

PROYECT TITLE:

Eco-evolutionary biodiversity dynamics in river-lake dendritic networks

MAIN AIM: Add value to existing Switzerland-wide fish metacommunity samplings by developing a theoretical framework to discriminate the most important mechanisms driving biodiversity and radiation dynamics at small and large spatial scales

MAIN TASKS: 1) Extend julia codes to a second model adding density-dependent biodiversity dynamics; 2) Use the ProjeLac database together with P/A lake-species matrix data, pairwise distance-altitude matrix data and phylogeny data; 3) Implement ABC methods by adding them to the julia or doing the ABC in r or similar languages.

POINTS TO DISCUSS: 1) Extending julia to a second model; 2) Checking the data from *ProjeLac*; 3) Use flow 30-year time series data (Philipp contribution)?; 4) Joining outputs from julia with r or similar to perform an ABC; 5) Draw dendritic networks; 6) Writing methods and results

Inferring radiations with Approximate Bayesian computation

Charles N. de Santana^{1,*}, Gregor Kalinkat^{1,*}, Jakob Brodersen¹, Ole Seehausen¹,
Carlos J. Melián¹ **AUTHORSHIP ORDER MAY CHANGE**

¹Department of Fish Ecology and Evolution
EAWAG, Swiss Federal Institute of Aquatic Science and Technology
Center of Ecology, Evolution and Biochemistry, Kastanienbaum, Switzerland

Keywords: α , β and γ -richness,
anagenesis speciation, cladogenesis speciation, metacommunity dynamics,

Type of Article: Article

Number of figures: X; Number of tables:

* Corresponding author: e-mail: charles.desantana@eawag.ch

Abstract

2

.

1 Introduction

4 **WRITTEN MOSTLY BY OLE: REFERENCES NOT INCLUDED IN THE .bib** Fresh-
water communities of fish in lakes are among the richest vertebrate species assemblages
6 on Earth (Vadebonceur et al. 2011; other refs), but they are also among those with
the steepest rates of species loss (Vonlanthen et al. 2011; Seehausen et al. 1997; other
8 refs) due to eutrophication, climate change, invasive species and the interaction of these
factors (Vorosmarthy et al 2010). Yet, surprisingly little theory and even less data is
10 available to predict how the effects of these environmental factors depend on the history
and mechanism of assembly of freshwater communities. Lake fish communities range from
12 little isolated entirely immigration-assembled communities to nearly entirely speciation-
assembled, with all intermediate combinations of both processes (Wagner et al. in press).
14 There are many reasons to expect that communities assembled by different mechanisms
can respond very differently to environmental change.

16 Freshwater bodies in general, and lakes in particular, support exceptionally high num-
bers of endemic species per unit area. Several examples have shown that diversity in
18 these systems can suddenly collapse when ecosystems are perturbed, and it does often
not recover after ecosystem restoration (refs for Lake Victoria; Swiss lakes; Laurentian
20 Great lakes; Lake of Mindanao). Very little is known about how environmental change
factors interact with each other and with the assembly history of species communities
22 to determine loss of diversity and consequences for ecosystems. Freshwater ecosystems,
and large lakes in particular, are strongly geographically isolated for organisms that lack
24 terrestrial or airborne dispersal phases such as fish. Large lakes have unique and diverse
habitats that provide resources for large and very diverse populations, but colonization of
26 lake-specific habitats like the pelagic and profundal zones often requires in-situ evolution
and speciation because river fish can rarely colonize such habitats and lakes are often

28 extremely strongly isolated from other lakes. Lakes hence function like isolated oceanic
islands where in-situ evolution is a major source of biodiversity (Hudson et al. 2011;
30 Wagner et al. in press).

