Multi-Agent Simulations of Evolution and Speciation in Cichlid Fish

Ross Clement
Department of Artificial Intelligence and Interactive Multimedia
Harrow School of Computer Science, University of Westminster
Northwick Park, Middlesex HA1 3TP, UK
Email: clemenr@wmin.ac.uk

ABSTRACT

An agent-based simulation has been built to model speciation in cichlid fishes in the Great Lakes of Africa. A real natural system has been chosen as the target of simulation, rather than a generalised system. This focusses research towards open problems in cichlid biology, and provides a library of field research to drive the design and parametrisation of the simulation. Visualisations of the end results of simulations are presented to confirm that the simulated fish actually do speciate. A further experiment suggests that the potential for organisms to adapt has strong effects on the competitive exclusion principle, but that this cannot be used to explain the patterns of cichlid species found on rocky reefs in African lakes. The simulation has been written in pure Java rather than a general agent-based modelling platform. The reasons for using Java and a number of alternative platforms are described.

INTRODUCTION

This paper describes the use of a purpose built multi-agent simulation system (TDLP – "The Digital Lake Project") to investigate open problems in the speciation of cichlid fish in the African Great Lakes. Previous simulation systems used to investigate speciation (e.g. Kondrashov & Kondrashov 1999; Dieckmann & Doebeli 1999; van Doorn & Weissing, 2000) attempt to find 'general' rules for speciation, by modelling highly abstract ecologies. In general, these simulations model three properties of individuals: trophic preference/adaptation (preferred foods), some sexual signalling phenotype (e.g. colour) , and a sexual preference (e.g. the preferred mate colour of a female). These simulations were typically used to support the contentious concept of *Sympatric speciation* (Via, 2001; Turelli *et al*, 2001). Sympatric speciation is a form of speciation where a single ancestor species will divide into two (or more) daughter species, without any division of the ancestor species into geographically isolated subpopulations. TDLP originally was an implementation of the aforementioned models which added geographical details, to allow the investigation of conditions under which sympatric speciation would, and would not, occur. It has since evolved into a much larger and more detailed system, intended to study speciation in a real natural ecology.

The cichlid fish have been chosen as the biological system to be modelled by TDLP to allow cichlid field biology to inform the choice of model form and parameters. Also, it is by no means certain that all organisms speciate by the same methods. E.g. Sturmbauer (1998) discusses possible effects of the specific environment found in the African Great Lakes on cichlid speciation. Comparing and contrasting what is known about cichlid speciation and the biology of the galaxid fish of New Zealand suggests that speciation occurs by different methods. Diandromous (living part of their lives in salt water and part in fresh water) Galaxids appear to speciate when they become land-locked. The strong effect of major geological events on galaxid evolution can be seen in genetic research such as that in (Waters *et al*, 2001). Cichlids do not appear to need such an obvious physical barrier for speciation to occur, with the best example being the cichlids of Lake Barombi-Mbo (Schliewen *et al,* 1994), where speciation has occurred in a small, smooth walled, crater lake. These cichlids are often quoted as being the most likely example of real-world sympatric speciation. Making te choice of cichlids allows Cichlid biology (e.g. Barlow, 2000) to be used to guide simulation design and choice of parameters, removing the ambiguity (and, we believe, insolvable problem) in choosing parameters for a 'generalised' system.. Additionally, these biological studies provide a benchmark with which to compare the performance of TDLP. A relatively recent survey of cichlid biology (and open problems) can be found in (Coleman, 2001).

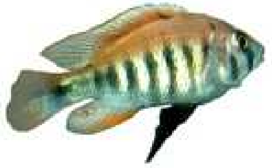

Figure 1: A Typical Cichlid from Lake Victoria[1]

Multi-agent system are frequently used in the study of ecological system (e.g. Mamedov & Udalov, 2002, Parrott & Kok, 2002). These systems are frequently presented as general systems useful for a wide variety of experiments. However, any such systems invariably make design choices for the users, with the most obvious example being that ecological simulation systems typically hard-code the names of species to which individual agents belong. This is clearly inappropriate for a system intended to model speciation, where the number and form of species will change over time, and pass through situations where species boundaries are not clear-cut.

---

[1] (c) 2003 M Pederson. (http://www.cichlidrecipe.com). Used by permission.

Proceedings 15th European Simulation Symposium
Alexander Verbraeck, Vlatka Hlupic (Eds.)

In the remainder of this paper, the current state of TDLP is described, and its use is illustrated by two experiments. First, we demonstrate the speciation (one ancestor species splitting into two) is occurring during 'typical' runs of the simulation, and show that the resulting groups of fish are what would be typically called 'different species'. To show the generality of the system, further experiments investigating the relationship between the speed by which fish can evolve, and the speed by which competitive exclusion occurs, are described.

CHOICE OF PLATFORM

TDLP is implemented in Java, the choice of language being made to allow the maximum degree of flexibility during the development and evolution of the project itself.

Considering potential agent-based platforms, we survey the systems of Mamedov & Udalov (2002), Parrot and Kok (2002). The system of Mamedov & Udalov (CENOCON) is "*a computer tool to build individual based models for simulation of population interactions*". Even at this high abstract level, there is a clear mismatch between the application domain of CENOCON, and that of TDLP. That is, modelling an ecology of (essentially) unchanging organisms in CENOCON, and modelling an ecology of changing (through evolution) organisms in TDLP. In CENOCON, animals are modelled by 38 individual properties, including detailed information about food preferences, growth parameters, body fat levels, gut capacity, and many others. These properties are all relevant to evolution, but the system as a whole lacks a genetic (or other) basis for allowing these to change over time. Animals in CENOCON are also labelled with a species name, which is entirely incompatible with the purpose of TDLP, where species are expected to emerge from a single-species population as a natural consequence of their ecology, and the environment they find themselves in.

