

# **ANALYSIS OF DEVELOPMENTAL GENETIC THEORIES USING EVOLUTIONARY PROGRAMMING**

M.A. Góngora

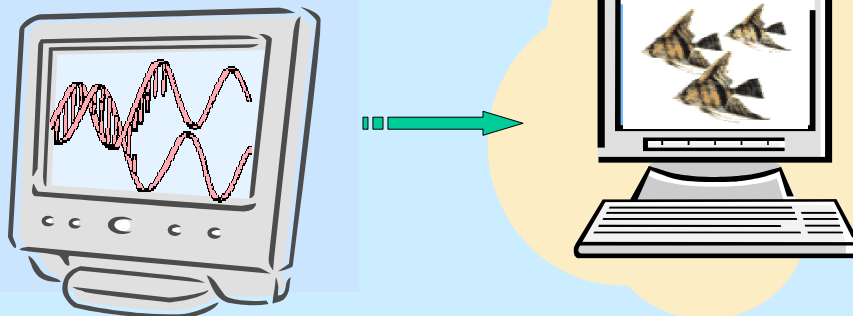
mgongora@dmu.ac.uk

## **DEVELOPMENTAL GENETICS**

- How a fertilised egg (a single initial cell) develops into a complex multi-cellular organism.
- The information in the DNA, organised in genes, is decoded for cell reproduction and the activation of differentiation processes.
- Ultimate goal: to determine the Genotype-Phenotype correlation.

## Developmental Genetics and Evolutionary Programming

- To be able to simulate a Genotype-Phenotype correlation for a given process.
- To be able to analyse a process.



## Developmental Genetics Theories

Criteria to choose what to analyse:

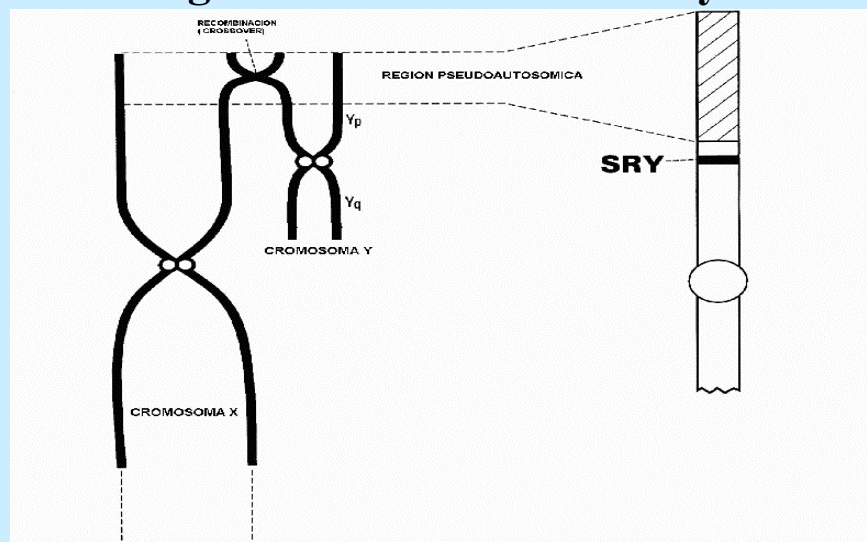
- Availability of information for analysis and for validation.
- Process as independent as possible (from other genes in the genome).

## Developmental Genetics Theories

SRY gene: the effects of some mutations in the gender phenotype of the individual

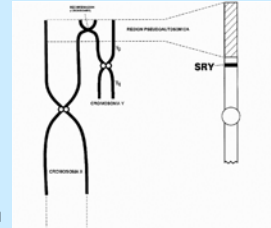
- Analysis involved a **single** gene
- Real clinical information available

## SRY gene simulation and analysis



## SRY gene simulation and analysis

- “Default human” is female
- Same DNA:
  - Without SRY gene → Female ♀
  - With functional SRY gene → Male ♂
- Mutated SRY gene (damaged) produced an ambiguous development in the transformation from female to male.



## SRY gene simulation and analysis

### Problem:

- To correlate the DNA sequence of mutated SRY genes with the physical characteristics of the individual that has it.

## **SRY gene simulation and analysis**

### **Resources:**

- Clinical data showing cases of SRY DNA and the physical characteristics of the individuals that have them.
- Lots of computing power and programming abilities.

## **SRY gene simulation and analysis**

### **Problem Configuration:**

- An artificial neural network has as inputs the DNA sequence of the gene.
- The outputs are the standardised list of the corresponding characteristics.

## **SRY gene simulation and analysis**

### **Inputs:**

- The DNA sequence is represented as a list of letters representing the amino acids it encodes (20 different possibilities). The SRY gene can have up to 250 of these.
- The input layer consists of 250 neurons, 0 represents no letter (no more DNA), 1 to 20 each valid letter.

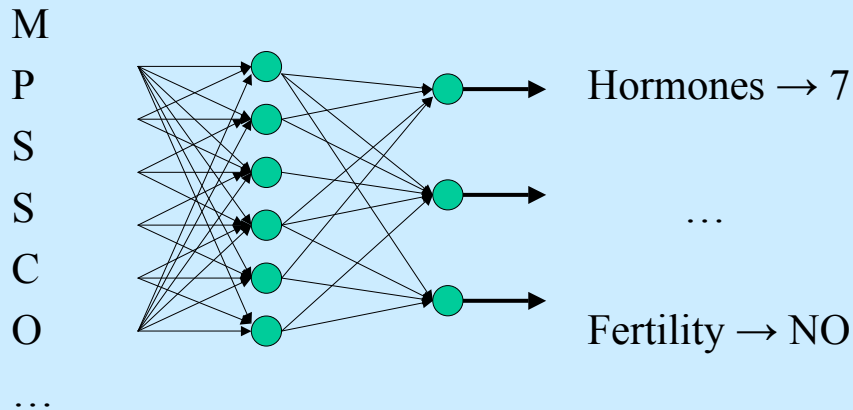
## **SRY gene simulation and analysis**

### **Outputs:**

- From the clinical data, 11 characteristics were defined as the standard clinical description of the gender phenotype for this study.
- The output layer consists of 11 neurons, each one having the standardised levels to describe each characteristic.