Theory suggests that the assembly of communities by in-situ evolution and by dis-
32 persal from a regional pool make very different predictions for richness and abundance.
For instance, whereas the species-area curve in dispersal-assemblies typically has a slope
34 around 0.3, it is much steeper (near 1) for communities assembled by speciation (Rosen-
zweig 2001; Wagner et al. in press). Theory and data suggest that locally evolved species
36 are often more abundant than others (Rosindell & Phillimore 2011). Yet, little is known
about how the topology of 3D networks change the assembly process of endemic and non-
38 endemic species in metacommunities. While clearly a questions of considerable relevance
to conservation in general, this is especially pressing for freshwater ecosystems. Their
40 heavy reliance on evolution for the building of species diversity is expected to influence
the ways these ecosystems respond to environmental change. One might for instance ex-
42 pect more dramatic responses in speciation-assembled communities because of the steeper
scaling of richness and abundance with area and the often much larger per area-richness
44 (Wagner et al. in press). Another important difference is that in speciation-assembled
communities even transient and reversible perturbation of niche structure can lead to
46 irreversible collapses of species diversity within just a few generations (Seehausen et al.
1997; Vonlanthen et al. 2011).

48 Here we develop a process-based modeling framework for island biogeography with
speciation in a 3D dendritic network of lakes connected by streams to predict the emer-
50 gence of species richness through immigration, speciation and in situ radiation. We focus
on fish but our model may be applicable more widely to lacustrine biota. However, fish
52 account for more than 60% of all endemic animal species in lakes globally (Vadebon-
ceur et al. 2011). At the same time fish represent critical links in aquatic food webs.

54 To parametrize our model and test its utility, we make it spatially explicit and apply it
to the network of large subalpine lakes of central Europe, a hotspot of freshwater fish
56 endemism in Europe (Kottelat Freyhof). Our empirical data includes >20 large lakes
and connected rivers located in three archipelagos of large subalpine lakes north (Rhine),
58 west (Rhône) and south (Po) of the central Alps. In each archipelago we have sampled
a similar range of lakes in terms of size, elevation but the archipelagos have distinct as-
60 sembly histories. The communities of all lakes assembled within the Holocene, i.e. in the
past 15000 years. But whereas on the South-side of the central Alps, there was rapid
62 recolonization by many lineages from nearby southern refugia with only few glacial relict
species surviving, northern and western lakes were probably colonized rapidly by just a
64 handful of cold water fish lineages that may have survived the Pleistocene in near refugia,
but then were colonized by all the other lineages only slowly from distant eastern (Black
66 Sea) refugia. They retained a diverse set of cold-water adapted glacial relicts, many of
which diversified into ecologically diverse lake endemics. As a consequence, lake fish com-
68 munities range from entirely immigration assembled to those with large contributions of
speciation. Locally evolved species fill ecological key roles in most northern and some
70 western lakes, but not in the southern lakes. We compare data with model predictions of
species richness, endemism and radiation.

72 **COMMENT ON NEW THEORY IN 2D DENDRITIC NETWORKS AND POPU-**
LATION DYNAMICS To generate theoretical expectations for diversity in each lake, we
74 developed an individual-based stochastic simulation model building on recent evolution-
ary extensions of island biogeography theory (Rosindell and Phillimore 2011; hereafter
76 RP). We are asking how richness scales with isolation and lake size, and how the assem-
bly process affects this scaling and also the abundances of species. Island biogeography
78 theory has been almost exclusively tested in terrestrial island systems. This is the first
evolutionary island biogeography model for freshwater systems. We wish it will help

80 understand the evolutionary and ecological dynamics of lacustrine species assembly in-
cluding the formation of endemism and radiations. We expect that our results will also
82 contribute to predicting and explaining variation in the susceptibility of ecosystems to
environmental change.

84 **MAIN RESULTS** Our modeling framework can be extended in several directions in
the future, including predicting intraspecific genetic and trait diversity and the genetic
86 distinctiveness of populations. With the increasingly wide access of Next Generation
Sequence methods for non-model organisms, population genomic data can now easily
88 be collected to test the predictions. Other possible extensions are to merge our island
biogeography model with whole-island regression and habitat-unit models in 3D (i.e.,
90 lake depth or an ecological gradient) to study the effects of habitat diversity and lake
size separately, with habitat occupation models (Gusian Rahbek 2011) and with trophic
92 theory of island biogeography (Gravel et al. 2011).

