

# Genetic Improvement

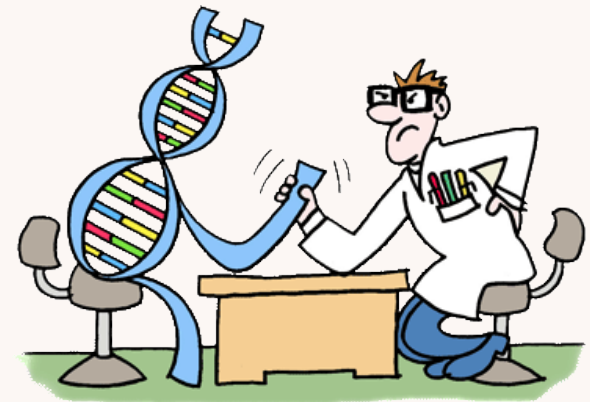
## GGGP

**EPSRC**

[EP/M025853/1](https://www.epsrc.ac.uk/EP/M025853/1)

W. B. Langdon

Computer Science, University College London



**Humies**

<http://www.human-competitive.org/>

Human-Competitive results

**\$10,000 prizes**

GECCO-2020 in Cancun, Mexico

Email 29 May [goodman@msu.edu](mailto:goodman@msu.edu)



GI: A Comprehensive Survey



**WIKIPEDIA**

Genetic Improvement

27.2.2020

# Genetic Improvement

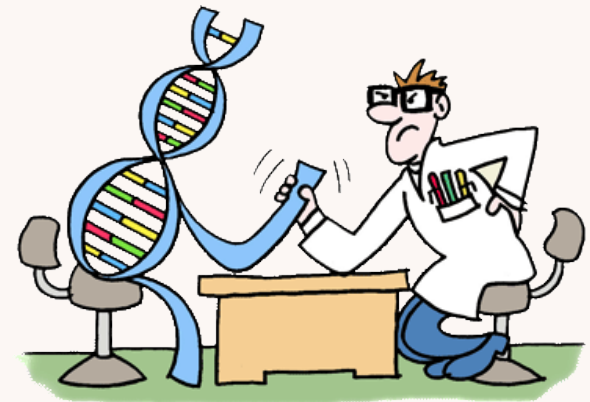
## GGGP

**EPSRC**

[EP/M025853/1](http://www.epsrc.ac.uk/EP/M025853/1)

[W. B. Langdon](#)

Computer Science, University College London



**Humies**

<http://www.human-competitive.org/>

Human-Competitive results

**\$10,000 prizes**

GECCO-2020 in Cancun, Mexico

Email 29 May [goodman@msu.edu](mailto:goodman@msu.edu)



[GI: A Comprehensive Survey](#)



**WIKIPEDIA**  
Genetic Improvement

# Genetic Improvement of Software

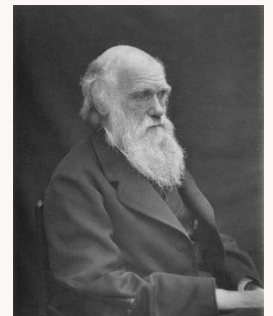
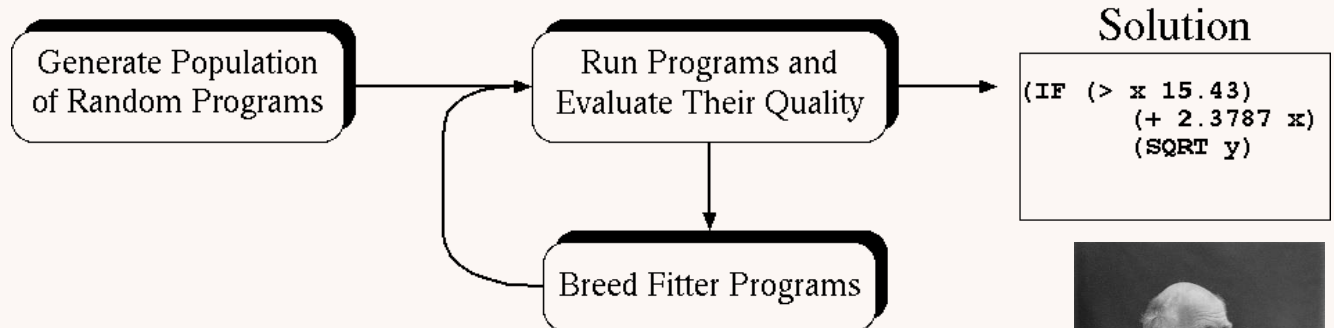
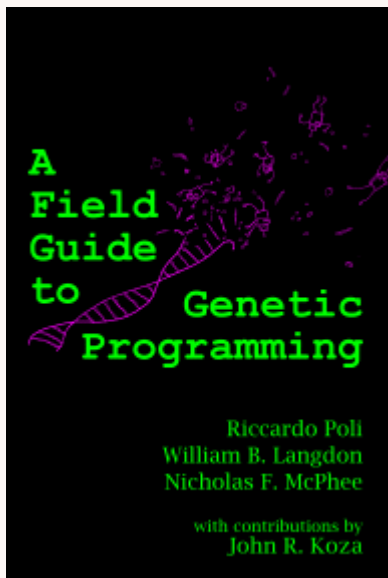
- What is Genetic Improvement
  - Optimisation, eg. Genetic Programming (GP), on existing code
- What has Genetic Improvement done
  - Technology behind automatic bug fixing
  - Improvement of existing code: speedup, transplanting, program adaptation, parallel, mobile energy reduction
  - Faster parallel DNA sequence analysis
  - **Better RNA folding predictions**
- Conclusions

# What is Genetic Improvement

# Genetic Improvement

Typically use GP to evolve a population of programs

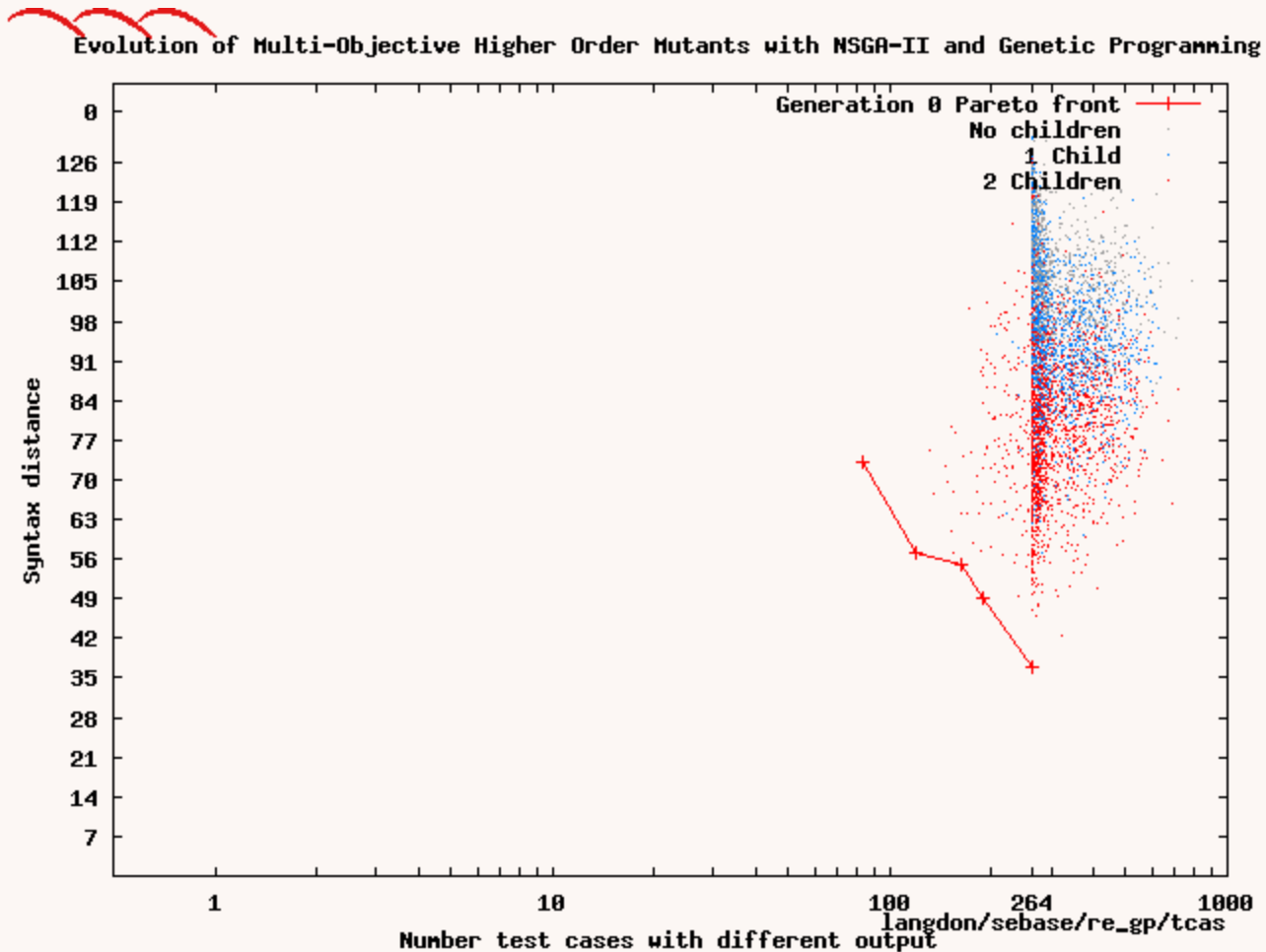
- Start with representation of human written code
- Programs' fitness is determined by running them
- Better programs are selected to be parents
- New generation of programs are created by randomly combining above average parents or by mutation.
- Repeat generations until solution found.



