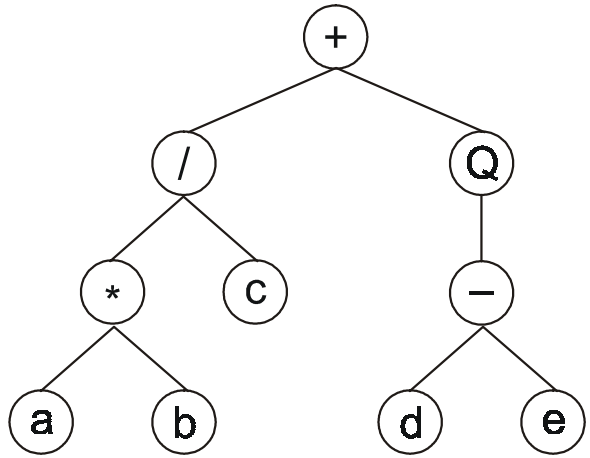
Gene Expression Programming
in
Problem Solving

Cândida Ferreira

candidaf@gene-expression-programming.com
http://www.gene-expression-programming.com

REPRESENTATION

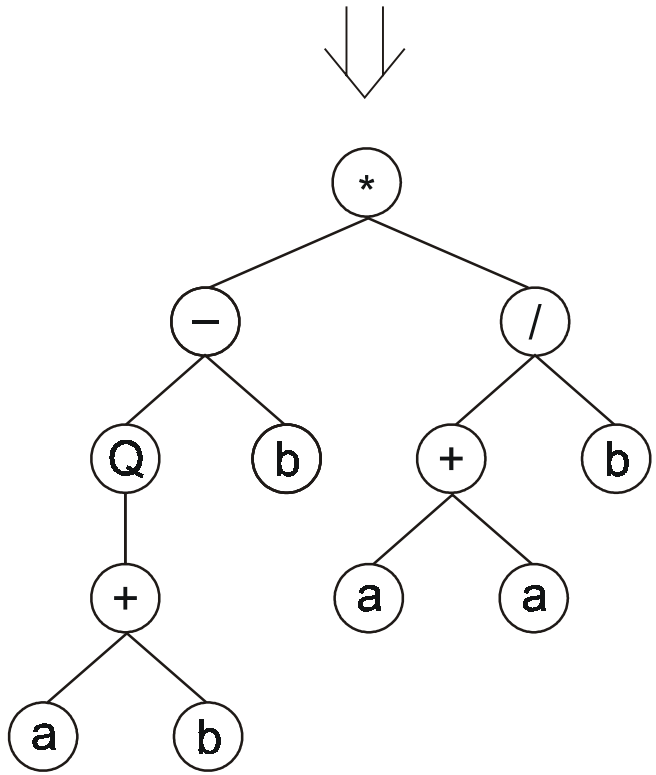
$$\frac{a \cdot b}{c} + \sqrt{d - e}$$



0123456789
+/Q*c-abde

K-Expressions or ORFs

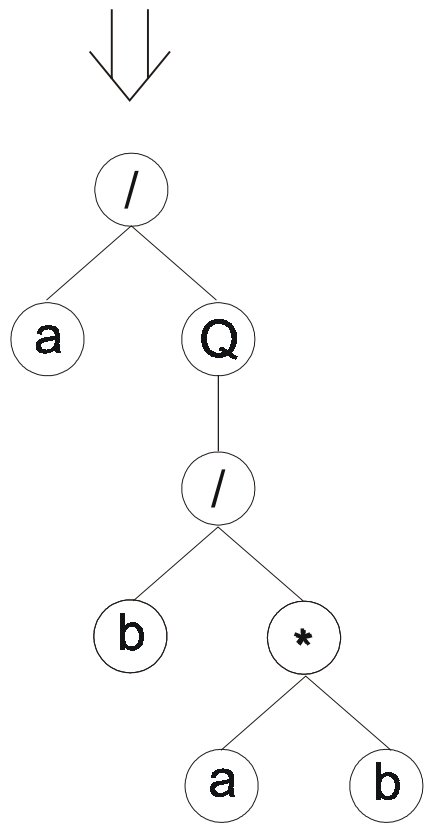
```
012345678901
*-/Qb+b+aaab
```



GEP Genes

```
012345678901234567890123456789 0
/aQ/b*ab/Qa*b*-ababaababbabbbba
```

