

DISCRETE MODELS FOR MOLECULAR EVOLUTION: SIMULATION AT THE POPULATION LEVEL

D.Yu. Sherbakov

**Limnological Institute SB
RAS
Irkutsk
RUSSIA**



MOTIVATION:

- Making the process of simulation of ecological or evolutionary process better understandable for a biologist
- Obtaining the results of a simulation in a form which would correspond directly to the outcome of an experiment

Logistic growth with competition and mutations – adaptive dynamics

$$\frac{dN(\mathbf{x}, t)}{dt} = r \cdot N(\mathbf{x}, t) \cdot \left[1 - \int C_{\mathbf{x}}(\mathbf{x} - \boldsymbol{\chi}) N(\boldsymbol{\chi}, t) d\boldsymbol{\chi} / K(\mathbf{x}) \right]$$

Here \mathbf{x} – ecological character defined by polygene

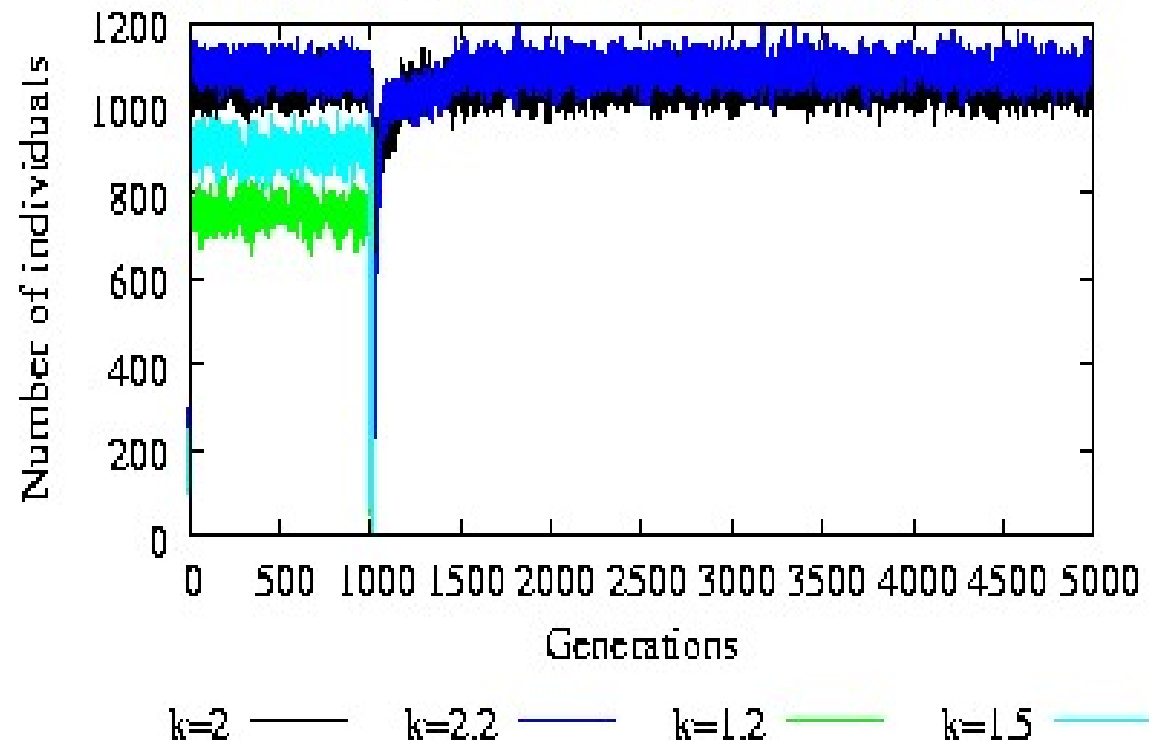
$$K(\mathbf{x}) = K_0 \exp\left[-\frac{(\mathbf{x} - \mathbf{x}_0)^2}{2\sigma_K^2}\right], \quad C_{\mathbf{x}}(\mathbf{z}) = \exp\left[-\frac{\mathbf{z}^2}{2\sigma_x^2}\right]$$

From: On the origin of species by sympatric speciation

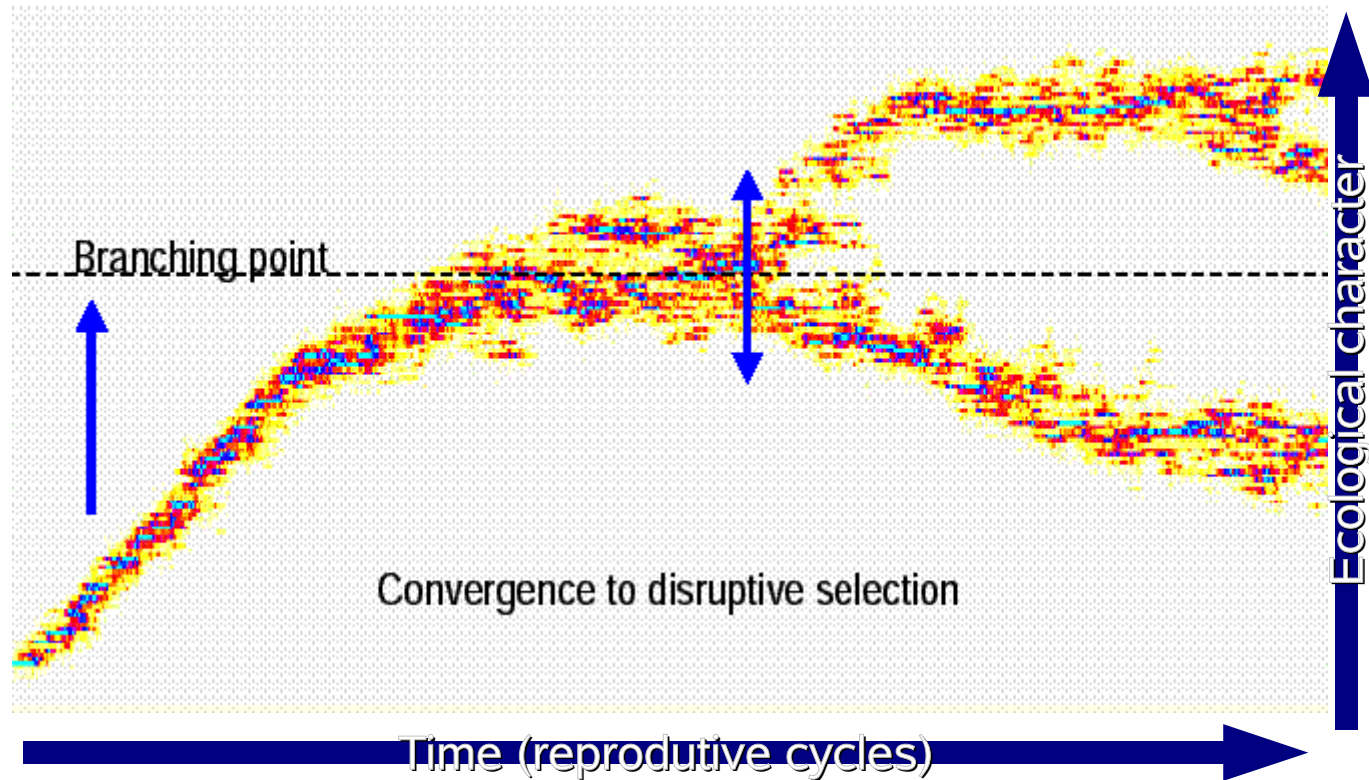
Ulf Dieckmann & Michael Doebeli, Nature, 1999

«Useless» simulation

Mask involves 4 aminoacids, the value jumps from 30 to 45



Individual-based representation



Useful simulation, the result is in comfortable format

```
ec1733 MAVRECGATHRTD
ec455 MAVRECGATDRTD
ec954 MAVRECGATHPYN
ec829 MAVRECGATHRLH
ec899 MAVRECGATDRWD
ec1617 MAVRECGATDRWD
ec672 MAVRECGATHPYN
ec1233 MAVRECGATHRTD
ec1274 MALRECGATQITD
ec654 MTVRECGADHWTS
ec463 MAVRECGATDRTD
ec1010 MTVRECGADHWTS
ec1128 MAVRECGATHRTH
ec1736 MAVRECGATQIQD
ec1201 MAVRECGATQITP
ec947 YAVRECGAVHRHD
ec210 MAVRECGATDRTD
ec1411 MAVRRIGATHRTY
ec316 MAVRECGATDRTD
ec955 MAVRECGAHHWQA
ec699 MAVRECGATQIQD
```

```
423 GTCCTTTGTCCCGOCTAAATATCTCGCTC
1176 GTCCTTTGTCCCGOCTAAATATCTCGCTC
498 GTCCTTTGTCCCGOCTAAATATCTCGCTC
169 GTCCTTTGTCCCGOCTAAATATCTCGCTC
291 GTCCTTTGTCCCGOCTAAATATCTCGCTC
1037 AACTCTGTCCCGOCTAAATATCTCCCTA
453 AACTCTGTCCCGOCTAAATATCTCCCTA
782 TCAGCTATCCCGOCTAAATATCTCCCTA
1739 AACTCTGTCCCGOCTAAATATCTCCCTA
1006 AACTCTGTCCCGOCTAAATATCTCCCTA
599 AACTCTGTCCCGOCTAAATATCTCCCTA
650 GAATTTGTCCCGOCTAAATATCTCCCTA
1299 AACTCTGTCCCGOCTAAATATCTCCCTA
1609 GAATTTGTCCCGOCTAAATATCTCCGTC
769 AACTCTGTCCCGOCTAAATATCTCCCTA
7 AACTCTGTCCCGOCTAAATATCTCCCTA
995 AACTCTGTCCCGOCTAAATATCTCCCTA
1121 GTCCTTTGTCCCGOCTAAATATCTCGCTC
336 GTCCTTTGTCCCGOCTAAATATCTCGCTC
1606 GTCCTTTGTCCCGOCTAAATATCTCGCTC
285 GAATTTGTCCCGOCTAAATATCTCCGTC
781 AACTCTGTCCCGOCTAAATATCTCCCTA
1796 AACTCTGTCCCGOCTAAATATCTCCCTA
```

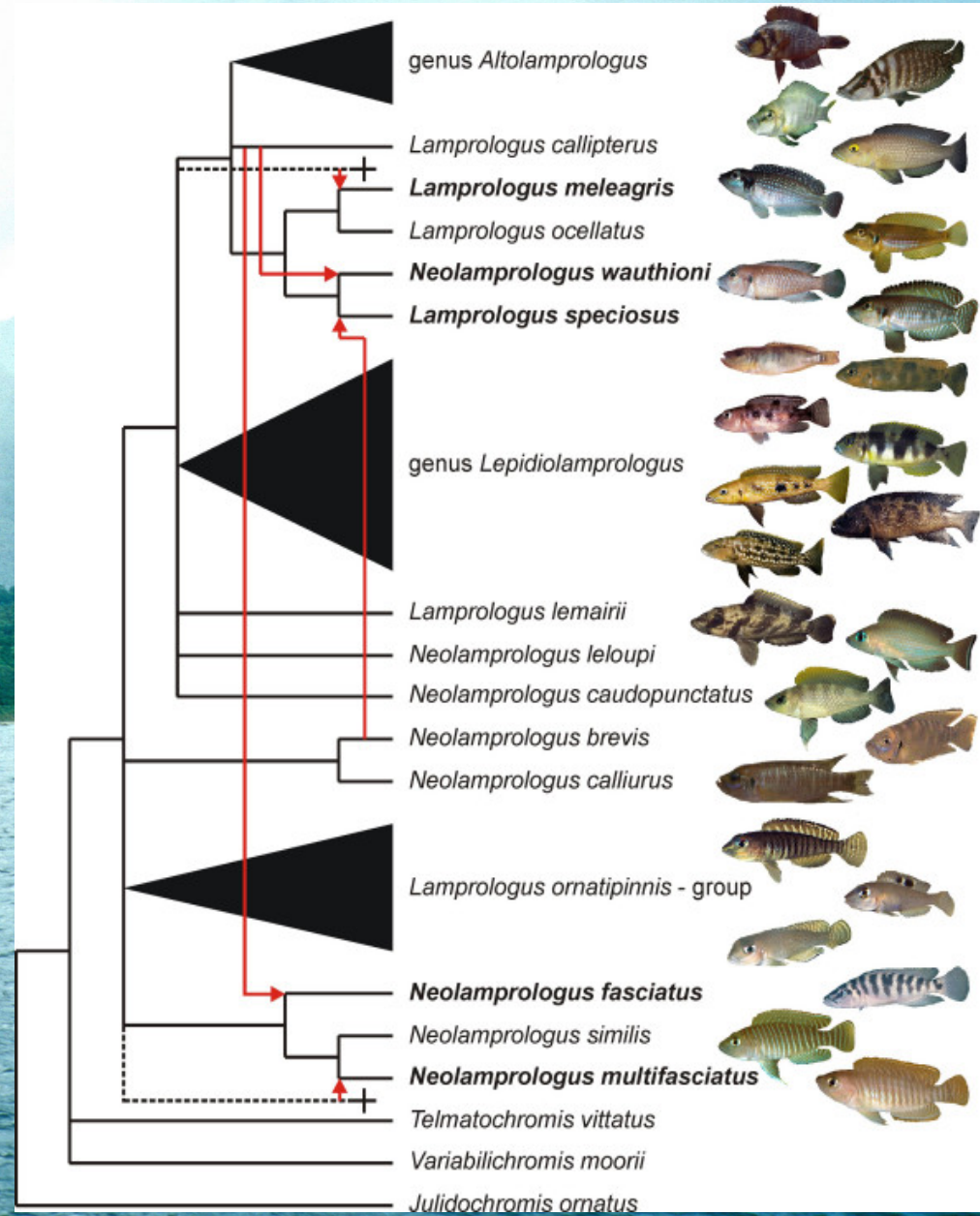

Motivation: explaining evolution of species flocks

Species flock is a group of species which evolved in confines of the same eco-system from a single ancestor

Species flocks are peculiar to ancient giant freshwater lakes like lake Baikal and Lake Tanganyika, where the most famous and species-rich is set of flocks of cichlids



