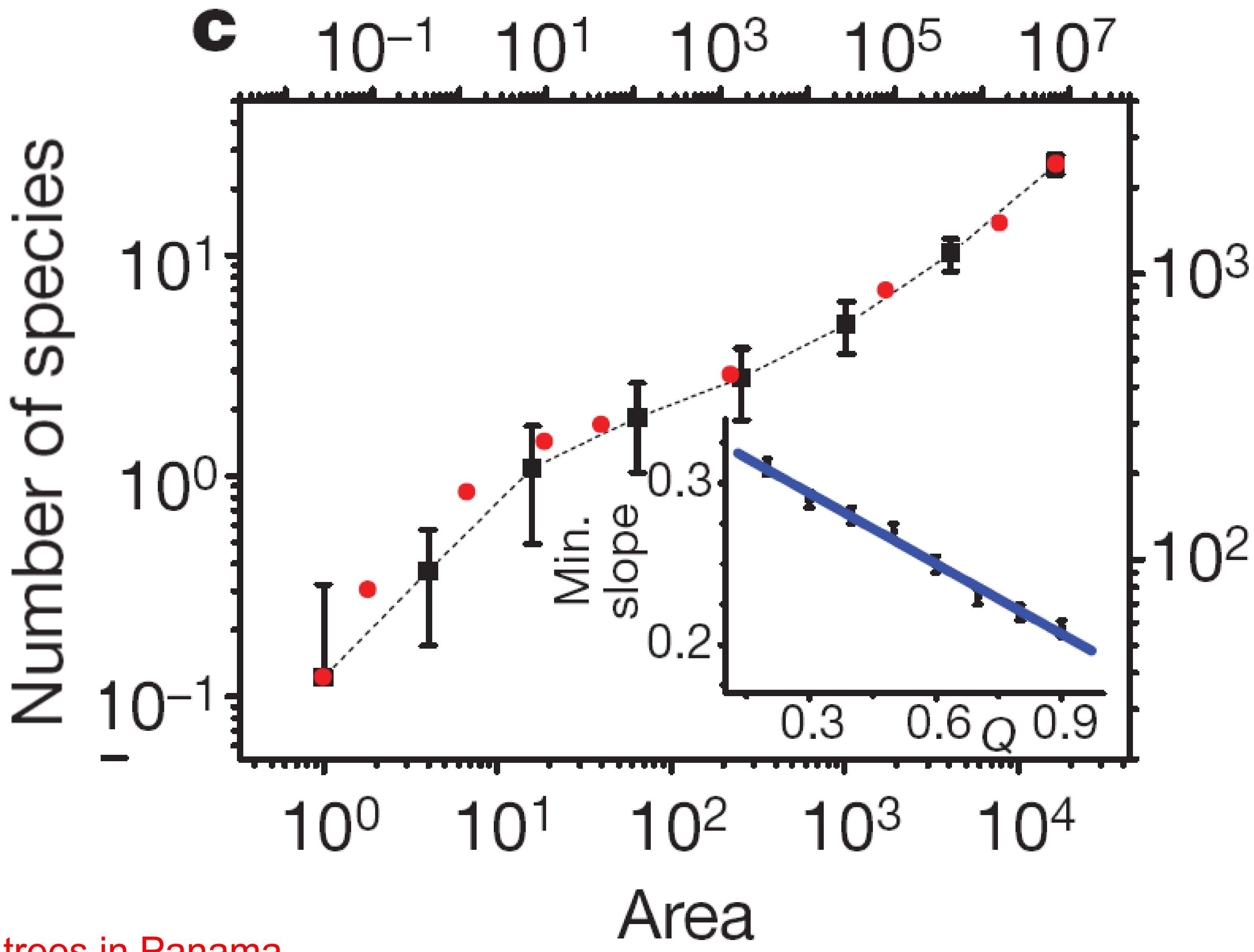

40
30
20
10
G
4
5
6
7
S

Trees in Panama Tropical Forest (BCI)

red circles
trees (BCI)

black points
G=20 S=5





red - trees in Panama

2. Partial barriers and Ring Species

Ayana B. Martins IB-USP
Yaneer Bar-Yam – NECSI – Boston
Marcus A.M. de Aguiar

Greenish Warblers Birds in the Tibetan Plateau

Irwin et al
Science (2005)

Genetic distances and geographic distances are strongly correlated.

We conclude that there is no break in gene flow through the ring of populations.

Genetic continuum seems to be unstable and ring might soon break down into separate species.