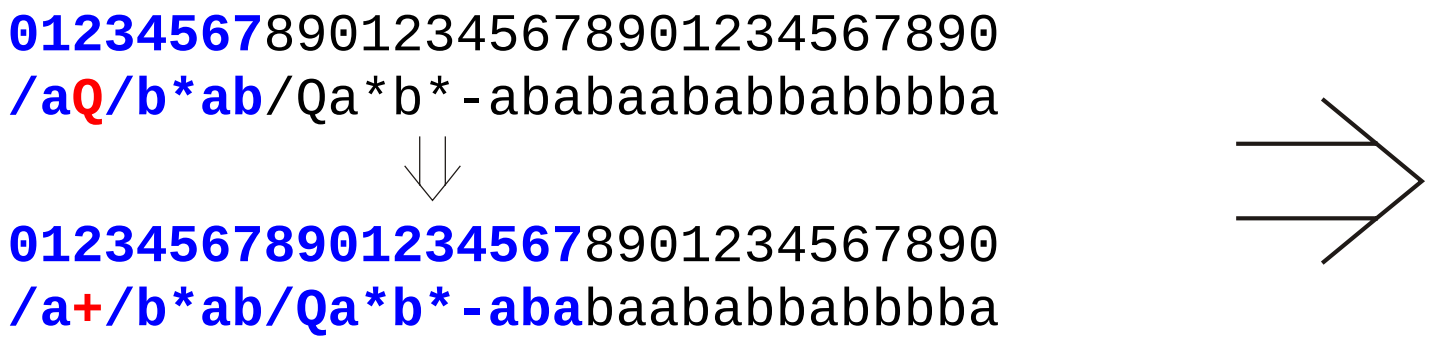
head tail

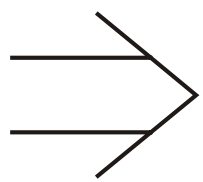


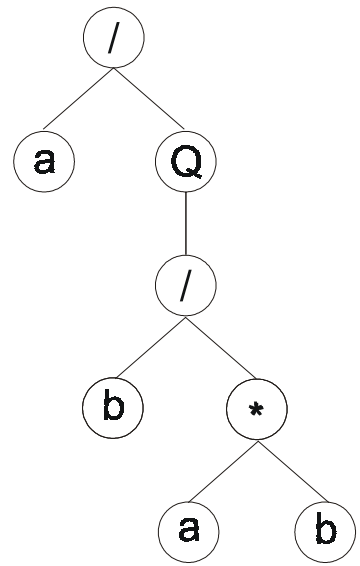
$t = h\,(n\text{-}1) + 1$

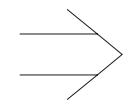
h - head
t - tail
n - max arity

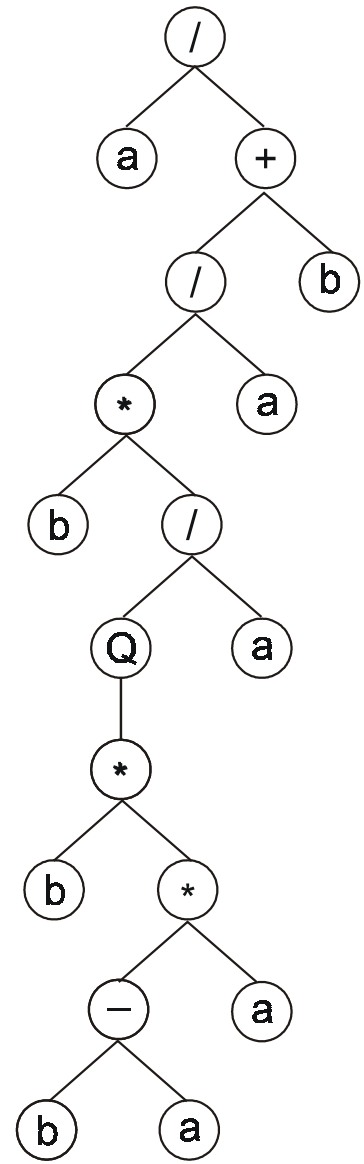
Plasticity of GEP Genes






/
a
Q
/
b
*
a
b

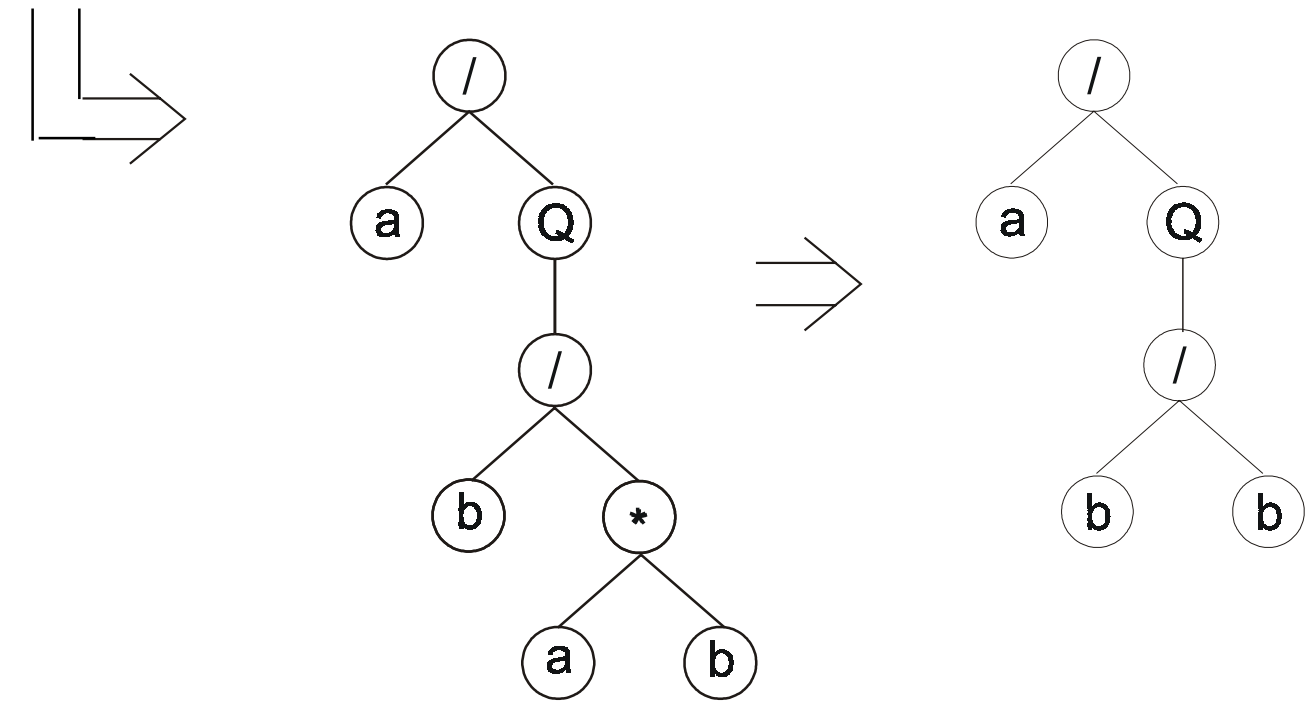



/
a
+
/
b
*
a
b
/
Q
a
*
b
*
–
a
b
a

Plasticity of GEP Genes

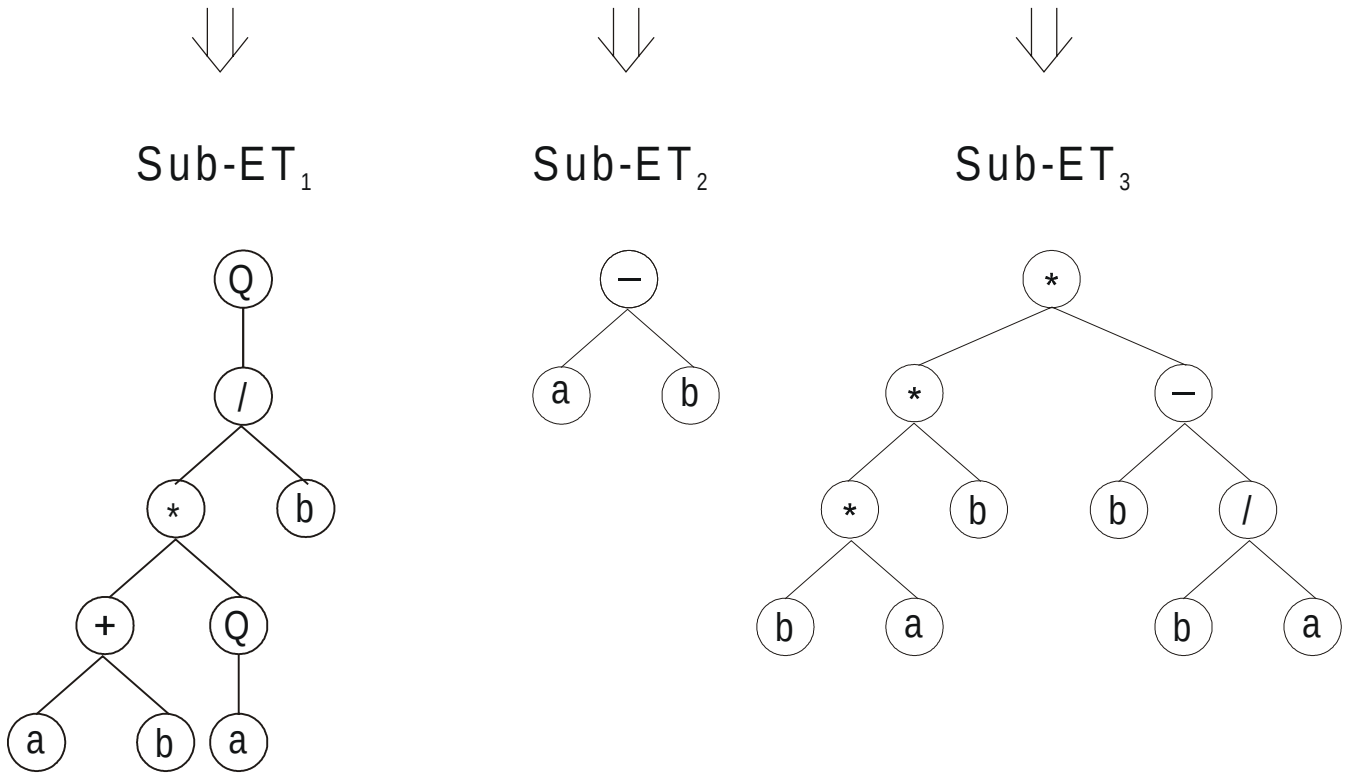
```
0123456789012345678901234567890
/aQ/b*ab/Qa*b*-abababbabbabbba
```

```
0123456789012345678901234567890
/aQ/bbab/Qa*b*-abababbabbabbba
```



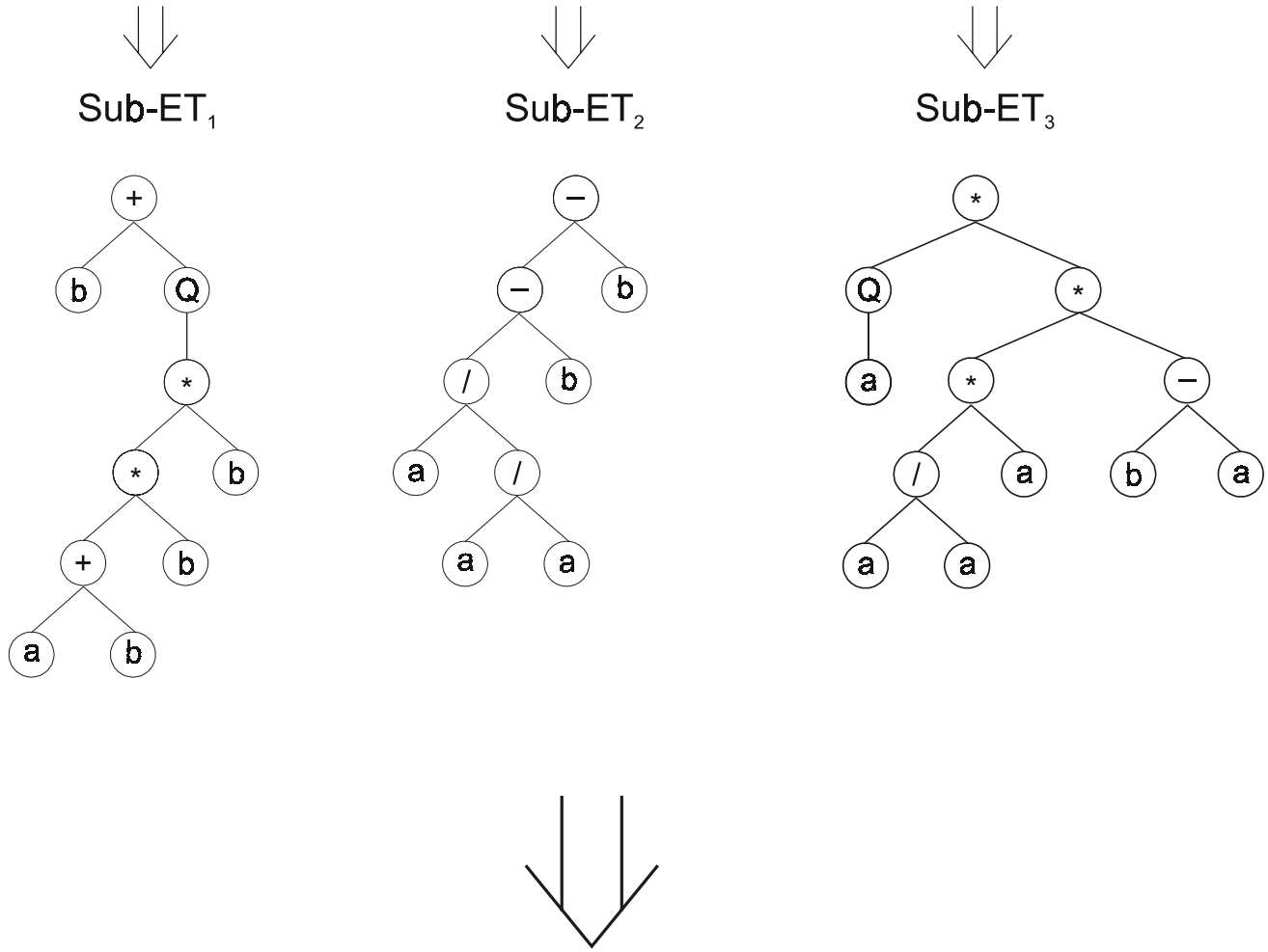
TRANSLATION

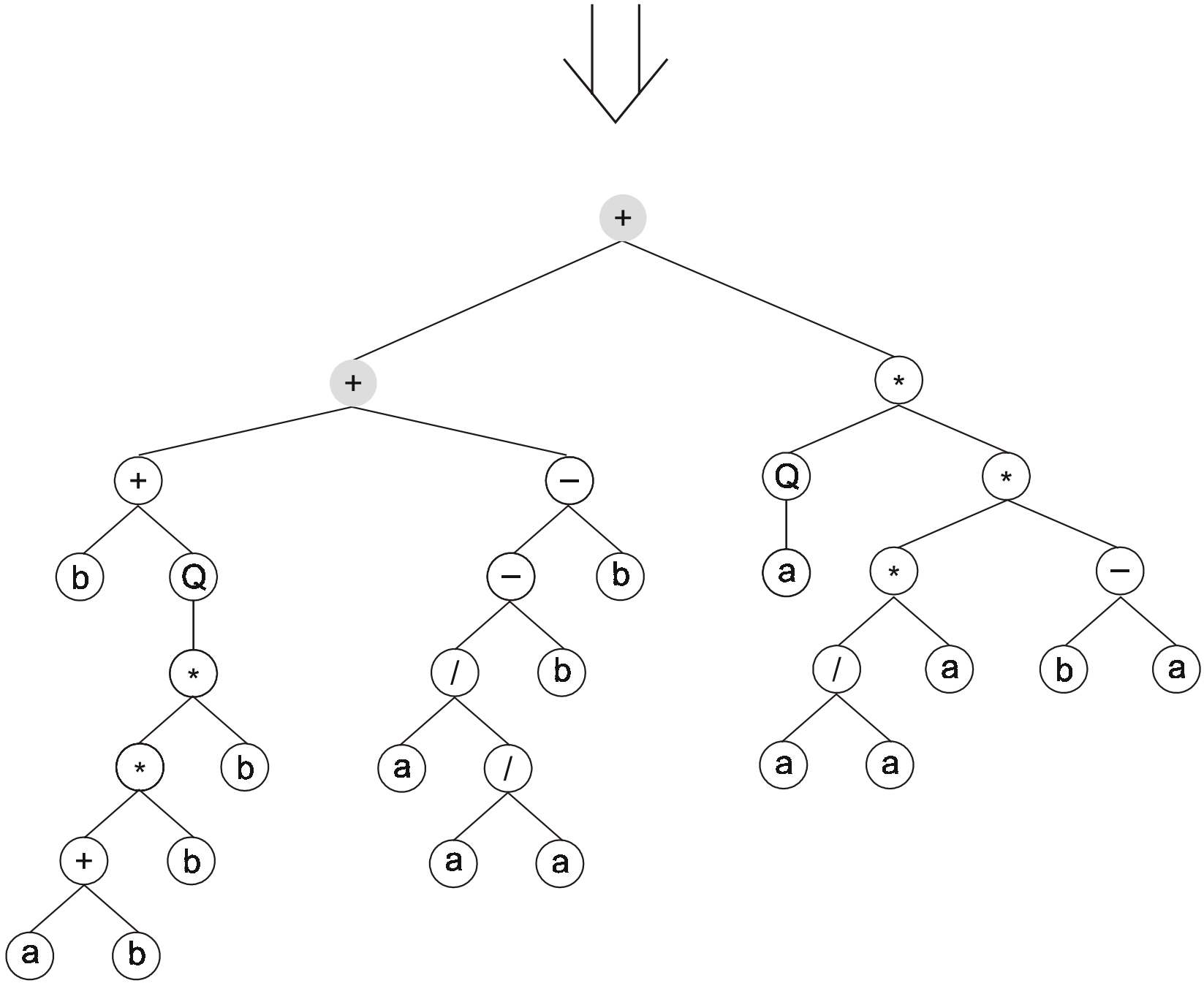
```
012345678901234012345678901234012345678901234
Q/*b+Qababaabaa-abQ/*+bababbab**-*bb/babaaaab
```



