

COMP 5660 Fall 2023 Assignment 2b

Matthew Freestone
maf0083@auburn.edu

18 November 2023

1 Green Deliverables

1.1 GP Parameters

See Table 1 for Green GP Parameters.

Parameter Name	Value
μ	200
λ	50
Mutation Rate	0.20
Depth Limit	7
Constant Terminal Range	[-5,20]
Bloat Control	Parsimony Pressure with Coefficient 0.5
Parent Selection Method	K-Tournament With Replacement ($k = 9$)
Survival Selection	K-Tournament Without Replacement ($k = 13$)
Recombination Method	Subtree Crossover
Mutation Method	Subtree Mutation
Tree Size Metric	Number of Nodes

Table 1: Green GP Parameter Values

1.2 Results

See Figure 1 for an Evals vs Population Mean and Population Max fitness. The plot also shows non-penalized fitness (without parsimony pressure).

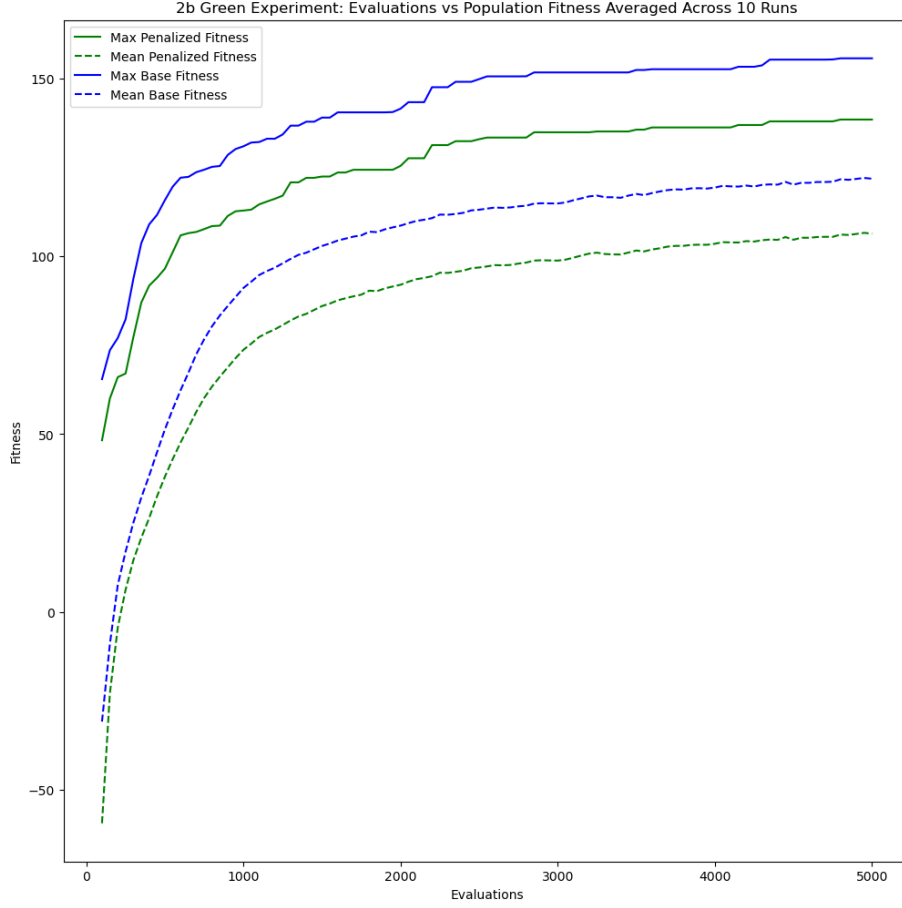


Figure 1: Green Evals vs Fitness and Penalized Fitness

1.3 Solutions

The best run from Green GP yielded a best score of 164.63 during run 1. The parse tree (Genotype) for the PacMan solution is shown below in the form of an in-order traversal that shows the underlying formula found.

$$(((G/((P \text{ RAND } (P * P)) \text{ RAND } ((P/P) \text{ RAND } (P * F)))) * G)/F)$$

The log of this game is saved in the 'data/2b/green' directory. The visualizer shows that this agent is much more decisive than the 2a agent, as it tended to balance both running from ghosts and collecting pills. This agent also did a better job prioritizing fruits over pills. The underlying formula is much more complicated than the best found from random search, which only had depth 2. This shows that, while the search space for larger trees scales exponentially,

there might be better solutions in larger trees.