## SRY gene simulation and analysis

### Inputs / Outputs



### Data configuration

- A sample of clinical data was used to “evolve” the connexions and weights (C&W’s) of the ANN.
- Thus: the chromosome consisted of a matrix of connexions and weights.
- The search space had more than  $11 \times 10^{850}$  possible combinations.
- The normal SRY gene (normal male) was included in this group

## **Data configuration**

- Other sample of data was used for validation
- The no-SRY gene (XY female) was included in this group
- When the ANN is “trained” with the G.A. the new data set is used to test how good it is in predicting the phenotype of unseen SRY sequences.

## **Population**

- The population consists of a number of individuals (chromosomes), each one representing a possible combination of all connexions and weights of the ANN.
- For different tests of the system, populations of 1,000 or 10,000 individuals were manipulated for a maximum of 100 to 50,000 generations (time-out part of the termination criteria).



## Fitness

- The fitness for each individual was measured using the sample clinical data used for development (24 SRY gene sequences having their corresponding phenotype characteristics)
- After the end each generation, every individual is tested using the sample data. The fitness is proportionate to the difference between the real characteristics and the ones produced by the ANN using the chromosome for C&W's

## Genetic Operations

The genetic operators manipulated the connexions and weights on each chromosome to produce the next generation's population.

- **Elitism:** Copied the best individual to the next generation. *Non-random operator.*
- **Reproduction:** Copied individuals to the next generation, with a probability of being copied proportional to their fitness.

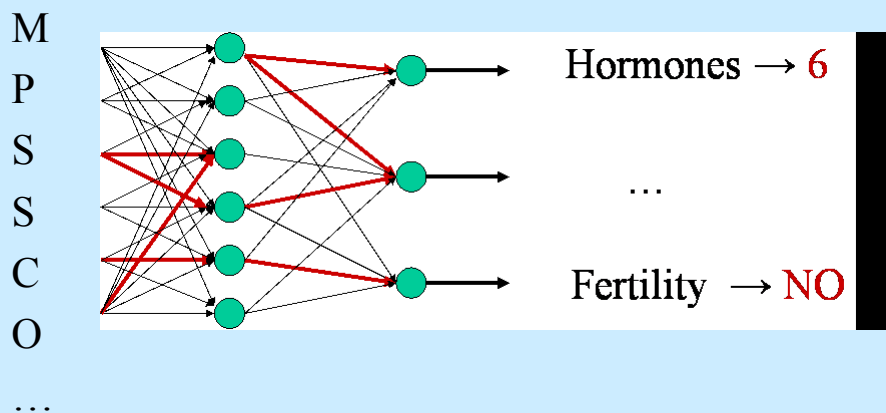
## Genetic Operations

**Using a probability to select proportional to the fitness:**

- **Crossover:** Produces an individual for the next generation, from 2 selected individuals. It takes part of one, part of the other and mixes them together.
- **Mutations:** Takes a selected individual and changes some of its data randomly to create a new individual.

## SRY gene simulation and analysis

### Inputs / Outputs

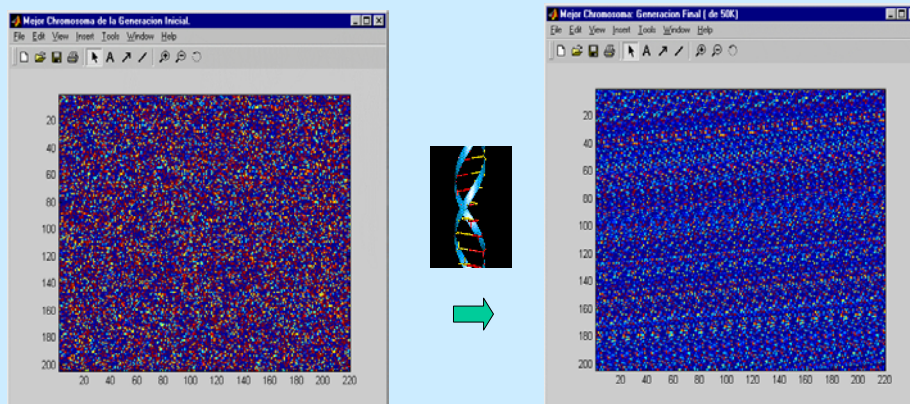


## SRY gene simulation and analysis

- After ending the evolution the best ANN had an average of 87% prediction.
- Thus: There was no individual solving the problem 100%.
- The termination criteria was **always time-out**.
- When validated it had an 85% prediction with the new data set.

## SRY gene simulation and analysis

Electronic *Signature* of C&W's matrix



## Conclusions

- High potential of Evolutionary programming in the analysis of genetic processes
- It is possible to predict basic Genotype-Phenotype correlations such as that for the SRY gene
- It is possible to simulate biological-type processes to analyse theories without in-vivo experiments
- Computer simulations can be used to pre-test experimental set-ups

## ANALYSIS OF DEVELOPMENTAL GENETIC THEORIES USING EVOLUTIONARY PROGRAMMING



Dr. M.A. Góngora

[mgongora@dmu.ac.uk](mailto:mgongora@dmu.ac.uk)