Parrot & Kok's unnamed generic modelling system (hereafter PK) has many similarities to CENOCON, in that individuals are modelled in terms of digestion, body composition, food preferences, and species are again fixed and labelled with a species name. Over 100 properties are used to model animals, though these include dynamic properties such as the direction an animal is moving at any particular time. Again, the possibility of the generic properties that define an organism changing over time is not allowed (as befits the aim of modelling ecologies).

TDLP combines both ecological modelling (as we wish to measure the effect of ecological factors on speciation and evolution) and a varying genetic model. In terms of the sophistication of the ecological factors modelled, both CENOCON and PK are superior to TDLP, and may well be used to guide the future development of TDLP in this area.

If high-level biological/ecological modelling systems are not appropriate for investigating African cichlid biology, then an alternative is to use a more general agent-based modelling framework such as RePast. In some ways, the arguments on whether to use a system such as RePast are due to learning curves, and the perceived (rather than actual) advantages of RePast. The central motivation of this research is to learn as much as possible about the biology of the cichlids of the African great lakes, not to work on agent-based modelling itself. Hence, the choice of implementation was between the known (Java) and the unknown (Agent tools), and driven by a high degree of motivation to dive straight into the biological details. RePast uses a similar method for performing time-based updates to TDLP, with event the name of the *step()* method being the same. However, agents in TDLP are stored at differing abstract levels. Hence, time does not pass by a high level co-ordination agent repeatedly calling the *step()* method of individual fish. An intermediate level class *(Population*), not clearly matching the definition of either an agent or a container class, is called by the true container class *(Environment)* once per time period. The *Population* class then calls a number of methods of individual *Fish* agents, requesting single steps in such features as metabolism, feeding, and breeding. While it is possible to implement structures like this in RePast (as the classes in RePast can be sub classed and used alongside any other Java classes), there is still insufficient confidence that RePast would be a significant advantage over pure Java. For example, the amount of code in TDLP concerning generic agent functions is a small percentage of the total. This differs from smaller and simpler models (which have important roles in biological and other simulations) where the proportion of generic code is higher.

An alternative to using Java would have been the adoption of a specialised agent-based modelling platform (Gilbert & Bankes, 1999) such as RePast (http://repast.sourceforge.net/). The decision to use Java rather than a system such as RePast was mainly driven by the author's long experience with Java, and lack of experience with specialist platforms. A further factor was a lack of confidence that existing platforms would be flexible enough to handle future developments in the simulation system. At the time the choice of implementation language was made, even the nature of these future developments (such as the inclusion of the simulation of behaviour) was unknown. Research in the field biology of African Cichlids (e.g. Kocher, 2003) throw up metaphorical 'curve balls' affecting the future course of TDLP. Work on BTExact's iCRM system for modelling the effects of policy on Customer Relationship Management (Baxter, 2003) initially used RePast, but later switched to Java to allow finer control on the simulation, including the user interface (Baxter, *priv. comm.*). It is expected that TDLP will be reimplemented from scratch in the near future, but at present there is not sufficient evidence to choose between the extremes of reimplementing in C or C++ to improve speed, or to use a dedicated agent platform to improve ease of simulation design and implementation. However, experience so far suggests that programming itself is not a bottleneck limiting experimentation.

Ginot *et al* (2002) describe MAS, another ecological multi-agent simulation system. They discuss the importance of making the agent-based simulation platform available to non-programmers. In the specific case of TDLP, the author of this paper has far greater experience and knowledge of programming (particularly in Java) than in cichlid biology. Hence, without an intended audience of non-programming biologists, another of the major advantages of generic platforms does not apply.

Given that no generic approach to agent modelling appears perfectly matched to the aims of TDLP, it might seem attractive to see TDLP as the first prototype of a new, generic agent-based platform for investigating speciation problems. This approach is criticised in the "Conclusions and Future Work" section of this paper.

TDLP simulates an ecology in discrete time steps roughly approximating one week each. A side project examined the possible role of social learning behaviour in partially dividing populations as a precursor to speciation (Clement, 2003a). Behaviour occurs (and can be learnt) in time periods much quicker than a week, and future version of TDLP may need to be rewritten to work on much finer grained time scales. Currently, the system is not spatially explicit other than modelling isolated populations in rocky reefs (as occurs in all of Lakes Victoria, Tanganyika, and Malawi). However, on these reefs, species tend to segregate according to water depth (e.g. Seehausen, 1996; Konings 2001). Models that investigate this characteristic of lake ecology will need to be investigated in the future, requiring more detailed spatial representations.

The system makes extensive use of Java interfaces (Sun Computers, 2003). Much of the code accesses abstract agents (such as a generic *FoodSource*) expressed by interfaces. An interface defines a set of methods (including their arguments and return types), and any class (here agent) written that includes these methods (and declares itself to do so), can be manipulated by the code. This allows agents to be substantially changed (or replaced by new agents), without forcing across-the board rewrites of the code, and allowing configuration files to load different types of agents fulfilling similar roles (e.g. *Constant food source* agents and *Regenerating food source* agents), with the remainder of the system (e.g. *Ecology* and *Fish* agents) being agnostic as to the inner workings of the *Food Sources.*

It is not intended that the system will become infinitely variable. There is a danger that in trying to build a general-purpose system, that research may enter an open-ended development phase without ever achieving results proposing interesting biological concepts. Focussing on one specific natural system helps avoid this problem.