Tanganyika



Tanganyika

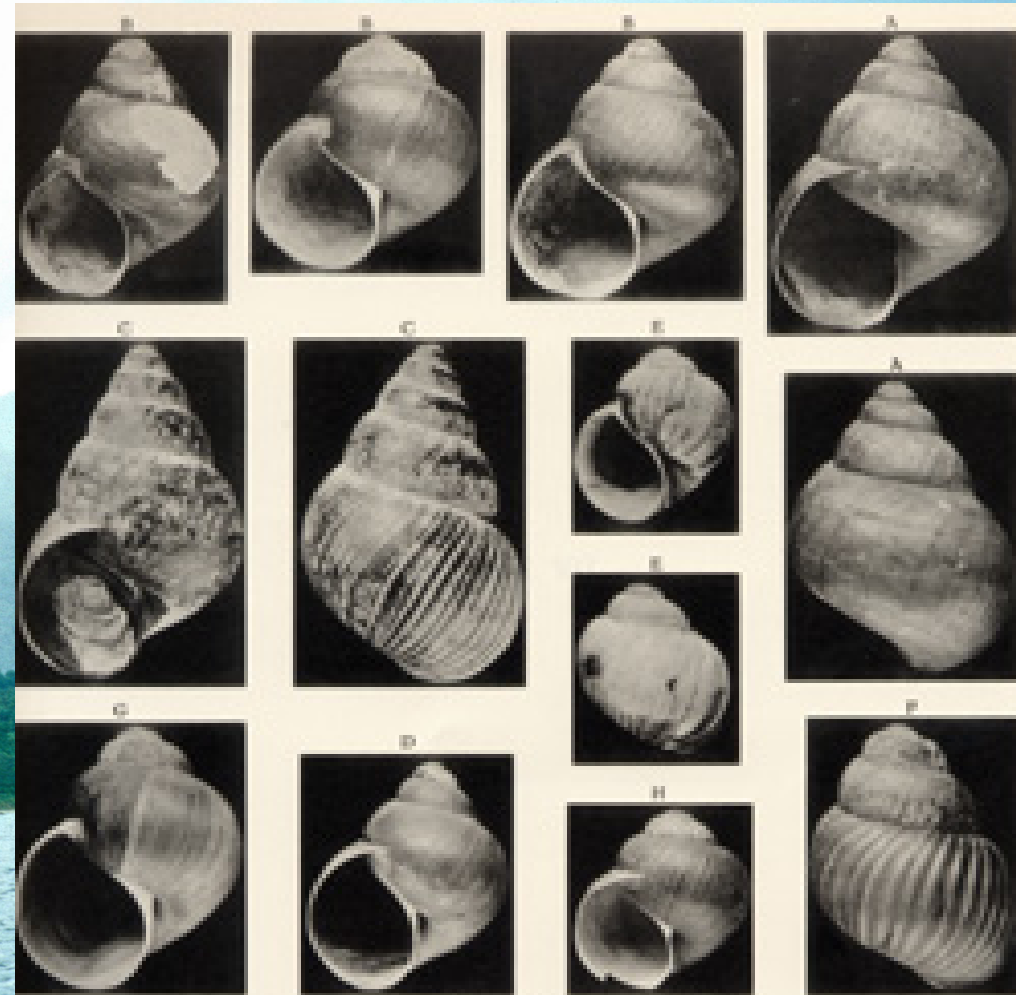


Fig. 1. - *Lantula (Melinisus) albana* (Bourcneri, 1927).

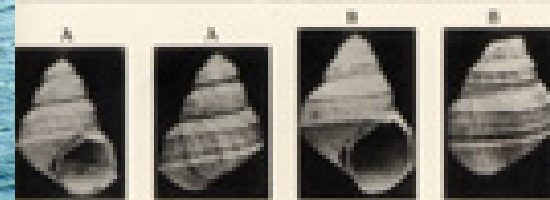


Fig. 2. - *Chrysina trilineata* Gussakov, 1905.

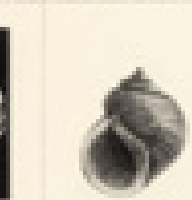


Fig. 3. - *Lantula* (*Chrysina*) *barberi* Gussakov, 1905.

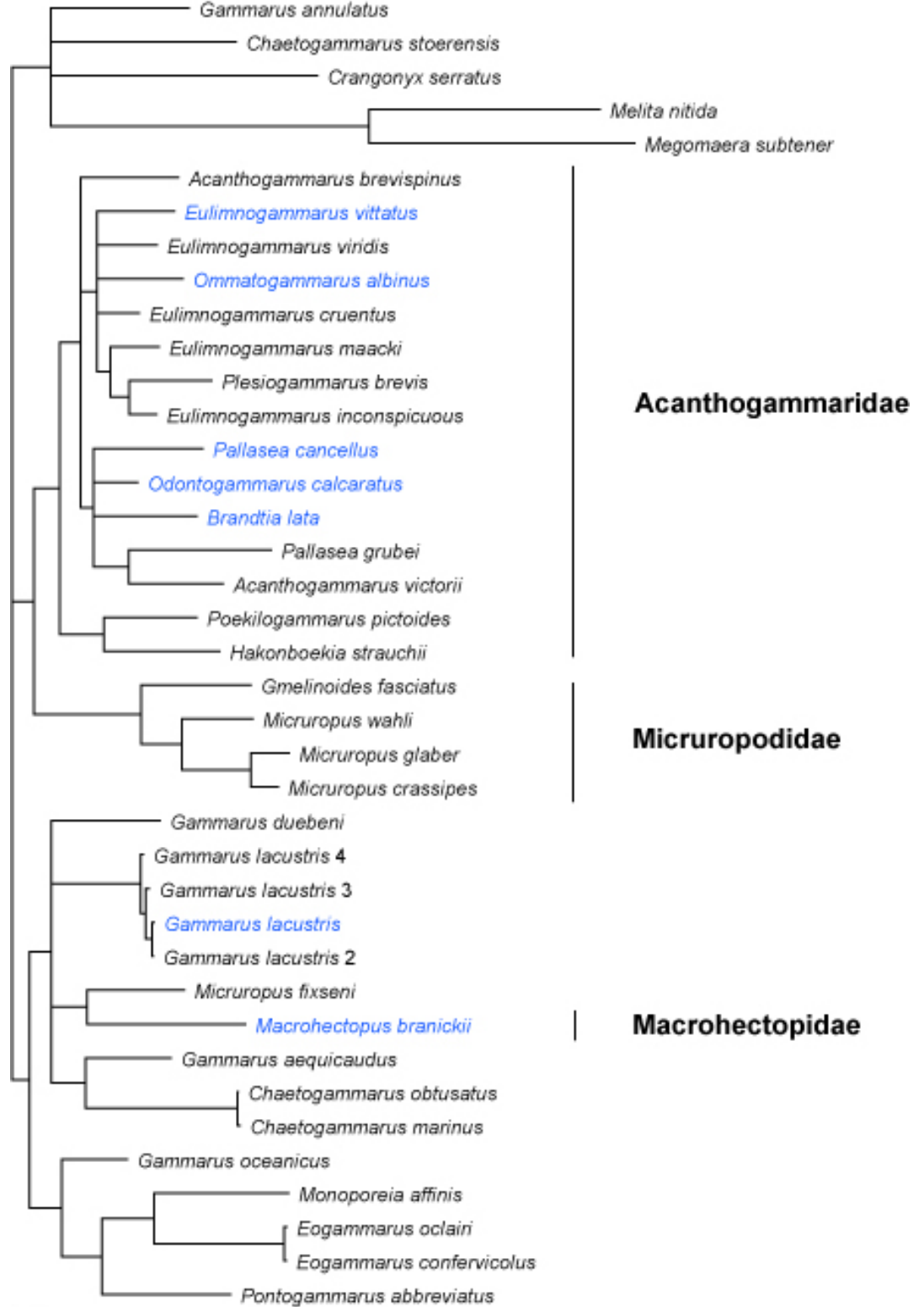
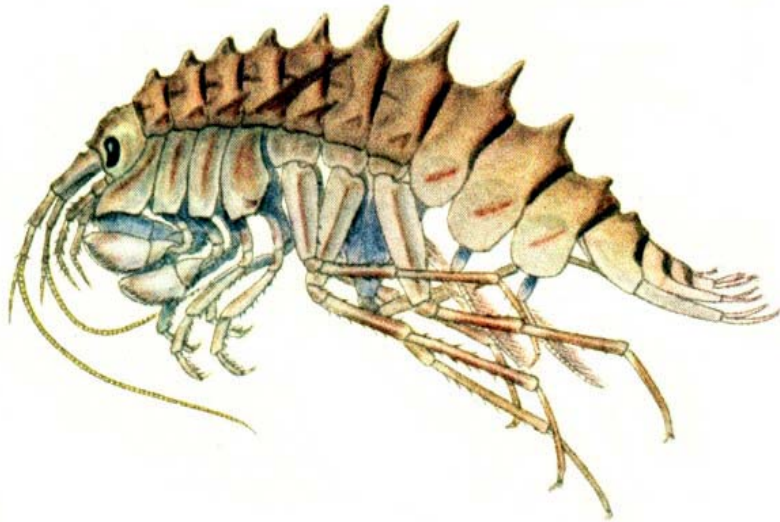


Fig. 4. - *Lantula* (*Chrysina*) *greenii* Tonnar, 1911.

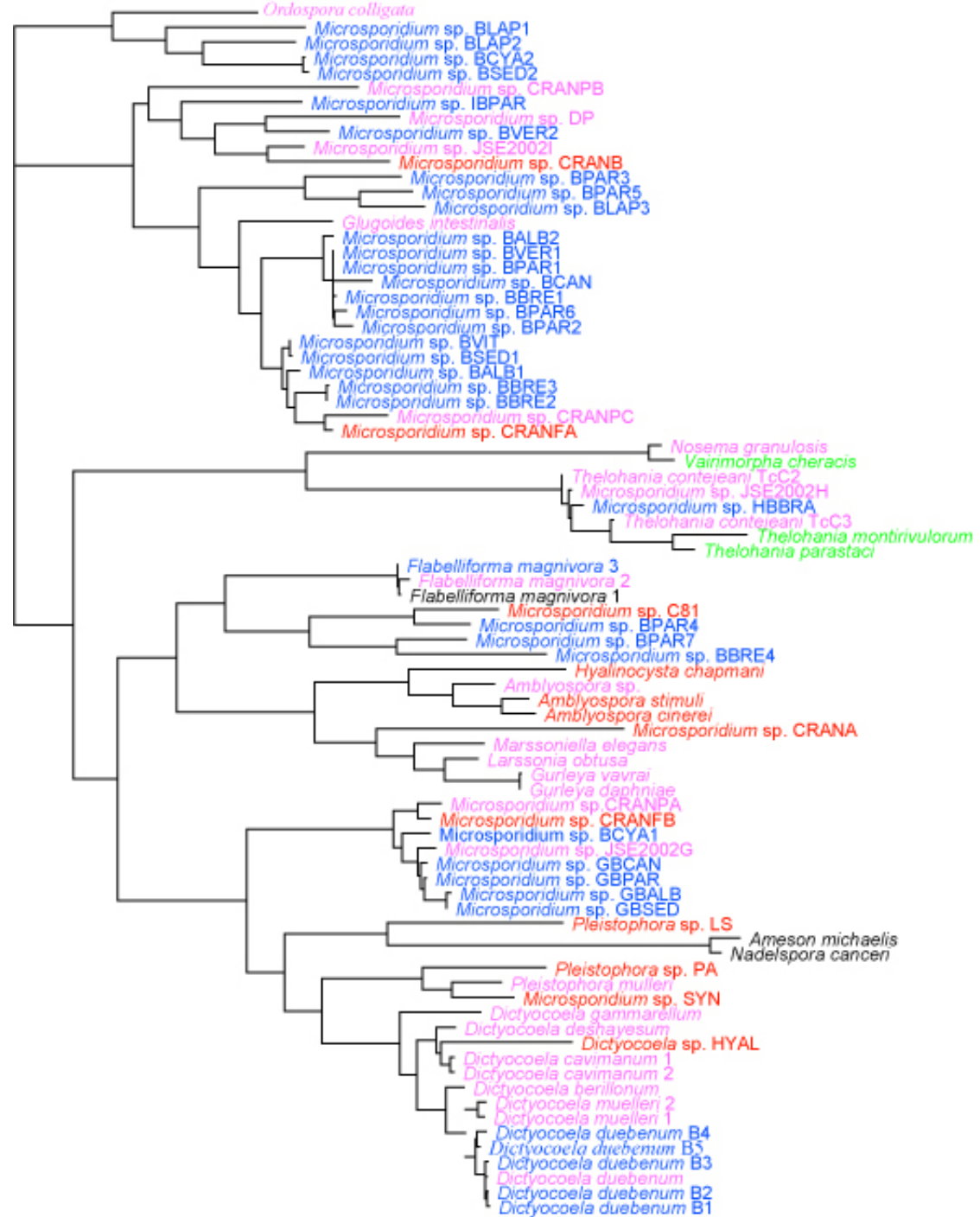
Baikal



Amphipods:



Microsporidia parasiting on Baikalian amphipods comprise the first fully parasitic species flocks. The main surprise here is absolute lack of evidence for any co-evolution between host and parasite. Another problem is co-existence of several parasite species on a single host species: this must cause ultimate extinction of the latter