Expression of Algebraic Multigenic Chromosomes

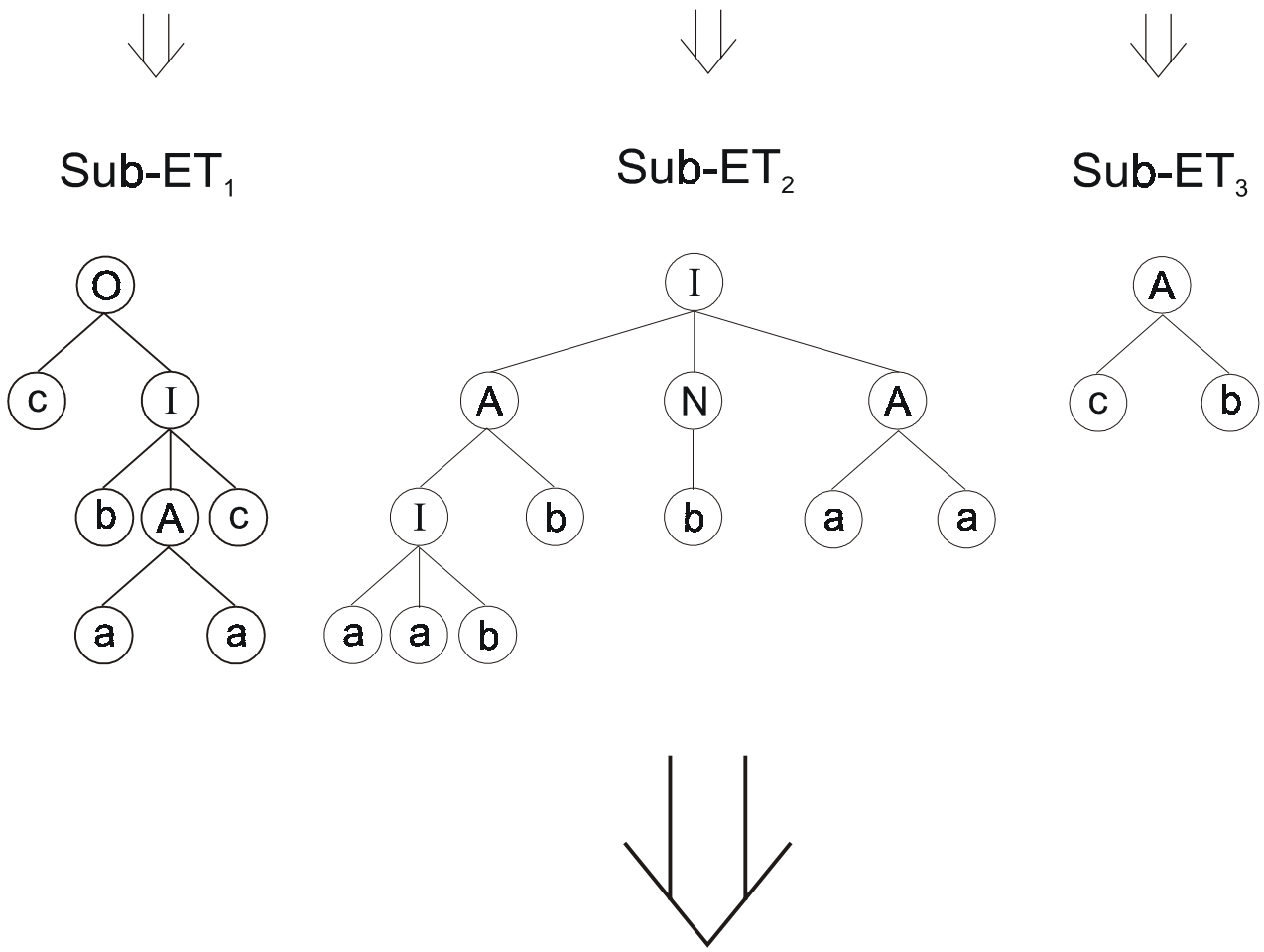
```
012345678901234012345678901234012345678901234
+bQ**b+bababbbb--b/ba/aaababab*Q*a*-/abaaaaab
```

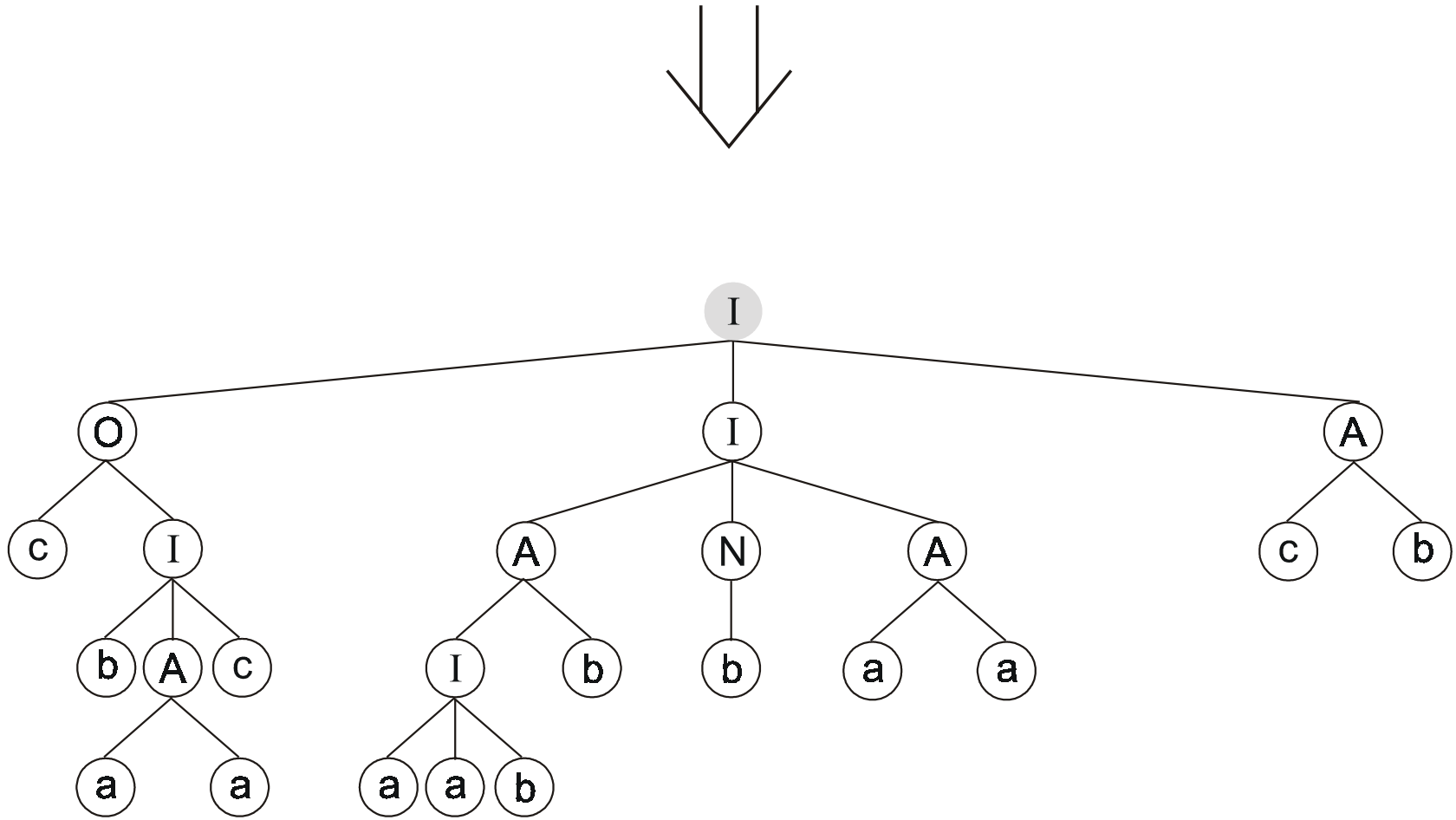



+
+
*
+
−
Q
*
b
Q
−
b
a
*
−
*
/
b
/
a
b
a
*
b
a
/
a
a
+
b
a
a
a
b

Expression of Boolean Multigenic Chromosomes

```
012345678901234501234567890123450123456789012345
OcIbAcaabcbccaaaIANAIbbaaaabaaabAcbcIcaaaacaccaa
```




I
O
I
A
c
I
A
N
A
c
b
b
A
c
I
b
b
a
a
a
a
b

Selection and Replication

```
Generation N: 0
*+-/a*aaaaaaa//+*aaaaaaaaa-[0] = 10.64033
/-/a//aaaaaaa+*+a/+aaaaaaa-[1] = 16.2117
*+a-+aaaaaaaa---///aaaaaaa-[2] = 13.81953
+a*/-aaaaaaaa**+a*aaaaaaaa-[3] = 18.32701
*-+a/-aaaaaaa/aa+a/aaaaaaa-[4] = 11.13926
+*//a/aaaaaaa---aa-aaaaaaa-[5] = 13.88255
*-*-*aaaaaaaa/-a///aaaaaaa-[6] = 7.777691
/++a-*aaaaaaa/+a*+-aaaaaaa-[7] = 13.14786
//+*aaaaaaaaa*+-/--aaaaaaa-[8] = 7.713599
-**+-/aaaaaaa*//aa/aaaaaaa-[9] = 8.73985

Generation N: 1
*+a-+aaaaaaaa---///aaaaaaa-[0] = 13.81953
/-/a//aaaaaaa+*+a/+aaaaaaa-[1] = 16.2117
*-+a/-aaaaaaa/aa+a/aaaaaaa-[2] = 11.13926
+*//a/aaaaaaa---aa-aaaaaaa-[3] = 13.88255
+a*/-aaaaaaaa**+a*aaaaaaaa-[4] = 18.32701
-**+-/aaaaaaa*//aa/aaaaaaa-[5] = 8.73985
-**+-/aaaaaaa*//aa/aaaaaaa-[6] = 8.73985
//+*aaaaaaaaa*+-/--aaaaaaa-[7] = 7.713599
/++a-*aaaaaaa/+a*+-aaaaaaa-[8] = 13.14786
/-/a//aaaaaaa+*+a/+aaaaaaa-[9] = 16.2117
```