Visualisation of results is a particular problem in simulation. Species are a naturally fuzzy concept, in that there is no generally agreed definition of the concept of a 'species'. (See (Paterson, 1993) for a discussion on defining species.) And even when individual species concepts are applied, real organisms may not fit these categories exactly, let alone fit them exactly during the process of speciation. This creates problems for tracking the process of speciation over time. The research in this paper concentrates on 'final results', i.e. the set of species present when the biological system being modelled reaches an equilibrium. Work on detailed visualisation of the process (rather than just the result) of speciation (using fuzzy sets to represent species) is being undertaken in parallel to the main simulation system, and has been reported in (Clement, 2003b). Current research (this paper included) tends to use ad-hoc visualisation techniques, which often depend on properties of the individual simulation. E.g. if a food source is placed at an abstract co-ordinate (0.0) in a single dimensional 'trophic space', and another is placed at (1.0), then graphing the trophic adaptations of fish in a population can show lack of speciation (one group, presumably adapted near (0.5)), or speciation (two groups, one near (0.0) and one near (1.0)). However, these approaches do not adequately visualise the generation of large numbers of species, nor to co-ordinate systems higher than two dimensions.

The system is highly agent-orientated, over and above the typical 'individual-based' simulations often used in Biology. As well as data about individuals, the actions of the simulation are devolved to agents initialised according to a configuration file. E.g. an agent satisfying the '*Food Source*' interface supplies methods by which energy from that food source is supplied to fish in the local population. This is to allow radically different simulations to be performed by loading different types of agents when the system is configured.

The configuration file, like much of the system, reflects the system's growth from a much smaller program. Agent parameters are described in an ad-hoc language, with multiple 'files' included in a 'meta-file' capable of storing multiple experiments and sharing components between experiments. It is planned that this file format will be replaced by an XML format in the future.

The following sections describe the simulation system by describing each of the agent types in turn.

Fish

The basic living entity in the simulation is a fish (prey such as algae, or invertebrates, are modelled as the more general *Food Source* agent type). Each fish is modelled by a number of parameters defined by a 'lineage'. Lineage parameters are *lineage name* (basic type such as *cichlid, catfish, or cypriniform*), *sublineage* (a species name), *size at birth, maximum size, growth increment, typical lifespan, lifespan standard deviation* (hereafter *stddev*)*, sexual maturity time, breeding interval, initial energy supply, maximum energy (food) store, minimum (starvation) energy store, minimum energy reserves if breeding, typical brood size, brood size standard deviation.* Each of these properties is described in the following section:

- *Lineage name*. As described elsewhere, this allows the specification of species that can interbreed (same lineage name) and those that cannot (different lineage name).
- *Sublineage*. This parameter (e.g. species name) is not used in the simulation, other than during output of results.
- *Size at birth*. The size of fry when born. This is

specified in a generic 'unit', where 1.0 means a full-grown adult fish.

- *Maximum size*. The maximum size of an adult fish. Growth stops when this point is reached.
- *Growth increment*. The amount that a growing fish grows per time step.
- *Typical lifespan*. The average lifespan of a fish that dies of old age (which will be quite different from the real average lifespan when factors such as starvation are considered).
- *Lifespan standard deviation.* The standard deviation of the typical lifespan. These two parameters are used to choose random numbers from a normal distribution to assign lifespans to newly born fish. Fish reaching this age (having not died previously for other reasons) die of old age.
- *Sexual maturity time*. The age (in time steps) at which fish become sexually mature.
- *Breeding interval*. The interval between females breeding (contingent on sufficient energy being available)
- *Initial energy supply.* The amount of energy that a fish is born with. Removed from the mother's energy supply to prevent breeding being an (unrealistic) creator of energy in the model.
- *Maximum energy supply.* The maximum amount of energy that a fish can store.
- *Minimum energy store.* The minimum amount of energy that a fish needs to be storing before it starves to death.
- *Minimum energy reserves if breeding*. The minimum energy that a fish may have if it is to breed (females only, males only have to be alive).
- *Typical brood size*. The brood size of a fish. I.e. how many offspring are created when a female breeds.
- *Brood size standard deviation.* The standard deviation in brood sizes.

These properties control the ecological properties of the Fish themselves. These properties are typically identical across simulations, with generic values chosen for each, to approximate the small omnivorous rock dwelling cichlid *Cynotilapia afra.* E.g. parameters used in experiments for this paper are:

- *lineage* cichlid
- *sublineage* cynotilapia_afra
- *lifespan* 100
- *lifespan_var* 20
- *breedinginterval* 10
- *firstbreeding* 80
- *spawn* 10
- *maxsize* 1.0
- *initialsize* 0.03
- *growthincrement* 0.01
- *initialfood* 0.1
- *maxfood* 8.0
- *minfood* 1.4
- *minbreedfood* 4.0

In addition to these hard-coded, numerical, properties, the lineage also defines the plug-in models that control the three phenotypes (*trophic adaptation, sexual phenotype (male colour), (female) sexual preference.* These are the same properties modelled in the generic sympatric speciation experiments of Kondrashov & Kondrashov, and others. In the full modelling system, a number of plug-in models can be used for phenotypes, though the only models used at present are a *continuous phenotype model,* where the phenotype is represented directly as an *N*-dimensional real number, and the *genotype model*, where the model represents genes and chromosomes, potential mutations and their average frequencies, and a rule base to convert diploid genotypes to *N*-dimensional real number phenotypes. Gender of fish is also set by some model, which for all the experiments described in this paper is a *random* selection of gender with equal probabilities for male and female offspring. The parameters for cichlids used in experiments are:

- *trophicphenotype* continuous trophic.txt
- *sexualphenotype* continuous trophic.txt
- *preferredphenotype* continuous trophic.txt
- *gender* random

file trophic.txt

- *location* 0.5 0.5
- *var* 0.1

Indicating a single continuous phenotype (location (0.5,0.5), var 0.1, stored in a separate location) for initial values for the three phenotypes controlling *trophic, sexual,* and *preference* factors of the fish, and *gender* is random male or female.

