Genetic Algorithms and Genetic Programming
Lecture 2

Gillian Hayes

26th September 2006

School of informatics

Admin Reminder

Lecturer: Gillian Hayes, IPAB, School of Informatics

Email: gmh@inf.ed.ac.uk

Office: JCMB room 2107C, ext. 513440

Course Activities:

- Lectures: Tuesday 12:10 (JCMB LTB), Friday 12:10 (Daniel Rutherford Building LT1),
- Tutorials: Mon 10:00 (JCMB 6324), 15:00 (DHT FRN), Wed 13:00 (JCMB 6324), Fri 15:00 (JCMB 4310 and AT M3). Weeks 3–10

Admin Reminder

- Reading: From supplied course notes and set book (*An Introduction to Genetic Algorithms* by Melanie Mitchell, MIT Press 1998, £20.85 on Amazon, also available on MIT CogNet)
  See `http://www.lib.ed.ac.uk/resbysub/info/ebooks.shtml`

- Assignments: a single assignment worth 25% of the course mark, to be handed in at the start of Week 11.

- Exam: worth 75% of the course mark, taken at the end of Semester 2.

Syllabus Reminder

Part 1: Introduction

- Genetic Algorithms: biological inspiration

Part 2: Genetic Algorithms (GAs)

- The canonical genetic algorithm
- The schema theorem and building block hypothesis
- Formal analysis of genetic algorithms
- Methodology for genetic algorithms
- Designing real genetic algorithms

continued....

Syllabus Reminder

Part 3: Optimisation Problems

- Solving optimisation problems
- Swarm intelligence: ant colony optimisation (ACO)
- Adding local search: hybrid GAs and hybrid ACO
- Other methods: simulated annealing, tabu search

Part 4: Evolving Programs and Intelligent Agents

- Evolving programs: genetic programming
- Evolving controllers: neural networks and robots
- Evolving intelligence: agents that play games
- Evolving intelligence: programs that can plan

Genetics and Evolution

Aim of this lecture: to understand the biological inspiration behind GAs and understand where the curious terminology comes from:

- gene, allele, chromosome, genotype, phenotype, selection pressure, crossover, mutation, individual, population, generation, fitness . . .

Sources:

- Chapter 1 of Peter Ross's GAGP notes
- Richard Dawkins: The Selfish Gene, Chapter 3, Oxford University Press 1976, 1989
- Bruce Alberts et al.: Essential Cell Biology, 2nd ed., Garland, New York 2004

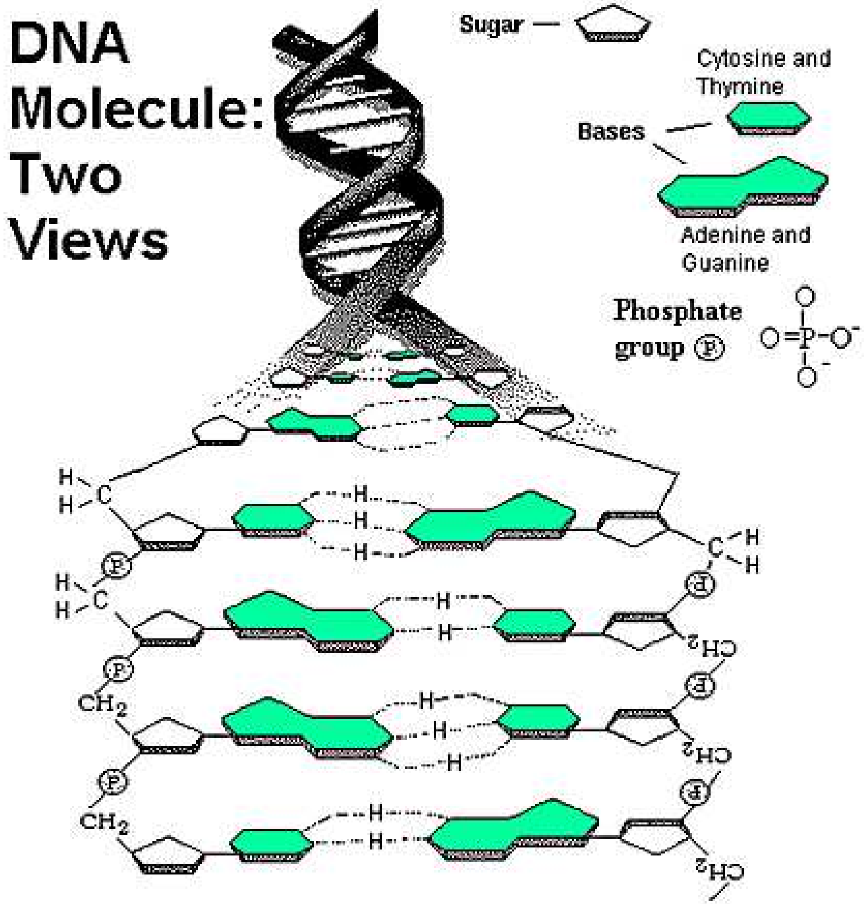
The Genetic Code

DNA = deoxyribonucleic acid

DNA is made up of a chain of simple molecular units. Each unit comprises a base, a sugar and a phosphate. The sugars and phosphates in many units link together in a chain with the bases sticking out. The bases in two chains attract one another resulting in a double helix structure.

There are just 4 kinds of **base** in DNA, labelled A, C, G and T (adenine, cytosine, guanine, thymine). C and G pair up, as do A and T.

. . . GATTACCA . . .
. . . CTAATGGT . . .


DNA
Molecule:
Two
Views
Sugar
Cytosine and
Thymine
Bases
Adenine and
Guanine
Phosphate
group

Chromosomes

Inside every human cell are 46 strands of DNA called **chromosomes**. There are 23 pairs:

Mother:	1a	2a	3a	4a	. . . 23a
Father:	1b	2b	3b	4b	. . . 23b

The 46 chromosomes are the instructions for making an entire human being. (This depends on a lot of other things happening just at the right time.)

Encoding Proteins

How does this work?

Sections of chromosome contain the instructions for building chains of amino acids – proteins. The proteins are the building blocks, regulation units and manufacturing units of the body:

e.g. lactase (enzyme), collagen (structure), haemoglobin (oxygen transport), actin (muscle contractions), CLOCK protein (circadian rhythm regulation).

Encoding:	3 DNA bases	$\rightarrow$	1 amino acid	AAA = lysine
	64 combinations	$\rightarrow$	20 amino acids	CCC = histidine
	– some redundancy			

A protein is made up of many amino acids strung together and folded up.

Genes

Sections of chromosome which encode a protein (the order in which to connect amino acids together) are often called a **gene**. (Plus sections to encode when the gene will be activated, i.e. when/where the protein is produced.)

Other meanings of gene:

- genetic material which encodes a trait
- a long-lived inheritable genetic unit

The 46 chromosomes in every cell build proteins which make a human body.

In fact, we could use only 23: the 1a and 1b chromosomes are alternative solutions to the same problems:

1a: . . . | Eyes are blue | . . .   1b: . . . | Eyes are brown | . . .

Sometimes one gene is dominant, the other recessive: in this case, eyes are brown. The "eyes are blue" gene is called an **allele** – a rival value.

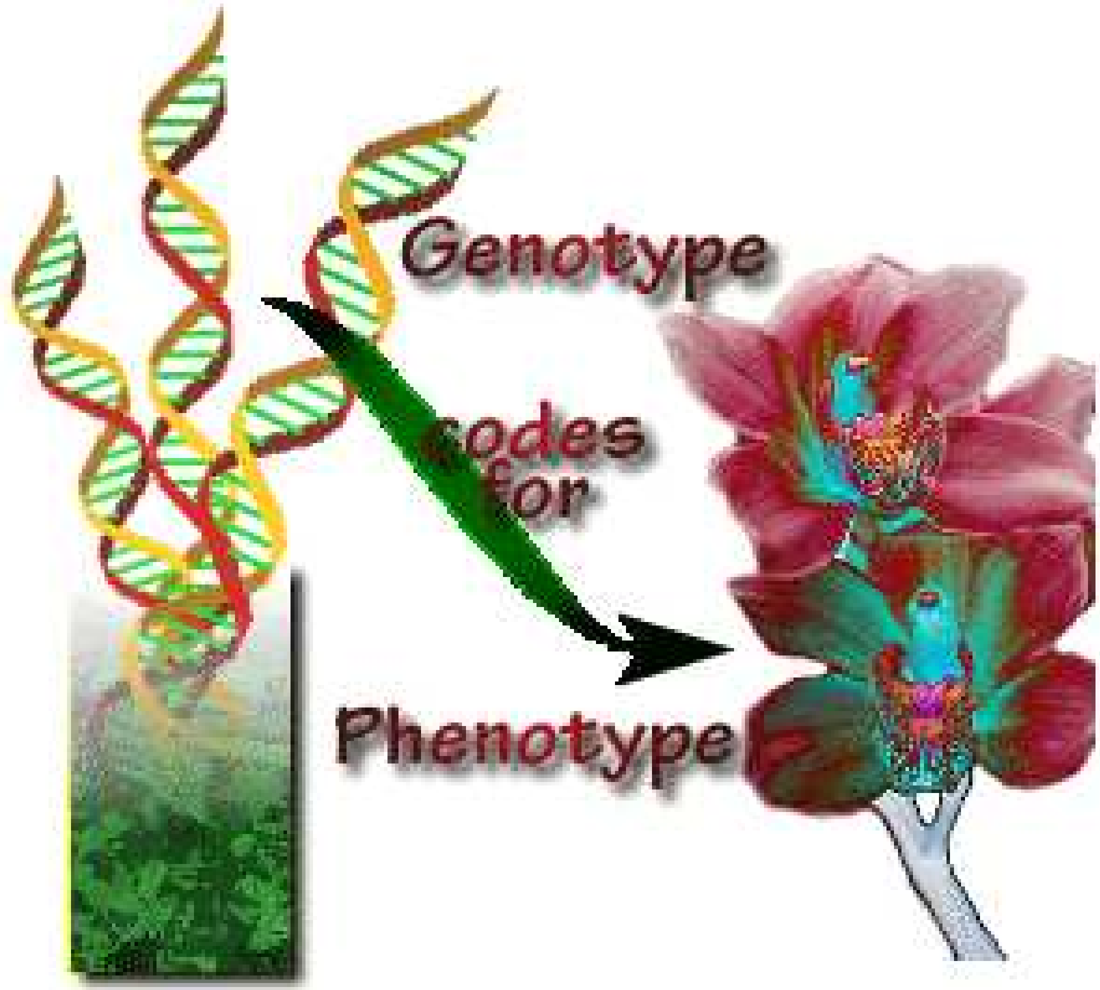
Sometimes the trait will be somewhere between the 2 values.

Sometimes it's more complicated!

Height vs. ADD (attention-deficit disorder)

Genotype and Phenotype

DNA code:	$\rightarrow$	human being
genotype	$\rightarrow$	phenotype


Genotype
codes
for
Phenotype

Selection

When individuals exist in populations, they **compete** for resources.

The fitter ones live, the less fit die.

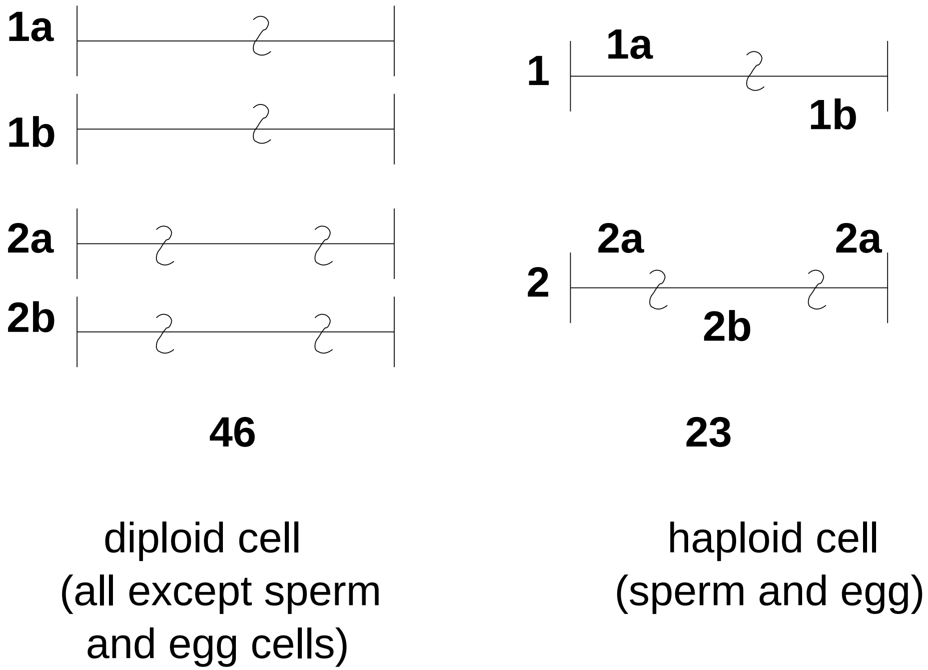
The survivors choose mates and produce offspring. The offspring are related to, but are not direct copies of, the parents:

- children inherit traits from parents
- offspring vary in their physical properties and behaviour

How can we explain this?

- sexual reproduction
- mutation
- crossover
- inversion

Crossover



Every haploid cell is a result of **different** recombinations of the 46 chromosomes.

When sperm and egg fuse, the 23 from the male and the 23 from the female are simply collected together to form a new collection of 46.

Mutation

During crossover, sometimes copying errors are made, with low probability:

. . . CGTATTCATGG . . .
. . . CGTA**C**TCATGG . . .

Also, sometimes strands of DNA become detached and flip end-over-end before reattaching – **inversion**:

. . . CGTATTCATGG . . .
. . . CGTA**ACTT**TGG . . .

So different amino acids are coded for. Crossover can be disruptive – it favours smaller units of inheritance (Dawkins).

Some (very large) parts of DNA seem to be there to protect genes from crossover (junk DNA, introns, bloat).

Selection

Selection governs which organisms live long enough to pass on their genes.

Selection pressure defines a fitness landscape which favours one type of creature over others. E.g. in landscapes that provide sparse food, creatures that store fat efficiently are more likely to survive.

$\rightarrow$ optimisation

Populations can also adapt to changes in their environment. E.g. if populations start farming rather than hunting/gathering they develop enzymes that can digest grains.

$\rightarrow$ adaptation

Genetics and Evolution → AI?

Evolution shows how complexity and solutions to the problem of survival arise from populations, selection and recombination.

In our search for AI, such ideas are appealing:

- intelligence as an emergent property
- optimisation of behaviours
- adaptation to changes
- learning of new behaviours
- searching for optimal solutions (or better than the rest?)
- Artificial Life

Using These Ideas

How does all this inspire us to build clever computer programs?

– See Lecture 1!

- find a hard problem (local maxima)
- encode solutions as a genotype
- map genotypes to phenotypes (optional)
- evaluate phenotype for fitness
- mate fit genotypes using crossover
- make a few mutations
- continue until you have your solution (or you are convinced that no improvement is possible)

The Main Issues

- How do I represent a solution?
- How should I rate a solution for fitness?
- How large should the population of solutions be?
- How much selection pressure should I apply?
- What form of crossover should I use?
- what form of mutation should I use?

Next lecture: The Canonical GA

DNA figure: Access Excellence Graphic Library.
Genotype–phenotype figure: Blamire's Science at a Distance.