0.1

- Russia
- Europe
- North America
- Australia



Baicalia carinata

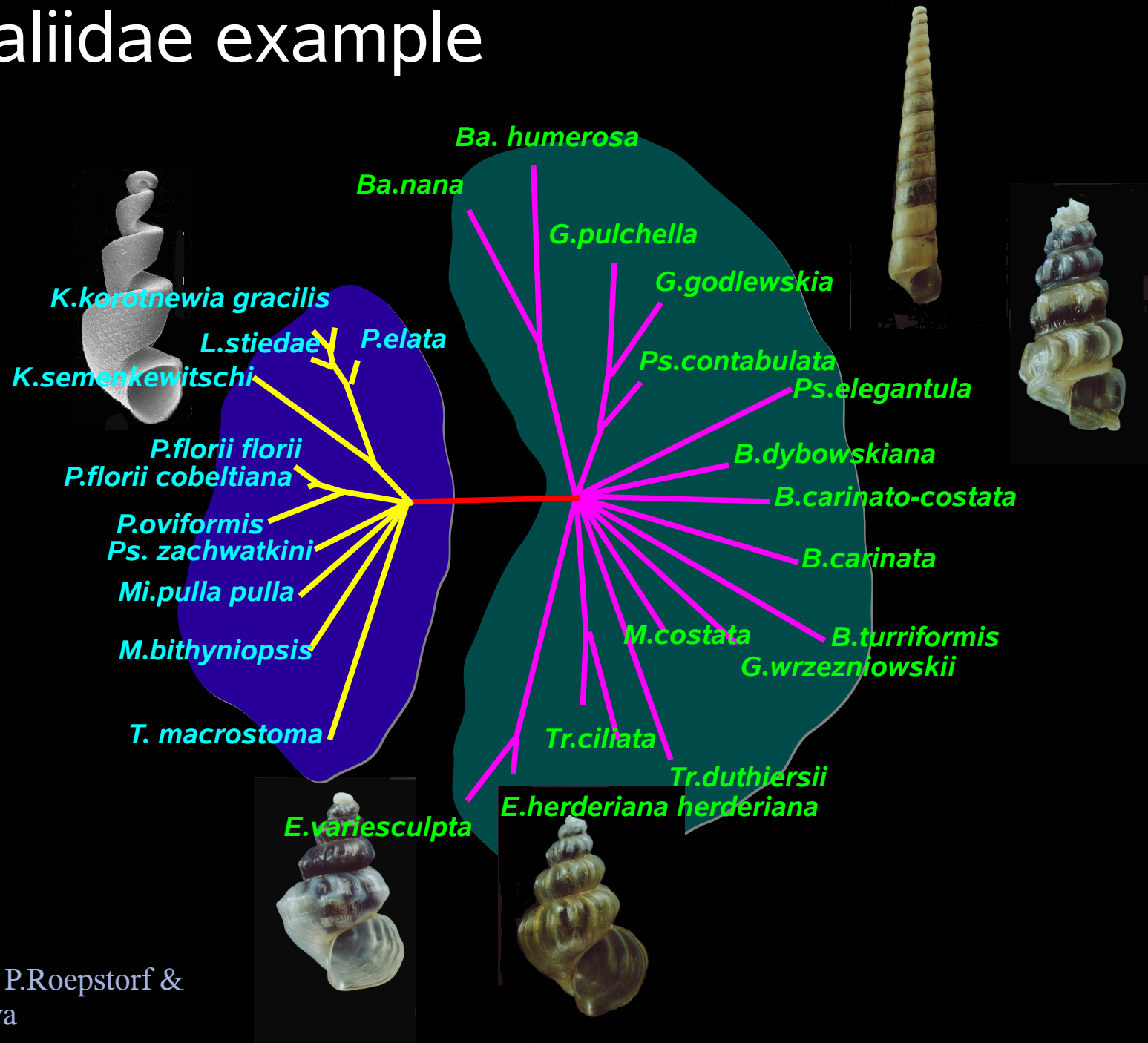


B. dybowskiana



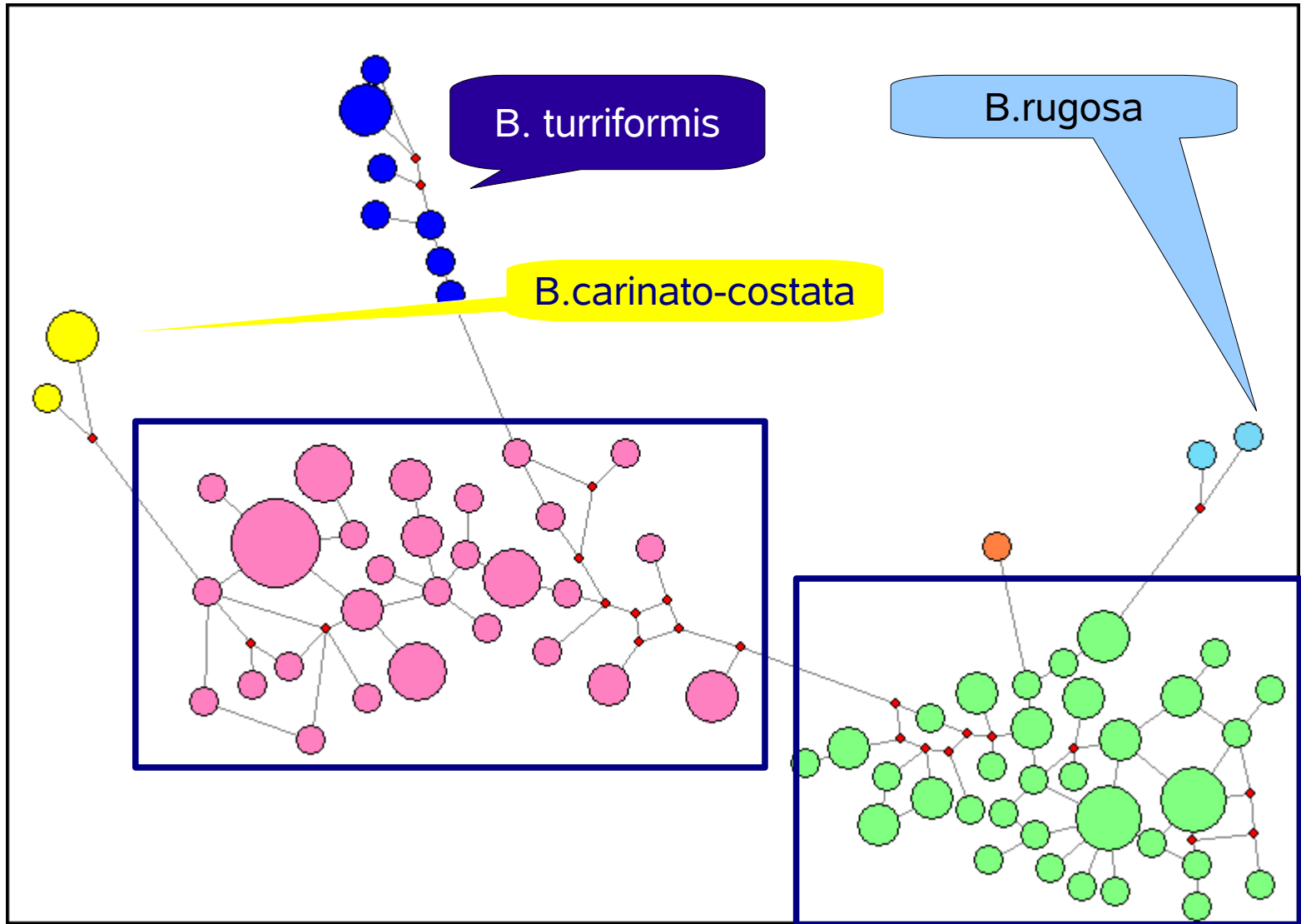
B. turiformis

Baicaliidae example

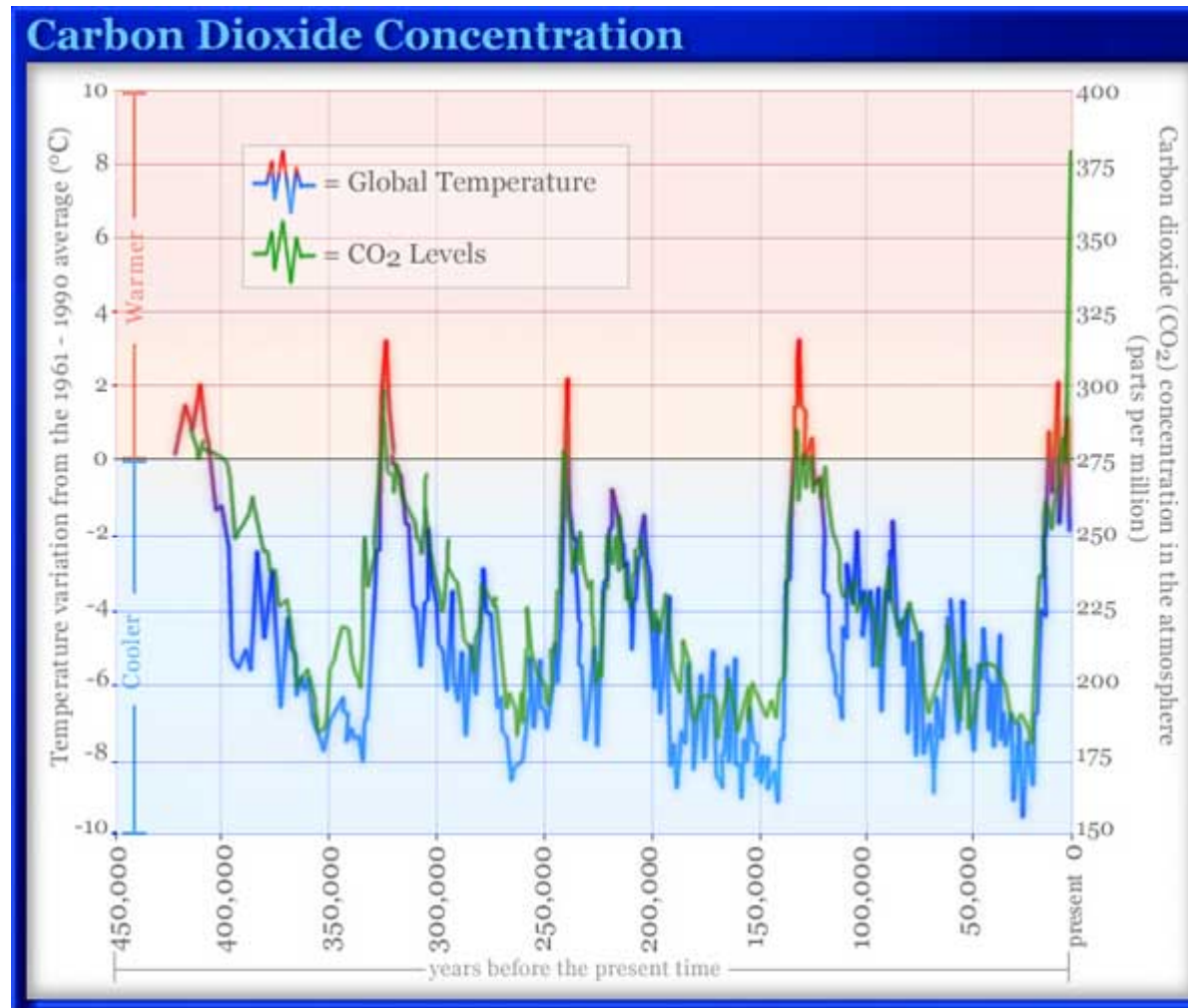


Photos by P.Roepstorf & T.Sitnikova

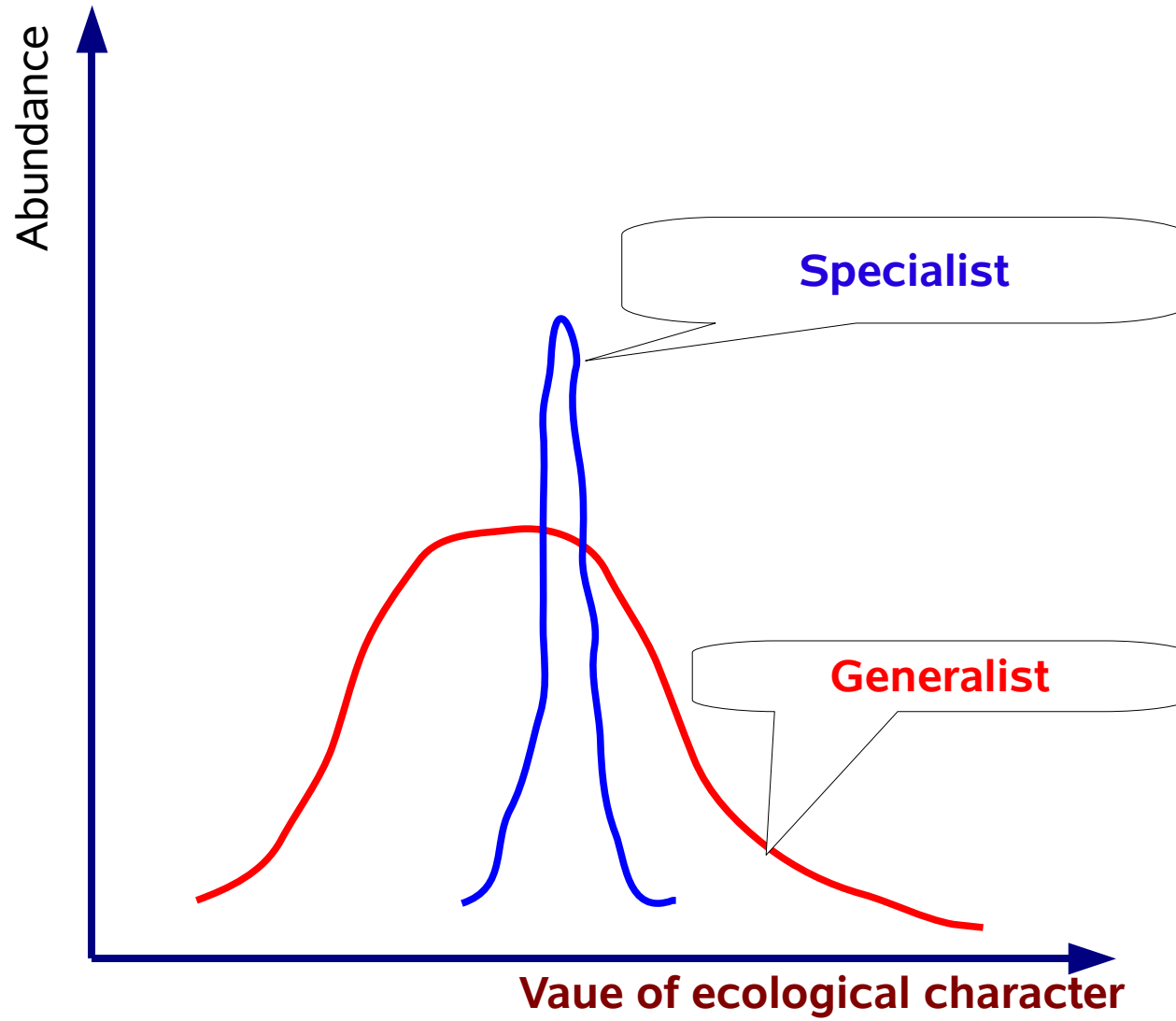
Some evolutionary relations within a flock cannot be represented correctly by a tree