Lineages, and sub-lineages, are important concepts for modelling African cichlids. Many African cichlids (especially those in the very young Lake Victoria, see (Seehausen, 2002)) are extremely closely related, and can easily mate and produce fertile offspring. It is thought possible (e.g. Goldschmidt, 1996) for several species of cichlids to merge into a single species, containing all the genetic variability of the original species. Hence, when creating an ecology, we may want to create closely related species which have different initial properties, but have the ability to interbreed, in certain circumstances (such as muddy water with poor visibility so that females cannot choose appropriate mates). This can be simulated by creating lineages with the same lineage name, but otherwise independent properties. On the other hand, it is sometimes desirable to create species that can never interbreed, no matter what the circumstances (e.g. *cichlids* and *catfish* in the competitive exclusion experiments described in this paper.)

The simulation program reads the model from a text configuration file, and creates the initial population. The simulation then proceeds from time 0, to a defined ending time. At each time step (approximating one week each), the population agent calls methods of the fish agents to simulate various life processes and events. At each step, each fish metabolises a unit of energy equal to its size (i.e. 1 unit of energy is sufficient to support one standard size fish for one week). If energy on hand falls below 0, then the fish starves. Otherwise, if the fish has reached the end of its natural lifespan, it dies from old age. If the fish survives, is female, is due to breed (has reached sexual maturity, and has reached the start of a new breeding period), and has sufficient energy reserves, then it breeds. The female chooses a mate from among the available (in the same location, of the same lineage, and male) mates according to the following probabilities:

$$p\,(fem\ chooses\ m) = \frac{f\,(|fem.preference - m.sexual|\,,\, fem.preference\,,\, \sigma_{fem.preference})}{\sum\limits_{m' \in Males} p\,(fem\ chooses\ m')} \tag{1}$$

where; *fem* is the female breeding, *m* and *m'* are potential (male) mates, and $_{fem.preference}$ is the stddev in the females sexual preference (i.e. how choosy the female is), |*f.preference* – *m.sexual*| is the Euclidean distance between the female's preference phenotype and the male's sexual phenotype (expressed as *N*-dimensional points), and *f(x, m, s)* is the standard function defining a normal distribution of mean m and standard deviation s. The phenotypes (and possibly genotypes) of offspring are chosen in a plug-in-dependent manner). E.g. for *continuous phenotype models*, the values of the phenotypes of the father and mother are averaged, and a random factor (generated at random from a normal distribution centred at 0, and with a standard deviation of *breeding sd*) added to represent variability in young. See (Rice, 1998) for an example of modelling evolution with continuous numeric phenotypes. The fish then broods young (holds the eggs, then young in its mouth, as is common in African cichlids, losing the energy expended in creating these young), and releases them into the general population. *Genotype models* create haploid gametes from each the mother and father, including crossover, and then creates diploid offspring as per standard diploid genetics. Note that almost all of this behaviour is performed within the *Fish* agent. However, *Fish* agents are aware of other fish in their own population, e.g. when choosing mates and producing offspring.

Assignment of energy from food sources to fish is a function of the food source agent, and is described in §2.2. The interface between fish objects and food source objects is quite minimal. The fish must expose an *N*-dimensional phenotype describing their trophic adaptation in an *N*-dimensional trophic space. This phenotype can be based on any underlying representation. Also, the fish is offered energy from a food source, and return the amount eaten (in case their stomachs become full, and the amount eaten is less than the amount available).

Food Sources

Food sources are represented by an interface, with the underlying food source being chosen in the configuration file. *Continuous food sources* supply a fixed amount of energy to fish in the population every time interval. Each food source is assigned a location (a *N*-dimensional point) in the 'trophic space' so that it can be evaluated how suitably adapted fish are for harvesting the food source. *Continuous food sources* also include a *random stddev*, indicating the degree to which a fishes success in harvesting food is due to luck. This factor was necessary to prevent unrealistic scenarios such as a population of fish growing, then all fish in the population dying at once when the fish received equal (and insufficient) amounts of energy. Other types of food source have been implemented, such as *Regenerating food source*, where the amount of energy available in each time period depends on the amount remaining from the previous time period. Only *continuous food sources* have been used in the experiments described in this paper.

An example food source is:

- *name* diatoms
- *amount* 40
- *foodsigma* 0.4
- *randomsigma* 0.1
- *location* 0.0 0.0

It is assumed that fish compete for the energy available from food sources, and that fish better adapted to the food source (both the fishes trophic adaptation and the 'location' of the food source in this space are *N*-dimensional numbers) obtain proportionally more energy from the food source. Available energy is given out according to the relative competitive advantage of the fish as follows:

$$competitive\,(fs\,,\, fish) = \frac{f\,(|fs.location - fish.trophic|\,,\, 0,\, \sigma_{fs.variance}) \times N\,(1.0,\, \sigma_{fs.random})}{\sum\limits_{fish' \in Fish} competitive\,(fs\,,\, fish')} \tag{2}$$

Note that *N*(1.0, $s_{fs.random}$) represents a normally distributed random number representing the degree to which a fish has been lucky or unlucky with respect to the food source in question. It is assumed that $s_{fs.random}$ is small, and that *N*(1.0, $s_{fs.random}$) will never generate a negative number.

Two types of food sources are available for us at present. *Constant food source(s)* assume that there is a constant supply of energy from the food (e.g. *diatoms*) at all times. This is not realistic, as constant grazing of fish on diatoms will reduce the number of diatoms, and restrict their ability to breed and produce more diatoms for the next time cycle. In reality predator-prey relationships are complex, and can produce cyclic, and unpredictable behaviour (e.g. random walks to extinction). Current experiments use *Constant food source(s)*, as is typical (explicitly, or implicitly) in much work on simulating evolution). However, a *Regenerating Food Source,* where rules relate the amount of a food source remaining at the end of one time period to the amount remaining at the next, has been implemented. Initial experiments with this food source type show that common patterns in predator/prey population dynamics can be reproduced.