Genetic Drift

(...)

```
Generation N: 8
/-/a//aaaaaaa+*+a/+aaaaaaa-[0] = 16.2117
/-/a//aaaaaaa+*+a/+aaaaaaa-[1] = 16.2117
/-/a//aaaaaaa+*+a/+aaaaaaa-[2] = 16.2117
/-/a//aaaaaaa+*+a/+aaaaaaa-[3] = 16.2117
/-/a//aaaaaaa+*+a/+aaaaaaa-[4] = 16.2117
/-/a//aaaaaaa+*+a/+aaaaaaa-[5] = 16.2117
/-/a//aaaaaaa+*+a/+aaaaaaa-[6] = 16.2117
/-/a//aaaaaaa+*+a/+aaaaaaa-[7] = 16.2117
/-/a//aaaaaaa+*+a/+aaaaaaa-[8] = 16.2117
/-/a//aaaaaaa+*+a/+aaaaaaa-[9] = 16.2117
```

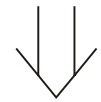
MUTATION

```
01234567890012345678900123456789
Q+bb*bbbaba-**--abbbaaQ*a*Qbbbaab
```

```
01234567890012345678900123456789
Q+bb/bbbabaQ**--abbbaaQ*+*Qbbbaab
```

IS TRANSPOSITION

```
012345678901234560123456789012345 6
-aba+Q-baabaabaabQ*+*+-/aababbaaaa
```



```
012345678901234560123456789012345 6
-aba+Q-baabaabaabQ*+a+Q*+ababbaaaa
```

RIS TRANSPOSITION

```
012345678901234560123456789012345678
*-bQ/+/babbabbba//Q*baa+bbbabbbbb
```

```
012345678901234560123456789012345678
Q/+*-bQ/babbabbba//Q*baa+bbbabbbbb
```

Gene Transposition

```
012345678901201234567890120123456789012
/+Qa*bbaaabaa*a*/Qbbbbbabb/Q-aabbaaabbb
```

```
012345678901201234567890120123456789012
/Q-aabbaaabbb/+Qa*bbaaabaa*a*/Qbbbbbabb
```

1-Point Recombination

```
012345678901234560123456789012345 6
+*-b-Qa*aabbbbaaa-Q-//b/*aabbabbab
++//b//-bbbbbbbbb-*-ab/b+bbbaabbaa
```

```
012345678901234560123456789012345 6
+*-b-Q/-bbbbbbbbb-*-ab/b+bbbaabbaa
++//b/a*aabbbbaaa-Q-//b/*aabbabbab
```

2-Point Recombination

```
012345678901234560123456789012345
*-+Q/Q*Qaaabbbbab QQab*++-aabbabaab
Q/-b-+/abaabbbaab/*-aQa*babbabbabb
```

```
012345678901234560123456789012345
*-+Q/+/abaabbbaab/*-aQa*-aabbabaab
Q/-b-Q*Qaaabbbbab QQab*++babbabbabb
```

Gene Recombination

```
012345678901201234567890120123456789012
/+/ab-aabbbbb-aa**+aaabaaa-+--babbbbaab
+baQaaaabaaba*-+a-aabbabbb/ab/+bbbabaaa
```

```
012345678901201234567890120123456789012
/+/ab-aabbbbb*-+a-aabbabbb-+--babbbbaab
+baQaaaabaaba-aa**+aaabaaa/ab/+bbbabaaa
```

Solving a Simple Problem

Test Function:

$$y = 3a^2 + 2a + 1$$

Fitness cases

a	f(a)
-4.2605	46.9346
-2.0437	9.44273
-9.8317	271.324
2.7429	29.0563
0.7328	4.07659
-8.6491	208.123
-3.6101	32.8783
-1.8999	8.02906
-4.8852	62.8251
7.3998	180.071

SETTINGS

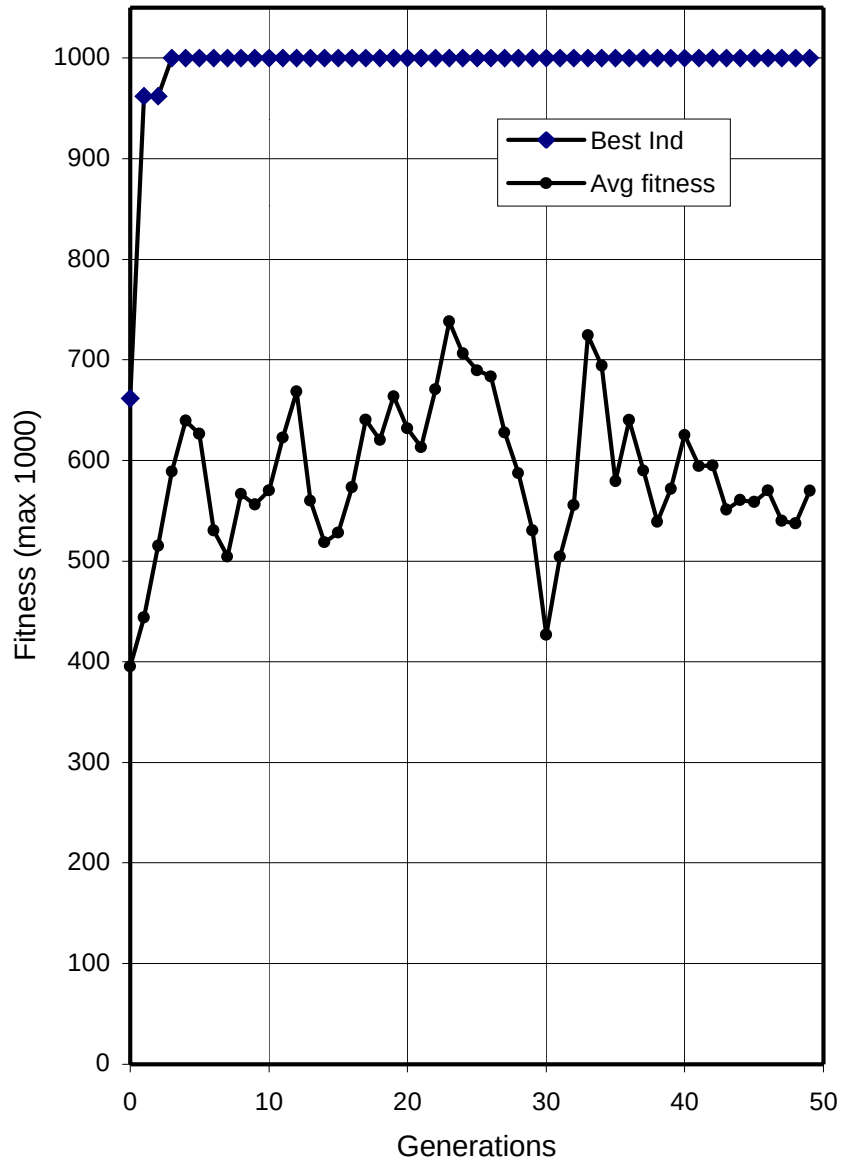
FITNESS FUNCTION:

$$f_i = \sum_{j=1}^{C_t} \left(M - \left| C_{(i,j)} - T_j \right| \right)$$

GENERAL SETTINGS

Number of generations	50
Population size	20
Number of fitness cases	10 (Table 1)
Function set	+ - * /
Gene length	13
Number of genes	3
Linking function	+
Chromosome length	39
Mutation rate	0.051
1-Point recombination rate	0.3
2-Point recombination rate	0.3
Gene recombination rate	0.1
IS transposition rate	0,1
IS elements length	1,2,3
RIS transposition rate	0.1
RIS elements length	1,2,3
Gene transposition rate	0.1
Selection range	100
Precision	0.01