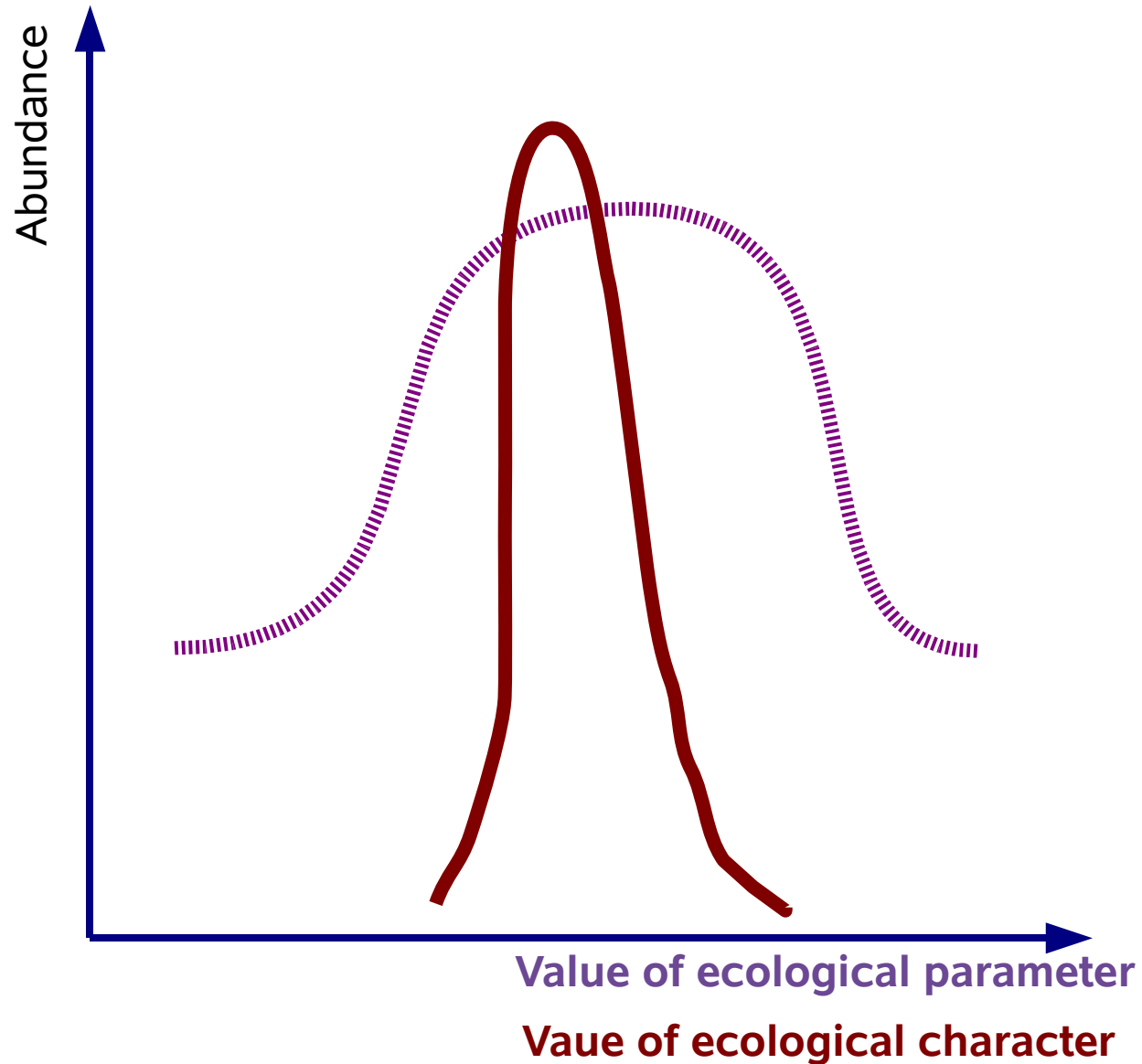
Shackleton cycles: evolution in changing environment



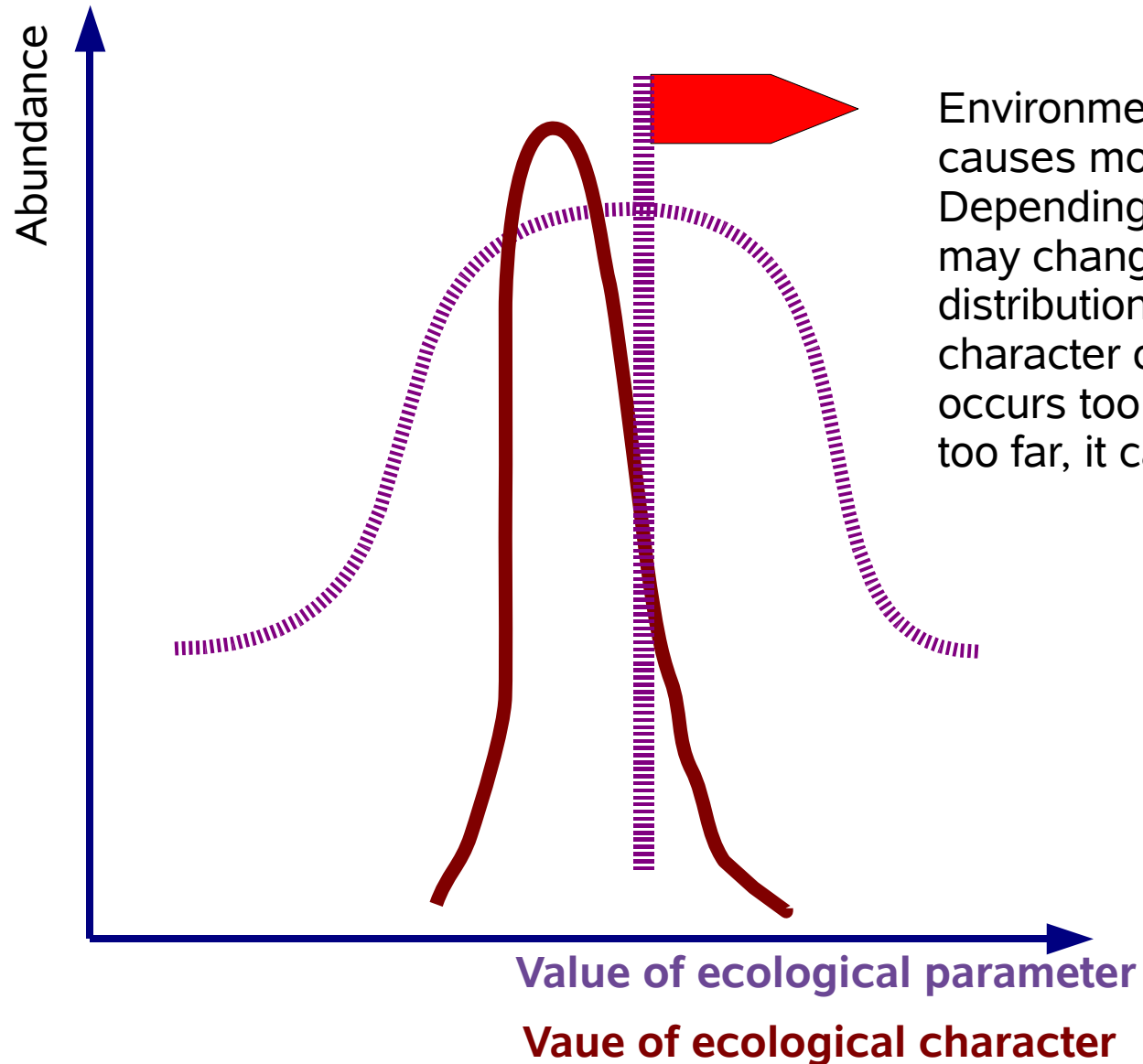
Specialists vs. generalists



Specialists vs. generalists

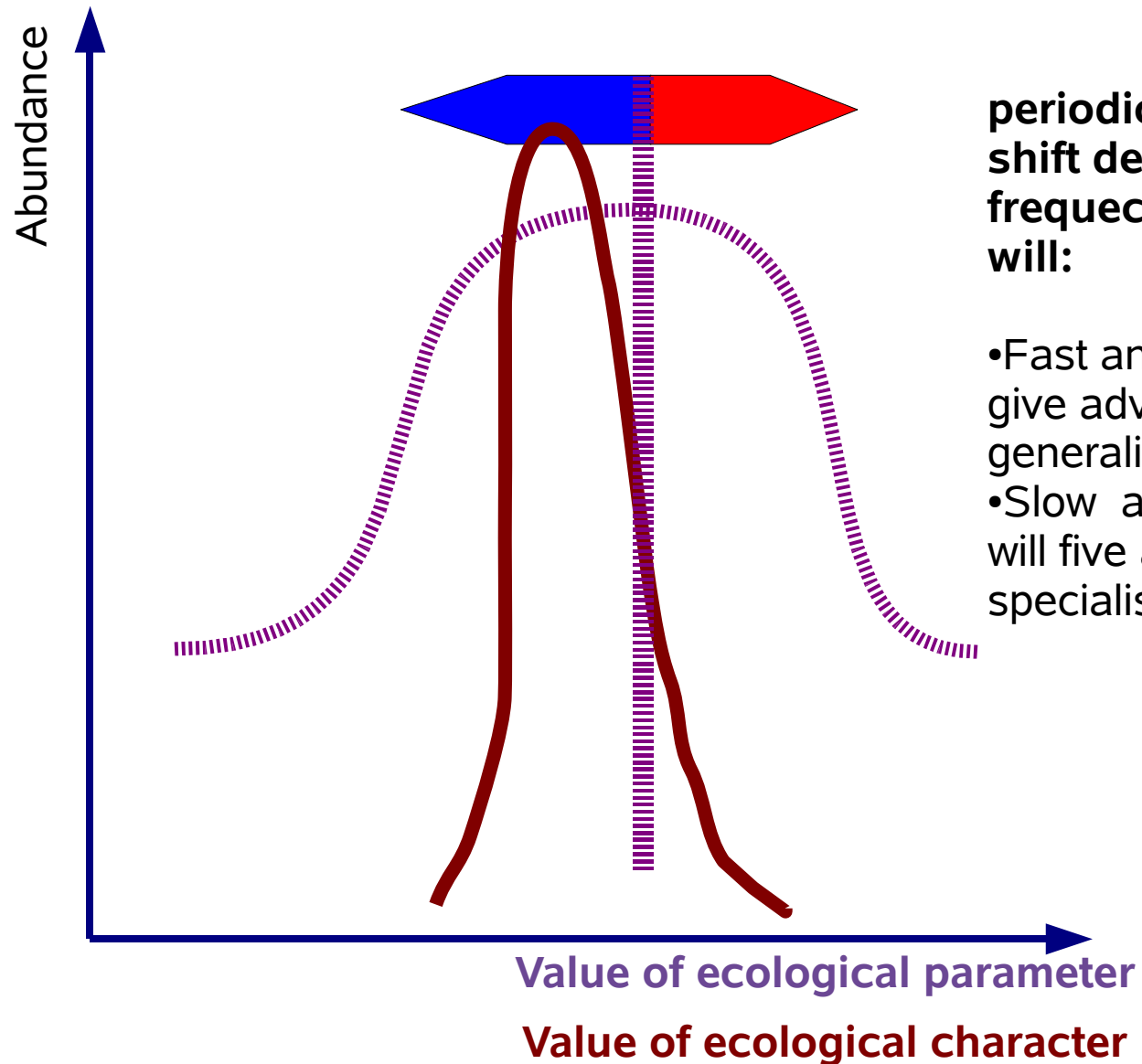


Specialists vs. generalists



Environmental shift causes moving selection. Depending on the speed, it may change the distribution of ecological character of a species. If it occurs too fast and goes too far, it causes extinction

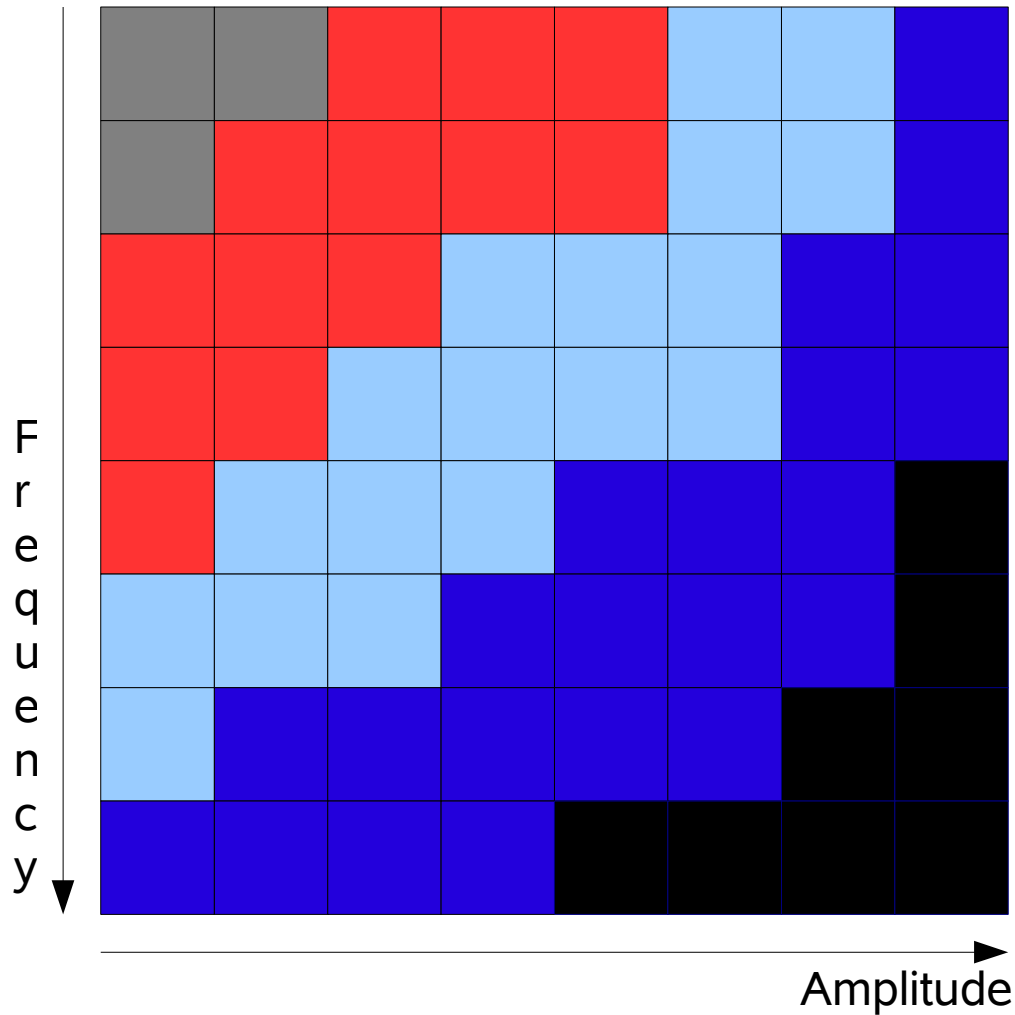
Specialists vs. generalists



periodical environmental shift depending on frequency and amplitude will:

- Fast and wide sweeps will give advantage to generalists;
- Slow and moderate sweeps will give advantage to specialists (narrow niche)