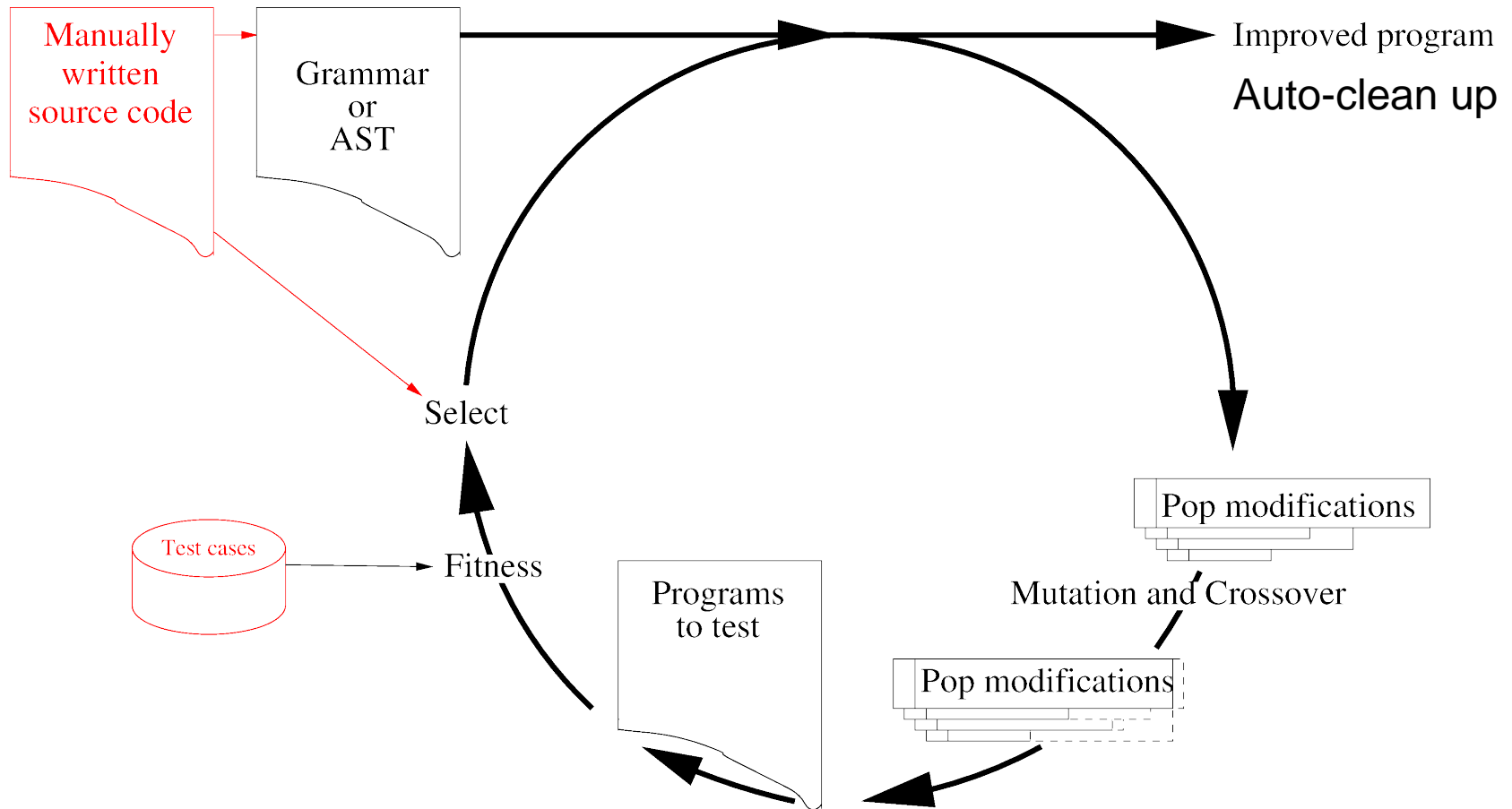
Charles Darwin 1809-1882

Free PDF    Free E-book kindle

# Evolving population of programs



# Typical GI Evolutionary Cycle



Many types of mutation.

Eg replace line of C++ code with another from the same file.

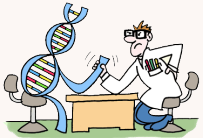

# GI Automatic Coding

- Genetic Improvement does not start from zero
- Use existing system
  - Source of non-random code
    - Use existing code as test “Oracle”.  
(Program is its own functional specification)
    - Can always compare against previous version
    - Easier to tell if better than if closer to poorly defined goal functionality.
- Testing scales (sort of).

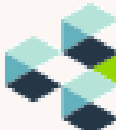


# What has Genetic Improvement done

# GP Automatic Bug Fixing (APR)

- Run code: example to reproduce bug, a few tests to show fixed code still works.
- Search for replacement C statement within program which fixes bug. Fault location tool
- Real bugs in real programs (mostly C/C++ or Java).
  - Multiple prizes and best papers, including: 
  - 1<sup>st</sup> prize Human-Competitive [\[ICSE\]](#) Gold [Humie](#)
- In daily use: Iceland health clinic [\[GI-2017\]](#) Python Facebook [SapFix](#)  Mark Harman

# GI to Speed up human written programs

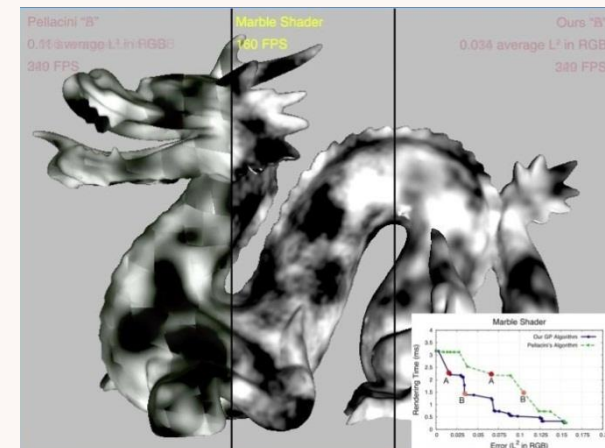
- Bowtie2, 70 times faster [[IEEE TEVC 2015](#)]
- GPGPU BarraCUDA [[BioData Mining](#)]
  - In use since 2015. 4000 downloads from [SF](#)
  - On real data speed up to 3 times ([arXiv.org](#))
  - Commercial use by [Lab7](#) (in BioBuilds [2015](#))
  - Ported by IBM to their Power8
  - [Cambridge Epigenetix](#)   
GTX 1080 21x faster than bwameth (twin core CPU)
- [Microsoft Azure GPU cloud](#)
- RNAfold CUDA, AVX, improved predictions