EVOLUTIONARY DYNAMICS

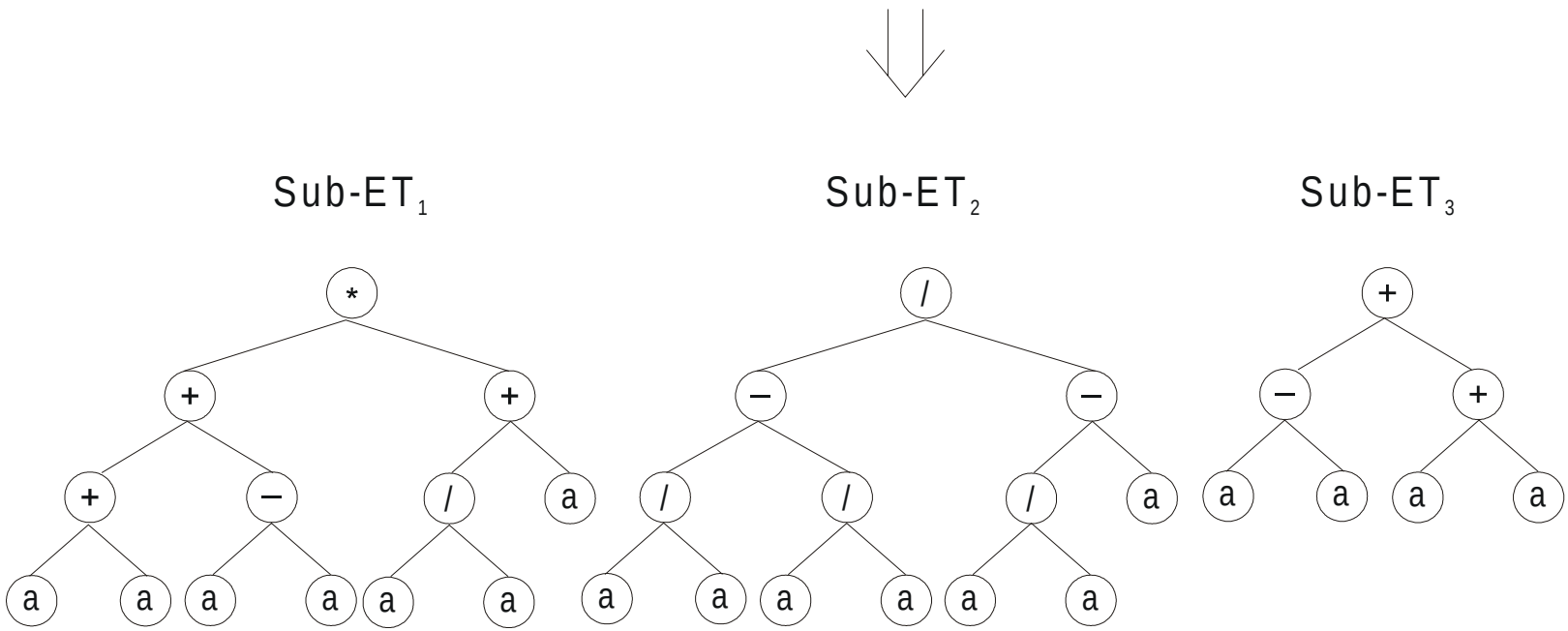


INITIAL POPULATION

```
Generation N: 0
01234567890120123456789012012345678901 2
+**/*/aaaaaaa/+a/a*aaaaaaa/a-*a+aaaaaaa-[ 0] = 577.3946
--aa++aaaaaaa+-/a*/aaaaaaa/--a-aaaaaaaa-[ 1] = 0
/***/+aaaaaaa*+/+-aaaaaaaa++aa/aaaaaaaa-[ 2] = 463.6533
-/+/++aaaaaaa+-//+/aaaaaaa+-/a/*aaaaaaa-[ 3] = 546.4241
++a/*aaaaaaaa+-+a*-aaaaaaa-a/-*aaaaaaaa-[ 4] = 460.8625
*+*a-*aaaaaaa*a/aa/aaaaaaa//+*a/aaaaaaa-[ 5] = 353.2168
*/**+aaaaaaaa+a/**+aaaaaaa----+/aaaaaaa-[ 6] = 492.6827
*aa-+-aaaaaaa+a/-+/aaaaaaa***/-*aaaaaaa-[ 7] = 560.9289
+/-*//aaaaaaa*+*//+aaaaaaa-/**+*aaaaaaa-[ 8] = 363.4358
--a+*/aaaaaaa+a++--aaaaaaa+a+aa+aaaaaaa-[ 9] = 386.7576
+-*-**aaaaaaa*/-+**aaaaaaa*+--++aaaaaaa-[10] = 380.6484
/a-**/aaaaaaa/-a/a/aaaaaaa+/a/-*aaaaaaa-[11] = 0
+--+//aaaaaaa+*+/*-aaaaaaa/*-a-+aaaaaaa-[12] = 551.2066
-a/+a/aaaaaaa*/--/aaaaaaaa*-+/a+aaaaaaa-[13] = 308.1296
/+/-+-aaaaaaa+-a/aaaaaaaaa**+-*-aaaaaaa-[14] = 0
//-*+/aaaaaaa//*a+aaaaaaaa/a++a*aaaaaaa-[15] = 489.5392
*a-a*-aaaaaaa+*+-a/aaaaaaa*/*aa*aaaaaaa-[16] = 399.2122
-a++*/aaaaaaa+/aa-*aaaaaaa---/**aaaaaaa-[17] = 317.6631
--a/*aaaaaaaa++*+-aaaaaaaa+-/*+-aaaaaaa-[18] = 597.8777
*+++-/aaaaaaa/--///aaaaaaa+-+aaaaaaaaaa-[19] = 661.5933
```

Best of Generation 0

```
012345678901201234567890120123456789012
*+++-/aaaaaaa/--///aaaaaaa+-+aaaaaaaaaa
```



$$y = \left(2a^2 + 2a\right) + \left(0\right) + \left(2a\right)$$

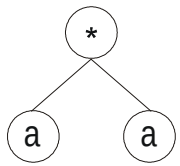
GENERATION 1

```
Generation N: 1
012345678901201234567890120123456789012
*+++-/aaaaaaa/--///aaaaaaa+-+aaaaaaaaaa-[ 0] = 661.5933
-a++*/aaaaaaa+//a--aaaaaaa---/**aaaaaaa-[ 1] = 0
+-*-**aaaaaaa*/-+**aaaaaaa*+--++aaaaaaa-[ 2] = 380.6484
+-*-**aaaaaaa*/-+**aaaaaaa*/*a**aaaaaaa-[ 3] = 356.9471
+-+aaaaaaaaaa*+++-/aaaaaaa/--///aaaaaaa-[ 4] = 661.5933
*aa-+-aaaaaaa+a/++/aaaaaaa***+-*aaaaaaa-[ 5] = 567.9289
*a-a*-aaaaaaa+/*-a/aaaaaaa*+-*++aaaaaaa-[ 6] = 449.802
*aa-+-aaaaaaa+a/-+/aaaaaaa*+--++aaaaaaa-[ 7] = 961.8512
/***/+aaaaaaa*+/+-aaaaaaaa-a/-*aaaaaaaa-[ 8] = 470.5862
+--+//aaaaaaa+*+/*-aaaaaaa/*-a-+aaaaaaa-[ 9] = 551.2066
*+++-/aaaaaaa-//--/aaaaaaa+-+aaaaaaaaaa-[10] = 0
--+a*-aaaaaaa++a/*aaaaaaaa-a/-*aaaaaaaa-[11] = 487.3099
-a++*/aaaaaaa+/aa-*aaaaaaa---/**aaaaaaa-[12] = 317.6631
++a/*aaaaaaaa+-+a*-aaaaaaa++aa/aaaaaaaa-[13] = 451.464
+--+/-aaaaaaa+a/**+aaaaaaa----+/aaaaaaa-[14] = 493.5336
*/-a++aaaaaaa+/aa-*aaaaaaa---/**aaaaaaa-[15] = 356.4241
+/-*//aaaaaaa*+a//+aaaaaaa-/+*+*aaaaaaa-[16] = 493.9218
*/**+aaaaaaaa+*+/*aaaaaaaa***/-*aaaaaaa-[17] = 448.4805
+-*-**aaaaaaa*/-+**aaaaaaa*+--++aaaaaaa-[18] = 380.6484
++a/*aaaaaaaa+-+a*+aaaaaaa--/-*aaaaaaaa-[19] = 380.8585
```

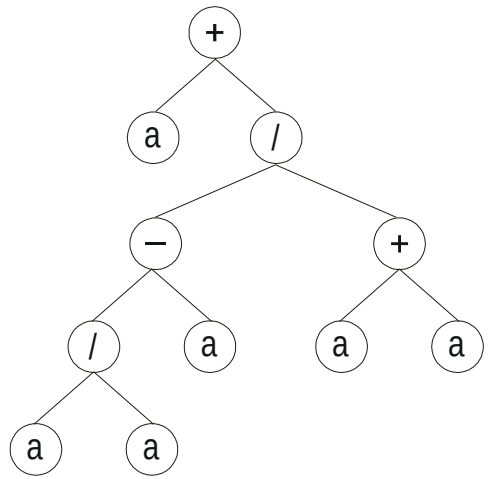
Best of Generation 1

```
012345678901201234567890120123456789012
*aa-+-aaaaaaa+a/-+/aaaaaaa*+--++aaaaaaa
```

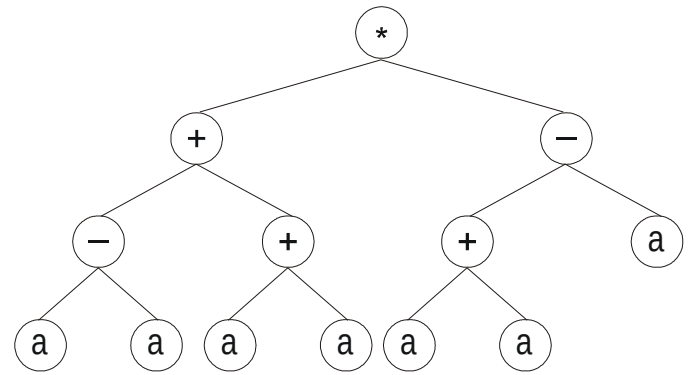
Sub-ET$_1$



Sub-ET$_2$



Sub-ET$_3$



$$y = \left(a^2\right) + \left(a + \frac{1-a}{2a}\right) + \left(2a^2\right)$$

GENERATION 2

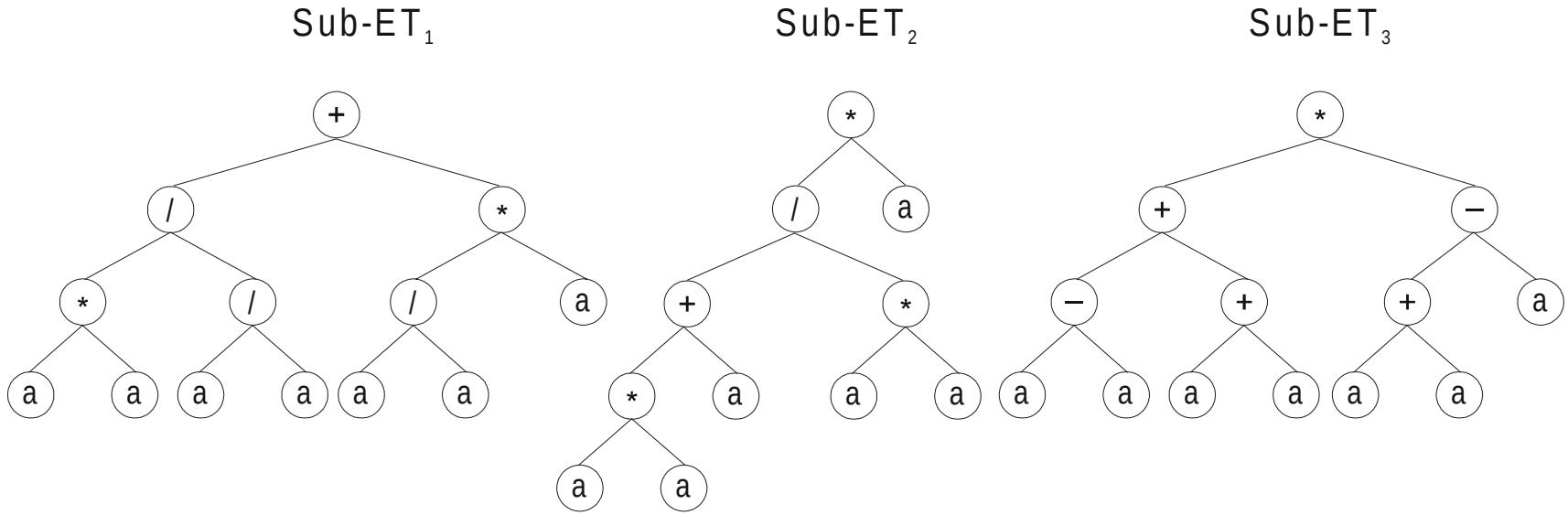
```
Generation N: 2
012345678901201234567890120123456789012
*aa-+-aaaaaaa+a/-+/aaaaaaa*+--++aaaaaaa-[ 0] = 961.8512
*/**+aaaaaaaa*/-+**aaaaaaa***/-*aaaaaaa-[ 1] = 446.2061
+-*-**aaaaaaa*+a//-aaaaaaa-/+*+*aaaaaaa-[ 2] = 323.1036
+--+//aaaaaaa+*+/*-aaaaaaa/*-*-+aaaaaaa-[ 3] = 551.2066
*aa-+-aaaaaaa+a/++/aaaaaaa***+-*aaaaaaa-[ 4] = 567.9289
++a/*aaaaaaaa*/-+-*aaaaaaa*+--++aaaaaaa-[ 5] = 0
+-*-**aaaaaaa+*+/*aaaaaaaa*/*a**aaaaaaa-[ 6] = 386.6484
++a/*aaaaaaaa+-+/*-aaaaaaa+aa++aaaaaaaa-[ 7] = 466.1533
+-*-a*aaaaaaa*/-+**aaaaaaa*a*a**aaaaaaa-[ 8] = 194.0452
/***/+aaaaaaa*+/+-aaaaaaaa-a--*aaaaaaaa-[ 9] = 541.4829
+-*-+*aaaaaaa+-+a*-aaaaaaa***/-*aaaaaaa-[10] = 346.2235
--*+*-aaaaaaa*aa-+-aaaaaaaaa/-+/aaaaaaa-[11] = 467.0862
*/-+**aaaaaaa+-*-*+aaaaaaa*/*a**aaaaaaa-[12] = 672.877
*aa+*/aaaaaaa+a/-+/aaaaaaa*+--++aaaaaaa-[13] = 961.8512
*+++/+aaaaaaa*++/+-aaaaaaa-a/-*aaaaaaaa-[14] = 395.858
/***-/aaaaaaa/--///aaaaaaa+-+a-aaaaaaaa-[15] = 467.0862
*aa-+-aaaaaaa+a/++/aaaaaaa***+-*aaaaaaa-[16] = 567.9289
+-+aaaaaaaaaa*+++-/aaaaaaa/--///aaaaaaa-[17] = 661.5933
+/-*//aaaaaaa*/a+**aaaaaaa*+--++aaaaaaa-[18] = 903.8886
*/**+aaaaaaaa+*+/*aaaaaaaa+/aa/aaaaaaaa-[19] = 423.885
```