Populations

A population is just a set of fish, and the actions performed by a population agent typically just call methods of the fish in that population. E.g. a population may call the *breed*() method of all fish (male and female, mature and immature) and let the fish themselves decide what will and will not happen. Configuration parameters control the lineages of fish present in the population, as well as the total initial number of fish. This is modelled by a single parameter for the total number of fish, and weights for each lineage present. E.g. if there are 60 fish, and weights of 0.4 for *cichlids* and 0.6 for *catfish,* then 24 *cichlids* and 36 *catfish* will be present in the initial population. The population sizes for the experiment are described on a per-location (roughly corresponding to a rocky reef) basis.

Locations

A location represents one reef (rocks, not coral). It is modelled as a set of food sources, a population, and a set of probabilities describing the likelihood that a fish will migrate to another location during a single time period. Like population agents, location agents basically perform few operations of their own. Their main role is to coordinate food sources with populations, and marshal migrating fish. Sturmbeyer (1998) describes possible effects that reefs, and their environment (e.g. changes in water level exposing and covering reefs) may have on speciation, and hence representing this environment is important. However, experiments reported in this paper have not made use of this feature of TDLP.

- location=lionscove
- populationSize=40
- foodSources=lionsfoods.txt
- lineages=lineage.txt

- location=nkhalireef
- populationSize=40
- foodSources=nkhalifoods.txt
- lineages=lineage.txt

Note that both locations make use of the same 'lineage' file, meaning that both locations start with identical populations.

Fish can migrate from one reef to another, providing that probabilities are defined describing the probabilities of each fish making a transition for each time step.

- from lionscove to nkhalireef 0.03
- from nkhalireef to lionscove 0.03

Experiment One: Speciation

In order to make claims about the progression of speciation in a simulated system, it is of course necessary to demonstrate that the changes occurring in simulated populations is actually speciation. This has been questioned after previous conference presentations of the system, and The system was run with parameters approximating the following ecology. Two clearly distinct food sources (0.0, 0.0) and (1.0, 1.0) were present, with the initial population of fish being placed exactly half-way between them (0.5, 0.5). Trophic, sexual, and preference values were represented by two dimensional co-ordinates, to allow easy graphing of the distributions of the final population. Fish were given ample opportunity to adapt (in terms of the variability and malleability of their phenotypes), and the simulation was run for 40,000 time steps (a time step is intended to represent one week meaning that the entire simulation lasts roughly 770 years). Phenotypes of the final population were extracted from the simulation trace, and visualised by conversion (normalisation of coordinates, and plotting of numbers identifying fish) to an *xfig* format file. (*xfig* (http://www.xfig.org) is a drawing and diagram creation tool for Unix computers which uses a simple text-based format for storing diagrams.) The three plots for the three phenotypes given in Figures 2, 3, and 4.

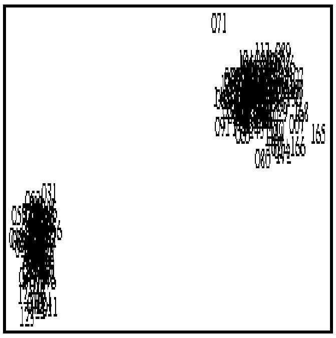

Figure 2: Sexual Phenotypes

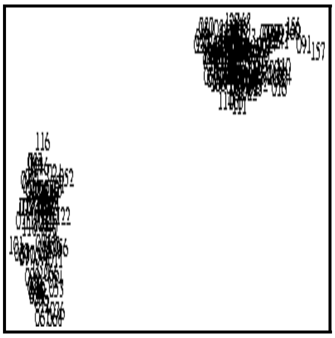

Figure 3: Preference Phenotypes

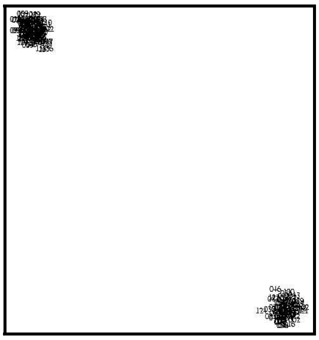

Figure 4: Trophic phenotypes

In order to confirm speciation, it is necessary to show that individual fish are not randomly assorted into the groups, but that there is a strong covariance (or high mutual information) between membership in the groups. Arbitrarily labelling groups for the three phenotypes (*sexual*, *preference*, and *trophic*): *As, Bs, Ap, Bp,* and *At, Bt,* and counting the membership of all eight combinations of groups given in Table 1. These results clearly show that group membership is consistent across all three phenotypes. In particular, fish belonging to the 'high' group for sexual preference (*As)* all belonged to the 'high' group for sexual phenotype (*Ap)*, showing a preference for intra group mating. Similar results occur for *Bs,* and *Bp,* and these groups also segregate according to trophic adaptation. It is felt that these results confirm that speciation has occurred in the simulation.

These figures all show clear separation of each of the phenotypes into two groups (of fish) consistent with speciation. It is particularly interesting to note that clustering around food sources is much tighter than the (emergent covariance) for colour and sexual preference phenotypes.

	ApAs	*ApBs*	*BpAs*	*BpBs*
At	96	0	0	0
Bt	0	0	0	136

Table 1: Group membership

Experiment Two: Competitive Exclusion

The competitive exclusion principle (see e.g. Begon *et al,* 1996) states that two species competing for the same resources cannot co-exist. Eventually one of the species will go extinct leaving the other species to dominate the environment. Cichlids in African lakes appear to contradict this principle as rocky reefs (and other environments) appear to support large numbers of species eating more or less identical diets. Parallel work (Clement, 2003b) investigated the role behaviour may play in dividing up resources among apparently competing species. The experiments reported here investigate the possibility that species that have lost the ability to adapt (due to inbreeding among small populations) may take much longer to exclude other species from an environment (or be excluded).

Food sources were placed at single dimensioned co-ordinates (0.25), and (0.75). These food sources were given *foodsource stddevs* of 0.3, sufficiently large to make the preferred strategy of fish a generalist adapted to (0.5). Two different lineages of fish were created (labelled *cichlid*, and *catfish*, the labels preventing any interbreeding or the collapse of the two species into one). *Cichlids* were given an initial trophic adaptation of (0.25), exactly specialised for the first food source, and *Catfish* were exactly specialised for the second (0.75) food source. In the first experiment, both groups of fish were given a *breeding stddev* (equal for all three phenotype) of 0.1, allowing rapid adaptation. After simulation, the time at which competitive exclusion was complete (one of the lineages completely disappeared with the death of the last fish from that lineage) was recorded. Each experiment was repeated 10 times, and the average time of competitive exclusion was calculated. Experiments for *breeding stddevs* of 0.08, 0.06, 0.04, 0.02, 0.01, and 0 were performed. The results of all experiments excluding *breeding stddev* of 0 are plotted in Figure 5.

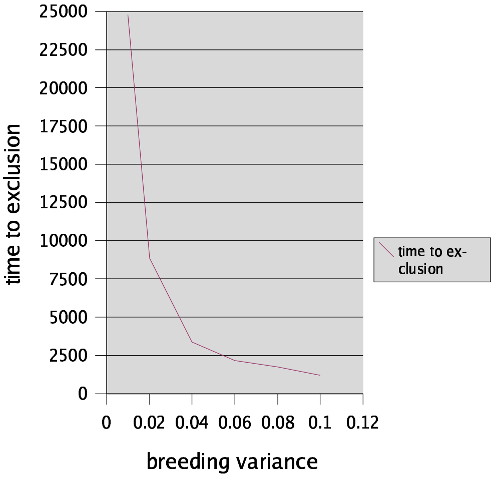

Figure 5: Competitive Exclusion Experiments

At a *breeding stddev* of 0 (i.e. the fish have no ability to adapt) competitive exclusion does not occur, even when the simulation was run to 400,000 steps (approximately 7700 years).

CONCLUSIONS AND FUTURE WORK

The (virtual) fish simulated in TDLP clearly undergo a process that closely parallels speciation in the real world, producing species groups that are differentiated in terms of diet, and have developed a method (emergent covariance between colour and preference) of maintaining different species within a single location. The model only allows pre zygotic barriers to speciation (i.e. fish of different species do not mate and produce zygotes), rather than post-zygotic barriers (e.g. where matings may occur but the resulting offspring are infertile), based on colour male colour and female choice. This is acceptable, as this matches the situation for real African mouth brooding cichlids, where fish tend not to mate outside their species, but would produce viable offspring if they did. Difficulties remain, particularly in visualising the process of speciation, rather than the result, of speciation. Every aspect of the simulation is far from a perfect match to the natural analogues, though this will always be necessary due to the near infinite complexity of real natural environments (especially lakes with the surface area of Switzerland). However, the simulation appears sufficiently realistic for investigation of conjectures concerning cichlid speciation and evolution.

The investigation of the effect on adaptability on competitive exclusion suggests that the ability to adapt increases the likelihood of competitive exclusion occurring. Adaptive organisms can evolve towards a global optima for harvesting a number of food sources, forcing out other species that might also gravitate towards the same optima. However, if the species (fish) are initially specialists in different food sources, and cannot adapt, then competitive exclusion will not occur.
While appearing interesting in itself, this result does not help explain the situation found in the African lakes. Cichlids are plainly quick at adapting to new food sources, as shown by the young Lake Victoria, where in ~14,000 years one or two generalist species appear to have evolved to fill nearly every available niche (Seehausen, 1996), though recent work (discussed in (Kocher, 2003)) suggests that Lake Kivu may be involved in the seeding of Lake Victoria, and therefore the evolutionary system being modelled may be much more complex than originally believed. One of the original conjectures on which this experiment was based was that the mixing of two fish with a very small ecological difference might lead to a stable system where the two species monopolise their own specialist food sources, removing the evolutionary pressure towards the single optimum, and exclusion of one of the species. Very similar African cichlids do co-exist in a single geographic location, and something prevents these species from excluding each other. But, it does not appear to be the factor conjectured in the exclusion experiment.

The concept may help explain other concepts. E.g. one frequently asked question about cichlids in the African Great Lakes is why cichlids dominate these lakes, to the expense of other fish families such as catfish and the carp families. Future experiments are planned to investigate whether differing relative speeds of adaptation will strongly affect the likelihood of one species winning the competitive exclusion race.

Using a real natural system as the basis for TDLP has many advantages over abstract systems. E.g. the number of parameters and model types that need to be considered can be reduced to only those that apply to the natural system in question. E.g. the use of male colour and female choice as method by which matings occur.

Furthermore, by choosing African cichlids as the natural system, well-known open problems in cichlid biology (such as the speed, and scale, of speciation, as well as the unexpected coexistence of species adapted to apparently identical niches). One of the outcomes of the exclusion experiment emphasises this point. The pattern found, where the ability to adapt sped competitive exclusion, is a potentially valid concept in biology. But, it isn't needed (or helpful) in explaining the open question of why so many cichlid species co-exist, with no observable ecological differences. In fact, it adds to the mystery. Hence, the further consequences of this concept (apart from those mentioned above which may help explain the dominance of cichlids over other species in these lakes) do not need to be investigated.

One major problem with an agent-based simulation such as TDLP is the huge number of potential experiments that can be performed. Hence, it is necessary to carefully pick and choose what experiments are to be performed in order to prevent huge amounts of time being given over to small variations of the same experiment. The concentration on a real system (with real open problems) helps provide this focus. Hence, in this research we are attempting to rein in the temptation to attempt to model all and every possible influence on cichlid speciation. E.g. Kooi & Kooijman (2000) show how invading (predator) species can stabilise the populations of potential prey species, reducing the tendency towards competitive exclusion. Predation is a very important factor in cichlid ecology, but it has not yet been modelled. The reason for this is that in African lakes, the primary natural predators of cichlids are other cichlids. Since these cichlids evolved from common ancestors, modelling this reality will introduce considerable complexity into TDLP. The separation between sexual factors (colour) and trophic factors (food sources) disappears, a predator may adapt itself to better see a prey species, by increasing the number of photo receptors in the eye which detect the wavelength of the prey species breeding colouration.