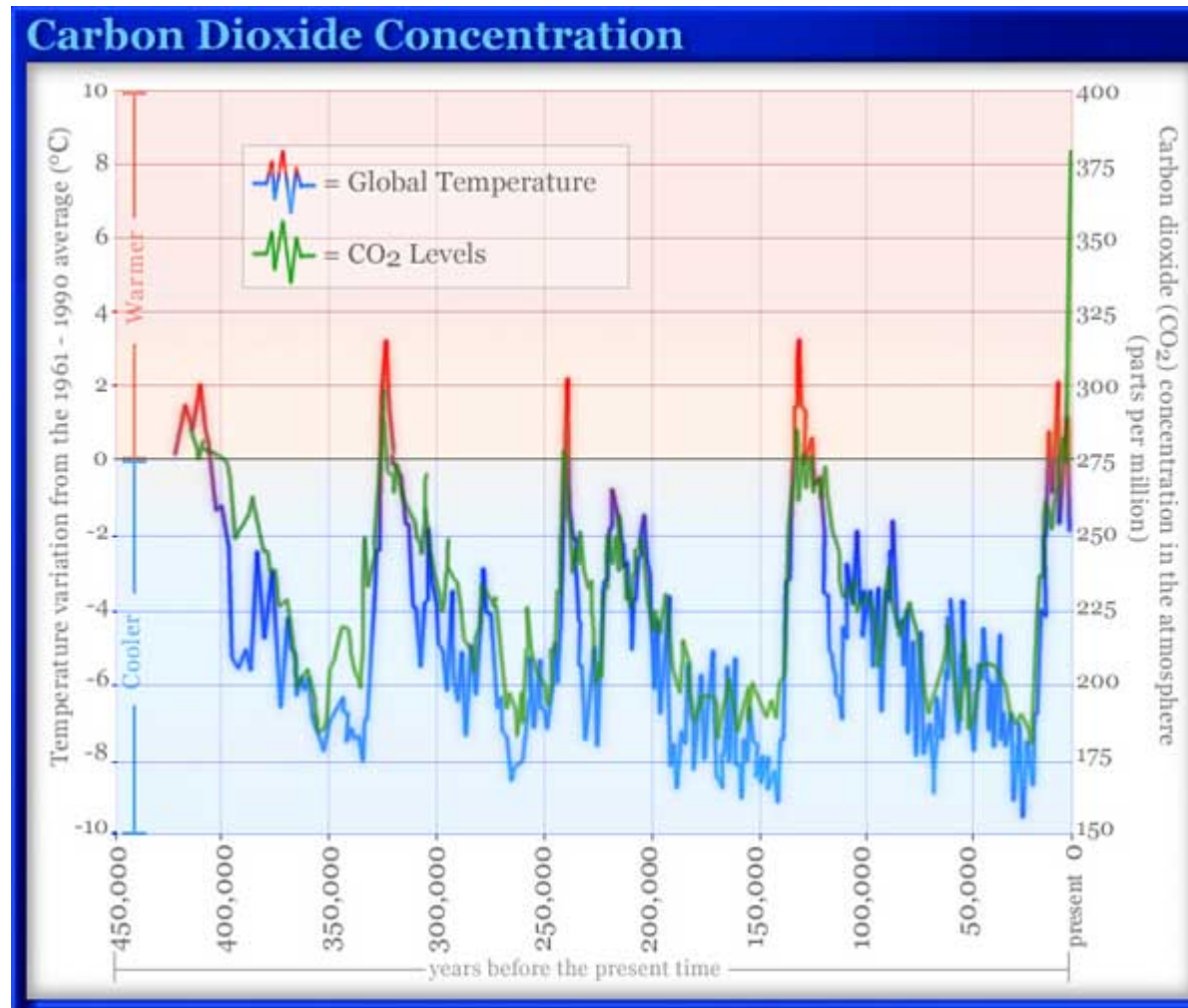
Specialists vs generalists



In Lake Baikal there are no strict specialists, at least close to surface, where impact of environmental changes is more dramatic.

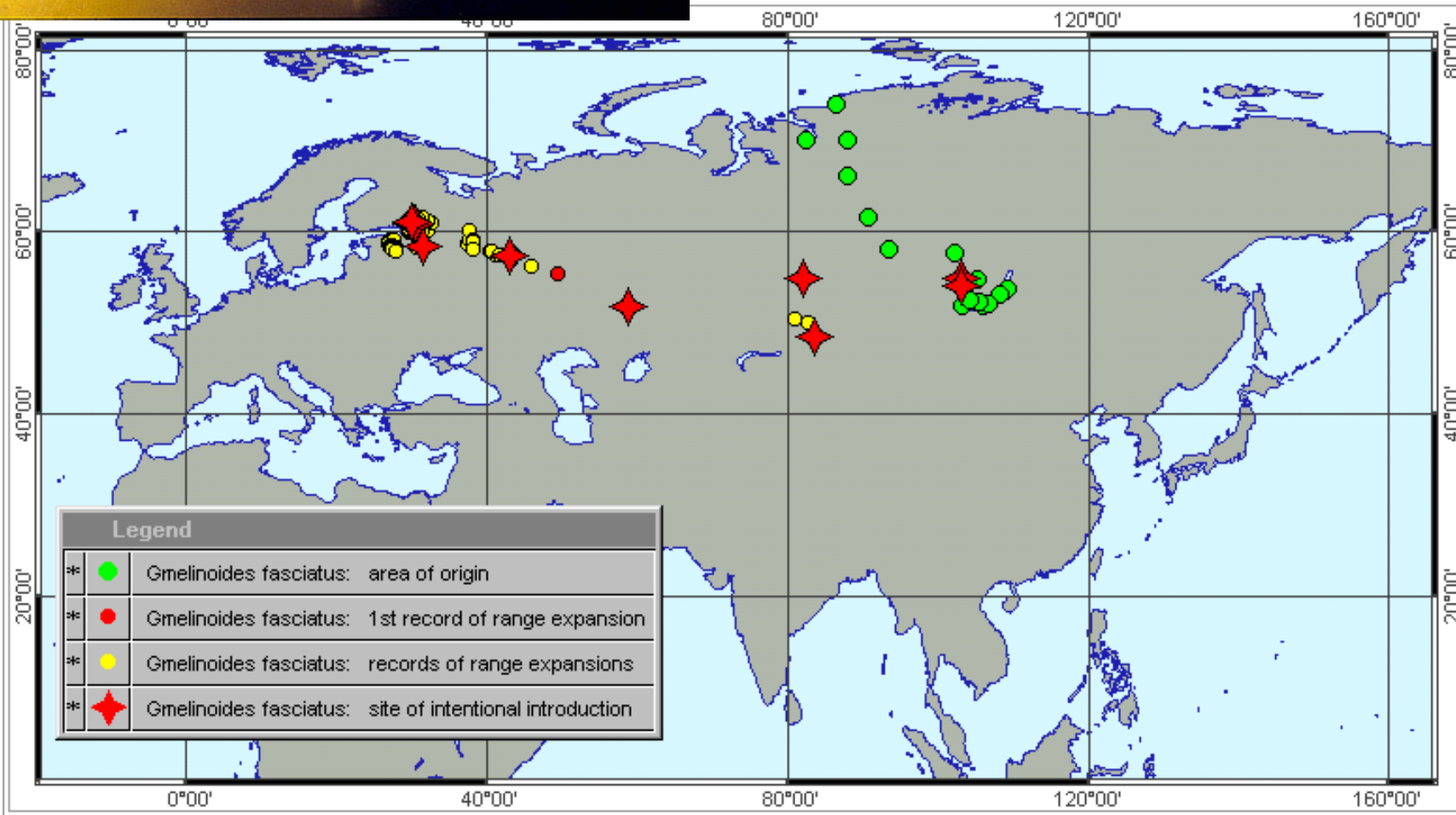
Are there specialists in the abyssal zone? It is not yet known, but this zone is relatively young.

Shackleton cycles: evolution in changing environment





Subdivision in *Gmelinoides fasciatus*
(Environmental changes are not the only cause of species subdivision)

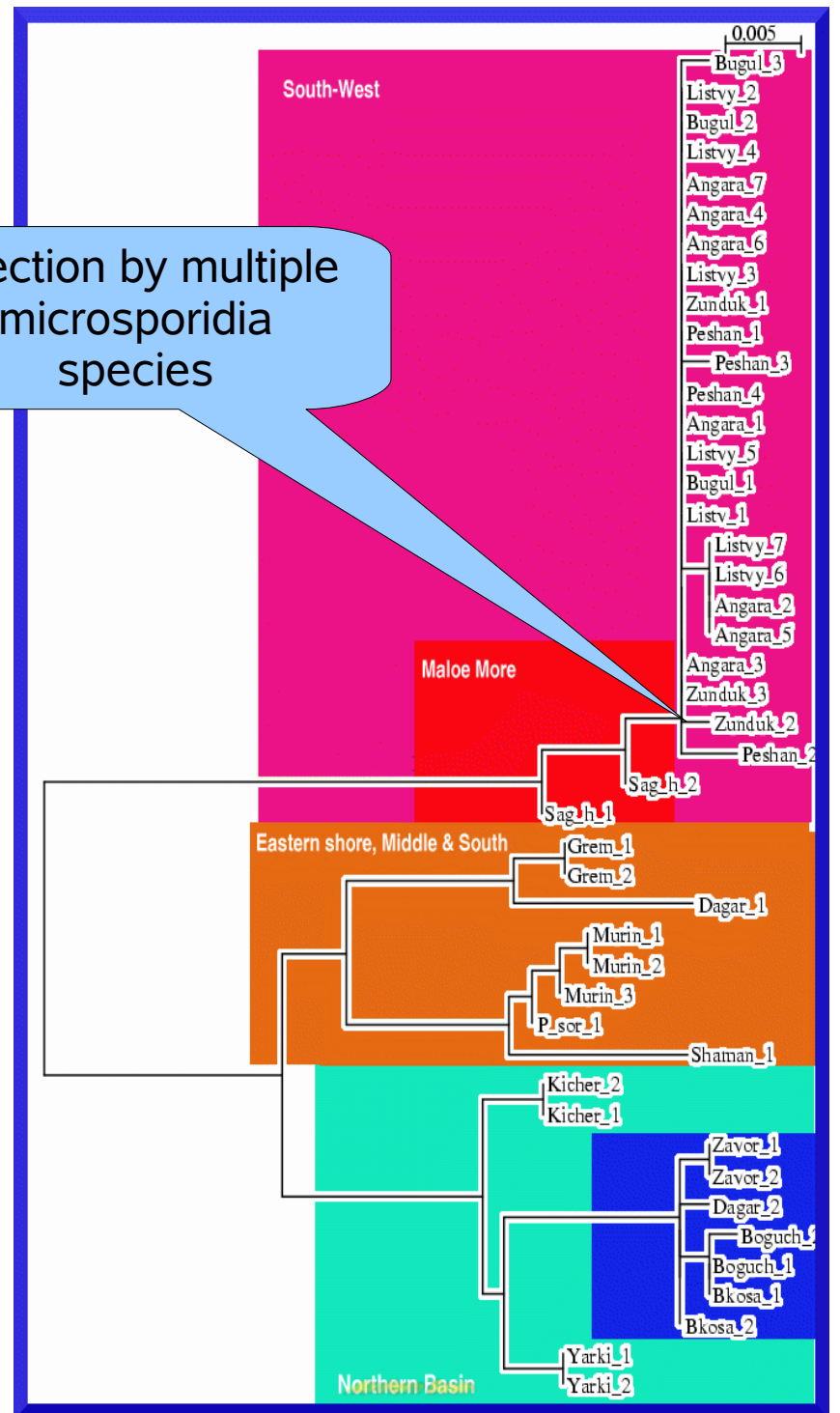




Are different populations separate species?
How could this pattern occur?



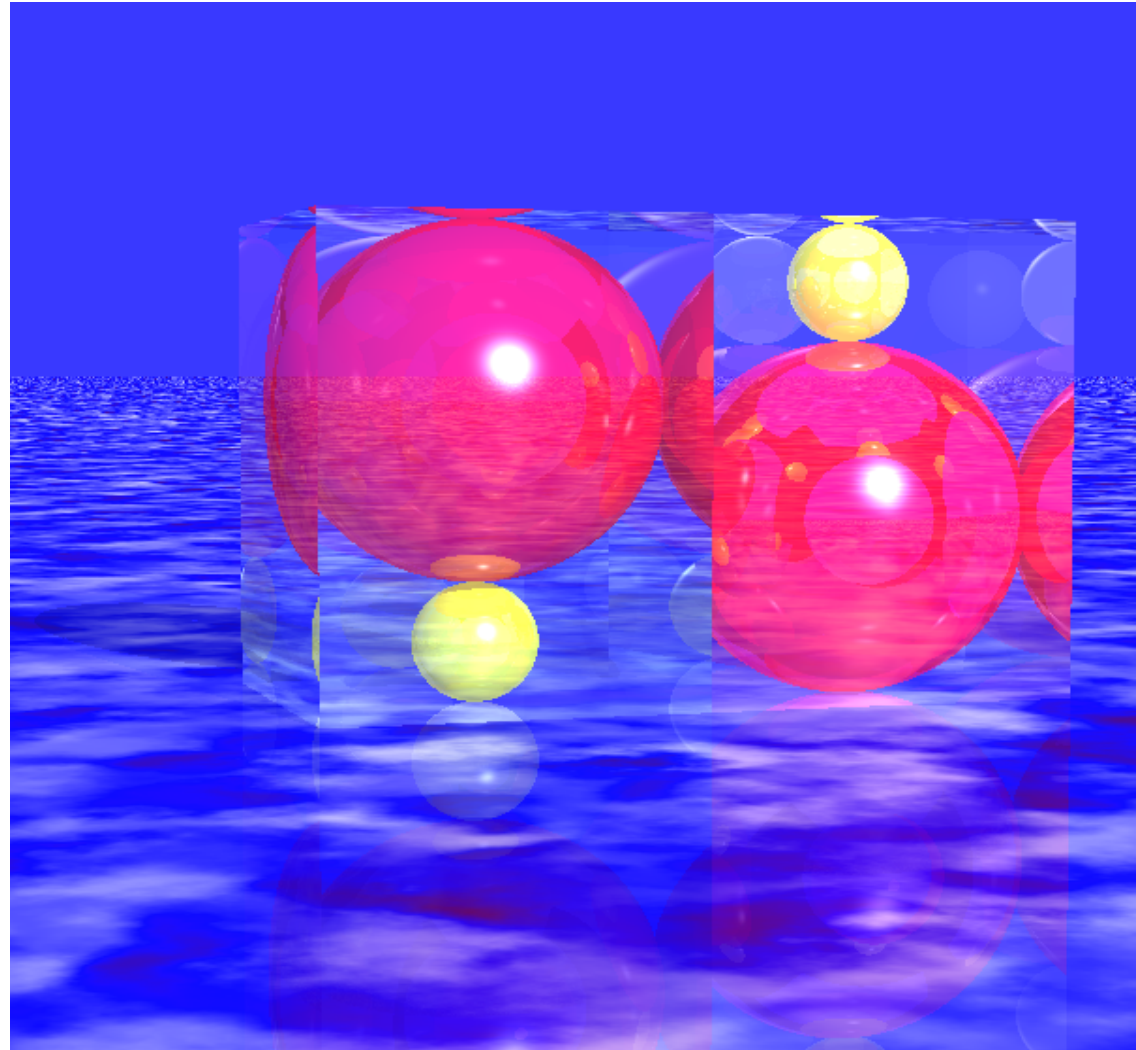
Infection by multiple
microsporidia
species



Coordinated evolution of amino acid sequences

Some features of proteins preserved in evolution involve more than single amino acid residue