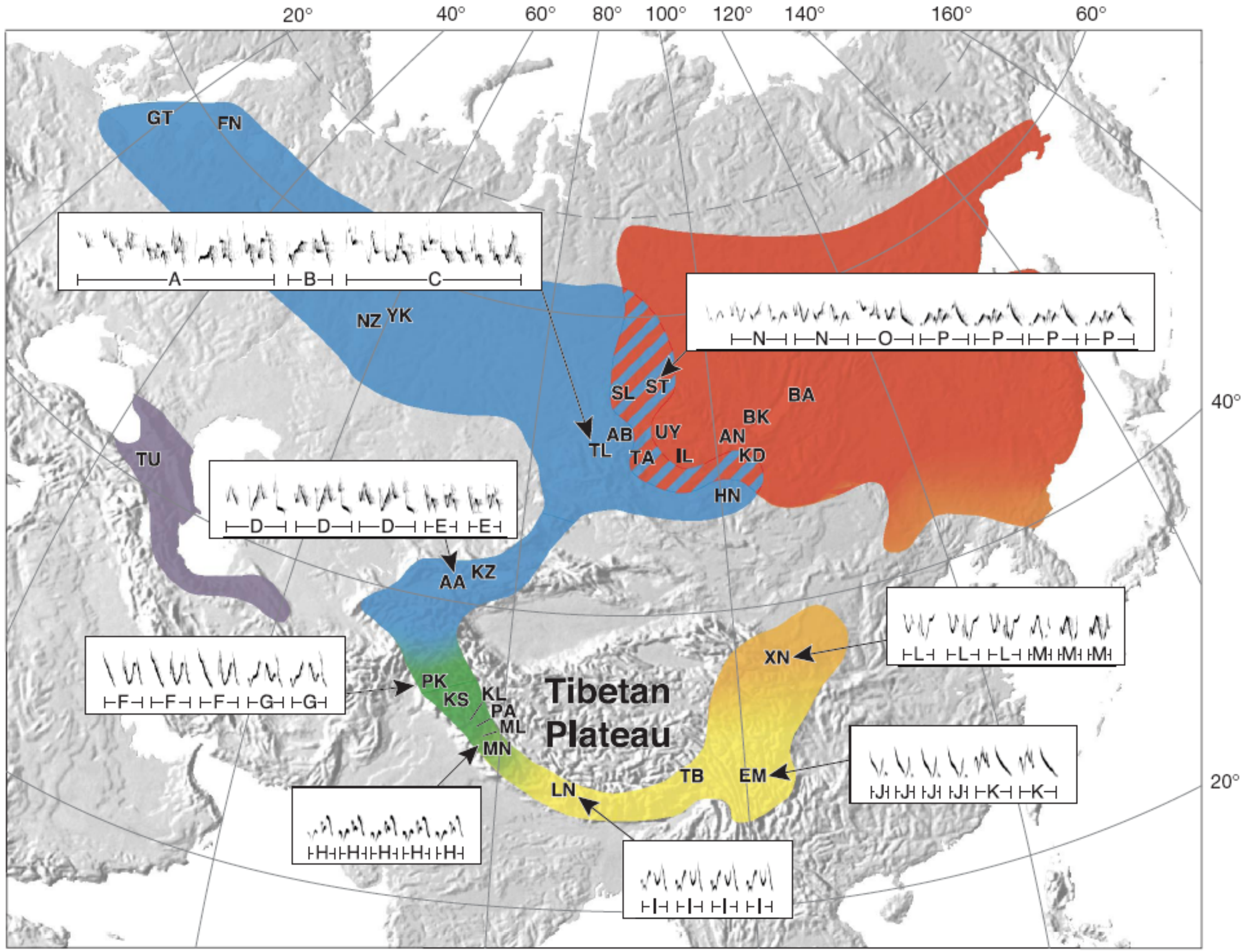
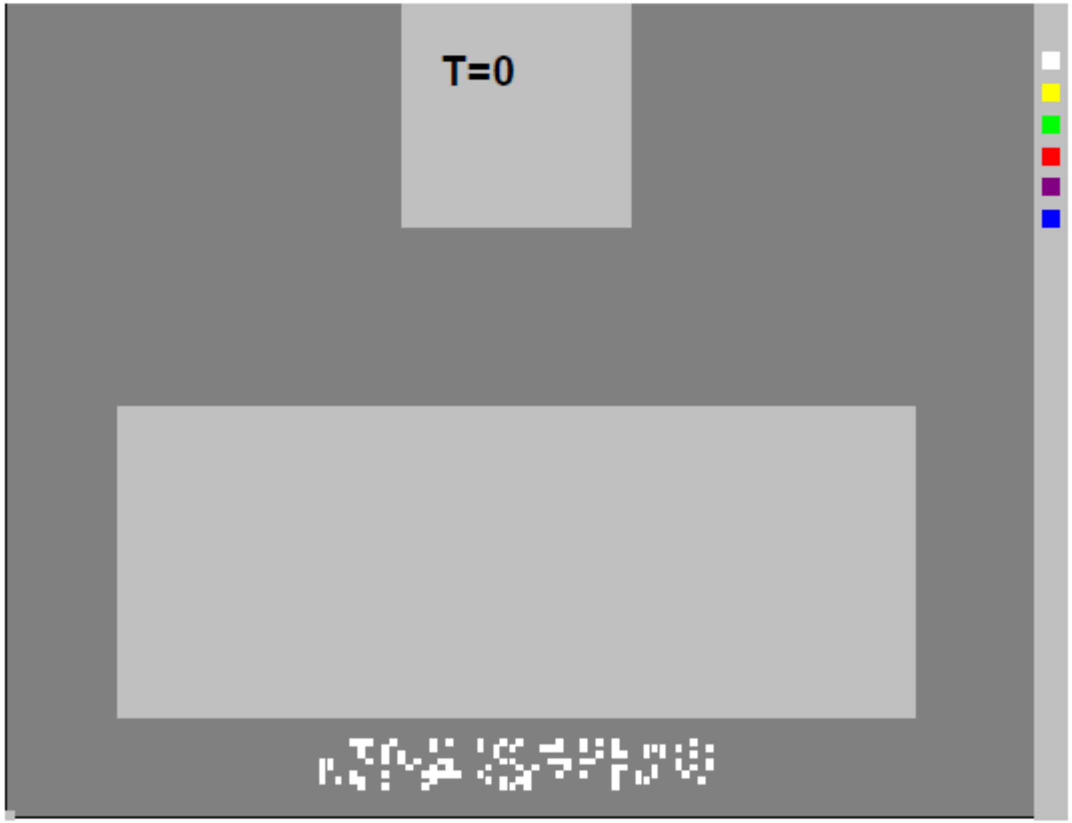
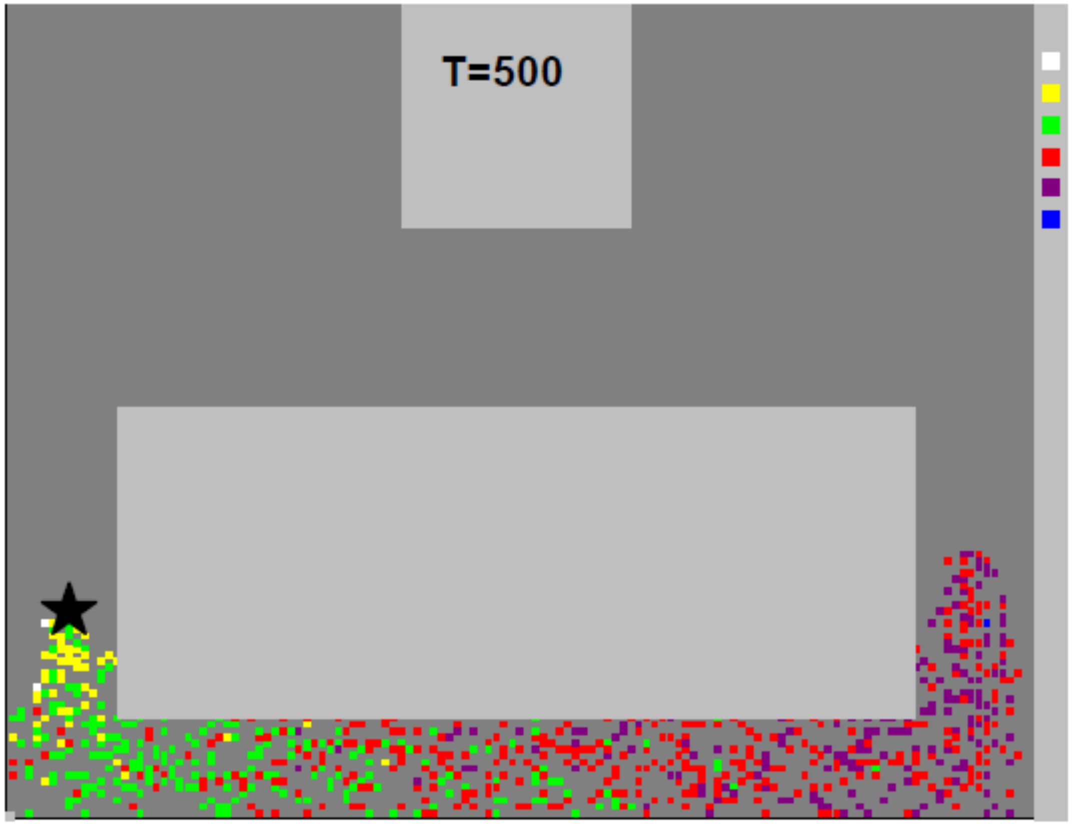
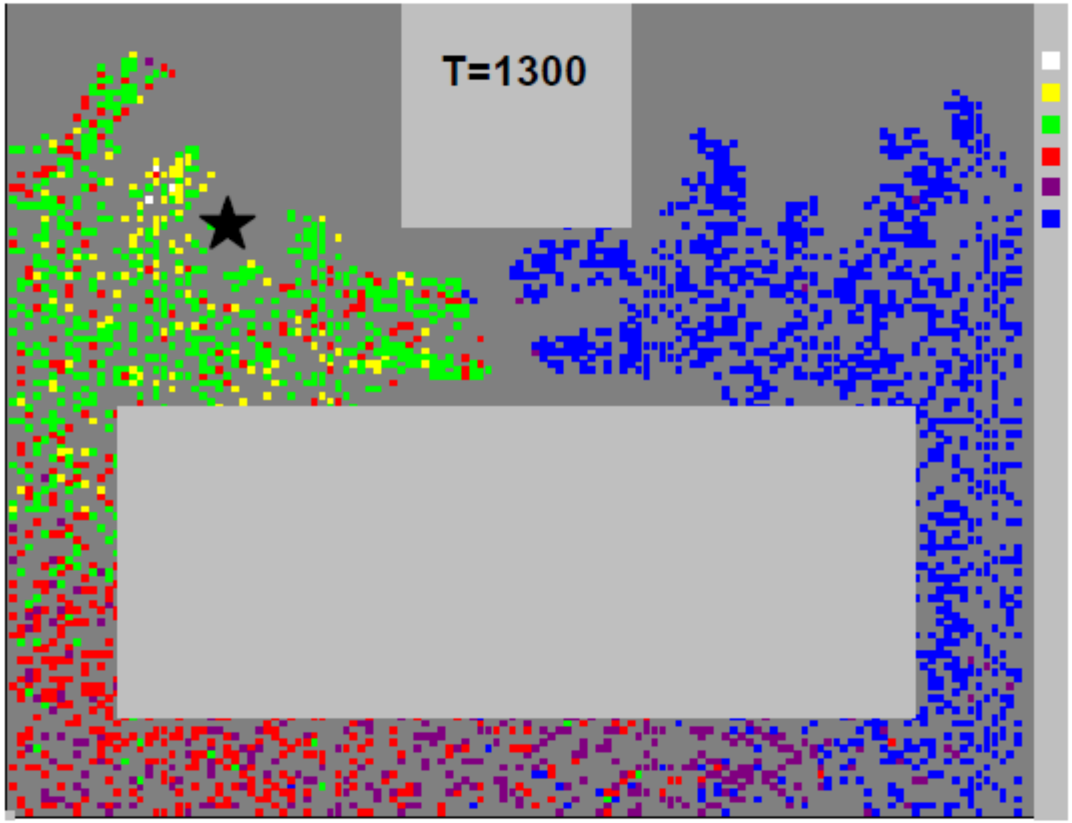
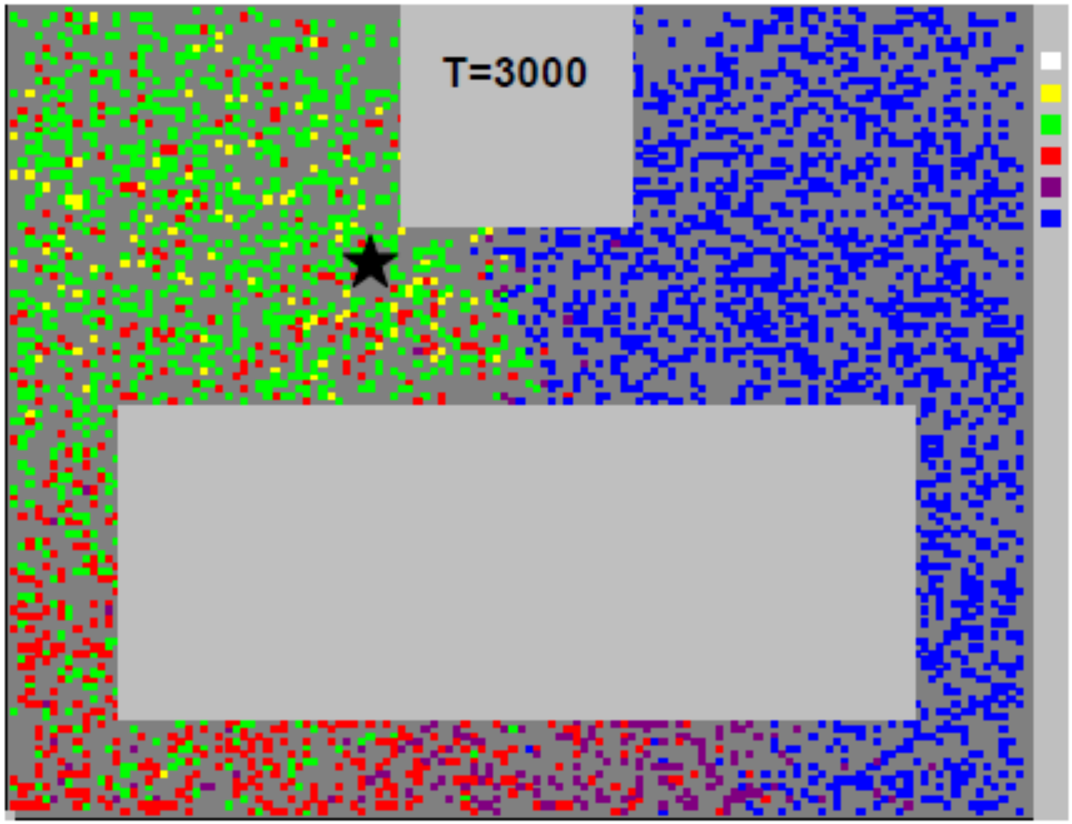


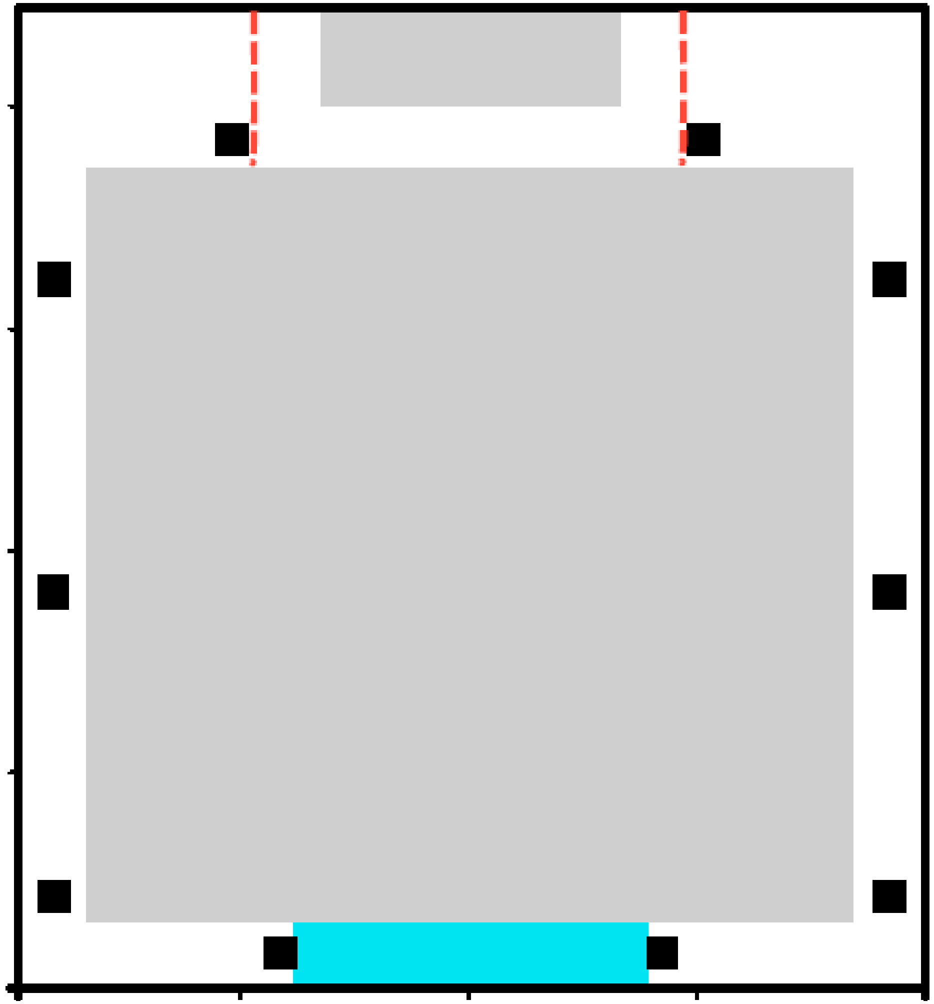
Figure 1 Geographic range of the greenish warbler species complex, along with research sites and representative song spectrograms. Different colours illustrate the ranges of six taxa commonly considered to be subspecies of *Phylloscopus trochiloides*[4]: purple, *nitidus*; blue, *viridanus*; green, *ludlowi*; yellow, *trochiloides*; orange, *obscuratus*; red, *plumbeitarsus*. Colours grade together in regions where Ticehurst[4] described gradual change between subspecies. Research sites are indicated by their two-letter designation. Also shown are representative song spectrograms (horizontal axis is time, vertical is frequency, darkness is amplitude) from eight locations[7]. Letters and brackets below the spectrograms indicate distinct song units. Song structure (for example, length of each unit, repetition of units, frequency range) differs between *viridanus* and *plumbeitarsus*, but there is a gradient in song around the southern side of the ring[7].