2 Methods

94 **START WITH A BRIEF INTRODUCTION OF THE AT LEAST TWO MODELS TO
TEST: WHY DO WE EXPLORE THESE TWO SCENARIOS? HOW DO THEY DIF-
96 FER AND WHY? HOW ARE WE GOING TO COMPARE THEM USING ABC? (See
a flow chart in Fig. 2)**

98 **2.1 Two scenarios**

EXPLAIN THE TWO SCENARIOS IN DETAIL

100 **FROM HERE METHODS ARE COMMON TO THE TWO SCENARIOS**

2.2 3D dendritic networks

We generate landscapes consisting of randomly located sites with a spatial location given by x_i , y_i and z_i with z_i representing the altitude (Figure 1 **SHOW A 3D DENDRITIC NETWORK**). The hydrological distance between each pair of nodes is given by the Euclidean distance. We distinguish two types of nodes. Nodes that represent lakes with population dynamics and nodes that represent streams without population dynamics. The third dimension, z_i represents elevation randomly sampled from a uniform distribution with minimum and maximum elevation ranging between 10m and 1000m, respectively.

2.3 Population dynamics

We run an individual-based model based on demographic stochasticity on 3D dendritic networks using two speciation modes. This will enable us to predict the effect of isolation in 3D dendritic network on the number of endemic species, radiation zones and species richness. Individuals organisms die at a constant rate, thus species' populations can dwindle the population size and become extinct via the process of demographic stochasticity or ecological drift (ref). The smaller the population size, the greater the probability of imminent extinction. The probability to choose a lake to have demographic events (i.e., births and deaths) is the same across all the lakes. Thus we assume that all the lakes have the same volume or area.

Each lake i is at its carrying capacity with the number of individuals in each lake fixed. With every death of an individual on a lake, a gap opens. With probability $1 - m - \nu - \mathcal{K}$ this gap will be filled by offspring from another living individual on the lake, and with probability m , ν , or \mathcal{K} the gap will be filled by an immigrant, from a regional migration event, or from a cladogenesis initiation event, respectively (see below section "Dispersal dynamics"). We model two distinct modes of speciation; cladogenesis and anagenesis.

DESCRIBE HOW WE MODEL CLADO- AND ANAGENESIS. SPECIFICALLY WE

HAVE TO DESCRIBE THAT CLADOGENESIS IS CORRELATED WITH ANAGENESIS IN BOTH SCENARIOS

2.4 Dispersal dynamics

We frame our hypothesis in dendritic networks using elevation and the hydrological distance between each pair of nodes (i.e., lake-stream, stream-stream, stream-lake). We start by describing the simplest scenario in a 2D network. Dispersal rate between two nodes is a function of their hydrological distance. This leads to the dispersal rate of species k from lake j to the lake receiving the migrant, the lake i

$$m_{ij}^k = m \left(\frac{1}{d_{n,j} + \sum_{n=1}^{\mathcal{N}-1} d_{n+1,n} + d_{i,n+1}} \right), \quad (1)$$

where $d_{n,j}$, $d_{n+1,n}$, and $d_{i,n+1}$ are the hydrological distances between lake j and node n , node n and node $n+1$, and node $n+1$ and lake i , respectively, and m is the intensity of dispersal rate. An extension of equation (1) to a 3D network considers dispersal rate not only as a function of the distance, but also as a function of elevation distance between lakes or nodes in the network. This leads to the dispersal rate of species k from lake j with elevation e_j to lake i with elevation e_i

$$m_{ij}^k = m \left(\frac{1}{d_{n,j}^x + \sum_{n=1}^{\mathcal{N}-1} d_{n+1,n}^x + d_{i,n+1}^x} \right), \quad (2)$$

with x

$$\begin{cases} x = 1, & e_{em} \geq e_{im}, \\ x = \frac{(1+c(e_{im}-e_{em}))}{1}, & e_{im} > e_{em}, \end{cases} \quad (3)$$

where c is the upstream migration cost between two nodes, e_{em} and e_{im} represent the elevation of the emigration (i.e., e_j , e_n , e_{n+1}) and immigration (i.e., e_i , e_n , e_{n+1}) nodes.

We note that equation (2) is the same than equation (1) when $c = 0$, that is, when upstream migration cost is not present as part of the dynamics (i.e., from 3D to a 2D network). We consider downstream migration has not cost and use the same hydrological distance between two nodes (i.e., $x = 1$). At a given upstream migration cost, the larger the difference in elevation between two nodes the lower the dispersal rate between these two nodes (Figure 3).