At present, repeating Kooi & Koojman's studies is a major future subproject. Kooi & Koojman use mathematical modelling to support their results. Our intended repetition of their results will be as an agent-based implementation, most probably using RePast. This study is expected to answer the following questions. (i) Whether agent-based and mathematical modelling investigations of the same properties will give the same results. (ii) If the results agree, the degree of sensitivity of the agent-based model to changes in model parameters. (iii) Throw further light on the effects of predator-moderated competitive exclusion. And, (iv) allow the comparison of implementing agent-based models of speciation in both pure Java, and an agent-modelling platform.

As mentioned when discussing the choice of platform, it is interesting to discuss whether TDLP should be rewritten as a generic tool for agent-based simulations of speciation problems. One of the great difficulties of building a system such as TDLP is that a great deal of biology needs to be studied and incorporated to make the simulation results useful. In creating a general—purpose tool, the required amount of biological knowledge that would need to be incorporated (or at least understood) grows to seemingly unmanageable levels. TDLP contains many short-cuts which are not even applicable across all African cichlids (e.g. that young are cared for only by the female, and that mate choice is female-driven). Furthermore, Genetic models assume sexual reproduction only, which is not universal over all organisms. As can be seen by studying the Galaxids of New Zealand, even among fish, the factors underlying speciation may be quite different. If another researcher were to attempt to use TDLP for research into the speciation of snails, or stick insects, even a single mismatch between the biological system and the facilities provided by TDLP (such as asexual reproduction), could make TDLP unusable. Because development and research time is not unlimited, a broadening of the intended audience of TDLP could result in a system which is theoretically capable of a wide range of tasks, but is ideal for none.

Within the biological research surveyed, the most advanced agent-based simulation platforms found are those in Ecology. Research has progressed to a point where review papers on individual-based (agent) simulation are common (e.g. Berec, 2002). But, the most advanced of those found (CENOCOM, PK, MAS) tended to be publications describing the agent-based systems themselves, not papers describing new results found using these simulations. This may be a property of the age of these systems (typically very recent), or a symptom of a larger problem where the providers of agent-based simulation platforms being a different group of individuals from the potential consumers of these systems.

In terms of the future choice of generic agent-based platform versus low-level programming language, an important factor that has not yet been considered is repeatability of experiments. While plain Java is a convenient language for the implementation of TDLP, it may be very inconvenient for any non-programming researcher who might wish to repeat or further investigate any of the experiments performed with it. E.g. a large number of the biological details (such as Female-only mate preference) are not stated in the (text) configuration file, but buried within 14,000 lines of Java code. Due to the way TDLP developed, many classes have redundant methods included, which would make the mechanisms involved less than obvious for any observer. If a generic agent platform of tool were used, there would be a (at least partial) separation of biological model parameters, and genetic modelling tool code, which should improve accessibility of the model itself.

Compared to much ecological modelling, one of the major problems modelling speciation is the implicit nature of the factor being investigated. Particularly in African cichlids, species membership can often be a 'fuzzy' concept, with considerable argument as to what is a separate species, and what is a race. At present a considerable amount of work is being put into extracting, and visualising the process of speciation. This is based upon a clustering of Fish from a stable, final population, and the mapping of these clusters onto a set of potential ancestor species. (Clement, 2003b). From traces of these ancestor species (represented by fuzzy sets), phylogenetic trees can be extracted and plotted (Clement, 2003c). This work is in a fairly advanced state of development, and appears to be moderately well solved.

The bulk of the 'remaining' work for TDLP (though in reality, the 'remaining' work is infinite), is in improving the biological accuracy of the simulation. In addition to the inclusion of predation, other improvements planned are the inclusion of more, and more accurately modelled, food sources. One potential future project is the modelling of cichlid digestive tracts (and teeth) in considerable detail. At present TDLP only models a fish's suitability for a food source as a single (plastic) co-ordinate in a *N*-dimensional space. Hence a herbivore might be at a location of (0.3, 0.9), while an insectivore might be at (1.2, 0.1). Optimal locations for food sources (such as algae, and worms) is also given as points in this space, and therefore 'suitability' for a particular food sources decreases according to distance between these points. This is an abstract model approximating real-world properties of the fish and the environment. E.g. herbivore cichlids often have a longer intestine than carnivores, due to the greater difficulty of digesting plant material compared to animal flesh. The optimal length of an intestine therefore depends on the diet of the fish. It isn't a case of longer is better due to the developmental and metabolic costs of a long intestine. The same applies to teeth, with different teeth being better adapted for different methods of harvesting different prey. E.g. the genus *Labidochromis* have pincer-like teeth for picking small food items (either plant or animals) from tight crevices in rocks (Konings, 2001). These teeth would be less effective for scraping algae from a rock. It may be interesting and informative to build a Fish agent which has a realistically modelled digestive tract, and the potential for this tract to vary under genetic (or other) control. As well as the tract itself, fish behaviour would also need to be modelled. This would allow experiments showing the effects on anatomy on evolution and speciation. In theory, a covariance should still emerge between digestive tract form, foraging (or hunting) behaviour, and some sexual lock (colour) and key (female preference) resulting in speciation.

It is as yet unknown exactly how big TDLP (in terms of agents modelled) will have to grow to accurately model speciation. All three great lakes are huge, and their cichlid populations are similarly huge. However, many rock living cichlids live in very constrained habitats (e.g. rocky reefs, sometimes the size of a small room). As well as the Great Lakes, there are other smaller lakes (such as Lake Albert, and in particular, Lake Barombi-Mbo), where populations are much smaller, but speciation still occurs. At present this research is continuing under an assumption that patterns and properties of speciation found in smaller populations will scale up to larger populations, but this remains to be proven.