- There is no obvious transition from one state to another: intermediate state is deleterious
- Numerous physical or/and chemical properties of amino acid residues may be conserved, as the consequences the values of these properties appear to be highly correlated in evolution
- More than two residues may take part in a group
- Correlated groups are quite common in proteins



Objectives:

- To elucidate the population mechanism allowing coordinated evolution of amino acid residues;
- Using computer simulation of microevolutionary events, to find the conditions favouring coordinated evolution
- To design statistical tests allowing one to study coordinated evolution or/and take it into account when necessary

Model

Organism

- Hermaphroditic, but during the same reproduction cycle may have only one gender
- Diploid and has maternally transferred («mitochondrial») DNA
- Has age limit measured in reproductive cycles
- Fitness depends on one protein coding gene and population density limited resource)
- Number of progeny is Poisson distributed

Mitochondrial marker

- Single vector $x_i \in A, C, G, T$
- No recombination
- Transferred from mother to all her progeny
- Does not affect fitness
- All substitutions (mutations) are equiprobable

```
423 G T C T T T G T C C C G C C T A A T A T C T G C T C
1176 G T C T T T G T C C C C C C T A A T A T C T G C T C
498 G T C T T T G T C C C C C C T A A T A T C T G C T C
169 G T C T T T G T C C C C C C T A A T A T C T G C T C
291 G T C T T T G T C C C G C C T A A T A T C T G C T C
1037 A A C T C T G T C C C C C C A A A T A T C T C C C T A
453 A A C T C T G T C C C C C C A A A T A T C T C C C T A
782 T C A G C T A T C C C C C C A A T T G C C T C C C T A
1739 A A C T C T G T C C C C C C A A A T A T C T C C C T A
1006 A A C T C T G T C C C C C C A A A T A T C T C C C T A
599 A A C T C T G T C C C C C C A A A T A T C T C C C T A
650 G A A T T T G T C C A C C C A A A T A T C T C C C T C
1299 A A C T C T G T C C C C C C A A A T A T C T C C C T A
1609 G A A T T T G T C C C C C C A A A T A T C T C C G T C
769 A A C T C T G T C C C C C C A A A T A T C T C C C T A
7 A A C T C T G T C C C C C C A A A T A T C T C C C T A
995 A A C T C T G T C C C C C C A A A T A T C T C C C T A
1121 G T C T T T G T C C C G C C T A A T A T C T G C T C
336 G T C T T T G T C C G C C T A A T A T C T G C T C
1606 G T C T T T G T C C C G C C T A A T A T C T G C T C
285 G A A T T T G T C C C C C C A A A T A T C T C C G T C
781 A A C T C T G T C C C C C C A A A T A T C T C C C T A
1796 A A C T C T G T C C C C C C A A A T A T C T C C C T A
```


Model

Nuclear gene

- Sequence consists of 20 «amino acids»
- Each substitution is equiprobable;
- Each residue has property, which changes from 0 to 19
- Each sequence consists of 2 parts: «the neutral one, where sequences does not affect fitness of the organism, and **non-neutral one**, where fitness is function of the sum of properties:

$$f = \sum (x_i - s_i)$$

```
ec1733 MAVRECGATHRTD
ec455 MAVRECGATDRTD
ec954 MAVRECGATHFPYN
ec829 MAVRECGATHRLH
ec899 MAVRECGATDRWD
ec1617 MAVRECGATDRWD
ec672 MAVRECGATHFPYN
ec1233 MAVRECGATHRTD
ec1274 MALRECGATQITD
ec654 MTVRECGADHWTS
ec463 MAVRECGATDRTD
ec1010 MTVRECGADHWTS
ec1128 MAVRETGATHRTH
ec1736 MAVRECGATQIQD
ec1201 MAVRECGATQITP
ec947 YAVRECGAVHRHD
ec210 MAVRECGATDRTD
ec1411 MAVRRIGATHRTY
ec316 MAVRECGATDRTD
ec955 MAVRECGAHHWQA
ec699 MAVRECGATQIQD
```

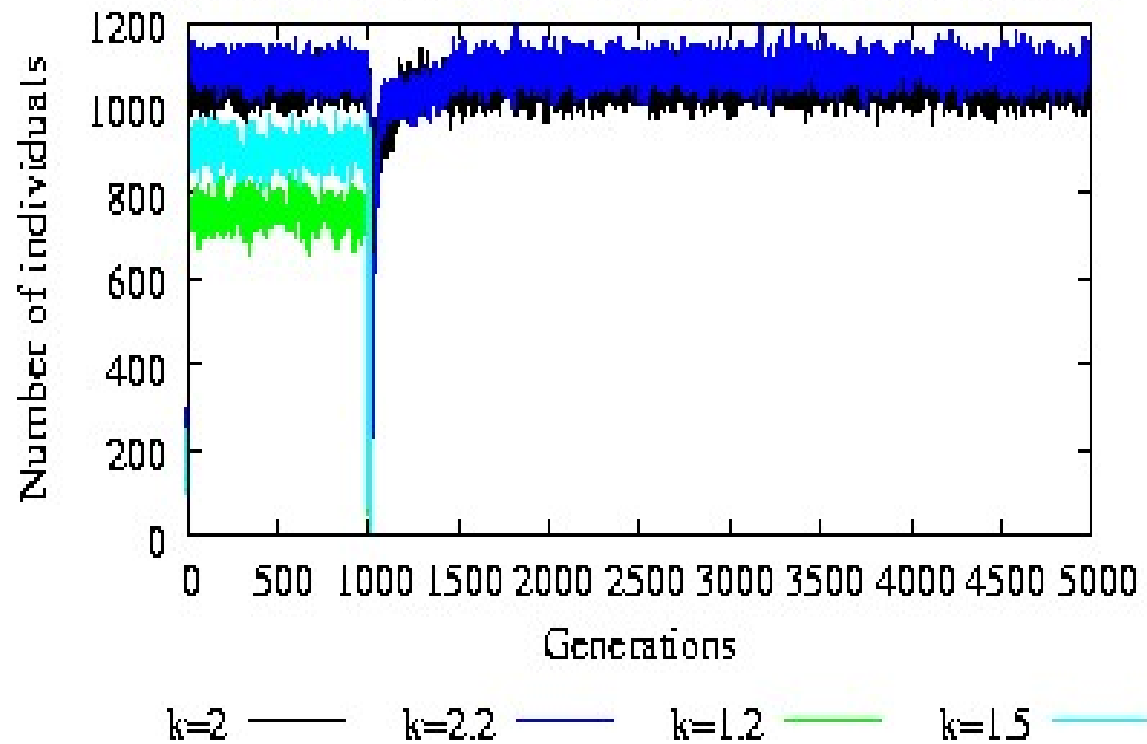
Interactions between organisms

- 1) The organisms who reached maximal age allowed die;
- 2) Each of the survivors determine it's gender for this round;
- 3) Density-dependent decrease of survival rate is introduced;
- 4) Pairs («families» are formed according to the random gender-choice made at stage (2);
- 5) Each pair «decides» how many children will they have. The number of children is Poisson-distributed;
- 6) Children are produced (if their number exceed the max. Number, their number is reduced proportionally), They inherit mother's «mitochondrial» marker and one «nuclear» allele from each parent;
- 7) Progeny fitness according to the group rule is calculated (Full dominance in this study)
- 8) New density dependent viability is calculated and the unlucky «die», the rest increase their age;
- 9) Stage (1) occurs. Life goes on.

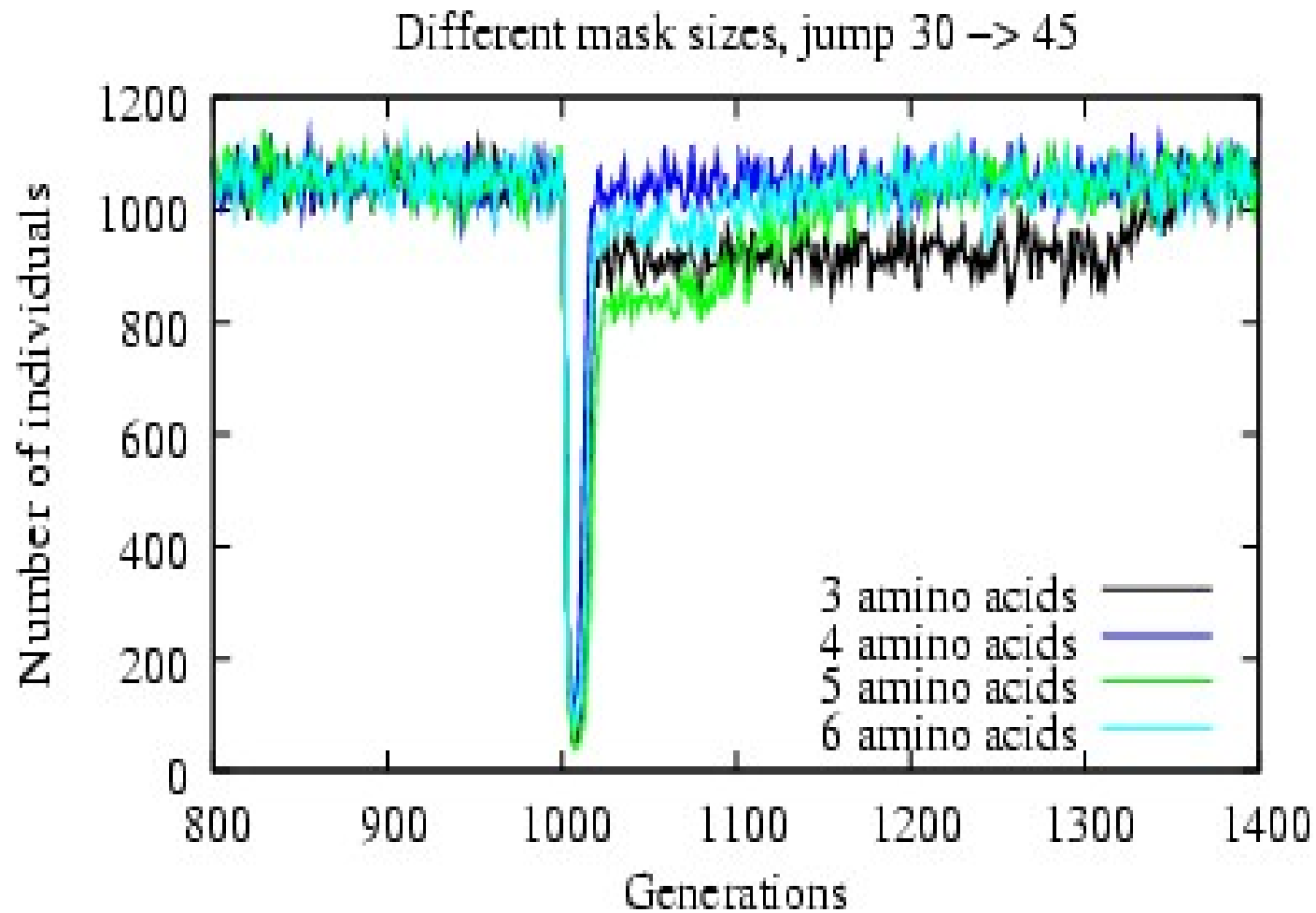
$$f = f \left(1 - \frac{n}{N} \right)$$

Mask size, number of kids etc.

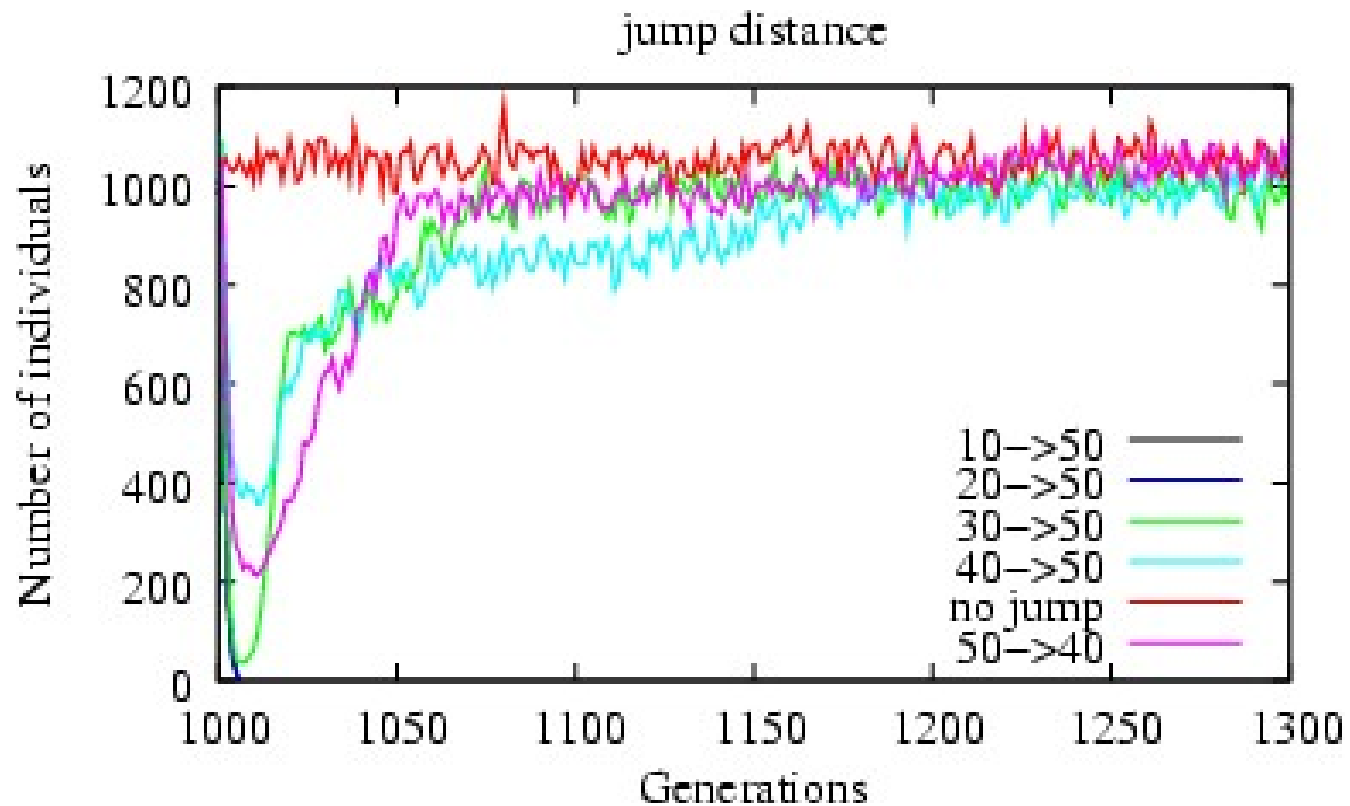
Mask involves 4 aminoacids, the value jumps from 30 to 45