2.5 Regional species pool

Dispersal dynamics between the regional species pool and the dendritic-network is represented throughout an entry point considered in each basin as the point at the lowest elevation. Species in the regional pool and species in the dendritic-network form two distinct sets. New species can emerge with very low probability, ν , from the regional species pool. We consider an extremely diverse regional species pool, containing an infinite number of species. Because of the infinite number of species in the regional pool, we assume that every immigration event introduces a new species. Immigration of a new species corresponds to (allopatric) speciation in the context of metacommunity models (?). SHOULD WE EXPLORE DIFFERENT DISTRIBUTIONS OF ABUNDANCE IN THE SPECIES REGIONAL POOL?

2.6 Speciation dynamics

2.6.1 Cladogenesis

To model cladogenesis we follow a version of protracted speciation (ref) under which speciation is a gradual process rather than an instantaneous event. Variant individuals appear at per capita rate \mathcal{K} and are each conspecific with their parent species. The probability \mathcal{K} of cladogenesis initiation is expected to be small; as this can be interpreted as implicitly encompassing gene flow' homogenizing influence on speciation. Variants may

170 give rise to offspring according to the previous demographic rules and if any offspring
survive after a transition time of τ generations has passed, then they are treated as a
172 new species. This way of modeling cladogenesis does not necessarily imply sympatric
speciation within a lake and it is equally consistent with intralake allopatric or parapatric
174 speciation. **SHOULD WE MENTION SOMETHING ABOUT CLADOGENESIS ON 3D
DENDRITIC NETWORKS?**

176 **2.6.2 Anagenesis**

We model anagenesis extending previous approaches (ref) to a 3D dendritic network **AND**
178 **ADD USTREAM COST TO ARRIVE TO HIGH ELEVATION LAKES (EQUATION 2).**

In our model the first immigrant to a lake is always regarded as being a variant but it is
180 not yet an endemic species. If the lake variant population survives for τ generations with
no further immigration of conspecifics, then it will be recognized as a new species. Prior
182 to τ generations elapsing, every further conspecific immigrant joining the lake population
retards anagenesis by increasing the remaining time to speciation completion by G (a
184 small pre-defined amount); in this way sufficiently high gene flow makes speciation near
impossible. We never allow the remaining waiting time to speciation be greater than τ .
186 Completed speciation events cannot be reversed; once a lake endemic is formed, further
immigration of the mainland sister species will not affect it, but will instead form a new
188 variant that can potentially become another endemic species. The complete model with
all the parameters used in explained graphically in Figure 2.

190 **2.6.3 Patterns in the data: α -, β - and γ -richness, isolation and phylogeny**

α -, β - and γ -richness

192 Isolation

We calculate isolation of lake i as the sum of all distances between the focal lake i and

all other lakes in the network. This leads to the isolation of lake i ,

$$I_i = \sum_{j=1}^S \left(d_{n,j}^x + \sum_{n=1}^{\mathcal{N}-1} d_{n+1,n}^x + d_{i,n+1}^x \right), \quad (4)$$

where $d_{n,j}$, $d_{n+1,n}$, and $d_{i,n+1}$ are the hydrological distances between lake j and node n , node n and node $n + 1$, and node $n + 1$ and lake i , respectively, and S is the number of lakes in the network.

Phylogeny

ASK OLE OR JAKOB FOR THE DATA

2.7 Simulations

THIS IS A VERY PRELIMINARY PARAGRAPH FROM THE THEORETICAL PAPER Simulations were done with one initial metapopulation containing only one species across S lakes with $S = 100$, and each lake i containing a fixed number of individuals, $J_i = 100$. Subsequently we ran simulations following equation (2). The number of generations per replicate was 10,000 and one generation, G , is 10000 iterations (SxJ_i). Results plotted in xx and xx were obtained after running 20 replicates per cost, c , considering a priori rates of dispersal from the regional species pool, ν , metacommunity-specific background dispersal, m , and probability of cladogenesis initiation, \mathcal{K} , following a uniform distribution with values in the range $\nu \in [10^{-4}, 10^{-3}]$, $m \in [2 \times 10^{-2}, 2 \times 10^{-1}]$, and $\mathcal{K} \in [10^{-5}, 10^{-4}]$, respectively. Local birth rates for each metacommunity, λ , were obtained as $\lambda = 1 - \nu - m - \mathcal{K}$. Transition time for cladogenesis, τ , and the number of generations retarding anagenesis, \mathcal{G} , follow 2 generations and 20 time steps ($20/G$ generations), respectively. For the simulations, we set each time step to have one birth-death event in a randomly selected lake. Thus, natural mortality was set to 1.