# Genetic Improvement to Reduce Resource Consumption

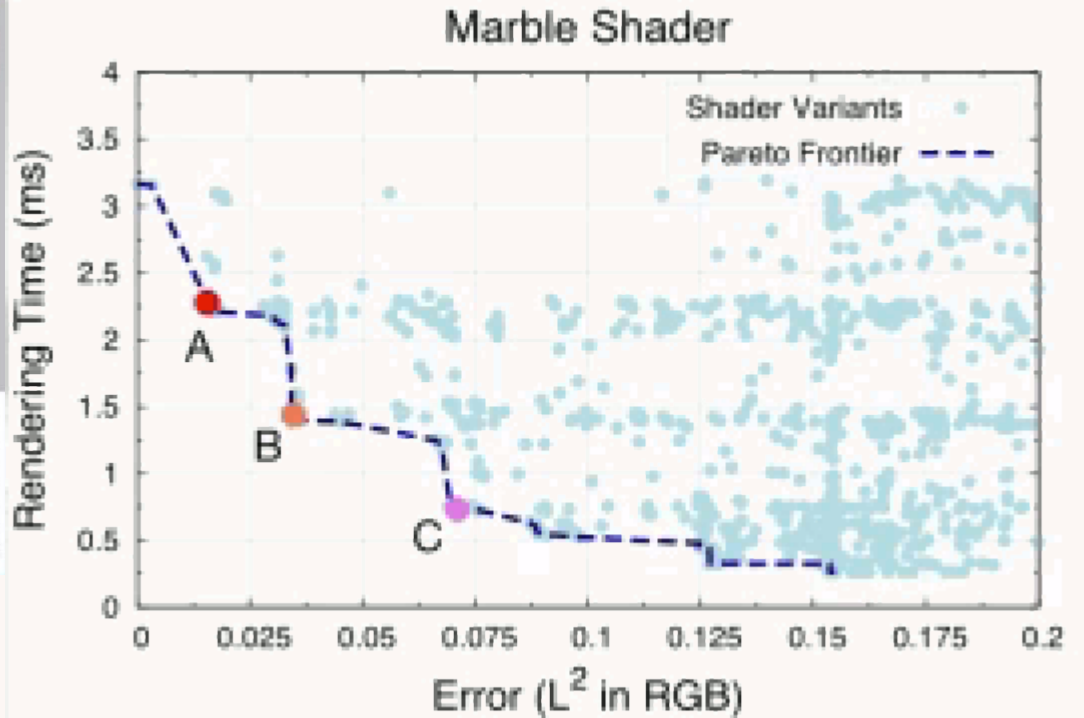
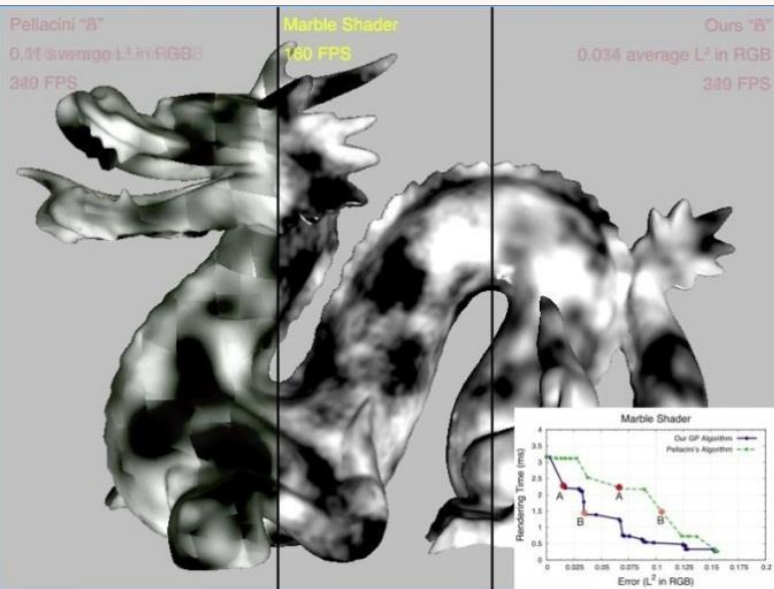
- Energy reduction [[GECCO 2015a](#),[SSBSE](#)] particularly for mobile computing [[GI-2017](#)]
- RAM memory reduction [[GECCO 2015b](#)]
- Reduce run time [[pknotsRG](#),[OpenCV](#), [RNAfold](#), [EuroGP2019](#)]
- Choose better library [[SSBSE-2017](#)]
- Improve library [[SSBSE 2014](#),[2016](#)]

# GI to Improve functionality

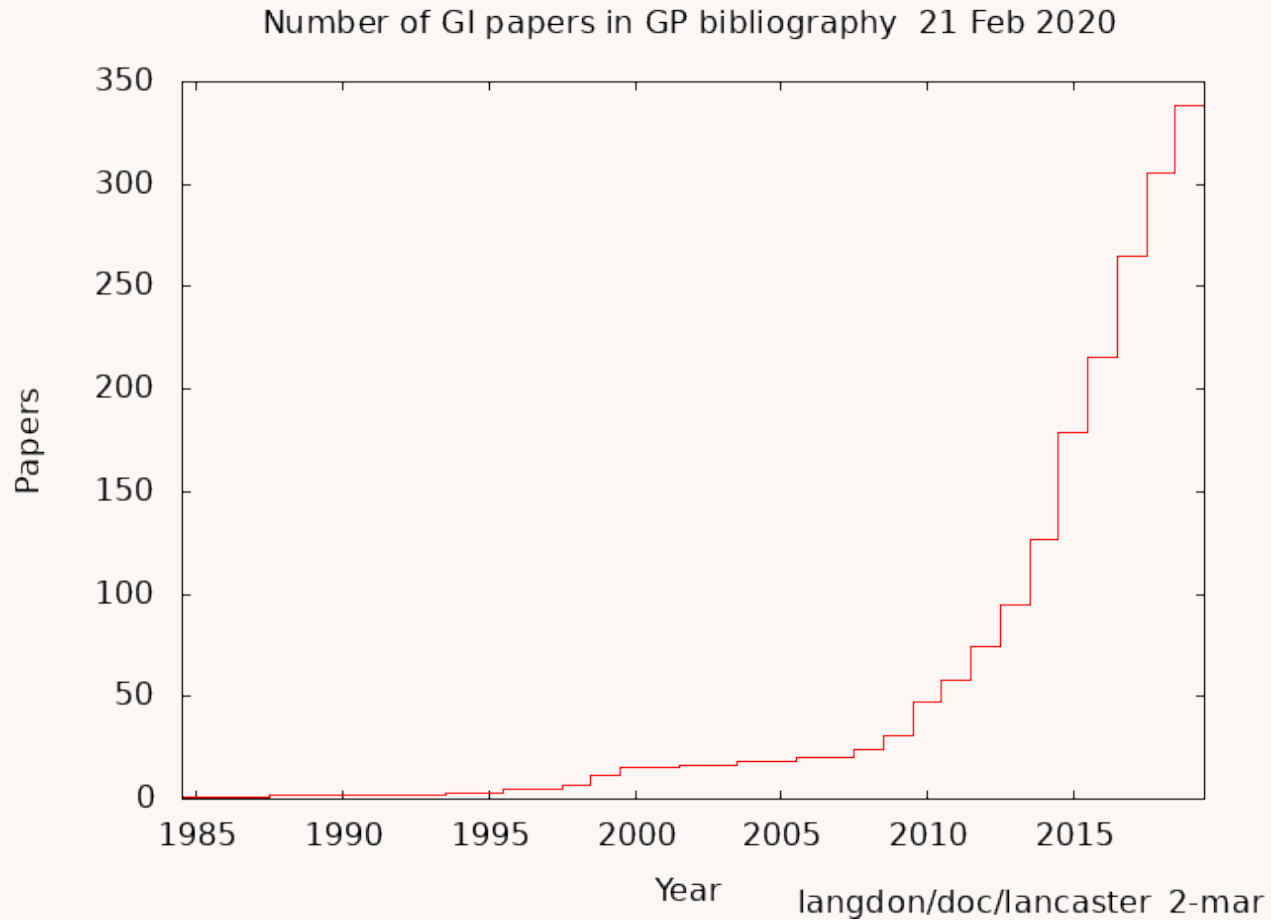
- Transplanting C++ [[Marginean SSBSE'15](#), [ISSTA'15](#)]  
E.g. graph layout into Kate, H.264 into VLC, awarded Gold [Humie](#), 26hours CPU v. 20days
- Autoporting
  - gzip to GPU [[CEC 2010](#)], RNAfold to SSE [[GI-2017](#)]
- Better RNA structure prediction
- Improving GPU shaders [[2011](#)]