Future work, in general, is based around three basic concepts; improve the biological accuracy (not necessarily flexibility) of the simulation, improve the visualisation of the results, and continue investigating conjectures concerning cichlid biology. Where this research will benefit from the use of agent-based platforms or tools, these will be used. But, in the immediate future, progress is likely to concern mainly the existing Java implementation.

ACKNOWLEDGEMENTS

This paper has benefited greatly from the comments and help of many individuals, including George Turner, Nicola Baxter, and the comments of a large number of anonymous referees both for this paper and for other papers.

BIOGRAPHY

Ross Clement was born in New Zealand. He completed his B.Sc. degree in Cellular and Molecular Biology, and a M.Sc. degree in Computer Science at the University of Auckland (New Zealand). He completed the degree of Doctor of Engineering in Systems and Information Engineering at the Toyohashi University of Technology (Japan) in 1991. After working as a research assistent, studying Genetic Algorithms for Transport Scheduling at the University of Leeds (UK). Since 1993 he has been a Lecturer, then Senior Lecturer in the Harrow School of Computer Science, of the University of Westminster (UK). His research interests are the application of Artificial Intelligence and Computer Science techniques to models and simulation of Evolution. His email address is clemenr@wmin.ac.uk

REFERENCES

Barlow, G. W. (2000). *The Cichlid Fishes: Nature's Grand Experiment in Evolution.* Perseus, MA.

Baxter, N. (2003). Intelligent Customer Relationship Management. OR45 Conference, Keele, UK. http://www.orsoc.org.uk/conf/or45/Handbook.doc

Begon, M., J. L. Harper, and C. R. Townsend. (1996). *Ecology: Individuals, Populations and Communities.* 3rd ed. Blackwell Science, Oxford, UK.

Berec, L. (2002). Techniques of spatially explicit individual-based models: construction, simulation, and mean-field analysis. *Ecological Modelling* 150: 55-81.

Clement, R. P. (2003a). Plausible roles for Social Learning in the Speciation and Evolution of Cichlid Fish. *Journal of the AISB* (to appear).

Clement, R.P. (2003b). Visualising Speciation in Cichlid Fish. *Proceedings of the 17th European Multisimulation Conference*, Nottingham.

Clement, R.P. (2003c). Visualising Speciation in Artificial Cichlid Fish. Submitted to *Artificial Life.*

Coleman., R. M. (ed.) (2001). Cichlid Research: State of the art. *Journal of Acquariculture and Acquatic Sciences* 9.

Dieckmann U, Doebeli M (1999). On the Origin of Species by Sympatric Speciation. *Nature* 400: 354-357.

van Doorn, G. & F. Weissing F. (2001). Ecological versus sexual selection models of sympatric speciation. *Selection* 2: 17-40

Gilbert, N., & Bankes, S. (2002). Platforms and methods for agent-based modelling. *PNAS* 99 suppl. 3: 7197-7198.

Ginot, V., Le Page, G., & Souissi, S. (2002). A multi-agents architecture to enhance end-user individual-based modelling. *Ecological Modelling* 157: 23-41.

Goldschmidt, T. (1996). *Darwin's Dreampond: Drama in Lake Victoria.* MIT Press.

Kocher, T. D. (2003). Fractious phylogenies. *Nature* 423: 489-491.

Kondrashov, A. & Kondrashov, S. (1999). Interactions among quantitative traits in the course of sympatric speciation. *Nature* 400: 351-354.

Konings, A. (2001). *Malawi cichlids in their natural habitat 3rd edition.* Cichlid Press.

Kooi, B. & Kooijman (2000). Invading species can stabilize simple trophic systems. *Ecological Modelling* 133: 57-72.

Mamedov, A. & Udalov S. (2002). A computer tool to develop individual-based models for simulation of population interactions. *Ecological modelling* 147: 23-68.

Parrott, L. & Kok, R. (2002). A generic, individual-based approach to modelling higher trophic levels in simulation of terrestrial ecosystems. *Ecological Modelling* 154: 151-178.

Paterson, H. (1993). *Evolution and the Recognition Concept of Species.* John Hopkins University Press.

Rice, S. H. (1998) The evolution of canalization and the breaking of von Baer's laws: modeling the evolution of development with epistasis. Evolution 52: 647-657.

Schliewen, U, Tautz, d, Pääbo, S. (1994). Sympatric speciation suggested by monophyly of crater lake cichlids. *Nature* 368: 629-632.

Seehausen, O. (2002). Patterns in fish radiation are compatible with Pleistocene desiccation of Lake Victoria and 14,6000 year history for its cichlid species flock. *Proc R. Soc. Lond. B. Biol. Sci.* 269: 491-7.

Seehausen, O. (1996). *Lake Victoria Rock Cichlids.* Verduijin Cichlids.

Sturmbauer, C. (1998). Explosive speciation in cichlid fishes of the African Great Lakes: a dynamic model of adaptive radiation. *Journal of Fish Biology* 53 (Supplement A), 18-36.

Sun Computers (2003). *Creating Interfaces.* http://java.sun.com/docs/books/tutorial/java/interpack/interfaces.html

Turner, G. F. & Burrows, M. T. (1995). A model of sympatric speciation by sexual selection. *Proc. Royal Soc. Series B* 260 (1359), 287-292.

Turelli, M., Barton, N., & Coyne, J. (2001). Theory and speciation. Trends in Ecology and Evolution.16: 330-343.

Via, S. (2001). Sympatric speciation in animals: the ugly duckling grows up. *Trends in Ecology and Evolution* 16: 381-390.

Waters, J.M., Craw, D., Youngson, J.H. & Wallis, G.P. (2001). Genes meet geology: fish phylogeographic pattern reflects ancient, rather than modern, drainage connections. *Evolution* 55:1844-1851.