GENERATION 3

```
Generation N: 3
012345678901201234567890120123456789012
*aa+*/aaaaaaa+a/-+/aaaaaaa*+--++aaaaaaa-[ 0] = 961.8512
*aa-+-aaaaaaa+a/-+/aaaaaaa/--///aaaaaaa-[ 1] = 560.9289
*aa-+-aaaaaaa-++/+-aaaaaaa-a/-*aaaaaaaa-[ 2] = 558.2066
*+++/+aaaaaaa*+a/-+aaaaaaa++--++aaaaaaa-[ 3] = 569.0469
/+++/+aaaaaaa*++/+-aaaaaaa-a/-*aaaaaaaa-[ 4] = 699.5153
+-+aa/aaaaaaa++++-/aaaaaaa***+-*aaaaaaa-[ 5] = 466.1533
*aa-+-aaaaaaaaa--**aaaaaaa*+--++aaaaaaa-[ 6] = 957.9443
--++*-aaaaaaa*a+/*-aaaaaaa+aa++aaaaaaaa-[ 7] = 337.7807
*aaa*/aaaaaaa+a+-+/aaaaaaa*+-/++aaaaaaa-[ 8] = 953.9443
/***/-aaaaaaa*+/+-aaaaaaaa-a--*aaaaaaaa-[ 9] = 0
*aa-+-aaaaaaa+a/-+/aaaaaaa*/--++aaaaaaa-[10] = 560.9289
*aa-+-aaaaaaa+a/++/aaaaaaa/--///aaaaaaa-[11] = 567.9289
+-+a-aaaaaaaa/***-/aaaaaaa*+--++aaaaaaa-[12] = 676.0663
+/**//aaaaaaa*/a+**aaaaaaa*+--++aaaaaaa-[13] = 1000
*/-+**aaaaaaa+-*-*+aaaaaaa*/*a**aaaaaaa-[14] = 672.877
/***/+aaaaaaa/+*+/+aaaaaaa-a*/--aaaaaaa-[15] = 498.3734
+/-*//aaaaaaa*/a+-*aaaaaaa*+--++aaaaaaa-[16] = 0
--*+--aaaaaaa*/a-+-aaaaaaa/a/-+/aaaaaaa-[17] = 506.1233
++a/*aaaaaaaa+-a-+-aaaaaaa-a*-+/aaaaaaa-[18] = 815.7772
*+a//-aaaaaaa+a/-+/aaaaaaa-/+*+*aaaaaaa-[19] = 412.5237
```

Perfect Solution

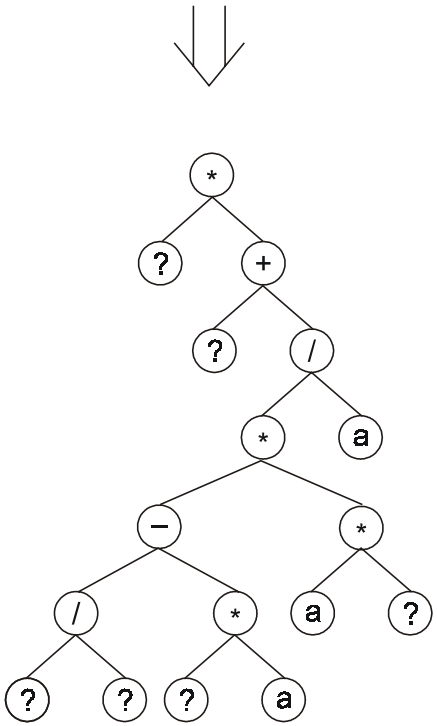
```
012345678901201234567890120123456789012
+/**//aaaaaaa*/a+**aaaaaaa*+--++aaaaaaa
```

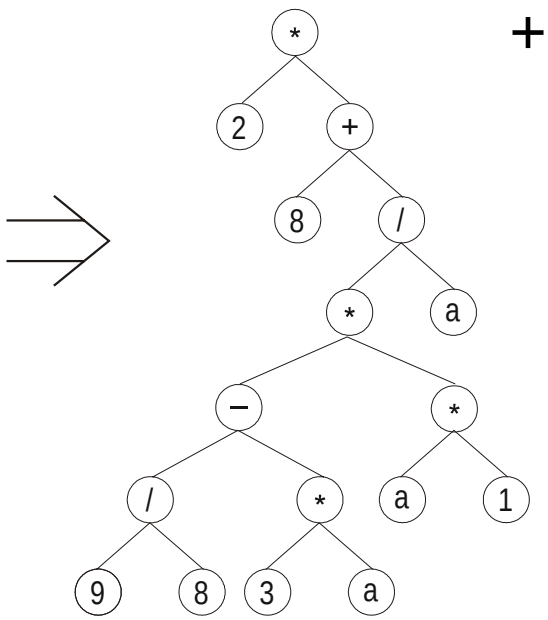


$$y = \left(a^2 + a\right) + \left(a + 1\right) + \left(2a^2\right) = 3a^2 + 2a + 1$$

Using Numerical Constants

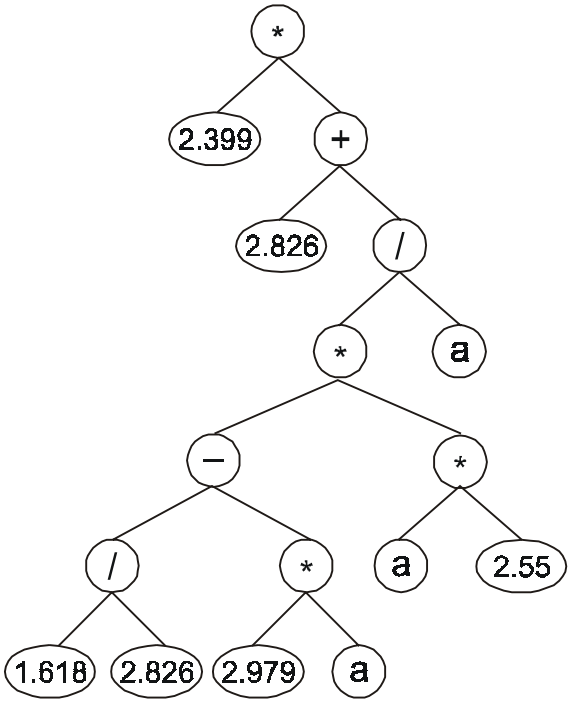
```
01234567890123456789012345678901234
*?+?/*a-*/*a????a??a??a281983874486
```





+

```
[0] = -2.829
[1] = 2.55
[2] = 2.399
[3] = 2.979
[4] = 2.442
[5] = 0.662
[6] = 1.797
[7] = -1.272
[8] = 2.826
[9] = 1.618
```



Finding The 'V' Shaped Function

Test Function:

$$y = 4.251a^2 + \ln(a^2) + 7.243e^a$$

Fitness function:

$$f_i = \sum_{j=1}^{C_t} \left(M - \left| \frac{C_{(i,j)} - T_j}{T_j} \cdot 100 \right| \right)$$

Fitness cases

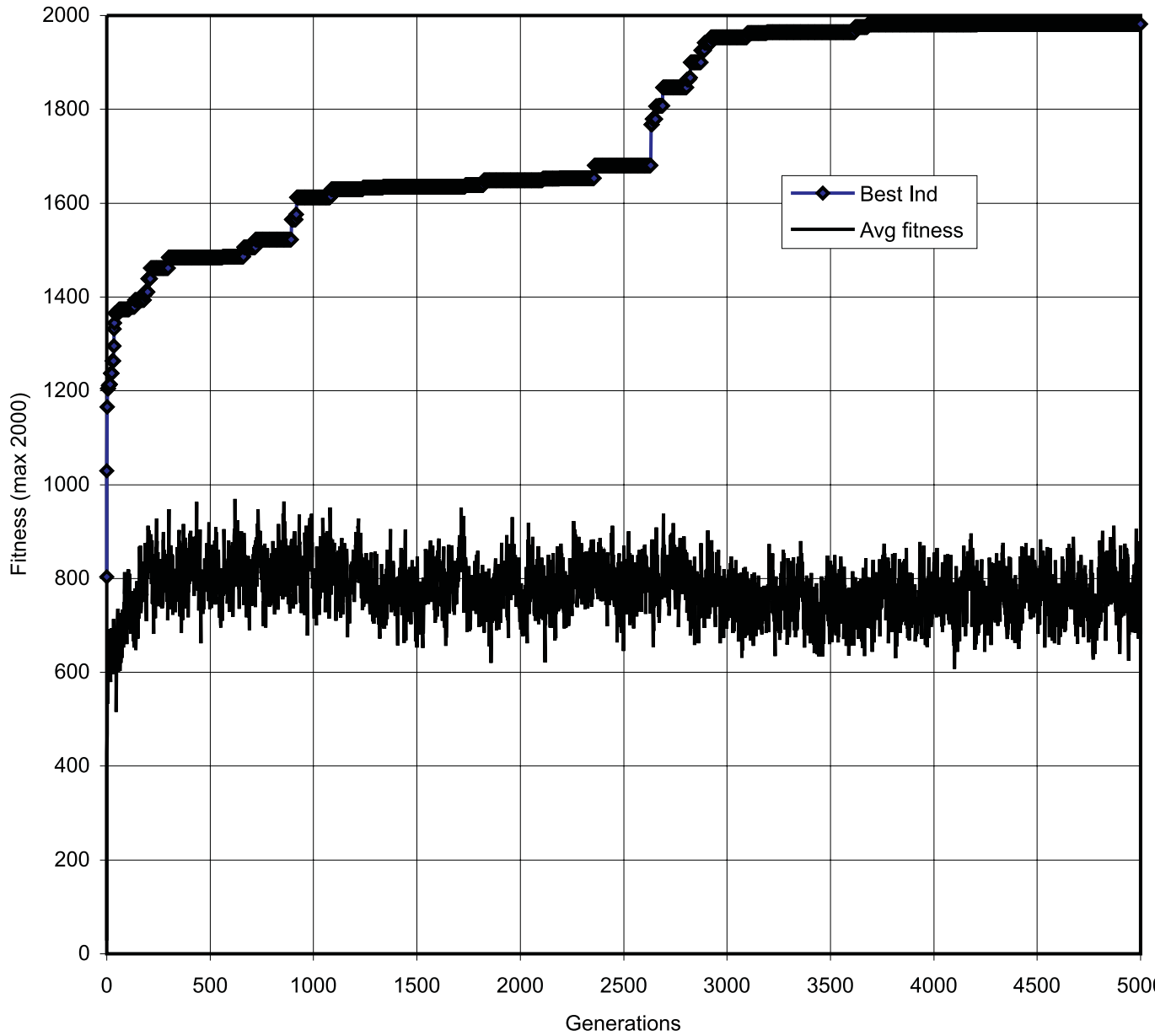
a	f(a)
-0.2639725157548009	3.194980662652764
0.0578905532656938	1.990520017259985
0.3340252901096346	8.396637039972868
-0.2363345775644623	3.070889769728257
-0.8557443825668047	5.879467636957033
-0.0194437136332785	-0.7753263223284588
-0.1921343881833043	2.834702257744086
0.5293079101246271	12.21547266421373
-0.007889741187284598	-2.498039834186359
0.4389698049506311	10.40717348588088
-0.1075592926980396	2.094136356459081
-0.2745569943771633	3.239272780108398
-0.05953332196045281	1.197012847673475
0.3844929939583523	9.355807691898551
-0.8749230207363339	6.006424530013026
-0.236546636250546	3.071897290438372
-0.1678759417045577	2.674400531309863
0.9506821818220914	22.48196398441491
0.9469791595773622	22.37501611873555
0.6393399100595915	14.5701285332337