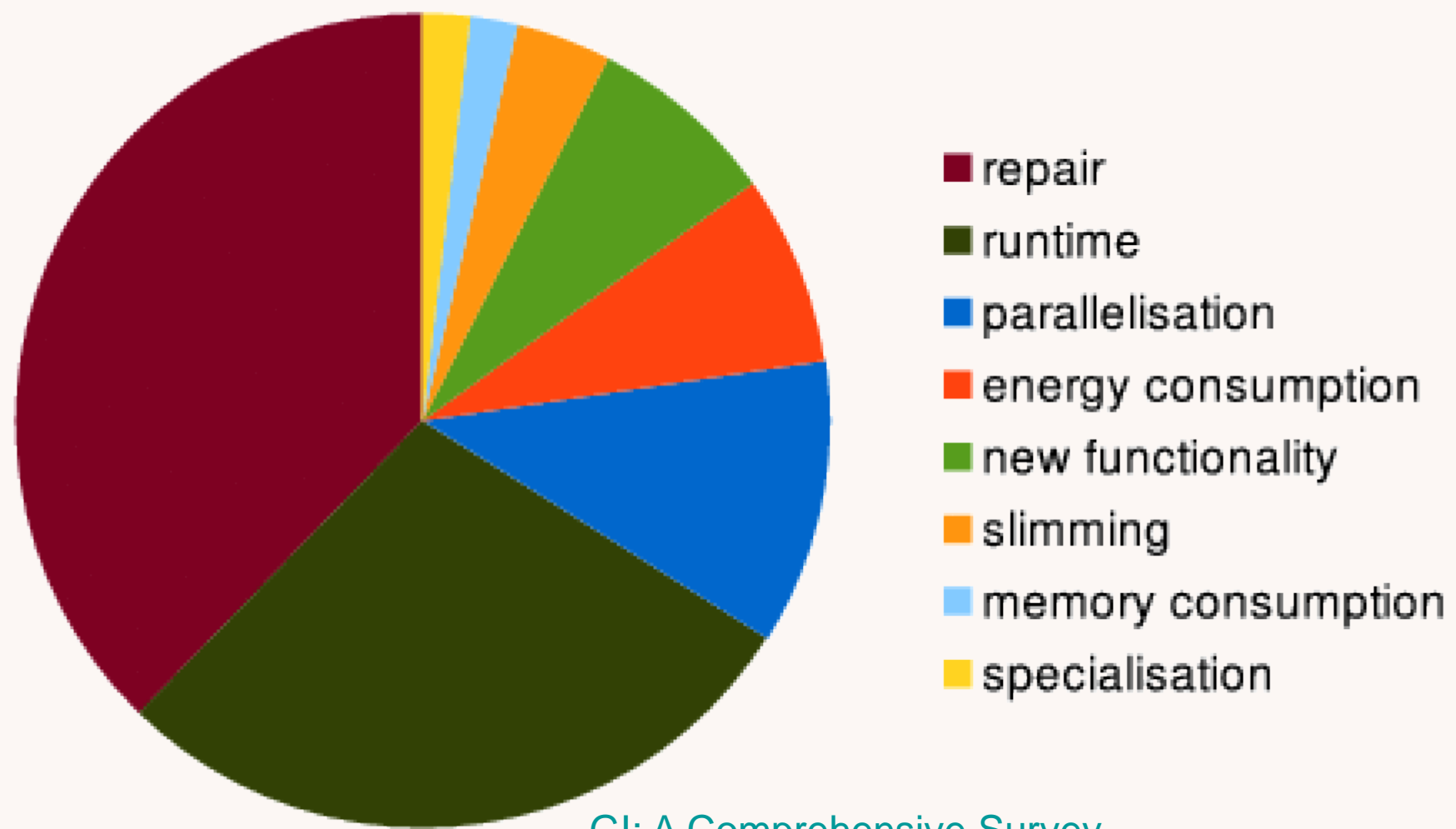
# GI Improving GPU shaders [2011]



# Genetic Improvement Papers



# Fig 3. software applications of empirical studies in core papers on genetic improvement



GI: A Comprehensive Survey





# Maintaining Embedded Constants

- Evolving better RNA predictions [EuroGP 2018](#)
  - RNAfold 7000 lines of code 50000 numbers
  - net 20% better prediction of RNA structures
  - In ViennaRNA package since 2.4.7
    - The GI code in ViennaRNA since 2.4.5 to 7 Feb 2020
      - downloaded 80,555 times
      - Online [server](#) used 1,096,867
- Evolving GNU sqrt into cube root [SSBSE-2018](#)  
[RN/18/05](#)  $\log_2$ ,  $x^{-1/2}$ , drcp

# What is RNAfold

- RNAfold is the state of the art prediction of how RNA molecule will fold up based on its sequence of bases.
- Open source program RNAfold 7100 lines of C source code.
- 51521 parameters (10 scalars+21 arrays)
- Training data  $\frac{1}{3}$  RNAstrand 4655 known structures  
(only use training sequences < 155 bases)

# Training Data RNAstrand

<http://www.rnasoft.ca/strand/>

RNAstrand contains known RNA secondary structures.  
 4666 secondary structures in total.  
 Example screen shot for PDB\_00865

www.rnasoft.ca/strand/show\_results.php?molecule\_ID=pdb\_00865&Submit=Search+RNA+STRAND+ID 150% Search

**RNA STRAND v2.0 - The RNA secondary STRucture and statistical ANALysis Database**


[ [Home](#) | [Search](#) | [Analyse](#) | [Submit structures](#) | [News](#) | [Help](#) ]

**General features for molecule pdb\_00865**  
 (click to expand/contract all tables)

Format: CT ▼ View the RNA sequence and secondary structure for molecule pdb\_00865

View the output of the RNA Secondary Structure Analyser for molecule pdb\_00865

Molecule ID [?]:	PDB_00865
Molecule name [?]:	SOLUTION STRUCTURE OF THE CENTRAL REGION OF THE HUMAN <span style="background-color: #e0e0e0;">GLUR- B R/G PRE-MRNA</span>
Source [?]:	<a href="#">RCSB Protein Data Bank</a>
Source ID [?]:	<a href="#">1YSV</a>
Reference [?]:	R.STEFL,F.H.ALLAIN. A NOVEL RNA PENTALOOP FOLD INVOLVED IN TARGETING ADAR2.. RNA V. 11 592 2005 ASTM RNARFU UK ISSN 1355-8382
Type [?]:	Synthetic RNA
Organism [?]:	SYNTHETIC
Validated by NMR or X-Ray [?]:	Yes
Method for secondary structure determination [?]:	NMR; ran through RNAview
Number of molecules [?]:	1
Length [?]:	27
Fragments used [?]:	No
Duplicated sequence [?]:	No other molecule in the database has the same sequence
Number of domains [?]:	1
Number of unpaired bases [?]:	3
Number of paired bases [?]:	24



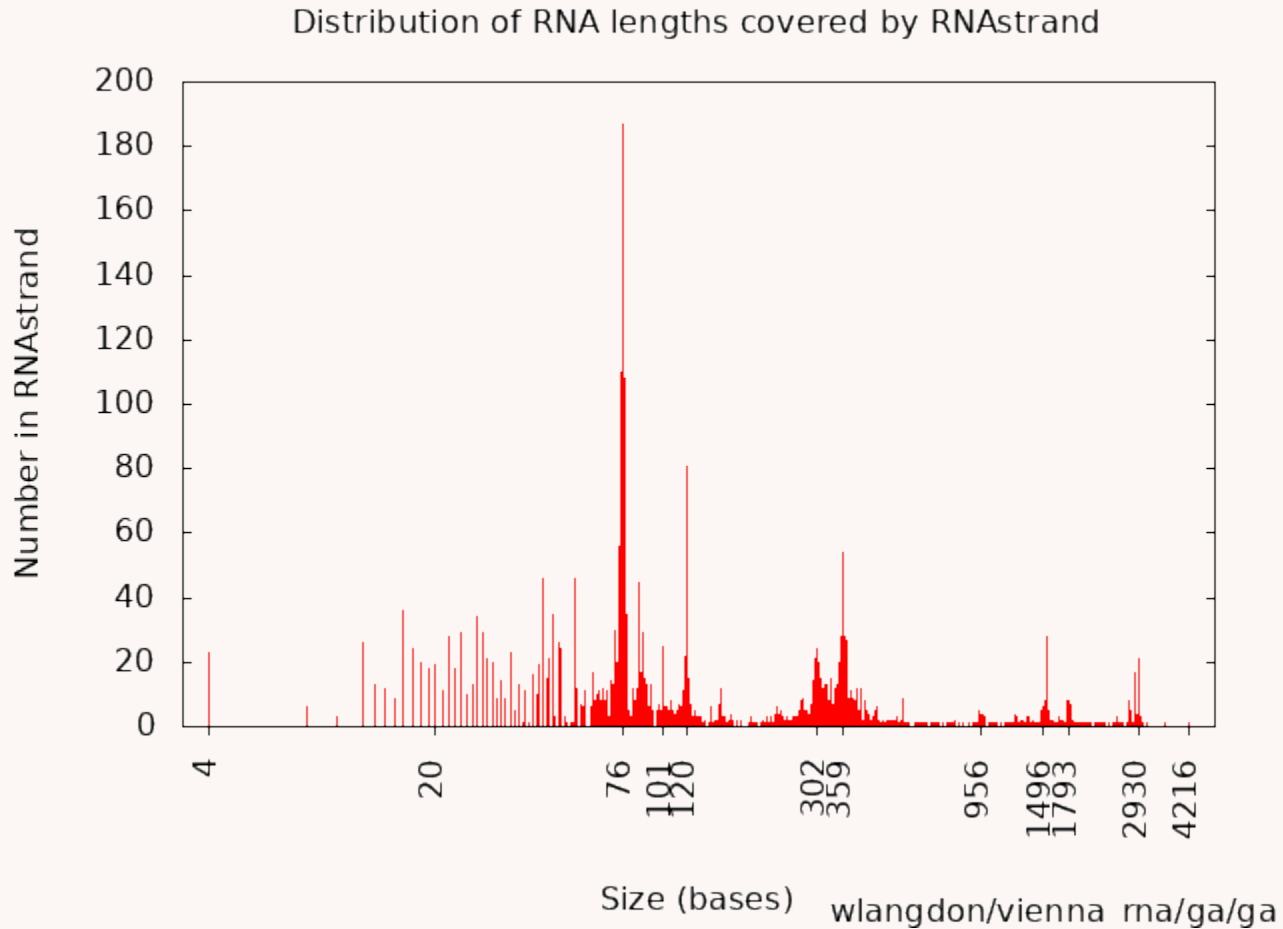
[?] [PS figure](#) [PDF figure](#)  
[Figure from original source](#)



Structure centre of human  
 GluR-B R/G pre-mRNA  
<https://www.rcsb.org/structure/1YSV>

Bit rot, broken images

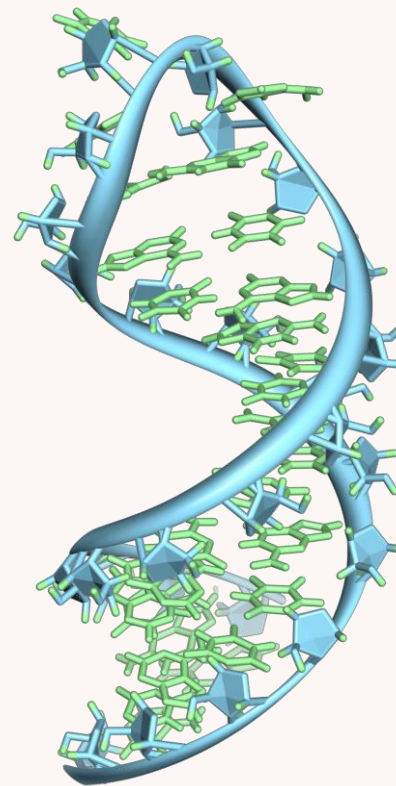
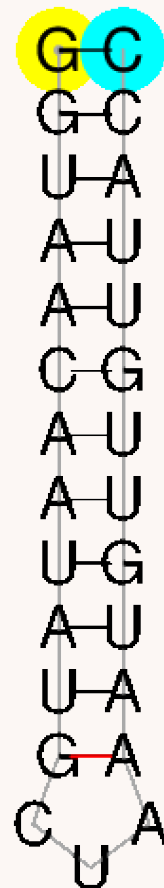
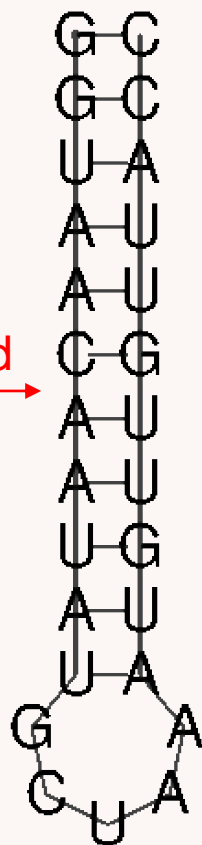
# RNAstrand



RNA sequence length, i.e. number of bases (log horizontal-scale)

# Compare RNAfold with RNAstrand

>PDB\_00865 → RNAfold  
 GGUAACAAUAUGCU  
 AAAUGUUGUUACC



Fasta text format  
 Input to RNAfold

Prediction  
 MCC 0.956018 RNAstrand

three D picture



# Nested brackets to connection matrix

PDB\_00865.ct\_ **rnafold**

```

GGUAACAAUAUGC UAAAUGUUGUUACC
G .....x
G .....x.
U .....x..
A .....x. .
A .....x. . .
C .....x. . . .
A .....x. . . . .
A .....x. . . . .
U .....x. . . . .
A .....x. . . . .
U .....x. . . . .
G .....x. . . . .
C .....x. . . . .
U .....x. . . . .
A .....x. . . . .
A .....x. . . . .
A .....x. . . . .
U .....x. . . . .
G .....x. . . . .
U .....x. . . . .
U .....x. . . . .
G .....x. . . . .
U .....x. . . . .
U .....x. . . . .
A .....x. . . . .
C .....x. . . . .
C .....x. . . . .

```

Prediction

PDB\_00865.ct\_ **rnastrand**

```

GGUAACAAUAUGC UAAAUGUUGUUACC
G .....x
G .....x.
U .....x..
A .....x. .
A .....x. . .
C .....x. . . .
A .....x. . . . .
A .....x. . . . .
U .....x. . . . .
A .....x. . . . .
U .....x. . . . .
G .....x. . . . .
C .....x. . . . .
U .....x. . . . .
A .....x. . . . .
A .....x. . . . .
A .....x. . . . .
U .....x. . . . .
G .....x. . . . .
U .....x. . . . .
U .....x. . . . .
G .....x. . . . .
U .....x. . . . .
U .....x. . . . .
A .....x. . . . .
C .....x. . . . .
C .....x. . . . .

```

Ground Truth

Non-standard  
G↔A pair

# Compare RNAfold & RNAstrand matrices

PDB\_00865.ct\_ **rnafold**  
GGUAACAAUAUGC UAAAUGUUGUUACC

.	TN	G	.....x
X	TP	G	.....x.
		U	.....x..
O	FN	A	.....x..
X	FP	A	.....x..
		C	.....x..
		A	.....x..
		A	.....x..
		U	.....x..
		A	.....x..
		U	.....x..
		G	.....O
		C	.....
		U	.....
		A	.....
		A	.....O
		A	.....x.
		U	.....x.
		G	.....x.
		U	.....x.
		U	.....x.
		G	.....x.
		U	.....x.
		U	.....x.
		A	.....x.
		C	.....x.
		C	x.....

PDB\_00865.ct\_ **rnastrand**  
GGUAACAAUAUGC UAAAUGUUGUUACC

G	.....x
G	.....x.
U	.....x..
A	.....x..
A	.....x..
C	.....x..
A	.....x..
A	.....x..
U	.....x..
A	.....x..
U	.....x..
G	.....x.
C	.....
U	.....
A	.....
A	.....x.
A	.....x.
U	.....x.
G	.....x.
U	.....x.
U	.....x.
G	.....x.
U	.....x.
U	.....x.
A	.....x.
C	.....x.
C	x.....

Prediction

Ground Truth

. TN = true negative



# Compare RNAfold with RNAstrand

PDB\_00865.ct\_rnafold

GGUAACAAUAUGC UAAAUGUUGUUACC

.	TN	G	.....x
X	TP	G	.....x.
		U	.....x..
O	FN	A	.....x...
X	FP	A	.....x....
		C	.....x.....
		A	.....x.....
		A	.....x.....
		U	.....x.....
		A	.....x.....
		U	.....x.....
		G	.....O.....
		C	.....
		U	.....
		A	.....
		A	.....O.....
		A	.....x.....
		U	.....x.....
		G	.....x.....
		U	.....x.....
		U	.....x.....
		G	.....x.....
		U	.....x.....
		U	.....x.....
		A	.....x.....
		C	.....x.....
		C	x.....

.	TN	705	true	negatives
X	TP	22	true	positives
O	FN	2	false	negatives
X	FP	0	false	positives
total 729 = 27 <sup>2</sup>				

## Matthews Correlation Coefficient

$$\frac{(TP \times TN - FP \times FN)}{\sqrt{((TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN))}}$$

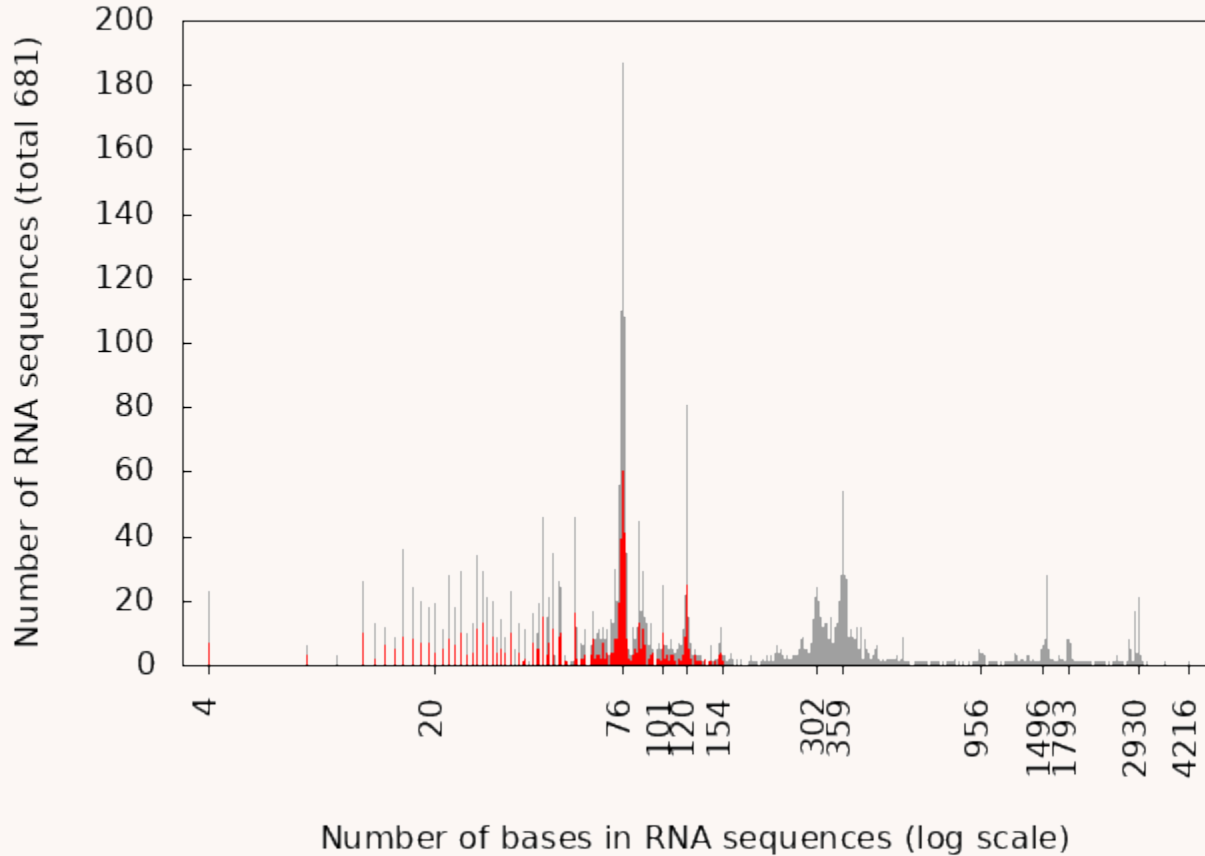
MCC = 0.9560

Product of true (TP×TN) minus product errors (FP×FN) normalised. MCC lies range -1 to +1

. TN = true negative Prediction

# Training RNA sequences

Distribution of RNA lengths used for training



681 =  $\frac{1}{3}$  data below 155 bases randomly selected for training

# Genetic Improvement Fitness Function

- Run RNAfold with modified internal data on 681 short training RNA sequences
- Calculate Matthews Correlation Coefficient for each prediction.
- Selection fitness is mean MCC over 681 predictions. (Select top 50% of population)
- Ignore data mutations which make no difference on training RNA examples

# 51521 RNAfold parameters

**Table 2.** 31 (10 scalars + 21 arrays) RNAfold parameters which can be optimised. Data structures marked<sup>E</sup> hold energy values which are always multiples of 10. (Mutation ensures they remain multiples of ten.) The original values of Tetraloop\_E<sup>E</sup> and Triloop\_E<sup>E</sup> are mostly zero <sup>a</sup> and so mutation of Tetraloop\_E<sup>E</sup> is limited to the first 15 elements and in Triloop\_E<sup>E</sup> to just the first element. NBPAIRS=7 and MAXLOOP=30.

noLP		mismatchM <sup>E</sup>	[NBPAIRS+1][5][5]
uniq_ML		mismatchExt <sup>E</sup>	[NBPAIRS+1][5][5]
dangles		dangle5 <sup>E</sup>	[NBPAIRS+1][5]
min_loop_size		dangle3 <sup>E</sup>	[NBPAIRS+1][5]
rtype	[8]	mismatchH <sup>E</sup>	[NBPAIRS+1][5][5]
gquad		stack <sup>E</sup>	[NBPAIRS+1][NBPAIRS+1]
special_hp		bulge <sup>E</sup>	[MAXLOOP+1]
pair	[21][21]	int11 <sup>E</sup>	[NBPAIRS+1][NBPAIRS+1][5][5]
noGclosure		int21 <sup>E</sup>	[NBPAIRS+1][NBPAIRS+1][5][5][5]
TerminalAU <sup>E</sup>		internal_loop <sup>E</sup>	[MAXLOOP+1]
MLintern <sup>E</sup>	[NBPAIRS+1]	ninio[2] <sup>E</sup>	
MLclosing <sup>E</sup>		mismatch1nI <sup>E</sup>	[NBPAIRS+1][5][5]
MLbase		int22 <sup>E</sup>	[NBPAIRS+1][NBPAIRS+1][5][5][5][5]
hairpin <sup>E</sup>	[31]	mismatch23I <sup>E</sup>	[NBPAIRS+1][5][5]
Tetraloop_E <sup>E</sup>	[200] (15)	mismatchI <sup>E</sup>	[NBPAIRS+1][5][5]
Triloop_E <sup>E</sup>	[40] (1)		

total 51521 int

<sup>a</sup> The energy contributions for Tetraloop and Triloop are only used under special circumstances. They represent tabulated exceptions of small hairpin loops that do not follow the values provided in hairpin. They are only used when the sequences in question match the corresponding patterns stored in the character arrays Tetraloop and Triloop.

# Genetic Improvement Representation

Variable length list of problem dependent mutations to data inside RNAfold.

**Replace mutation** > mismatchM -60>-40

Replace every element in array mismatchM whose value is currently -60 with -40

**Overwrite mutation** < mismatchH \*, 1, 2<-80

Overwrite eight elements in array mismatch (mismatchH[\*, 1, 2]) with -80

**Increment mutation** += mismatchH \*, \*, \*+=-90

Add -90 (ie subtract 90) from every element in array mismatch (ie mismatchH[\*, \*, \*])

**Creep mutation** Small change (<20%) to value of existing mutations

**Two point crossover**

# Append mutation

mismatchM -60>-40

## Child

mismatchM -60>-40 | *internal\_loop* 260>23

## Two point crossover

### 2 Parents

mismatchExt -50>-150 | *bulge* 280>480 | *int11* 220>190 | mismatchExt \*,1,1+=40

mismatch23I 0>90 | *dangle3* 2,0+=50 | *hairpin* \*<610

### Child

mismatchExt -50>-150 | *dangle3* 2,0+=50 | mismatchExt \*,1,1+=40

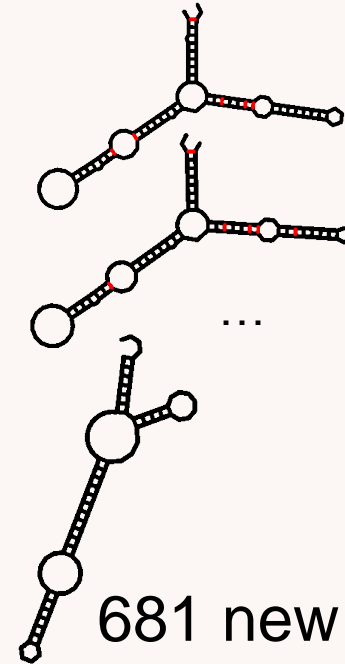
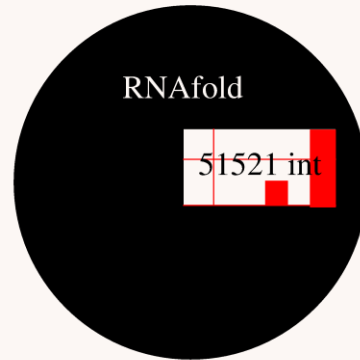
# Fitness of Mutated RNAfold

```
> CRW_00550  
NAUUUACGGCGGUC....  
GACAC
```

```
> CRW_00553  
NNNUUGGUGGCGGAG....  
CAAGC
```

...

```
> TMR_00272  
GGGGAUGAAUU....  
CACCA
```



681 short training sequences

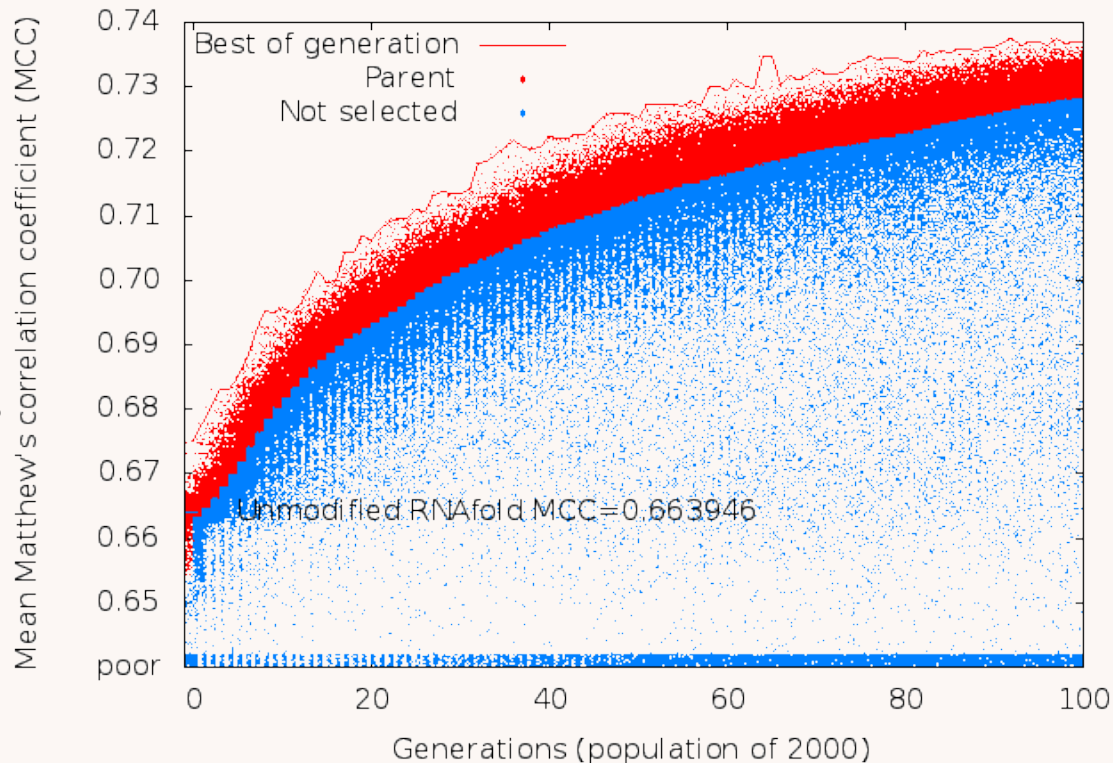
681 new predictions

- Mutate constants inside RNAfold and recompile
- Run mutated RNAfold on training RNA sequences
- Compare each new prediction with real structure
- Fitness mean Matthew's correlation coefficient on 681 training RNA molecules