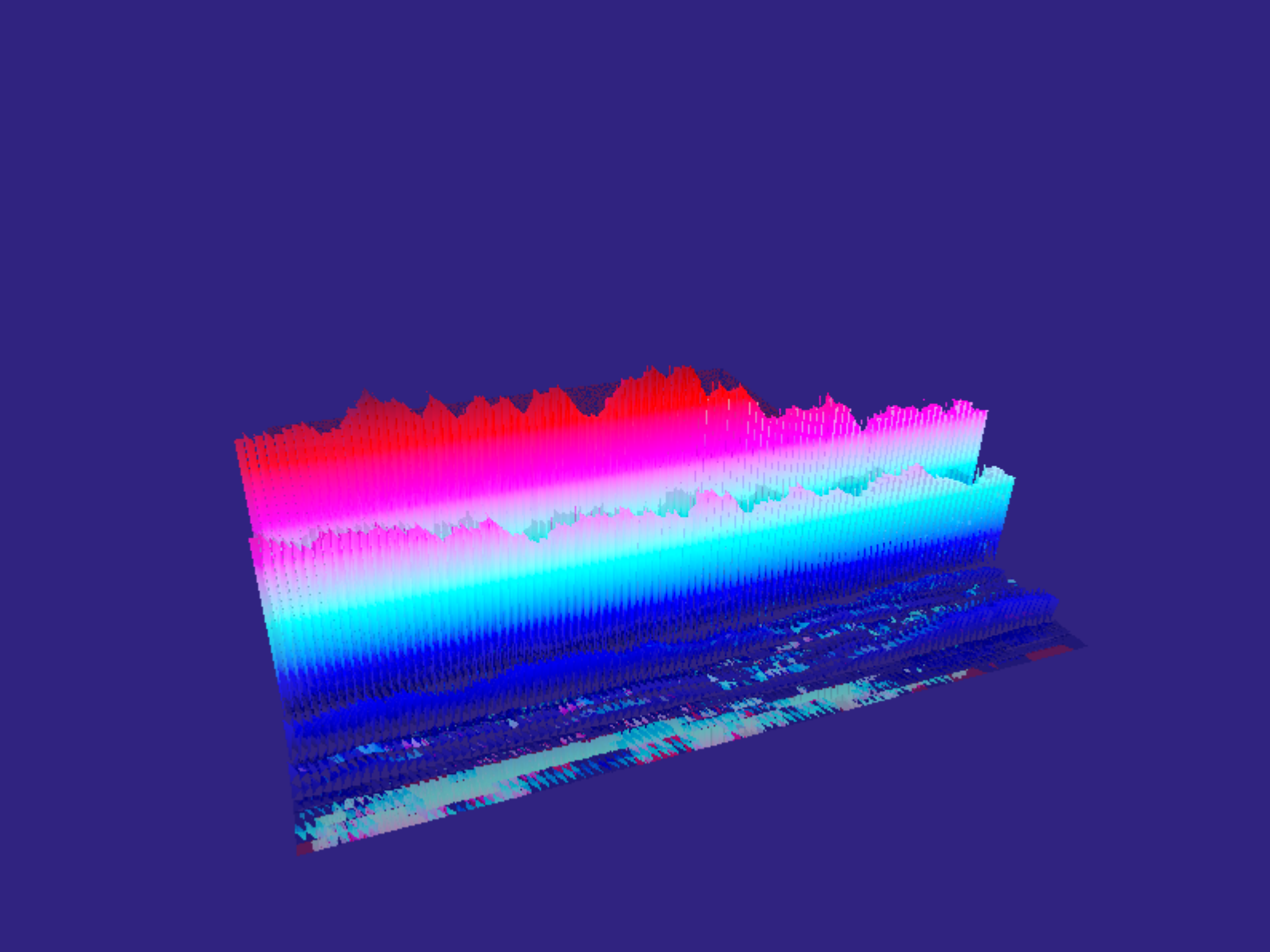


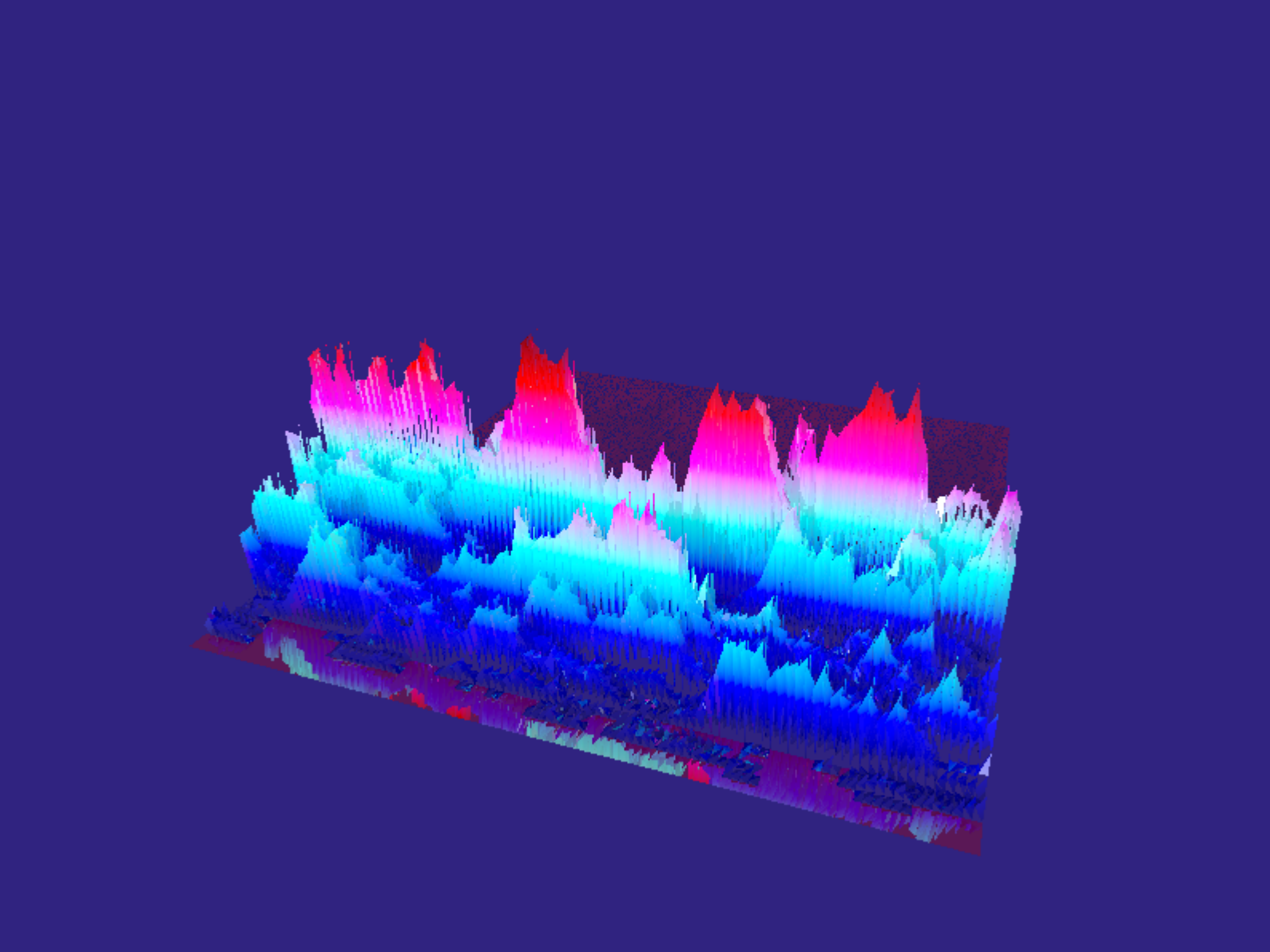
Group size influences relaxation curve after a jump



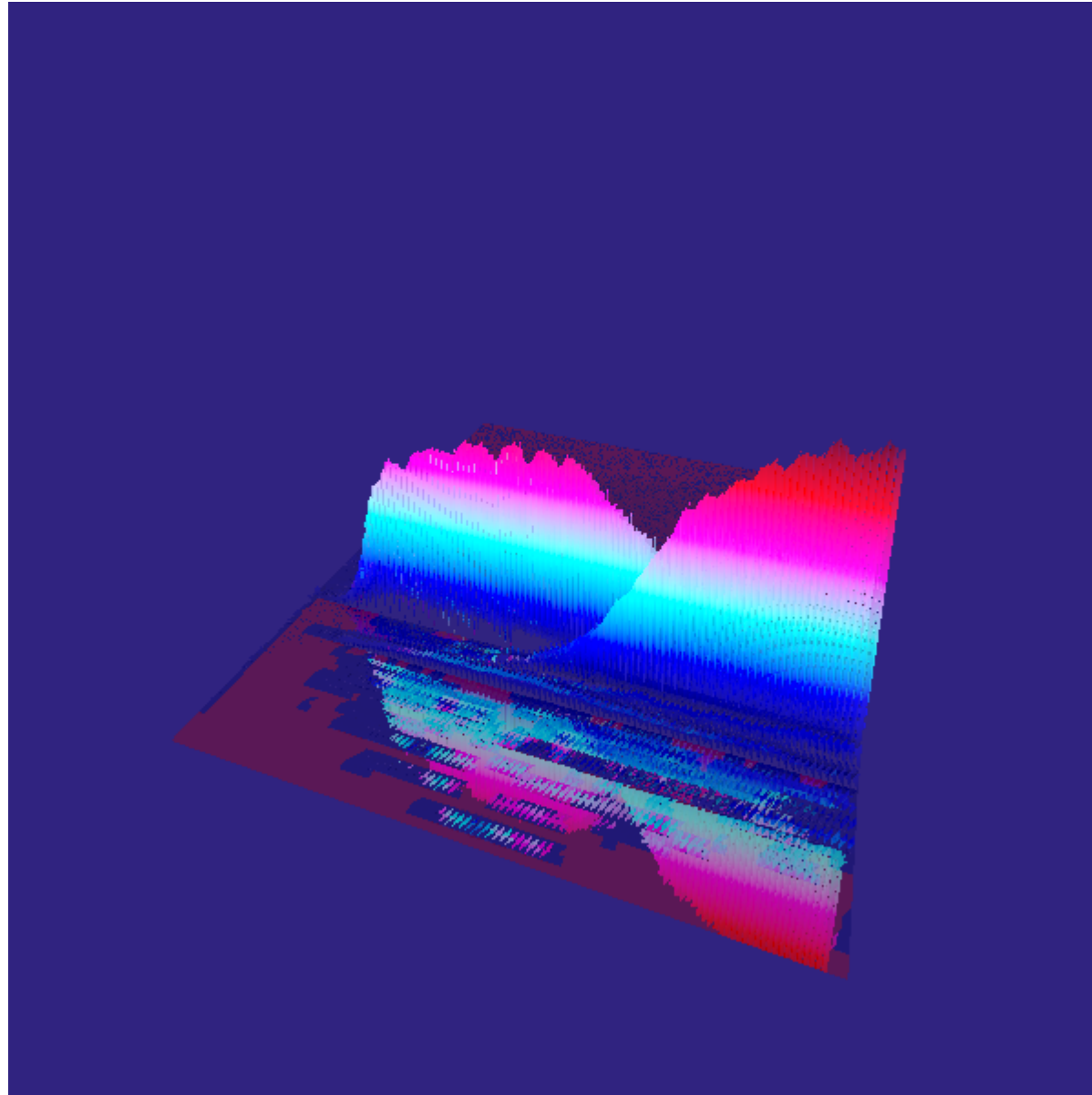
Relaxation curves depend on exact start and target values