The lack of constants is also interesting – the agent instead used terms like $P \times P$ to increase values rather than a constant.

1.4 Statistical Analysis

In this section, the data from the random initialization is compared to the base GP implemented in this report.

The statistics for each dataset are shown in Table 2.

2b data mean:	155.6980287929125
2b data stdv:	7.95118144540523
2a data mean:	111.56079734219269
2a data stdv:	17.073364697674794

Table 2: Green Statistical Data

After performing a Welch’s T-test, a p value was found of $5.788584 \cdot 10^{-6}$. With $\alpha = 0.05$, the null hypothesis of the methods yielding the same results can be firmly rejected, as $\alpha > p$.

Based on the means of the scores from each dataset, the results from the GP are clearly better.

2 Yellow Deliverables

2.1 GP Parameters

See Table 3 for Yellow GP Parameters.

Parameter Name	Value
μ	200
λ	50
Mutation Rate	0.20
Depth Limit	7
Constant Terminal Range	[-5,20]
Bloat Control	Parsimony Pressure with Coefficient 1.5
Parent Selection Method	K-Tournament With Replacement ($k = 9$)
Survival Selection	K-Tournament Without Replacement ($k = 13$)
Recombination Method	Subtree Crossover
Mutation Method	Subtree Mutation
Tree Size Metric	Tree Depth (longest path)

Table 3: Yellow GP Parameter Values

2.2 Results

See Figure 2 for an Evals vs Population Mean and Population Max fitness. The plot also shows non-penalized fitness (without parsimony pressure), as well as min, average, and max tree size in terms of depth.