SETTINGS

GENERAL SETTINGS

	With Random Constants	Without Random Constants
Number of runs	100	100
Number of generations	5000	5000
Population size	100	100
Number of fitness cases	20 (Table 3)	20 (Table 3)
Function set	+ - * / L E K ~ S C	+ - * / L E K ~ S C
Head length	6	6
Number of genes	4	5
Linking function	+	+
Chromosome length	80	65
Mutation rate	0.044	0.044
1-Point recombination rate	0.3	0.3
2-Point recombination rate	0.3	0.3
Gene recombination rate	0.1	0.1
IS transposition rate	0.1	0.1
IS elements length	1,2,3	1,2,3
RIS transposition rate	0.1	0.1
RIS elements length	1,2,3	1,2,3
Gene transposition rate	0.1	0.1
Rand. const. mut. rate	0.01	--
Dc specific IS transp. rate	0.1	--
Dc specific IS elements length	1,2,3	--
Selection range	100%	100%
Precision	0.01%	0.01%
Average best-of-run fitness	1850.476	1934.619

EVOLUTIONARY DYNAMICS



Model Evolved With Random Constants

Gene 0: **L*~*+/aa?a??a2132990**

A_0 = {0.565, 0.203, 0.613, 0.219, 0.28, 0.25, 0.48, 0.427, 0.821, 0.127}

Gene 1: **E-+-*?aaaaaaa7332660**

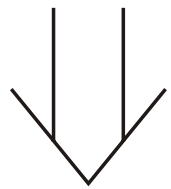
A_1 = {0.031, 0.046, 0.696, 0.643, 0.528, 0.417, 0.978, 0.811, 0.637, 0.988}

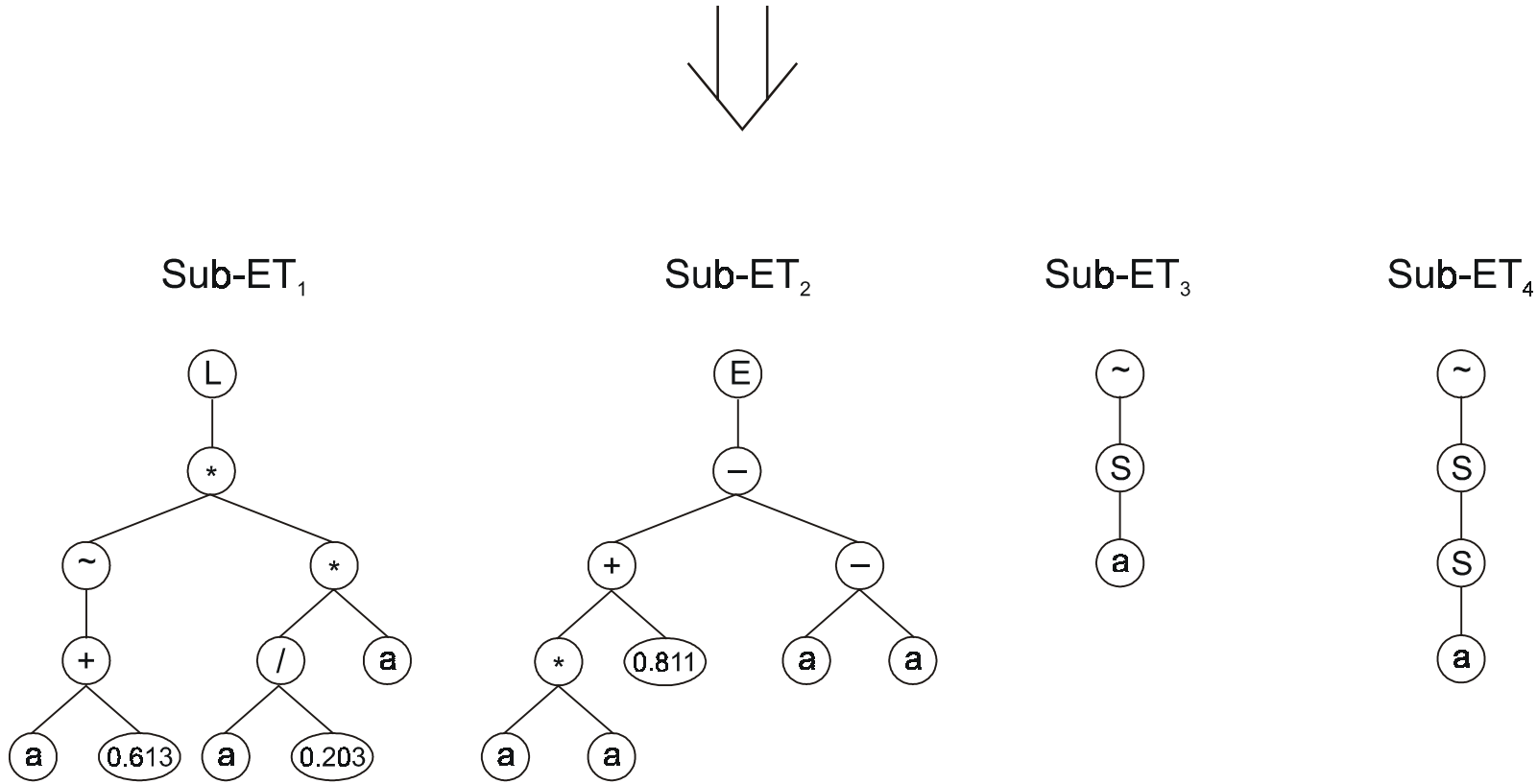
Gene 2: **~Saaa+??aa??a9109969**

A_2 = {0.515, 0.466, 0.254, 0.219, 0.425, 0.942, 0.306, 0.619, 0.821, 0.262}

Gene 3: **~SSaES?????aa5420661**

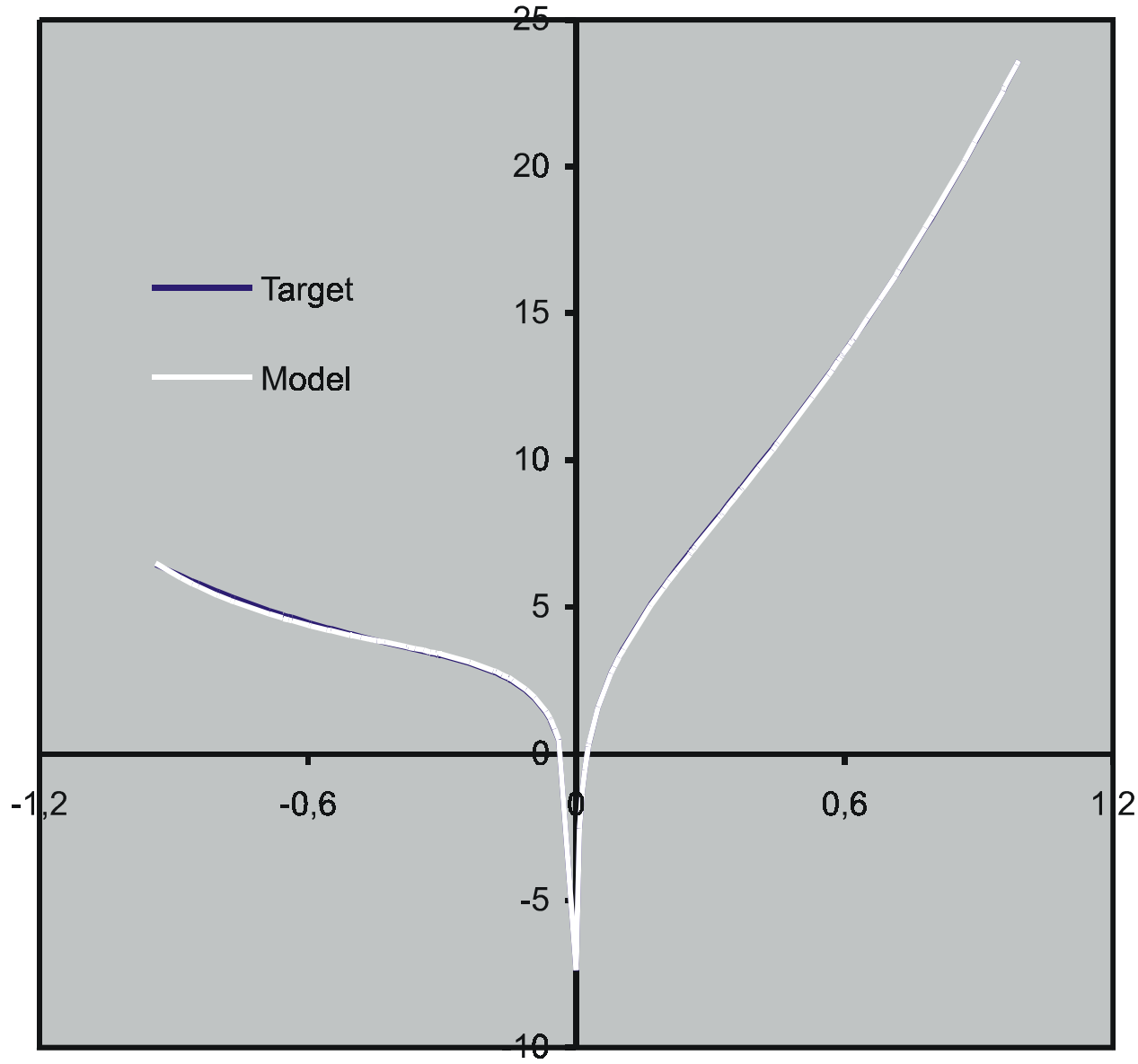
A_3 = {0.595, 0.547, 0.525, 0.219, 0.2970.387, 0.508, 0.695, 0.728, 0.415}