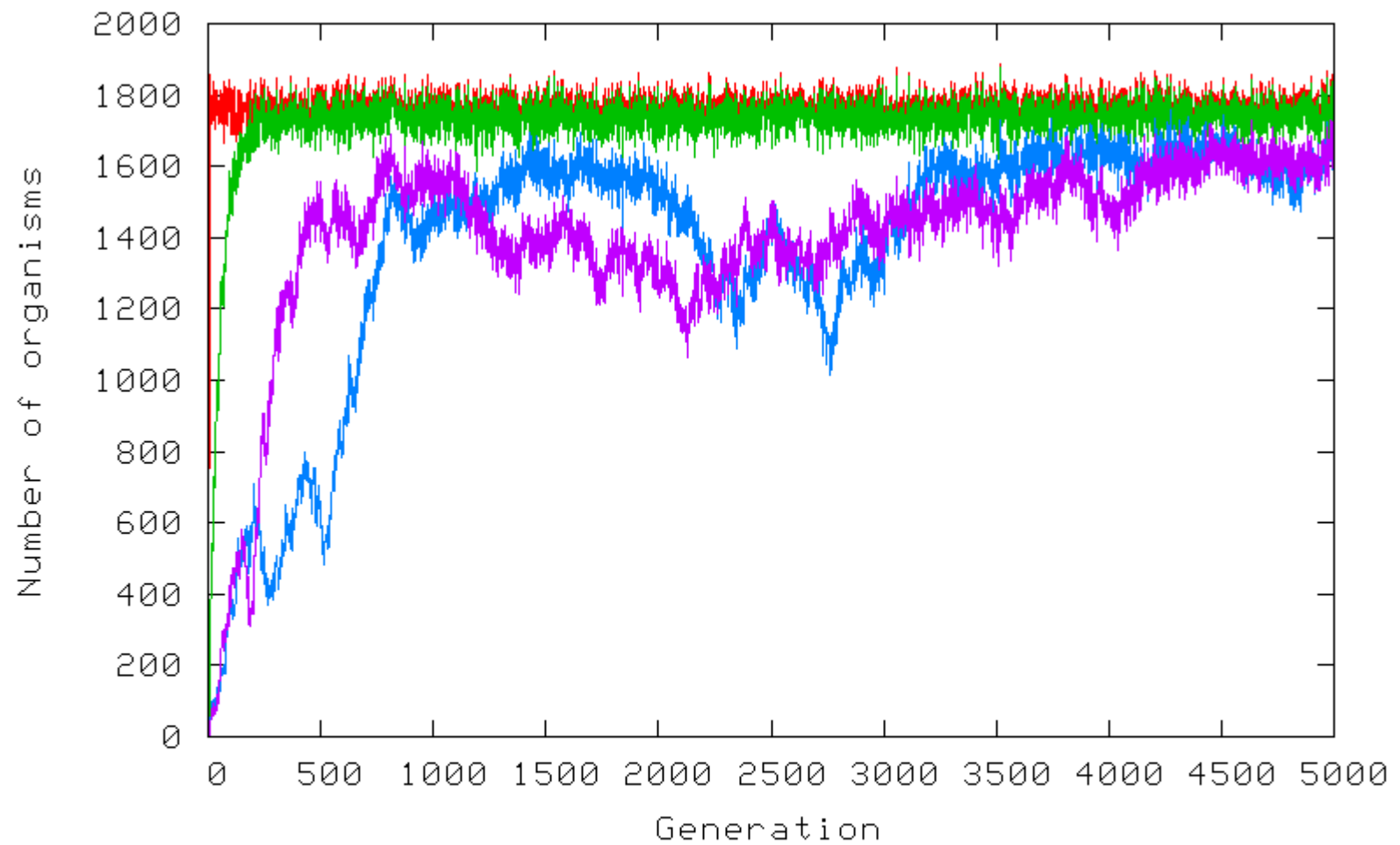




Optimum shift without coordinated evolution

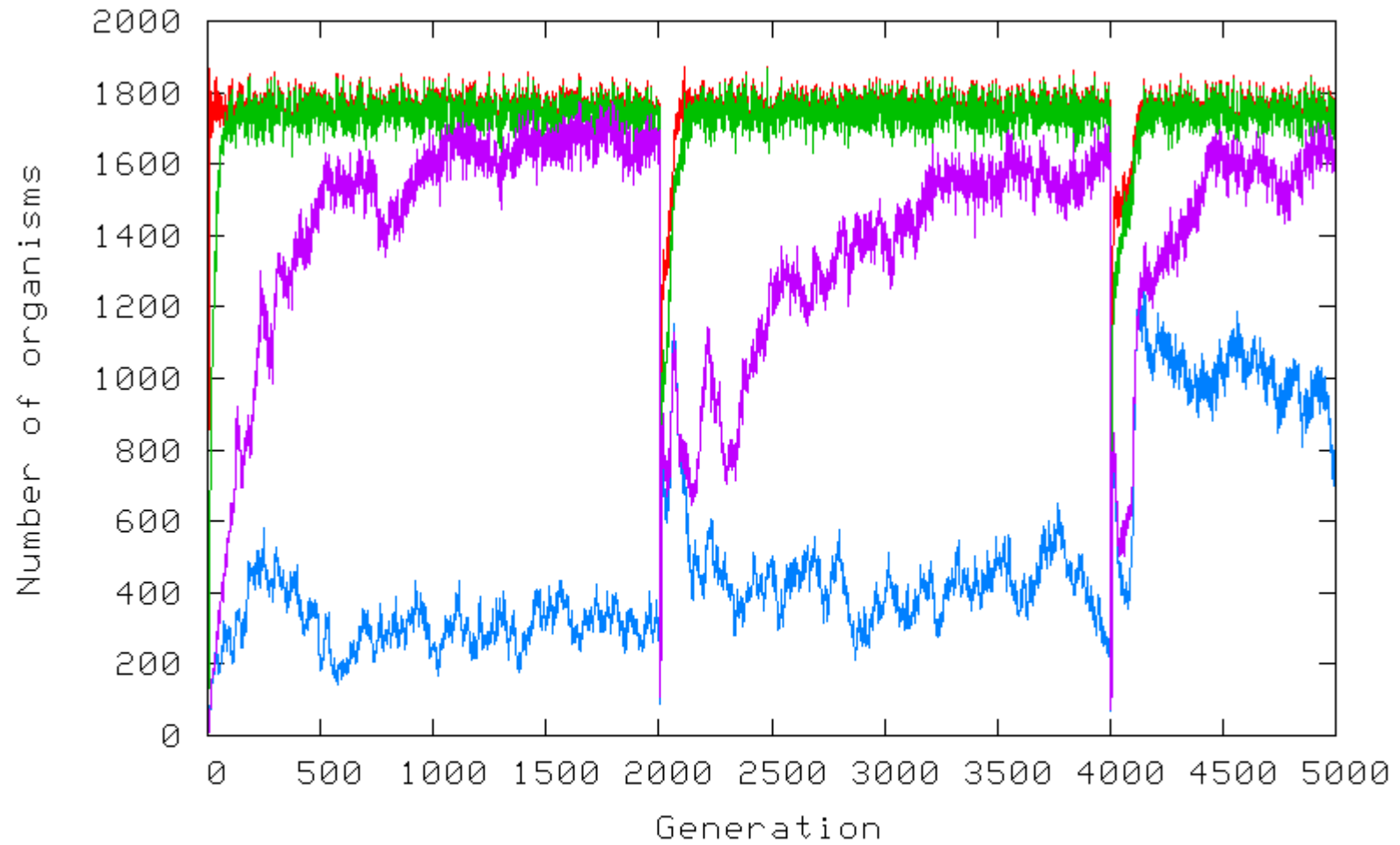


Two optimal values co-existing do not cause co-ordinated substitutions

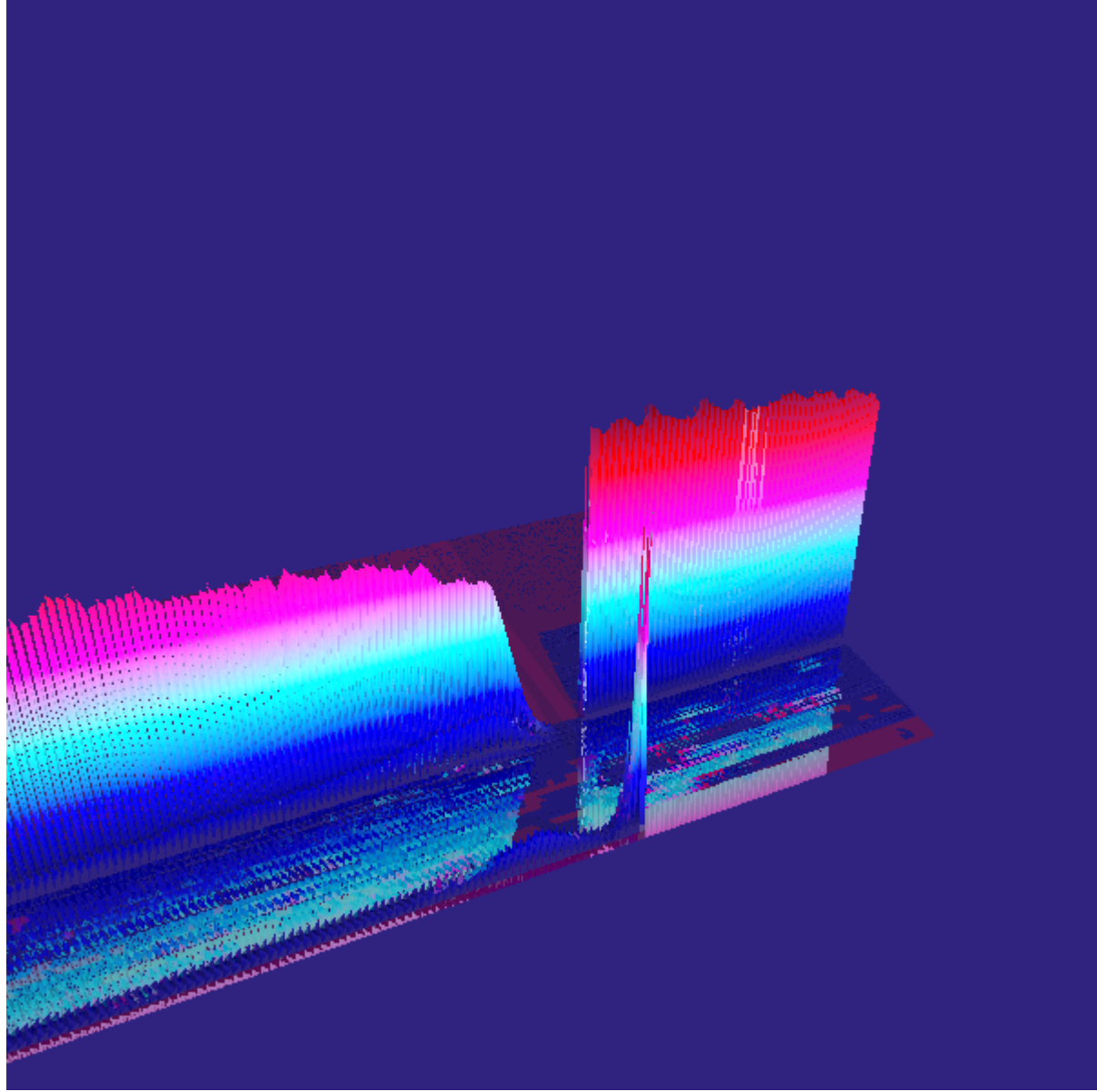


Number of organisms ———
Number of heterozygotes ———
Heterozygotes in C-group ———
Heterozygotes in N-group ———

Double jump



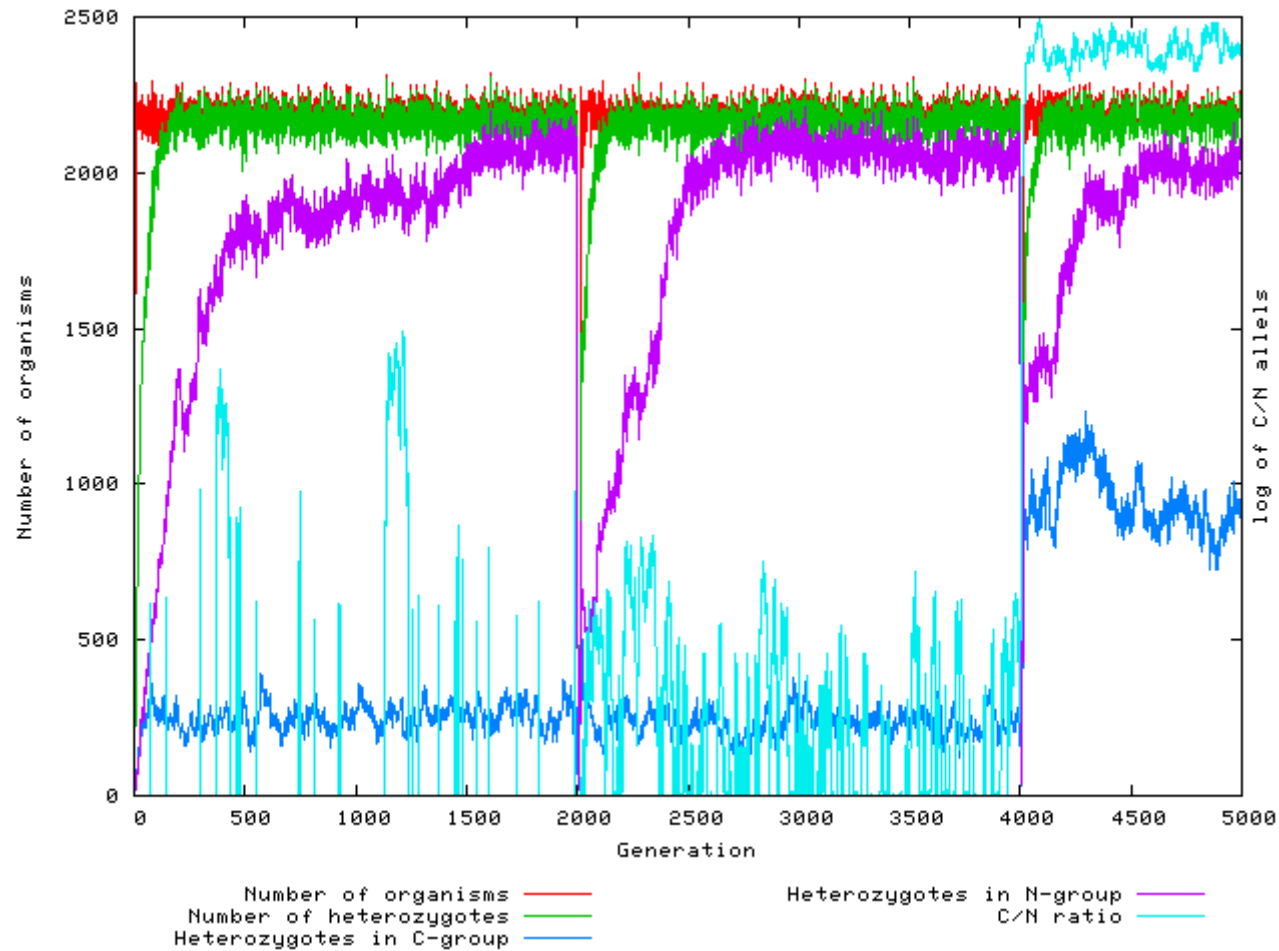
Number of organisms ————
Number of heterozygotes ————
Heterozygotes in C-group ————
Heterozygotes in N-group ————




```
sampleAA.phy
File ▾ Props ▾ Sites ▾ Species ▾ Foote
sel=3 1 Seq:23 P5s
STARTING I F Q R X X X X X X X X X X X X
Spec1119 K W M R I L M D Q G T G P C T
Spec1530 K W M R I L D D Q G T G P C A
Spec907 K W M R W N M C Q G Y V P L A
Spec34 K W M R W L Y D Q G O V P Q A
Spec1324 K W M R T L M D R S Y V P D A
Spec1742 K W M S W M Y D Q G Y V P L A
Spec228 K W M R W L Y D Q G O V P Q A
Spec652 K W M R W L Y D Q G O V P L A
Spec1176 K W M R W L Y D Q G O V P L A
Spec1837 K W M R W L M D L P Y Q P P A
Spec387 K W M R W L Y D Q G O V P L A
Spec1132 K W M R W R M L Q G Y P Y L A
Spec1689 K W M R W M Y D Q G Y V P L A
Spec1147 K W M R W R M L Q G Y P Y L A
Spec1864 K W M R I L M D G G T G P C T
Spec1883 K W M R W L M H Q P Y V P L A
Spec2026 K W M R W M Y D Q G Y V P L A
Spec344 K W M R W M Y D Q G Y V P L A
Spec104 K W M R W R M E Q G Y I P L E
Spec1207 K W M R W M Y D Q G Y V P L A
Spec713 K W M R W C M E Q G Y I P L E
Spec820 H W M R W L M D K G Y V M L A
Spec1871 K W M R W M Y D Q G Y V P L A
Spec727 K W M R W K Y D Y G Y V P D A
Spec373 K W M R W K M D Q G Y E P L E
Spec1073 K W M R G C M E Q G Y I P L E
Spec390 K W M R G C M E Q G Y I P L E
Spec1107 K W M R I L M W Q G T G P C T
Spec594 K W M R W L Y D Q G O V P L A
Spec1628 K W M R W L Y D Q G O V P L A
Spec305 K W M R W K Y D Q G Y V P D A
Spec995 K W M R W L M H Q P Y V P L A
Spec1223 K W M R W N M C Q G Y A P L A
Spec1834 K W M R W A M D Q G Y V S A Y
Spec1741 K W S R W L M D Q G Y V V M A
Spec1813 K W M R W L M D P P Y C P P A
```

Double shift of optimal value results in coordinated sunstitutions!

Repeated change of environment promotes coordinated substitutions



Very preliminary conclusions:

- *Number of group members must be more than 2. Coordinated substitutions become more frequent with the increase of number of group members.*
- *Periodic shift of optimum promotes coordinated substitutions*
- *New optimal value of a property must be reached at a single mutation*