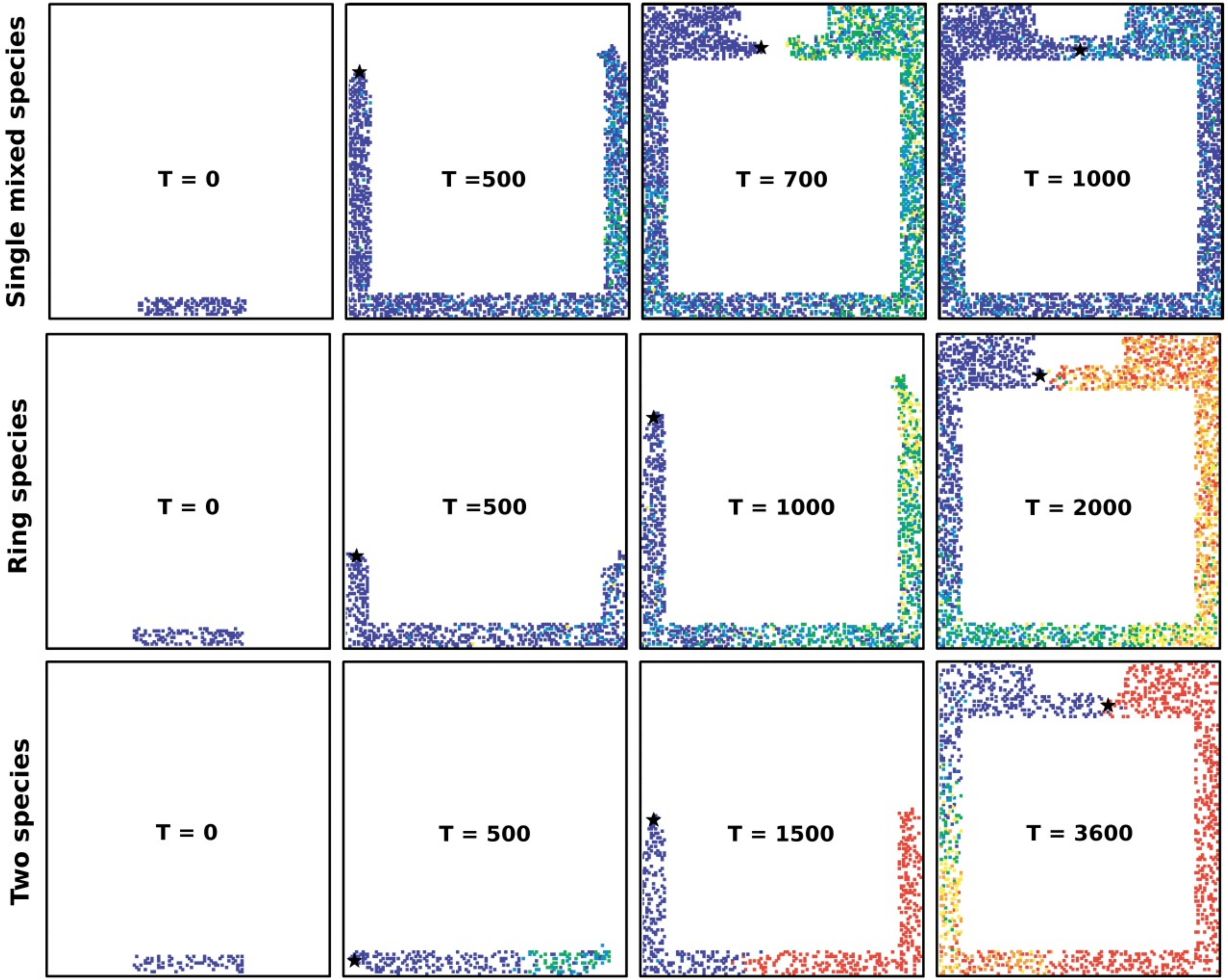

T=0


T=500

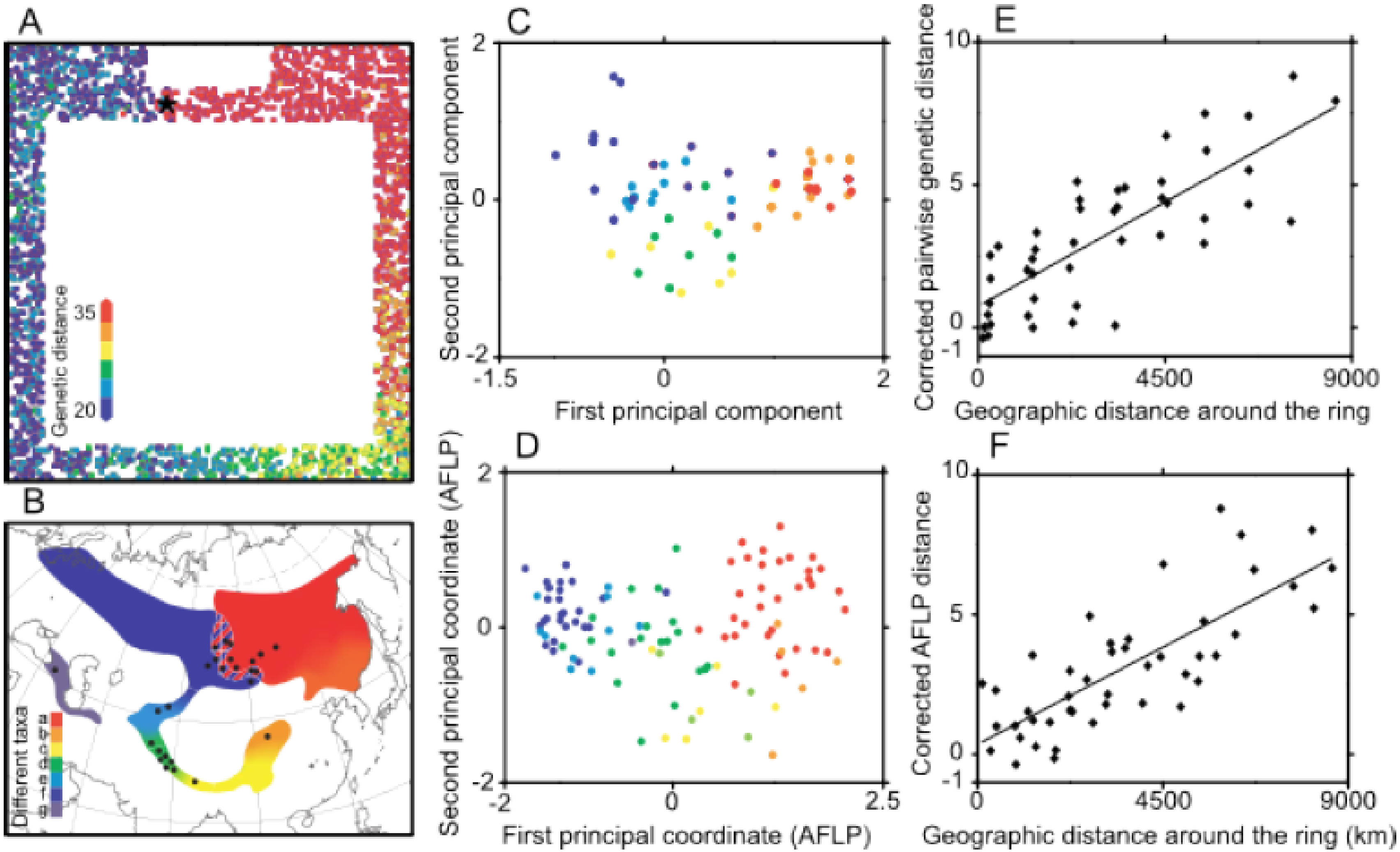

T=1300

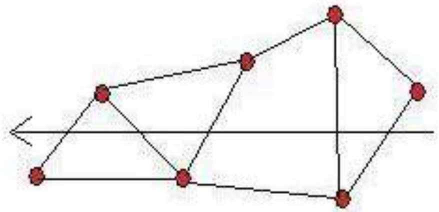

T=3000



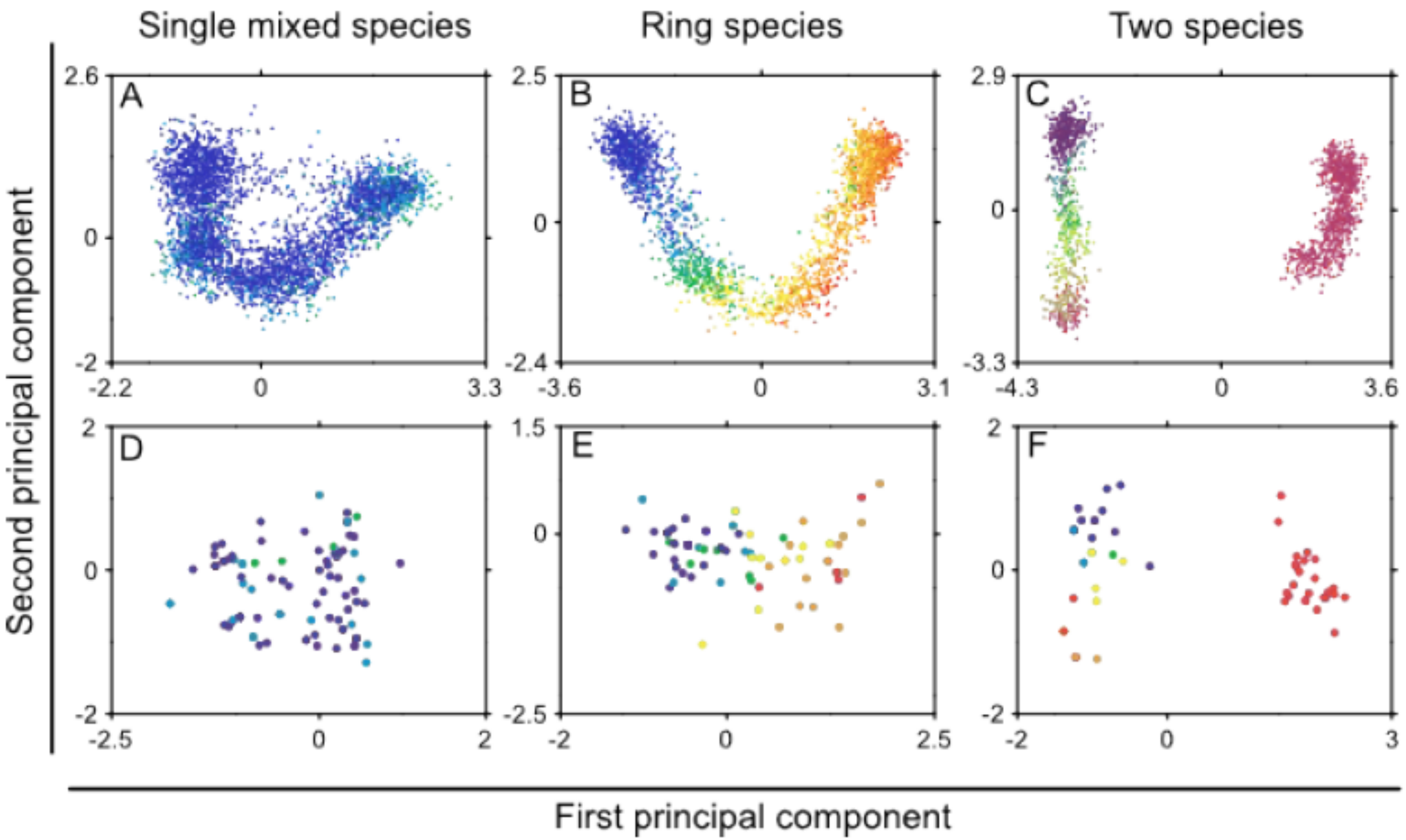


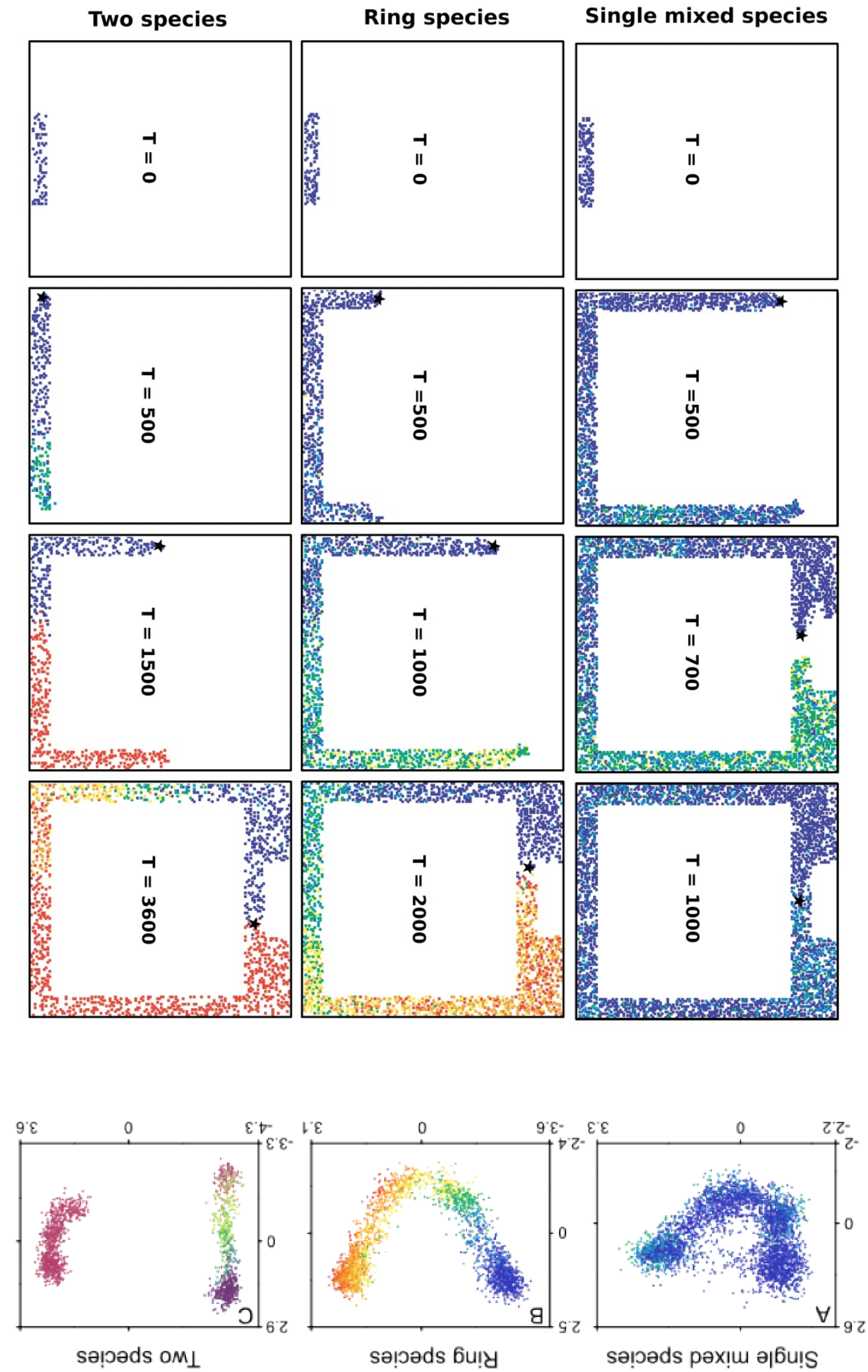
Single mixed species
Ring species
Two species
T = 0
T =500
T = 700
T = 1000
T = 0
T =500
T = 1000
T = 2000
T = 0
T = 500
T = 1500
T = 3600


A
B
C
D
E
F
Genetic distance
35
20
Different taxa
a
b
c
d
e
f
g
Second principal component
First principal component
2
0
-2
-1.5
0
2
Corrected pairwise genetic distance
Geographic distance around the ring
10
5
0
-1
0
4500
9000
Second principal coordinate (AFLP)
First principal coordinate (AFLP)
2
0
-2
-2
0
2.5
Corrected AFLP distance
Geographic distance around the ring (km)
10
5
0
-1
0
4500
9000