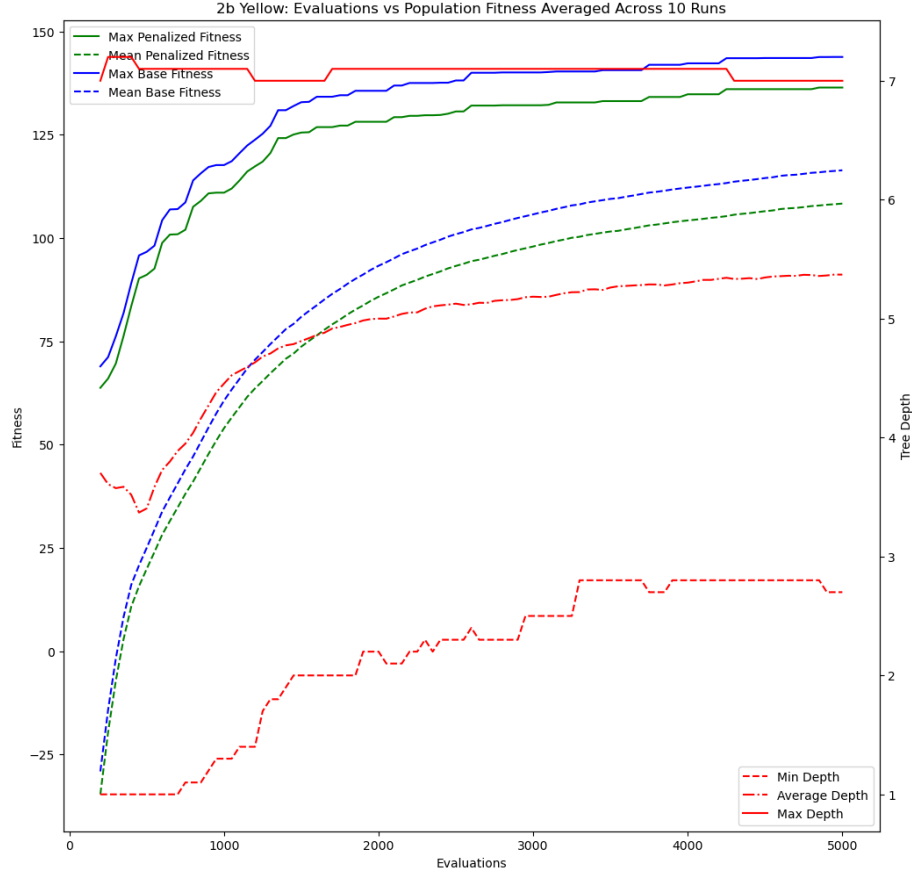


Figure 2: Yellow Evals vs Base Fitness, Penalized Fitness and Tree Size

2.3 Solutions

The best run from Yellow GP yielded a best score of 156.90 during run 6. The parse tree (Genotype) for the PacMan solution is shown below in the form of an in-order traversal that shows the underlying formula found.

$$((((P/F)/((W-15.378)/P)) \text{ RAND } G)*G)/((((P \text{ RAND } G) \text{ RAND } (G-(P/19.312))))/G)*(((F*P)*((13.989+G)+(P/F)))*F) \text{ RAND } F)))$$

The log of this game is saved in the 'data/2b/yellow' directory. Much like the Green agent, this agent was more decisive than random search's result,

but it seemed to have more random behavior than the Green model. This is likely a result of the size of the formula for Yellow and the amount of RAND operators present. This tree was also significantly larger than the Green tree, yet it performed worse, leading to a conclusion that some balance exists between tree size and performance, not a strictly positive correlation.

This agent made good use of constants, very much unlike the Green agent. With a deeper tree, perhaps it is easier to make use of constant values.

2.4 Statistical Analysis

In this section, the data from the Green GP is compared to the Yellow GP. The statistics for each dataset are shown in Table 4.

2b Yellow data mean:	143.81297895902546
2b Yellow data stdv:	6.624755712951387
2b Green data mean:	155.6980287929125
2b Green data stdv:	7.95118144540523

Table 4: Yellow Statistical Data

After performing a Welch’s T-test, a p value was found of 0.001993141849962456. With $\alpha = 0.05$, the null hypothesis of the methods yielding the same results can be rejected, as $p < \alpha$. However, the test indicates that the Green GP is actually significantly better, showing that, for this problem, using node count as a tree size metric might be more impactful than using tree depth.

3 Red2 Deliverables

In Red2, the performance of 3 PacMeni on the same board is measured, and all 3 agents use the same individual’s parse tree.

3.1 GP Parameters

See Table 5 for Red2 GP Parameters.

Parameter Name	Value
μ	200
λ	50
Mutation Rate	0.40
Depth Limit	7
Constant Terminal Range	[-5,20]
Bloat Control	Parsimony Pressure with Coefficient 0.5
Parent Selection Method	K-Tournament With Replacement ($k = 9$)
Survival Selection	K-Tournament Without Replacement ($k = 13$)
Recombination Method	Subtree Crossover
Mutation Method	Subtree Mutation
Tree Size Metric	Node Count

Table 5: Red2 GP Parameter Values

3.2 Results

See Figure 3 for an Evals vs Population Mean and Population Max fitness. The plot also shows non-penalized fitness (without parsimony pressure). Based on the shape of this, it would be reasonable to conclude that allowing more evaluations would likely lead to a continued improvement in average and best score.

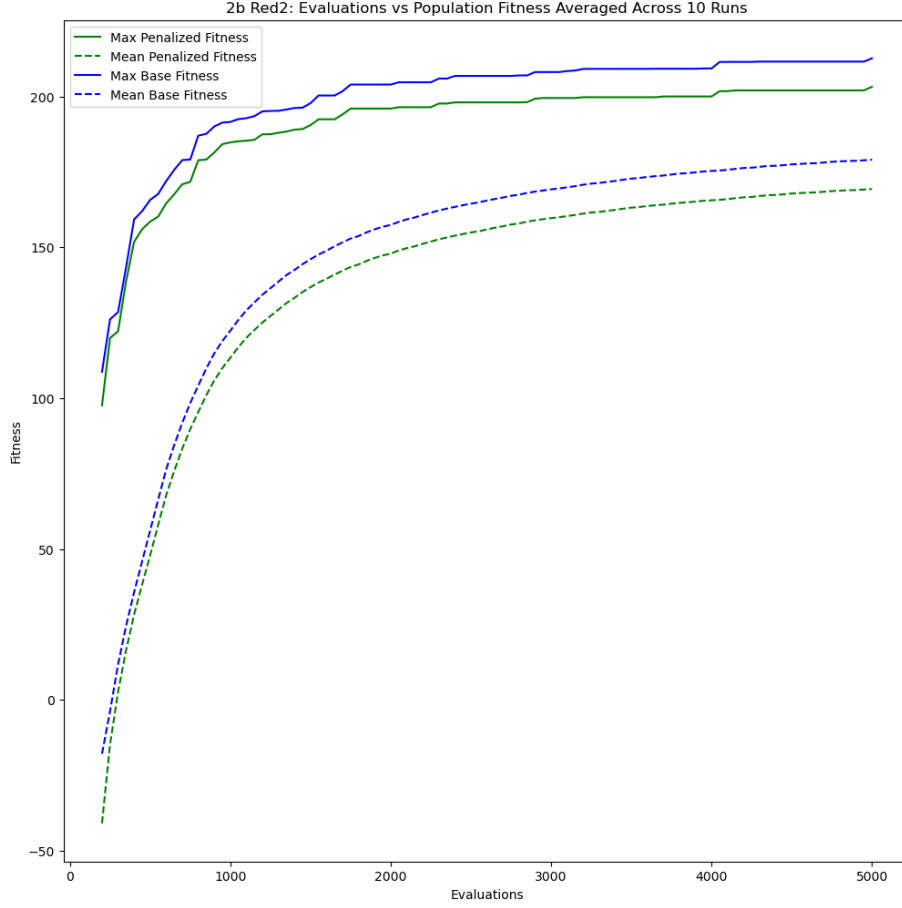


Figure 3: Red2 Evals vs Base Fitness and Penalized Fitness

3.3 Solutions

The best run from Red2 GP yielded a best score of 231.6 during run 3. The parse tree (Genotype) for the PacMani solution is shown below in the form of an in-order traversal that shows the underlying formula found.

$$(F \text{ RAND } (((P/F) - F) * (F - F)) \text{ RAND } P)) * (P * (F * F)))$$

The log of this game is saved in the ‘data/2b/red2’ directory. The log clearly shows the importance of having RAND operators in an environment where all agents use the same parse tree. When the PacMeni landed on top of each other, they did not copy each other’s moves exactly. There did not appear to be any particularly intelligent method to the movement of agents, but they did a good job of avoiding ghosts.

4 Red3 Deliverables

In Red3, the performance of 3 PacMeni on the same board is measured, and all 3 agents different parse trees. Individuals are randomly paired for evaluations, and an individual's fitness is determined by the average of the scores of the evaluations it participated in, after which it is adjusted for parsimony pressure.

4.1 GP Parameters

See Table 6 for Red3 GP Parameters.

Parameter Name	Value
μ	200
λ	50
Mutation Rate	0.40
Depth Limit	7
Constant Terminal Range	[-5,20]
Bloat Control	Parsimony Pressure with Coefficient 0.5
Parent Selection Method	K-Tournament With Replacement ($k = 9$)
Survival Selection	K-Tournament Without Replacement ($k = 13$)
Recombination Method	Subtree Crossover
Mutation Method	Subtree Mutation
Tree Size Metric	Node Count

Table 6: Red3 GP Parameter Values

4.2 Results

See Figure 4 for an Evals vs Population Mean and Population Max fitness. The plot also shows non-penalized fitness (without parsimony pressure). Based on the shape of this, it would be reasonable to conclude that allowing more evaluations would likely lead to a continued improvement in average and best score.