# GI RNAfold

- Pop 2000, 50% mutation 50% crossover
- Bloat removed. Best individual at gen 100  
2849 mutations, hill climbing(2), 42 left

Training fitness, 681 length <155BP RNA\_STRAND v2.0



2000\*101  
fitness evals  
<5 days  
2.1sec each



# Impact of 42 GI Changes

**Table 3.** Impact of the 42 components of the cleaned up evolved patches to 51 521 int parameters of RNAfold's dynamic programming model of RNA secondary structure. First column: components grouped by data structure (order in group is still significant). 2<sup>nd</sup> number of int changed. 3<sup>rd</sup> responsibility for fitness change (mutations build on each other, so isolated changes only give an indication of their importance). 4<sup>th</sup> again impact, this time on number of bonds changes across the whole training set. Last column describes changes with impact >2%. See also Sect. 3.3.

internal_loop *+=-40	29	-6.91%	667	Add 40 to internal_loop[2..30] ([0] and [1] are INF and so cannot be incremented)
MLintern *+10	8	-3.25%	437	MLintern[0..7] were all -90, now -80 except [3] is -150
MLintern 3<-150				
ninio[2] 80		-2.50%	501	Was 60 now 80
mismatch23I 70>10000000	108	-1.40%	131	
dangle5 *,*+60	40	-1.27%	101	
int22 260>80 int22 180>280 int22 *,*,2,*,*+10 int22 280>200 int22 200>10000000	10454	0.05%	37	
mismatchI *,*,0<100 mismatchI *,*,1+=-10 mismatchI 2,3,1+=-100 *,*,*+=-40	96	0.05%	617	
int11 *,*,*,*<200	1600	1.22%	1306	
int11 6,*,*,2+=-70				
dangle3 5,*,*=-80	5	1.28%	13	
mismatch1nI 70>110	125	1.89%	173	
TerminalAU 80		3.04%	759	Was 50 now 80
rtype 6<6 rtype 2+1	2	3.05%	1257	[2] 1<-2 and [6] was 5 becomes 6, page 14
mismatchExt *,*,*+80	200	3.90%	320	+80 is added to all elements, except 1 in 5 is set to -40
mismatchExt *,*,1<-40				
stack -100>60 stack -140>0 stack 2,2+=-20 stack *,*,4<-50	14	6.08%	2135	[0,4] 10000000<-50 [1,4] -140<-50 [1,7] -140<-0 [2,2] -340<-360 [2,4] -150<-50 [3,5] -140<-0 [4,1] -140<-0 [4,4] 30<-50 [4,6] -100<-60 [5,3] -140<-0 [5,4] -60<-50 [6,4] -100<-50 [7,1] -140<-0 [7,4] 30<-50
int21 230>260	1669	6.51%	287	283 values that were 230 replaced by 260. 161 values of 220 replaced by INF. And 1225 cases (of a possible 1600) where int21[*,*,*,3] is reduced by 70
int21 *,*,*,*,3+=-70				
int21 220>10000000				
bulge *+40	30	7.53%	635	All bulge[1..30] increased by 40. ([0] is INF and so cannot be incremented)
mismatchM -70>-130	142	10.70%	1227	15 cases where -70 is replaced by -130. 2 cases where -110 is replaced by -130. 20 cases where -60 is replaced by -40. 40 cases where [*.,0,*] is reduced by -170, 35 [*.,1,*] by -40, and 30 [*.,3,*] by -40
mismatchM *,*,*+20				
mismatchM *,*,1,*,*+=-40				
mismatchM -110>-130				
mismatchM *,*,0,*,*+=-170				
mismatchM -60>-40				
hairpin *<560	30	14.75%	1217	All hairpin[*] are set to 560 (Fig. 5)
mismatchH *,*,*,*+=-90	180	16.30%	1610	39 cases where mismatchH [*.,3] is set to -130. 8 cases mismatchH [*.,1,2] becomes -80 and 133 where other values in mismatchH are reduced by -90
mismatchH *,*,*,3<-130				
mismatchH *,*,1,2<-80				
Total:	14732			

14732 of 51521 (29%)  
changed

# Improving RNAfold parameters

[EuroGP-2018](#)

- RNAfold 7100 lines of C source code, 51521 parameters.
- Fitness correlation between prediction and true structure (Matthews Correlation, MCC).
- Post evolution tidy
- 14732 (29%) parameters changed
- Holdout set significant ( $p 10^{-16}$ ) increase MCC
- Also better constrained optimisation ( $p 10^{-15}$ )
- GI parameters [rna\\_langdon2018.par](#) shipped with ViennaRNA since 13 Jun 2018

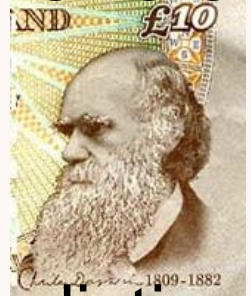
# Six impossible things before breakfast



- To have impact do something considered impossible.
- If you believe software is fragile you will not only be wrong but shut out the possibility of mutating it into something better.
- Genetic Improvement has repeatedly shown mutation need not be disastrous and can lead to great things.

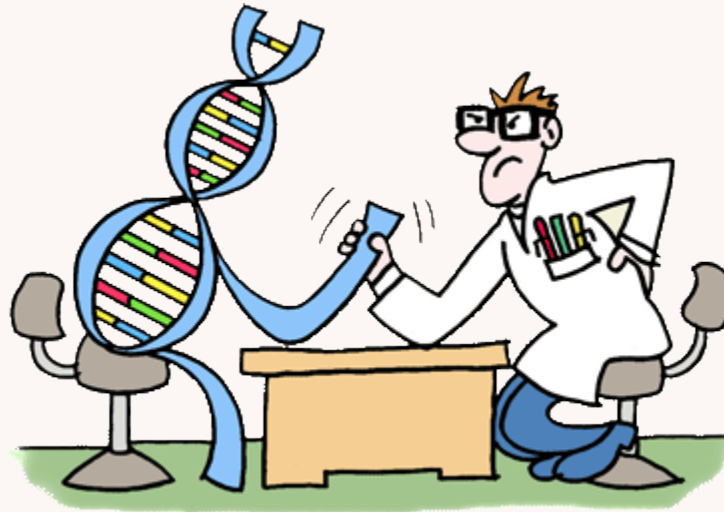
# Conclusions

- In a world addicted to software, maintenance is the dominant cost of computing.
- Genetic Improvement (GI) applies Darwinian survival of the fitness to software
  - bugfixing, software transplanting, performance improvement, faster answers or better answers.
- RNAfold on real data
  - No code changes
  - 50000 parameters 20% overall better prediction
- **Software is not fragile**  
**break it, bend it, Evolve it**