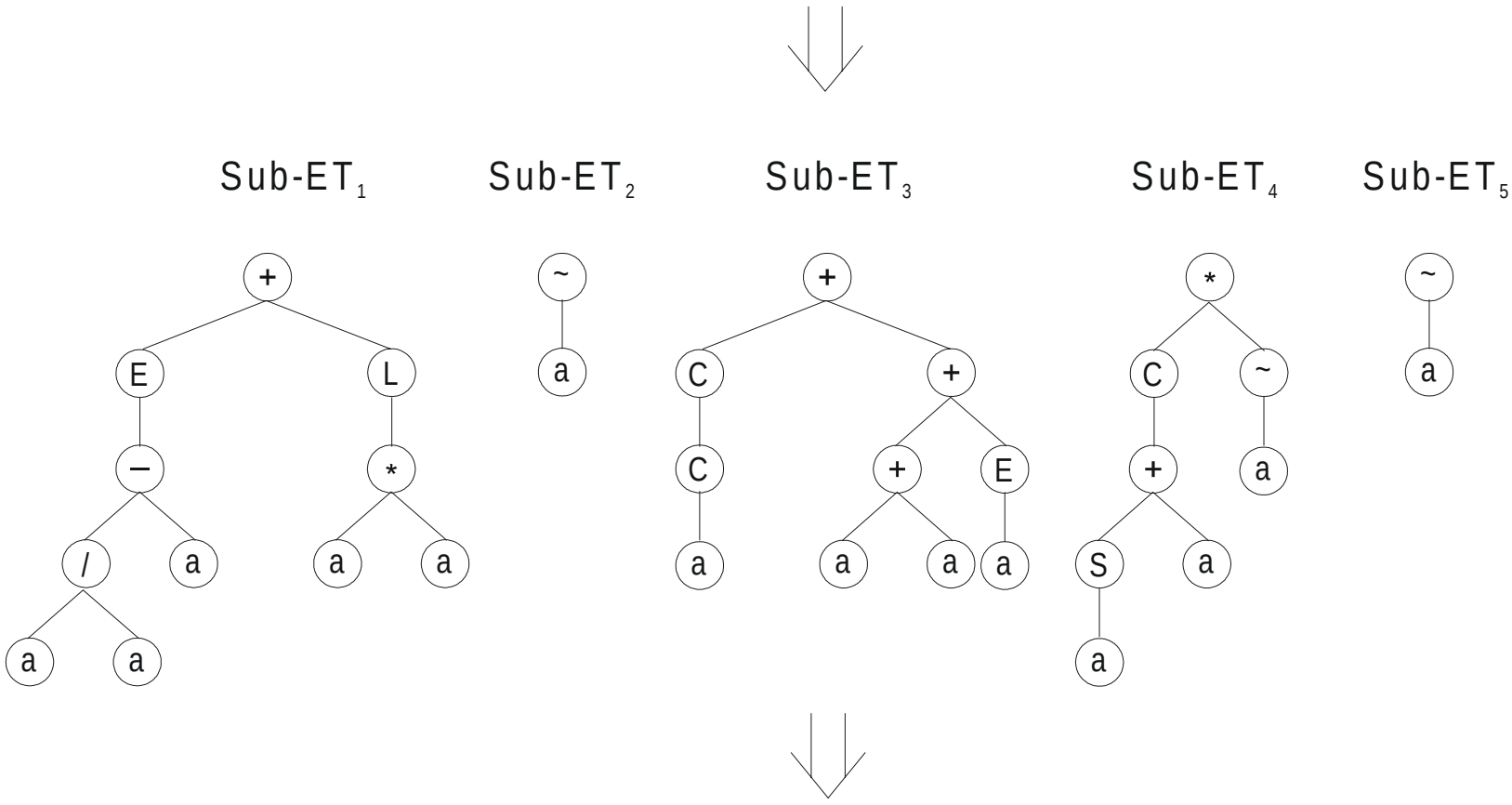
$$y = \left[\ln \left(10^{(a+0.613)} \cdot \frac{a^2}{0.203} \right) \right] + \left[e^{(a^2+0.811)} \right] + \left[10^{\sin(a)} \right] + \left[10^{\sin(\sin(a))} \right]$$

Plot Comparison



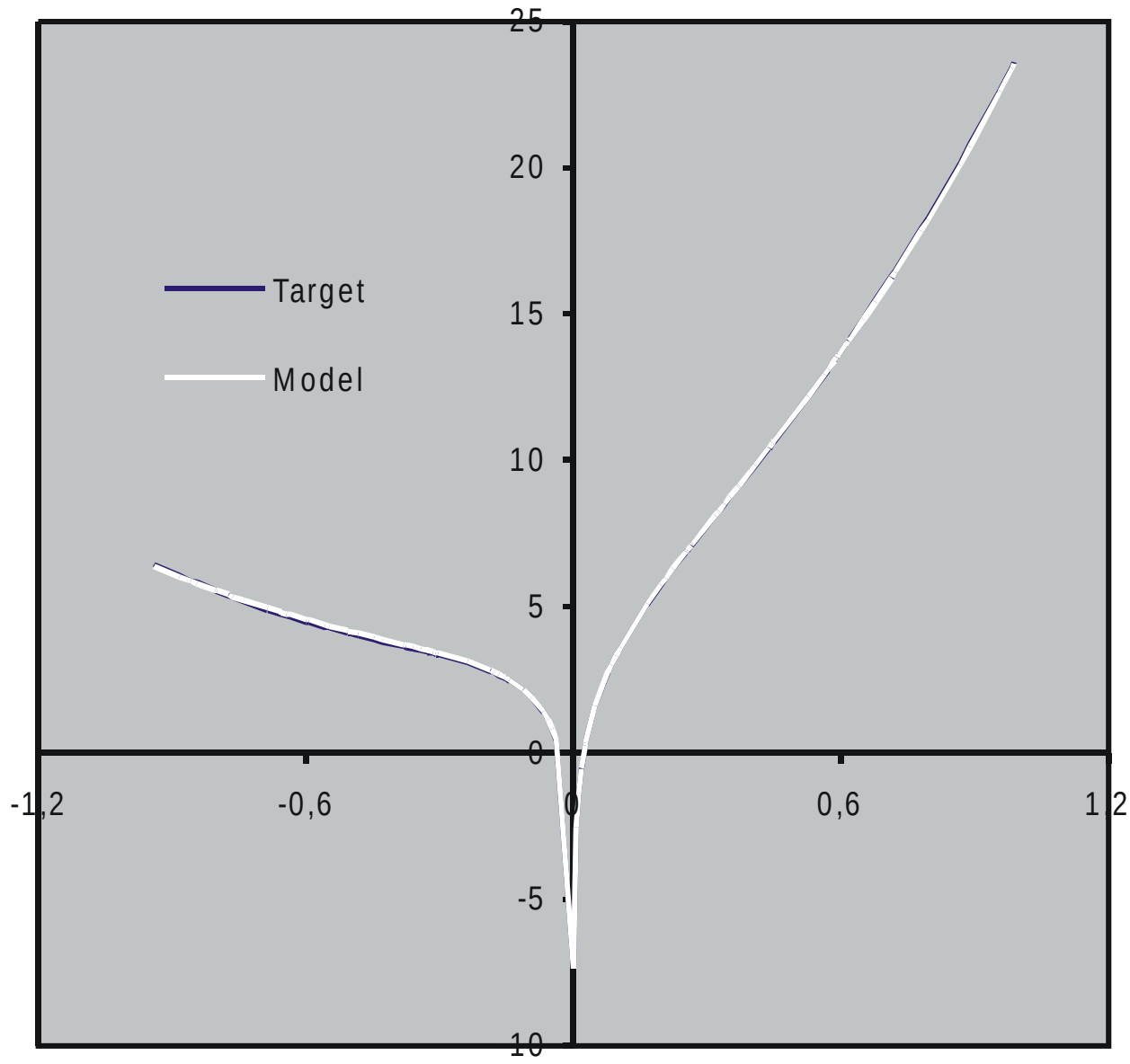
Model Evolved Without Random Constants

```
012345678901201234567890120123456789012012345678901201234567890120123456789012
+EL-*/aaaaaaa~a+E/Laaaaaaa+C+C+Eaaaaaaa*C~+aSaaaaaaa~a-L~+aaaaaaa
```



$$y = \left[e^{(1-a)} + \ln\left(a^2\right)\right] + \left[10^a\right] + \left[\cos\left(\cos\left(a\right)\right) + 2a + e^a\right] + \left[\cos\left(\sin\left(a\right) + a\right) \cdot 10^a\right] + \left[10^a\right]$$

Plot Comparison



FINDING A 5-PARAMETER FUNCTION

TEST FUNCTION:

$$y = \frac{\sin(a) \cdot \cos(b)}{\sqrt{10^c}} + \tan(d - e)$$

FITNESS FUNCTION:

$$f_i = \sum_{j=1}^{C_t} \left(M - \left| \frac{C_{(i,j)} - T_j}{T_j} \cdot 100 \right| \right)$$

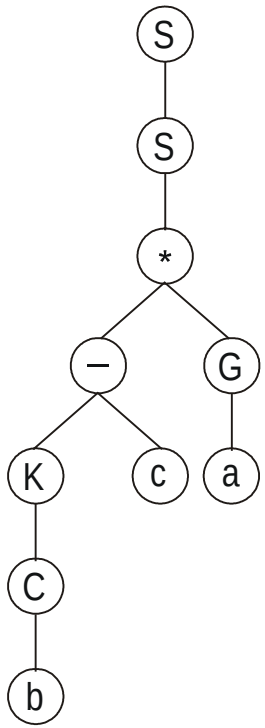
GENERAL SETTINGS

Number of generations	1000
Population size	100
Number of fitness cases	100
Function set	+ - * / Q K ~ S C G
Gene length	19
Number of genes	3
Linking function	+
Chromosome length	57
Mutation rate	0.044
1-Point recombination rate	0.3
2-Point recombination rate	0.3
Gene recombination rate	0.1
IS transposition rate	0,1
IS elements length	1,2,3
RIS transposition rate	0.1
RIS elements length	1,2,3
Gene transposition rate	0.1
Selection range	100%
Precision	0%

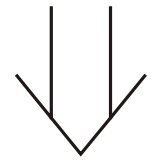
Evolved Model

```
012345678901234567801234567890123456780123456789012345678
SS*-GKcaCbbccbeabdbaC--SKaeGceadddabadG-de*add+adedabdeaa
```

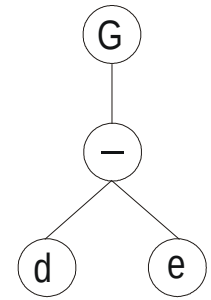
Sub-ET$_1$

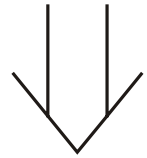


Sub-ET$_2$

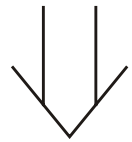


Sub-ET$_3$





$$y = \left[\sin\left(\sin\left(\left(\log\left(\cos\left(b\right)\right) - c\right) \cdot \tan\left(a\right)\right)\right)\right] + \left[a\right] + \left[\tan(d - e)\right]$$



```
double APSCfunction(double d[ ])
{
   double dblTemp = 0;
   dblTemp+=sin(sin(((log10(cos(d[1]))-d[2])*tan(d[0]))));
   dblTemp += d[0];
   dblTemp += tan((d[3]-d[4]));
   return dblTemp;
}
```