2.8 Approximate Bayesian computation: model-data comparison

MOSTLY FOLLOW THE OUTLINE FROM THE AM NAT

218	3	Results
	4	Discussion

Symbol	Explanation
\mathcal{N}	Number of nodes
\mathcal{S}	Number of lakes
\mathcal{E}_i	Elevation lake i
J_i	Number of individuals in lake i
m_{ij}^k	Dispersal from lake j to lake i for species k
\mathcal{K}	Probability of cladogenesis initiation
\mathcal{G}	Number of generations retarding anagenesis
τ	Number of generations to have cladogenesis speciation
m	Intensity of dispersal
ν	Dispersal rate from the regional species pool
c	Upstream cost
λ	Local birth rate of metacommunity
d_{ij}	Geographical distance between lake or node i and j

Table 1: Symbols used and parameter values

6 Figure Legends

Figure 1. 3D dendritic networks. FIGURE SHOWN ONE OF THE 3D EMPIRICAL DENDRITIC NETWORKS ANALYZED

Figure 2. Flow chart representing the ABC method to infer the mechanisms that best predict biodiversity patterns in 3D dendritic networks.

Figure 3. Upstream cost in 3D dendritic network. Probability to disperse (y-axis) from lake j to lake i as a function of distance (x-axis) given equal volume for lake i and j and with 500m and 10m elevation, respectively. Upstream cost increase from blue line ($c = 0.0001$) to red ($c = 0.001$), and to green ($c = 0.01$). At short distances, the larger the upstream migration cost, the lower the migration probability between these two lakes. At large distances the effect of the cost decreases and distances becomes more important.

7 Figures

234

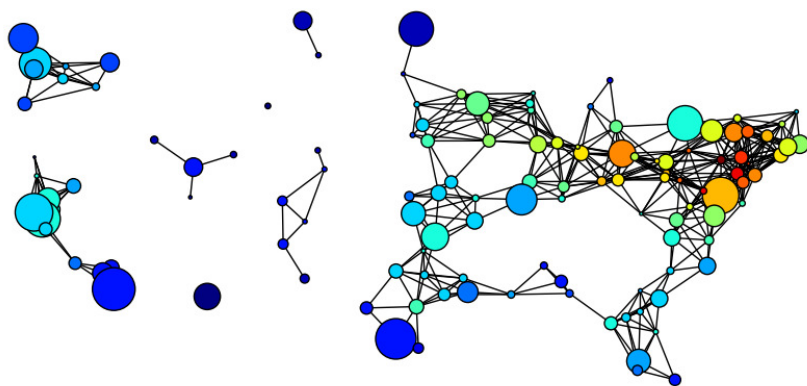


Figure 1:

ONE OF THE EMPIRICAL 3D DENDRITIC NETWORKS

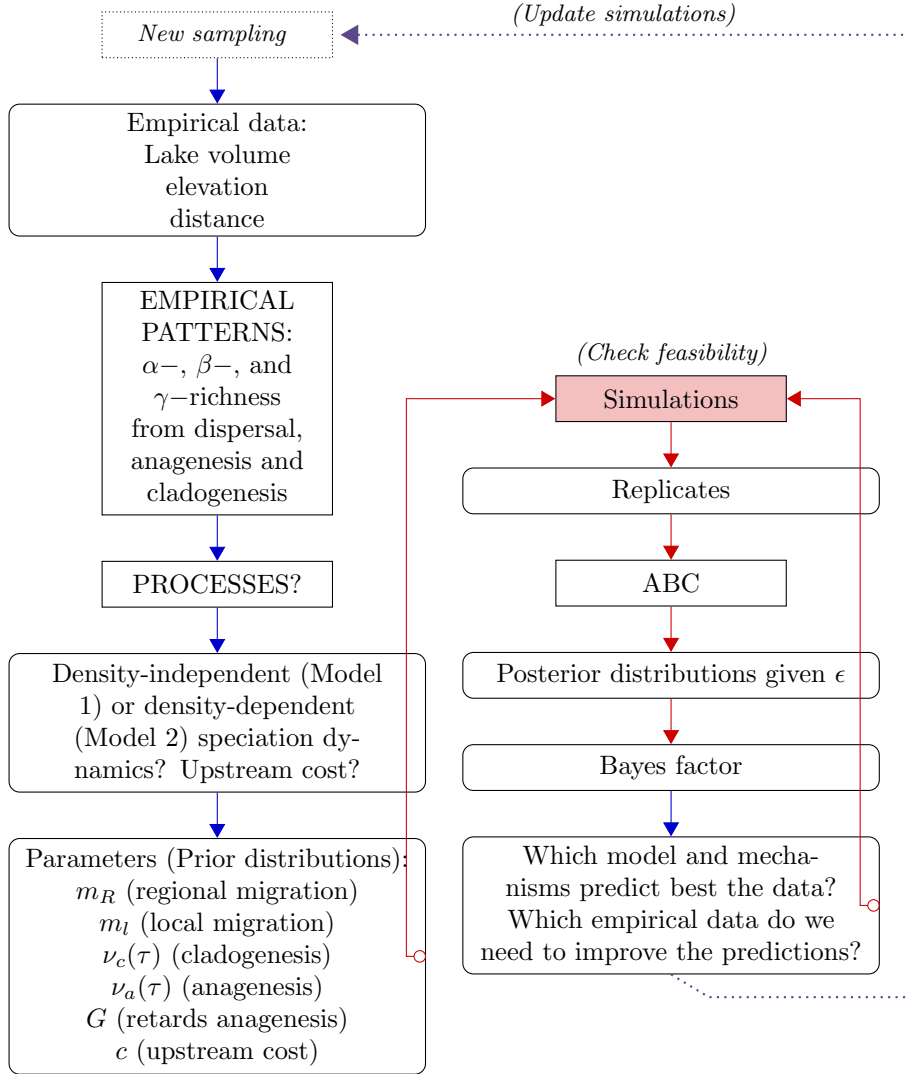


Figure 2:

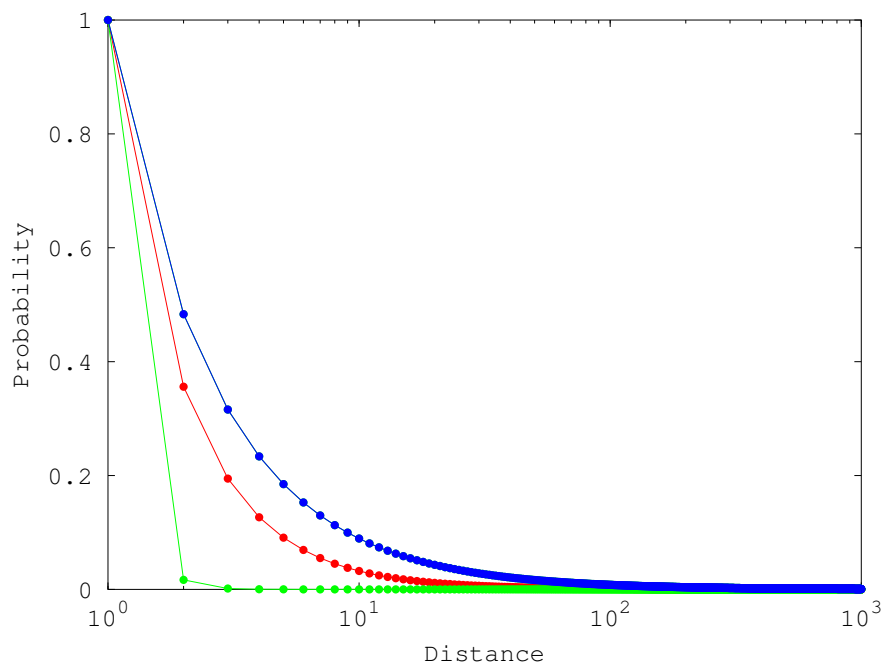


Figure 3: