

Gene Expression Programming

Jiří Kubařík
Department of Cybernetics, CTU Prague

Substantial part of this material is based on the article
Candida Ferreira: Gene Expression Programming: A New Adaptive Algorithm for Solving,
see <http://arxiv.org/ftp/cs/papers/0102/0102027.pdf>



<http://cw.felk.cvut.cz/doku.php/courses/a0m33eoa/start>

Gene Expression Programming

Gene Expression Programming (GEP) - genotype/phenotype genetic algorithm for creation of computer programs

- GEP uses **fixed length linear chromosomes** of specific structural organization of genes.
- The chromosomes are subjected to variation operators.
- The **linear chromosomes are expressed as nonlinear expression trees (ETs)** of different sizes and shapes that are evaluated and upon which the selection acts.
Any modification made in the genome always results in syntactically correct ETs (*given that the closure property holds*).
- GEP provides means for automatic defining and reusing functions.

GEP recalls to its analogy to the natural **gene expression**:

Gene expression is the process by which information from a gene is used in the synthesis of a functional gene product. These products are often proteins, but in non-protein coding genes such as rRNA genes or tRNA genes, the product is a functional RNA.

Wikipedia

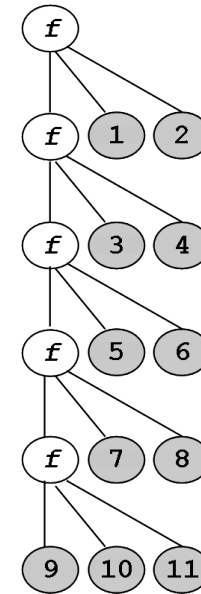
GEP: Size of the Tail Region

Consider

- Head length $h = 5$,
- The maximum arity of functions $n = 3$.

and assume that all of the symbols in the head part represent functions with arity $n = 3$.

What is the number of open nodes that must be filled in by terminal symbols?



The size of the tail t must be

$$t = h(n - 1) + 1$$

in order to ensure a sufficient number of terminal symbols even for the worst case scenario with only function symbols in the head part.

Note, the **length of the ORFs varies**, not the length of genes.

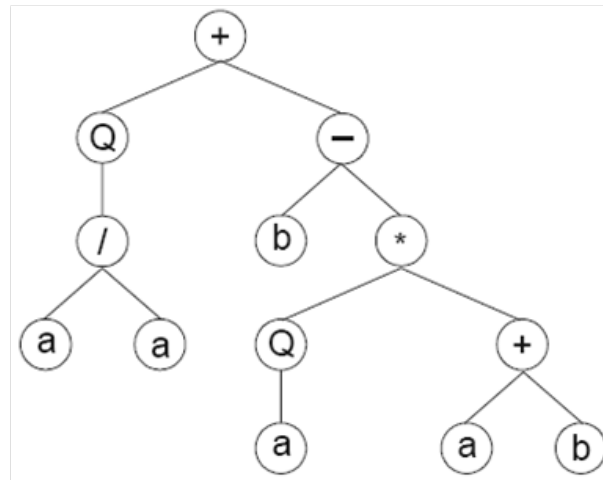
- the noncoding regions in genes allow modifications of the genome always producing syntactically correct programs.

GEP: Effect of Variable ORF Size

Suppose now the symbol at position 9 changed from 'b' into '+'.
The following gene will be:

```
0 1 2 3 4 5 6 7 8 9 0 1 2 | 3 4 5 6 7 8 9 0  
+ Q - / b * a a Q + a a b | a a b b a a a b
```

that codes for the following ET:



In this case the ORF ends at position 12.

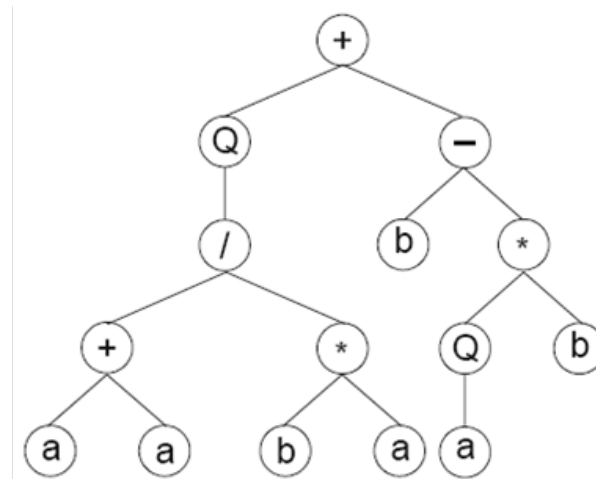
GEP: Effect of Variable ORF Size

Suppose now the symbols at positions 6 and 7 (in the original gene) changed into '+' and '*'.

The following gene will be:

```
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 | 5 6 7 8 9 0
+ Q - / b * + * Q b a a b a a | b b a a a b
```

that codes for the following ET:



In this case the ORF ends at position 14.

Despite its fixed length, each gene can code for ETs of different sizes and shapes!

Multigenic Chromosomes: Homeotic Genes

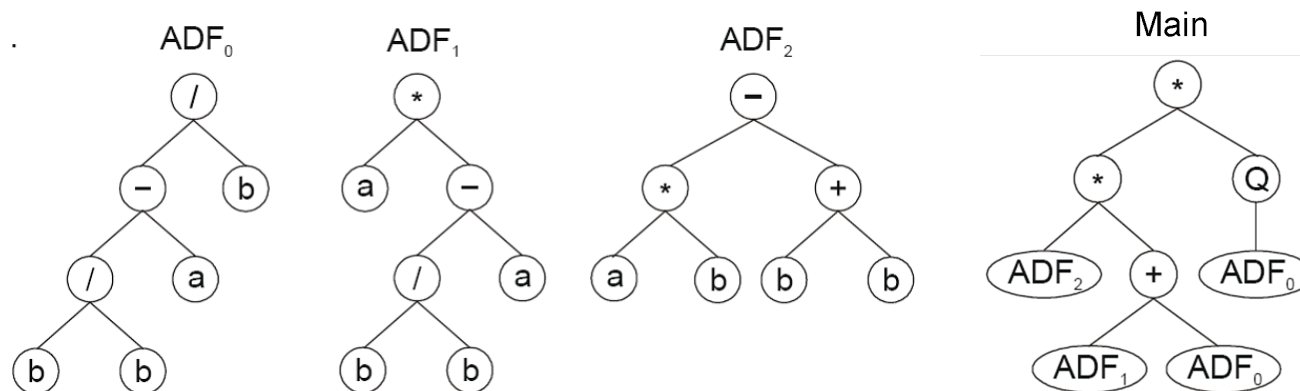
Example:

- the head length of the homeotic gene $h_H = 5$
- the head length of the conventional genes $h = 4$
- $F_H = \{+, -, *, /, Q\}$, $T_H = \{1, 2, 3\}$ denoting ADF_1 , ADF_2 and ADF_3
- $F = \{+, -, *, /\}$ and $T = \{a, b\}$

The following chromosome codes for three conventional genes and one homeotic gene (red)

01234567801234567801234567801234567890
 /-b/abbaa*a-/abbab-*+abbbaa**Q2+010102

that expresses as



GEP Genetic Operators: Recombination

Three type of recombination:

- 1-point recombination, 2-point recombination
- Gene recombination