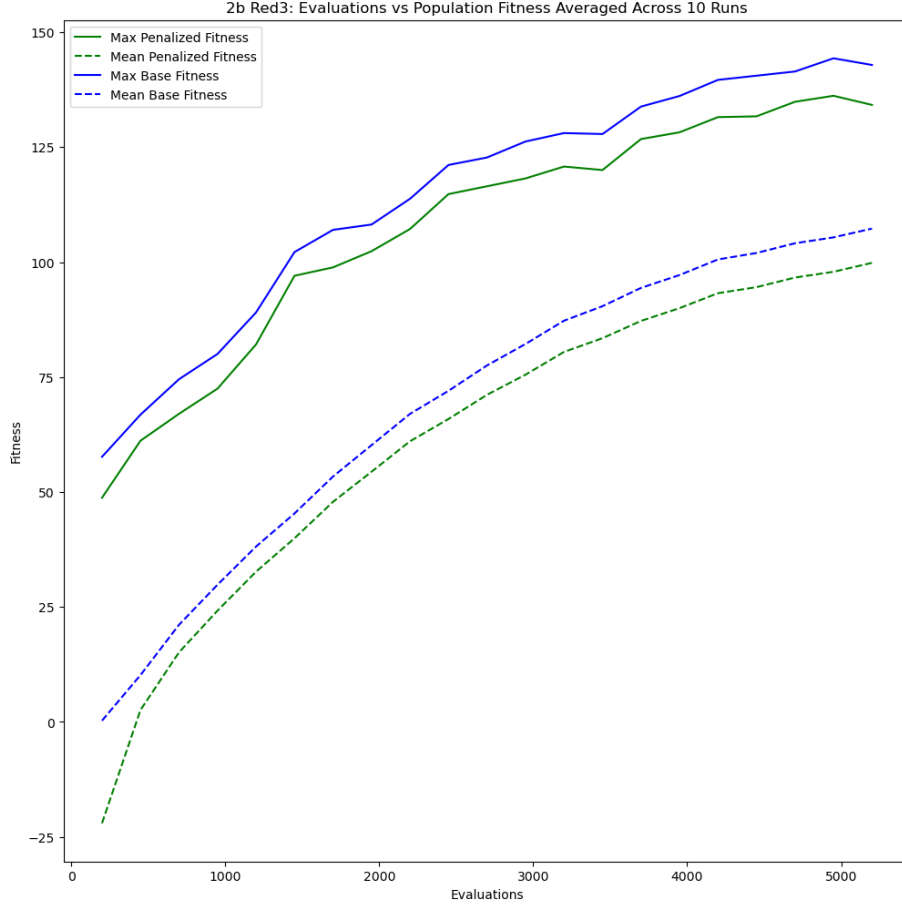


Figure 4: Red3 Evals vs Base Fitness and Penalized Fitness

4.3 Solutions

The best run from Red3 GP yielded a best single-game score of 214.4 during run 4. The parse trees (Genotype) for the PacMani solutions are shown below in the form of an in-order traversals that show the underlying formulas found.

1. $((P/G) \text{ RAND } G)$
2. $((G - (P/G))/((F * P) + ((F * P) + (((F * P) + (F \text{ RAND } -1.835)) * 11.610) \text{ RAND } G))))$
3. $((G - (P/G))/((F * P) + ((F * P) + (((P * W) \text{ RAND } F) * 11.610) \text{ RAND } G))))$

The log of this game is saved in the ‘data/2b/red3’ directory. It is almost immediately obvious that each PacMan is controlled by a different controller.

The simplest one more than likely controls the Red PacMan, which got firmly stuck in a corner, and was later killed by a ghost. As discussed previously, the problem more or less requires use of RAND.

The other agents were more successful; they are also likely related, with one probably being a crossover or mutation of the other, given that they share so many terminals and operators. These two performed better, which allowed the evaluation to have a high score. This success allowed the first solution to receive a higher fitness than deserved. Just as EC imitates evolution, it has now imitated group work.

5 Red4 Deliverables

In Red4, the performance of 3 Ghosts against one PacMan on the same board is measured, and all 3 agents use the same individual's parse tree.

5.1 GP Parameters

See Table 7 for Red4 GP Parameters.