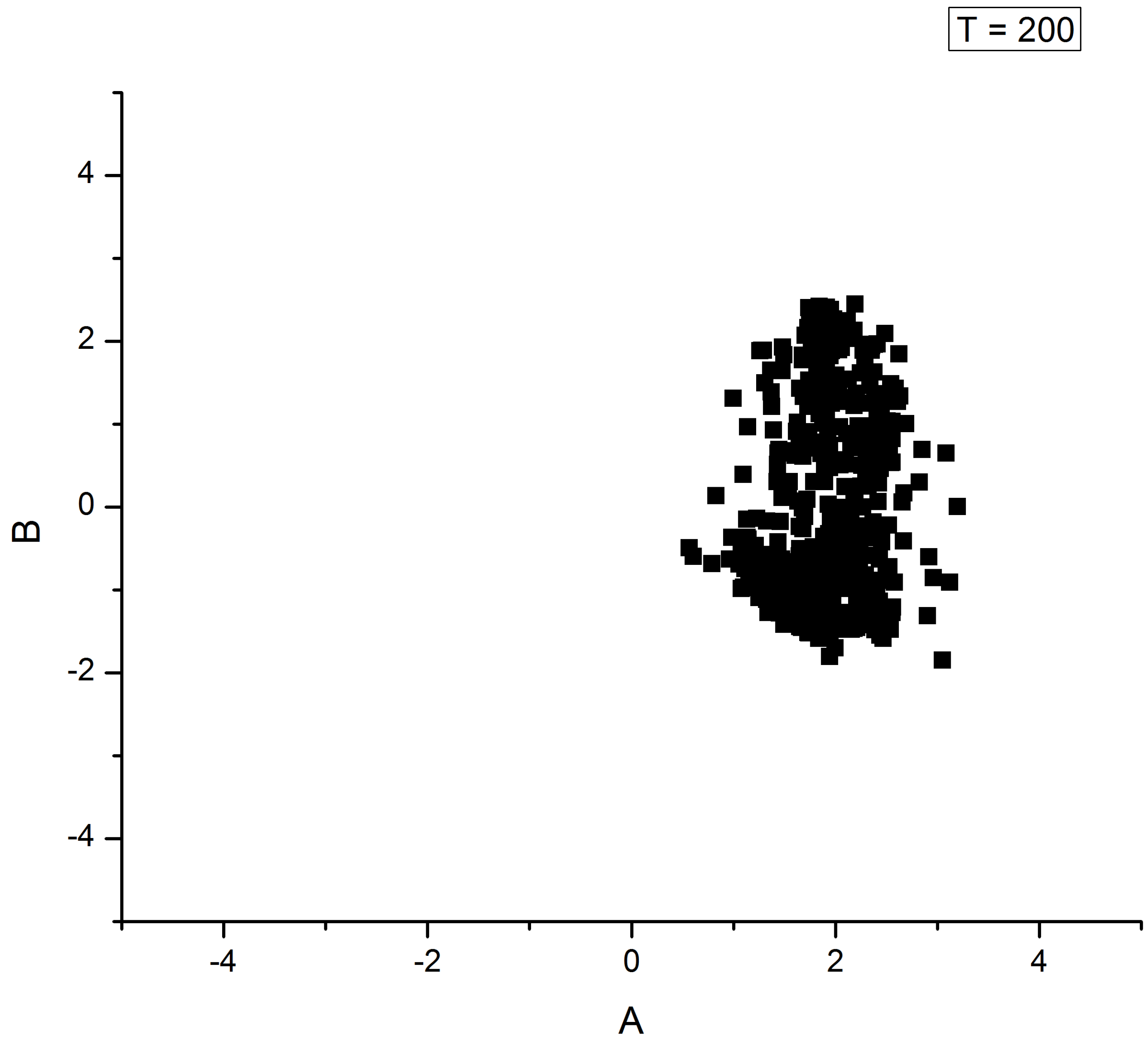
Principal Component Analysis

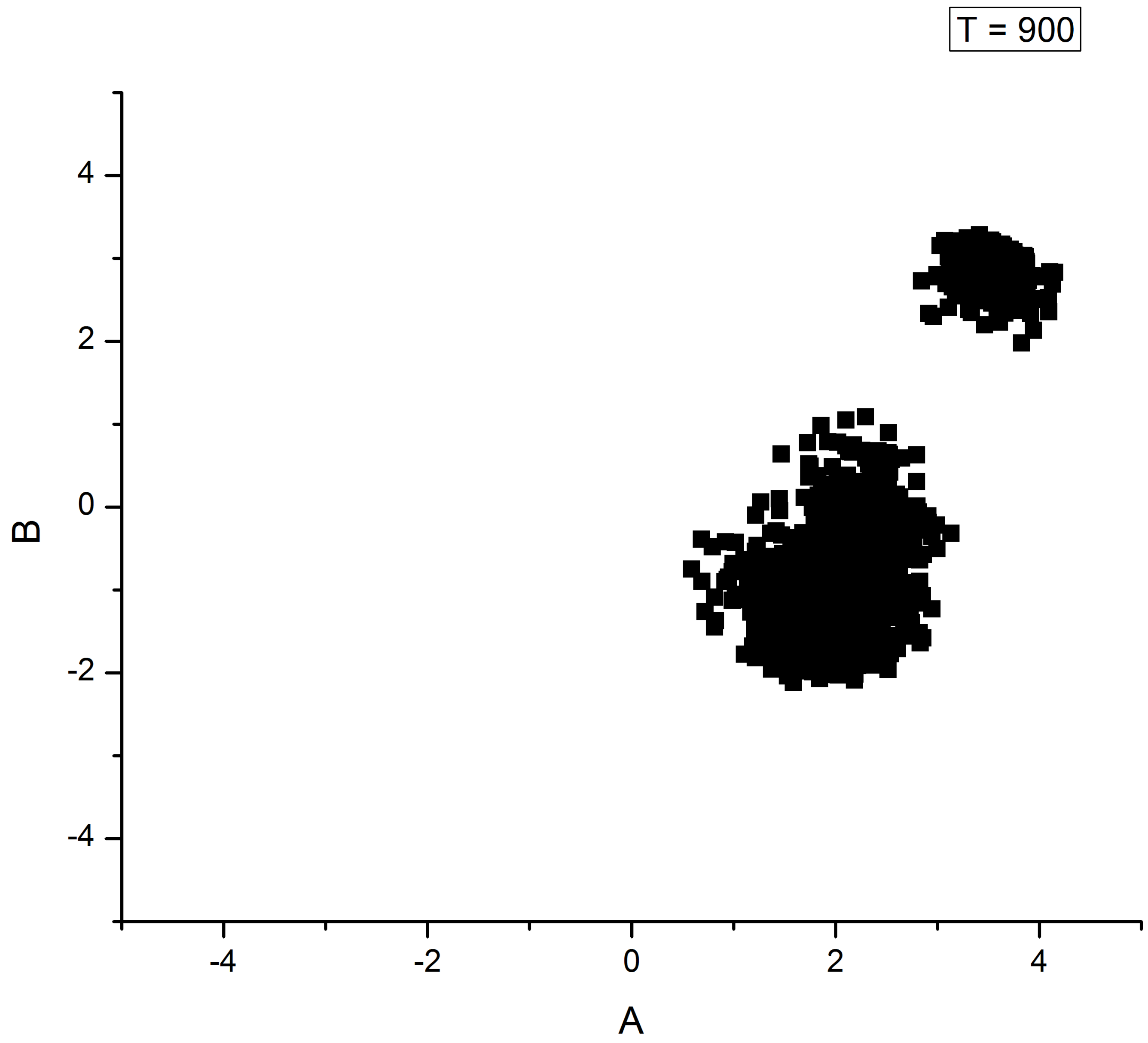


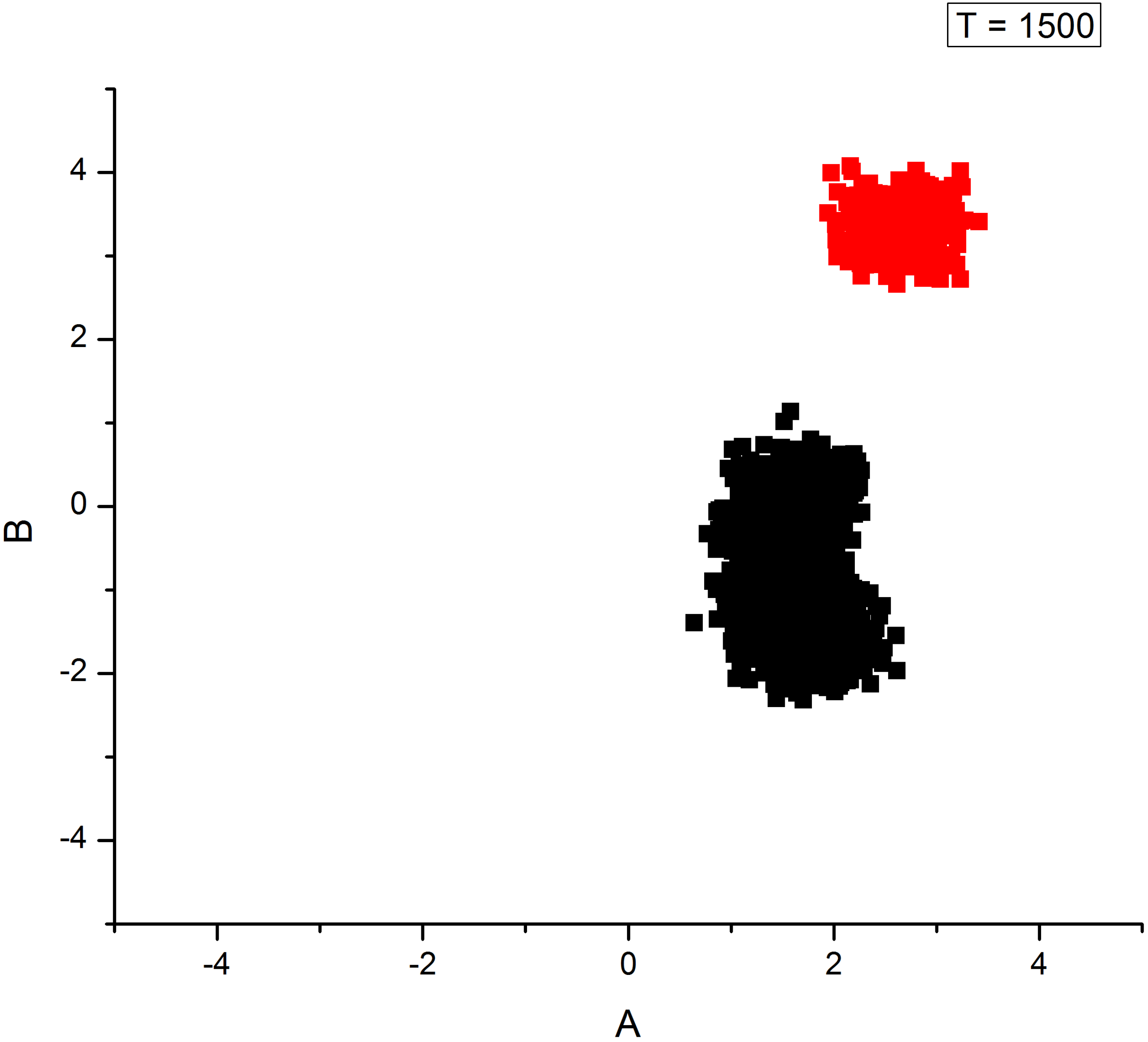

Single mixed species
Ring species
Two species
T = 0
T =500
T = 700
T = 1000
T = 0
T =500
T = 1000
T = 2000
T = 0
T = 500
T = 1500
T = 3600
A
B
C

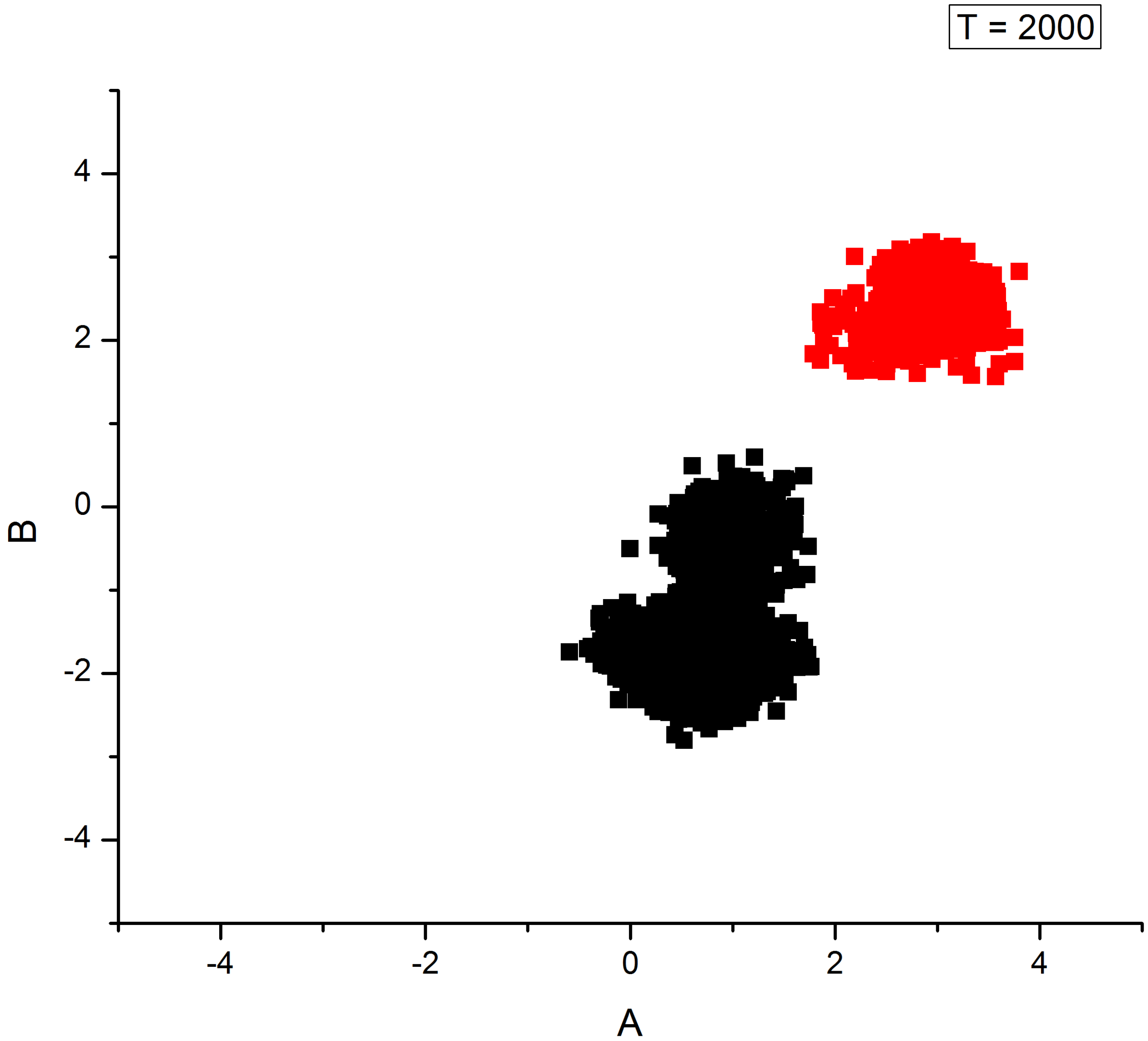
Darwin’s tree:

a PCA representation

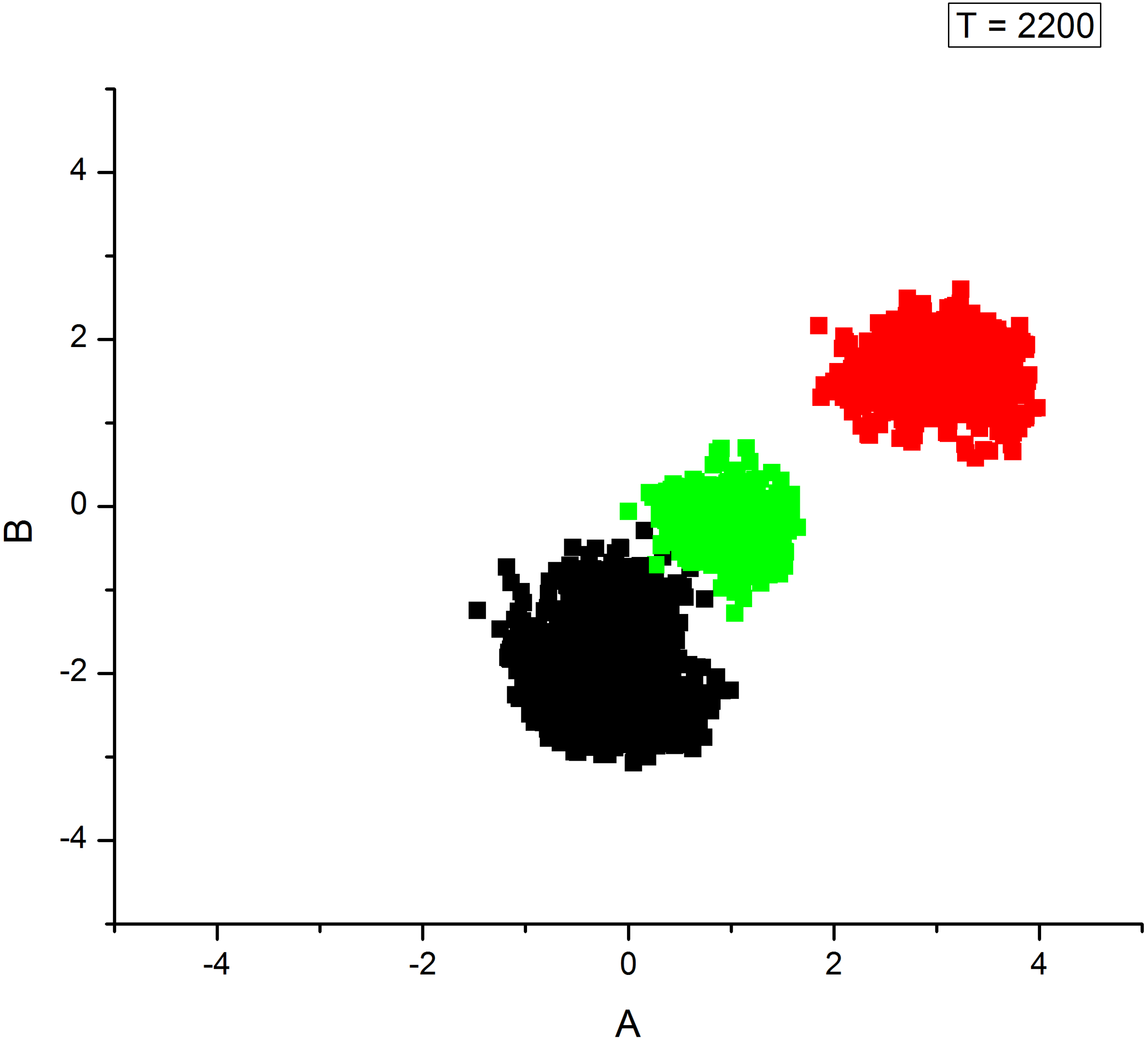

T = 200
4
2
0
-2
-4
B
-4
-2
0
2
4
A

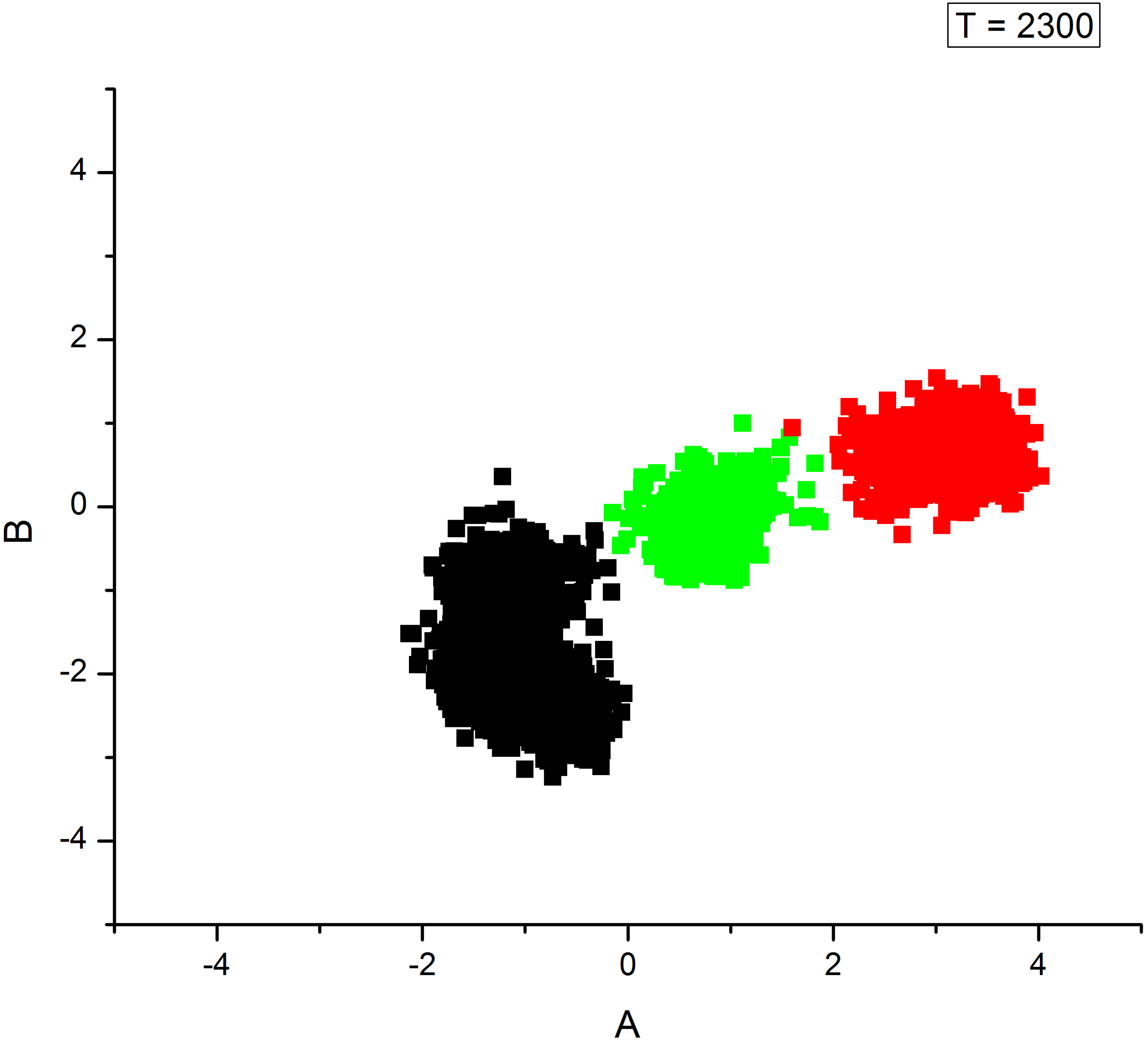

T = 900
4
2
0
-2
-4
B
-4
-2
0
2
4
A


T = 1500
4
2
0
-2
-4
B
-4
-2
0
2
4
A

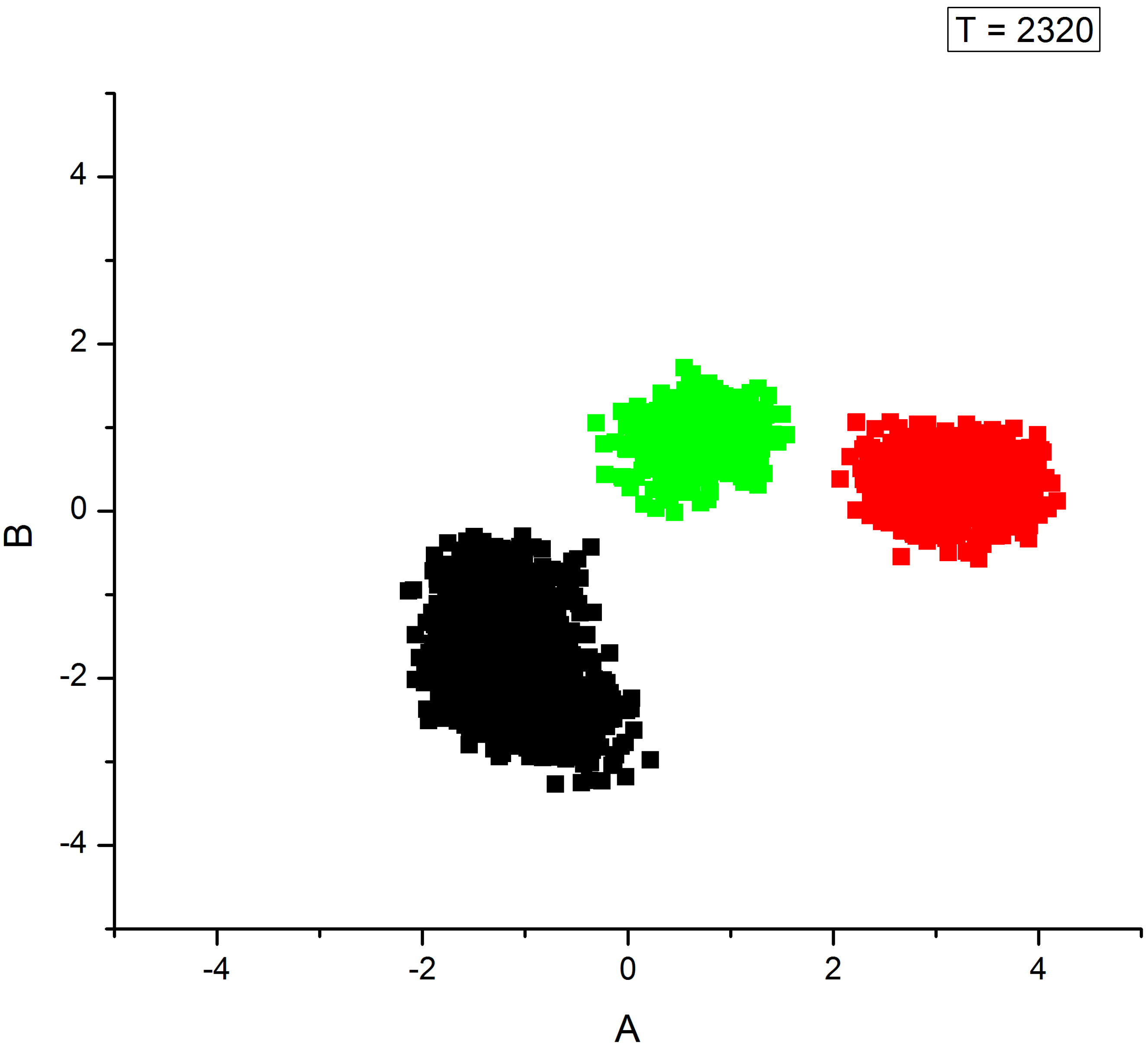


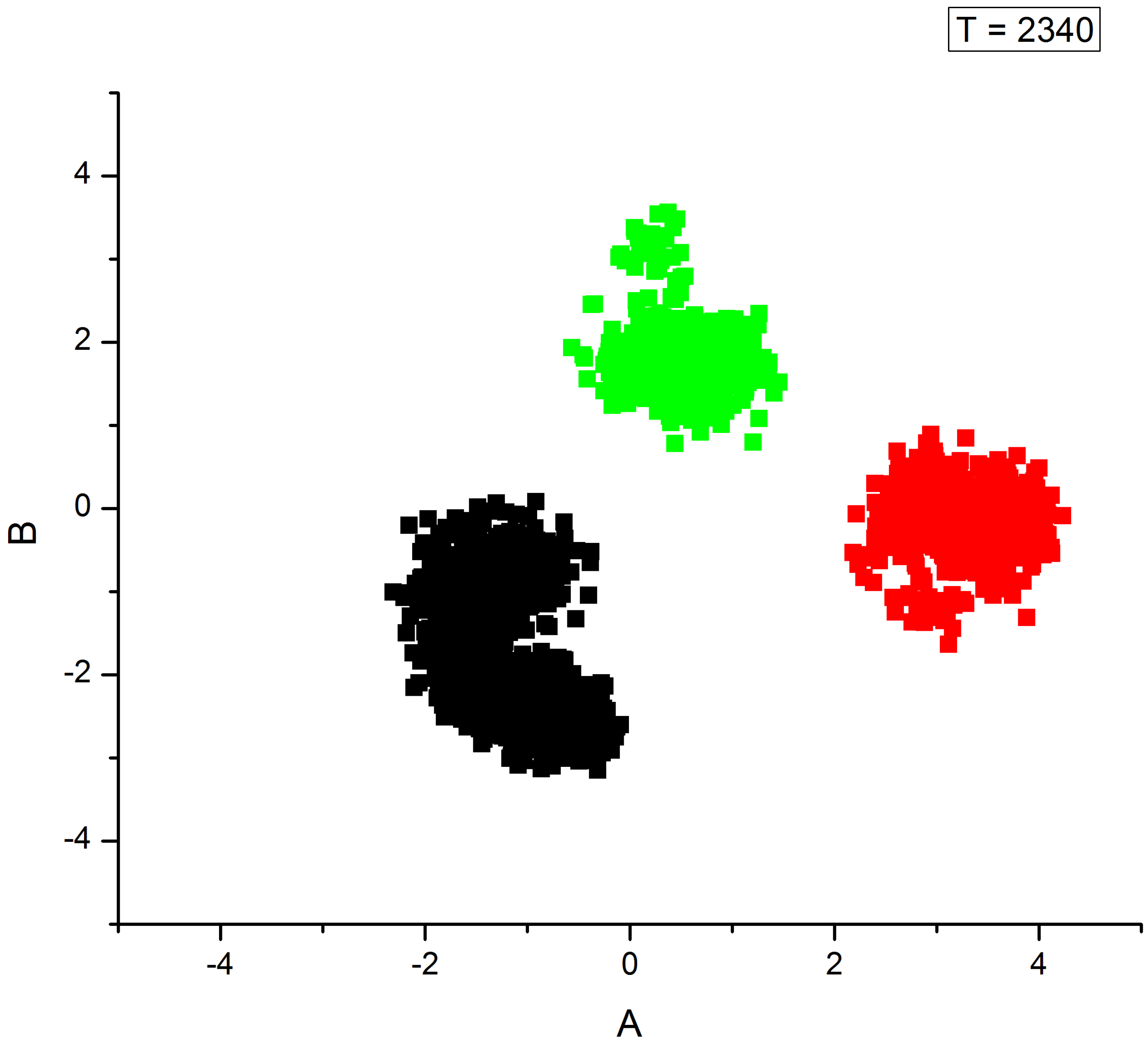
T = 2000
4
2
0
-2
-4
B
-4
-2
0
2
4
A

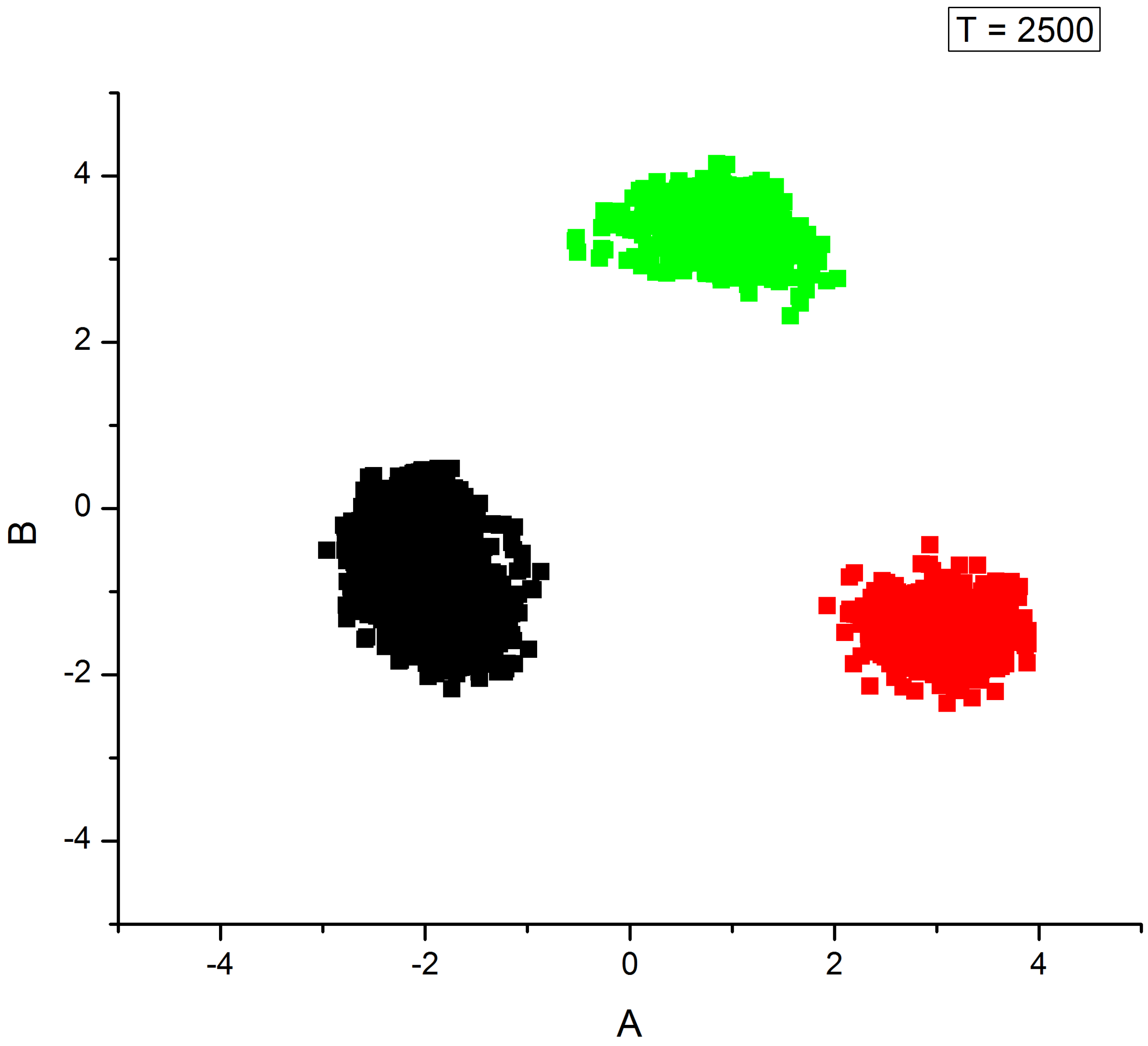

T = 2200
4
2
0
-2
-4
B
-4
-2
0
2
4
A



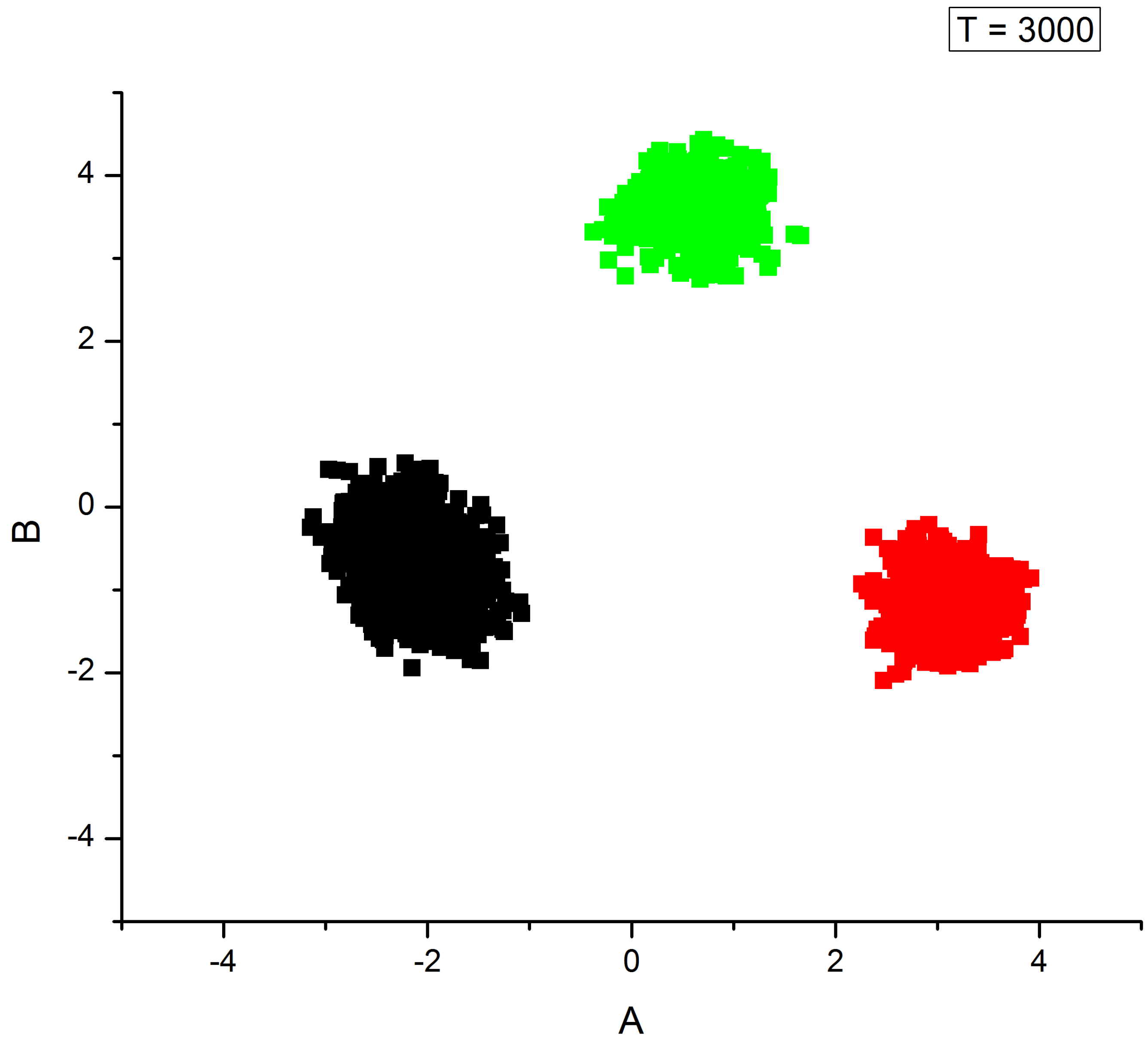
T = 2300
4
2
0
-2
-4
B
-4
-2
0
2
4
A

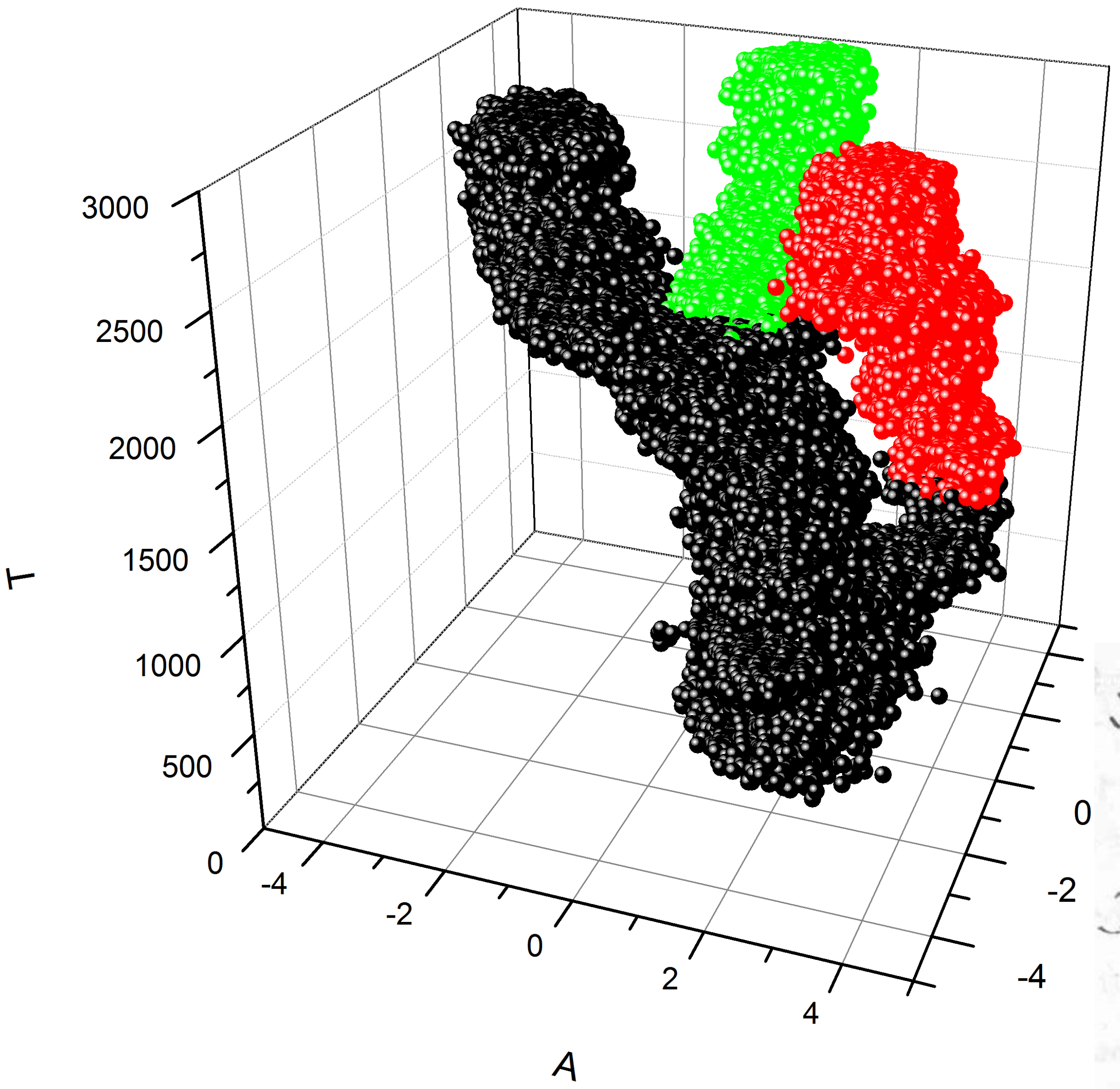

T = 2320
B
4
2
0
-2
-4
-4
-2
0
2
4
A

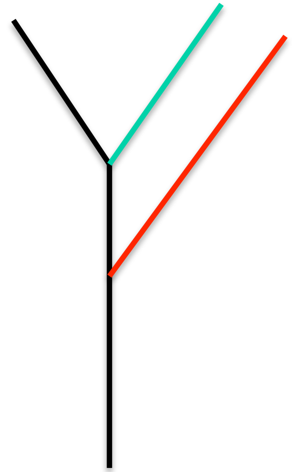

T = 2340
B
4
2
0
-2
-4
-4
-2
0
2
4
A

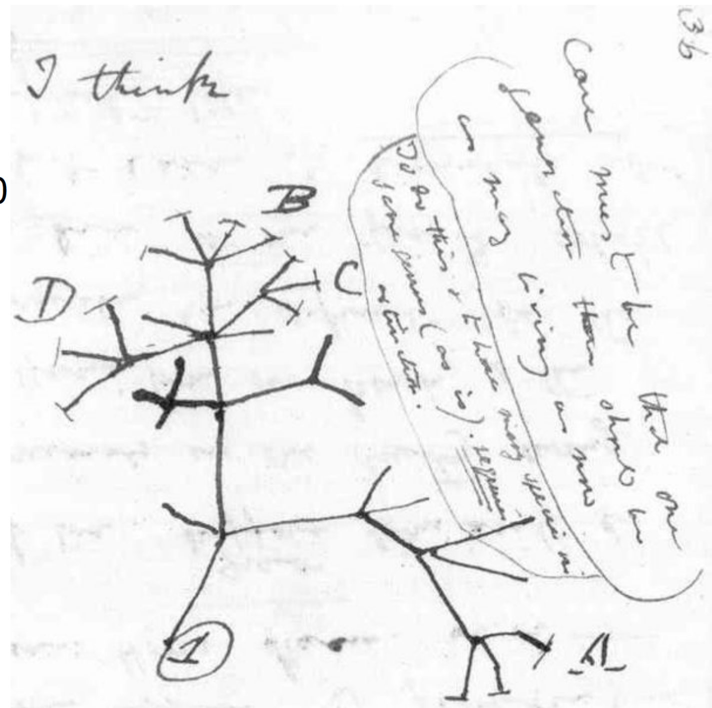


T = 2500
B
4
2
0
-2
-4
-4
-2
0
2
4
A


T = 3000
B
A
4
2
0
-2
-4
-4
-2
0
2
4


T
3000
2500
2000
1500
1000
500
0
-4
-2
0
2
4
A
0
-2
-4




I think
B
C
D
1
A

References:

Global patterns of speciation and diversity
M. A. M. de Aguiar, M. Baranger, E. M. Baptestini, L. Kaufman & Y. Bar-Yam
Nature (2009) **460,** 384

The Moran model as a dynamical process on networks and its implications for neutral speciation
M.A.M. de Aguiar and Y. Bar-Yam,
Phys. Rev. E84 (2011) 031901.

The role of sex separation in neutral speciation
Elizabeth M. Baptestini, M.A.M. de Aguiar and Y. Bar-Yam
J. Theor. Ecol. 6 (2013) 213.

Evolution and Stability of Ring Species
Ayana B. Martins, M.A.M. de Aguiar and Y. Bar-Yam
PNAS 110 (2013) 5080.

Theory

Simplifications:

- population is fully mixed (no space)
- populations is very large
- recombination gene by gene
- compute the frequency of each genotype

Notation:

$$i = (i_1, i_2, \ldots, i_B)$$

$$d(i, j) = \sum_{k=1}^{B} | i_k - j_k |$$

mating between i and j is allowed only if $d(i, j) \leq G$

frequency of genotype *i* in the population at time *t* is p^t_i

The general form of evolution equations is

$$p_i^{t+1} = \sum_{i',i''} c_\mu(i',i'';i)\; p_{i'}^t \; p_{i''}^t$$

where c_μ is the probability that parents with genomes *i'* and *i''* will give rise to an offspring *i* if the mutation rate is μ.

Calculate c_μ looking at one gene at a time. Let us forget about the restriction $d(i,j) \le G$ for now.

There are 3 cases:

(a) $i'_k = i''_k \neq i_k$.

In this case the allele transmitted to the offspring is $(1-i_k)$ and it contributes to p_i only if it mutates to i_k. Therefore it contributes a factor μ to the probability. We call α the number of pairs of genes satisfying this condition:

$$\alpha = \sum_{k=1}^{B} [1 - |i'_k - i''_k|] \; |i_k - i'_k|. \tag{8}$$

(b) $i'_k = i''_k = i_k$.

The allele transmitted is i_k if it does not mutate. It contributes a factor $(1-\mu)$ and the number of pairs of genes in this case is β:

$$\beta = \sum_{k=1}^{B} \left[1 - |i'_k - i''_k|\right] \left[1 - |i_k - i'_k|\right] \tag{9}$$

(c) $i'_k \neq i''_k$.

The allele transmitted is either i_k or $1-i_k$. It contributes a factor $\frac{1}{2}(1-\mu)+\frac{1}{2}\mu = \frac{1}{2}$. The number of pairs of genes of this type is

$$\gamma = \sum_{k=1}^{B} |i'_k - i''_k| = d(i', i''). \qquad (10)$$

It can be checked that:

$\alpha + \beta + \gamma = B$

and

$$\alpha = \frac{d(i,i') + d(i,i'') - d(i',i'')}{2}.$$

The equation becomes:

$$p_i^{t+1} = \sum_{i',i''} p_{i'}^t \, p_{i''}^t \, (1-\mu)^{B-\alpha-\gamma} \mu^{\alpha} \left(\frac{1}{2}\right)^{\gamma}.$$

with

$$\gamma = d(i', i'')$$

$$\alpha = \frac{1}{2}\left[d(i,i') + d(i,i'') - d(i',i'')\right]$$

$$\sum_i (1-\mu)^{B-\alpha-\gamma} \mu^{\alpha} \left(\frac{1}{2}\right)^{\gamma} = 1$$

Prohibiting mating between parents with d(i’,i’’) > G :

$$p_i^{t+1} = \frac{\sum_{i',i'',\gamma \leq G} p_{i'}^t \, p_{i''}^t \, (1-\mu)^{B-\alpha-\gamma} \mu^\alpha \left(\frac{1}{2}\right)^\gamma}{1 - \sum_{i',i'',\gamma > G} p_{i'}^t \, p_{i''}^t}.$$

Solutions:

1) $p_i = (1/2)^B$ - uniform distribution – is always a solution. It is stable only for sufficiently large values μ

2) For $\mu = 0$ the set of genomes of the form $(i_1, i_2, \ldots, i_G, 0, 0, 0, \ldots, 0)$

   where the first G alleles can be 0 or 1 but the remaining (B-G) alleles are all 0

   are solutions.

3) There are 2^G Binomial(B,G) solutions of this type.

Example 1: B=2 G=1; one allele fixed and one allele free

(0,0) and (0,1) - first allele fixed in 0, second allele free

(0,0) and (1,0) - first allele free, second allele fixed in 0

(1,1) and (1,0) – first allele fixed in 1, second allele free

Example 2: B=3 G=1; two alleles fixed and one allele free

(0,0,0) and (0,0,1) - first and second alleles fixed in 0, third allele free

(1,1,1) and (1,1,0) - third allele free, first and second alleles fixed in 1

If the population is structured is space, it might happen that the first solution emerges in a region and the second solution in a different region. These two solutions are reproductively isolated and correspond to 2 species.