**WIKIPEDIA**  
Genetic Improvement



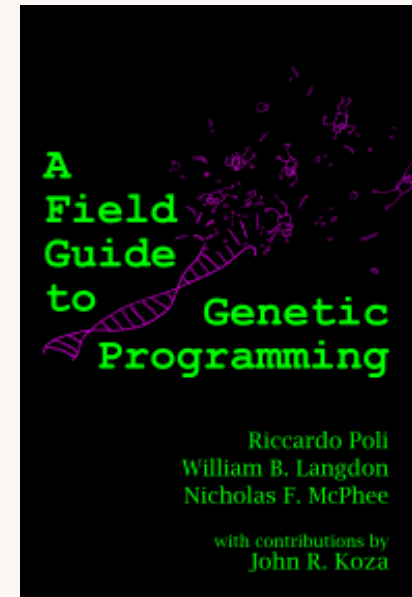
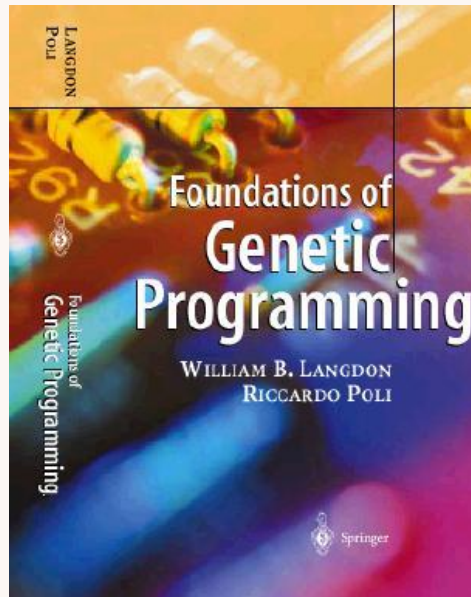
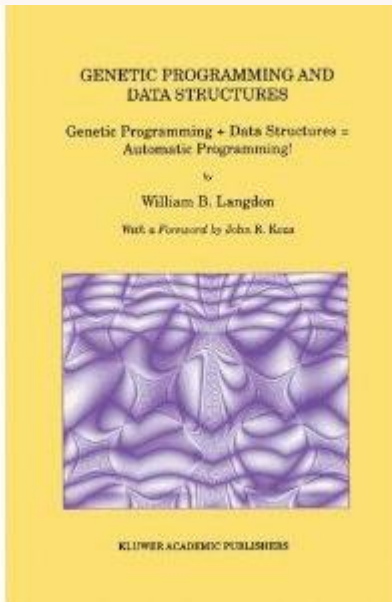
[Humies](#): Human-Competitive  
\$10,000 prizes. Finals in Cancun



W. B. Langdon

CREST

Department of Computer Science



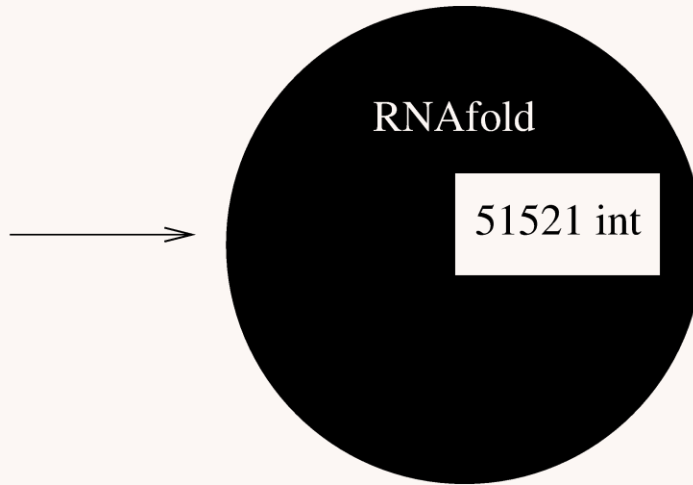
# Genetic Improvement of RNAfold

- Speed up via Intel SSE parallel instructions [GI 2017](#). Shipped since V2.3.5 2017-04-17
- GPU ViennaRNA Package [v2.3.0cuda](#)
- **Better predictions by evolving parameters**
  - On average better predictions of RNA folding.
  - Shipped since 2.4.7 2018-06-13
- AVX speedup in release 2.4.11 2018-12-17  
[EuroGP 2019](#)

# RNAfold

```
> CRW_01446
UUCAAACGAGGAAA.....
.....
.....
UGAAC
```

RNA sequence



RNA structure

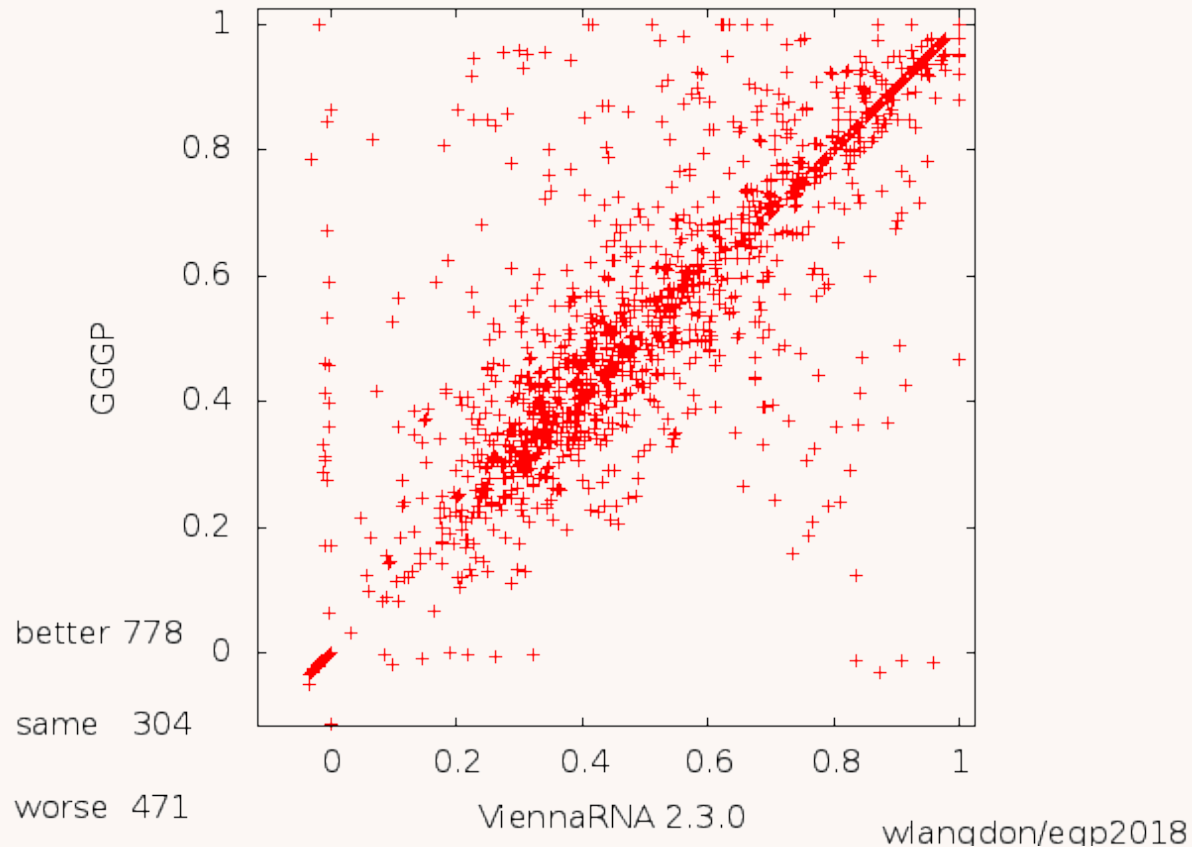


RNAfold reads RNA molecules base sequence.  
 Outputs prediction of how molecule will fold up.  
 Internally RNAfold uses 51521 parameters.



# Results $p < 10^{-17}$ on holdout

Matthews Correlation Coefficient of Prediction, holdout RNA\_STRAND



# Automatic Testing

- Hardware Improvement: Tetsuya Higuchi Analogue EHW chip for mobile phones
- Software quality continues to be dominated by the cost of manual effort
- Existing test suites are often run automatically
- Evolution can automatically create test cases (goal: code coverage) but still lacks knowledge of the correct answer (known as the test oracle problem).

# Demo count blue pixels

- Assumes Unix, tsch gcc
- [http://www.cs.ucl.ac.uk/staff/W.Langdon/ftp/gp-code/opencv\\_gp.tar.gz](http://www.cs.ucl.ac.uk/staff/W.Langdon/ftp/gp-code/opencv_gp.tar.gz)
- `gunzip -c ftp/gp-code/opencv_gp.tar.gz | tar xvf -`
- **README.txt**

Simple blue example of  
Genetic Improvement

[opencv\\_gp.tar.gz](http://www.cs.ucl.ac.uk/staff/W.Langdon/ftp/gp-code/opencv_gp.tar.gz)

[RN/18/06](#)



# GP and Software

- Genetic programming can automatically re-engineer source code. E.g.
    - hash algorithm
    - Random numbers which take less power, etc.
    - mini-SAT
  - fix bugs (5  $10^6$  lines of code, 16 programs) [EuroGP 2014](#)
  - create new code in a new environment (GPU) for existing program, gzip [WCCI 2010](#)
  - 70 speed up 50000 lines of code [IEEE TEC](#)
  - 7 times speed up for stereoKernel GPU [EuroGP 2014](#)
- 3D NMR Brain scans [GECCO 2014](#)

# The Genetic Programming Bibliography

<http://gpbib.cs.ucl.ac.uk/>

**13491** references, [12000 authors](#)

**Make sure it has all of your papers!**

E.g. email [W.Langdon@cs.ucl.ac.uk](mailto:W.Langdon@cs.ucl.ac.uk) or use | [Add to It](#) | web link



Part of gp-bibliography 04-40 Revision: 1.1794-29 May 2011

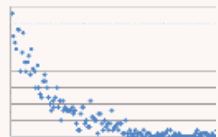
A web form for adding your entries.  
Co-authorship community. Downloads

Downloads by day



A personalised list of every author's  
GP publications.

[blog](#)



Your papers