Parameter Name	Value
μ	200
λ	50
Mutation Rate	0.40
Depth Limit	7
Constant Terminal Range	[-5,20]
Bloat Control	Parsimony Pressure with Coefficient 1
Parent Selection Method	K-Tournament With Replacement ($k = 9$)
Survival Selection	K-Tournament Without Replacement ($k = 13$)
Recombination Method	Subtree Crossover
Mutation Method	Subtree Mutation
Tree Size Metric	Node Count

Table 7: Red4 GP Parameter Values

5.2 Results

See Figure 5 for an Evals vs Population Mean and Population Max fitness. The plot also shows non-penalized fitness (without parsimony pressure). Based on the shape of this plot, we can see that the Ghosts quickly found a near perfect strategy against a randomly moving PacMan, and regularly achieve fitness near 100, which corresponds to a -100 score in the game.

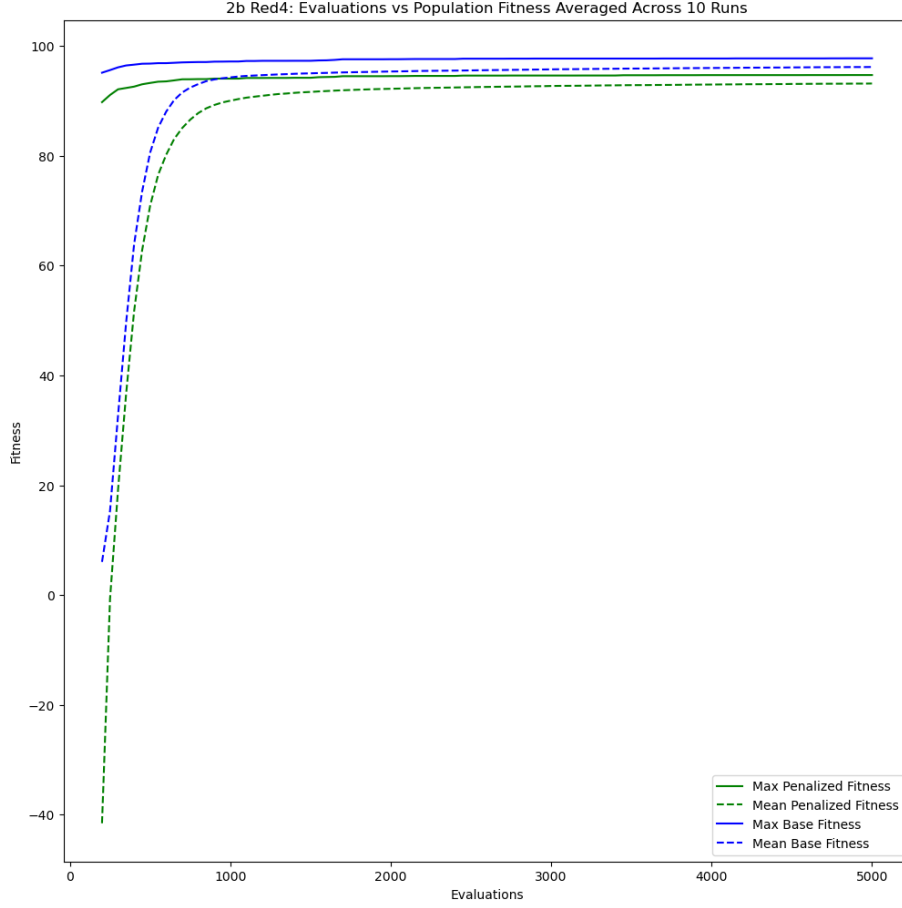


Figure 5: Red4 Evals vs Base Fitness and Penalized Fitness

5.3 Solutions

The best run from Red2 GP yielded a best score of -97.79 during run 1. The parse tree (Genotype) for the Ghosts is shown below in the form of an in-order traversal that shows the underlying formula found.

$$((G - M) - M)$$

The log of this game is saved in the ‘data/2b/red4’ directory. The game replay shows all 3 ghosts head directly toward PacMan. The bottom-left ghost reaches the player quickly, ending the game. The other two got stuck in corners, which is a behavior that more intelligent PacMani could possibly exploit. The formula found boils down to a simple policy: prioritize going for PacMan first, and staying away from other ghosts second.

6 Red5 Deliverables

In Red5, the performance of 3 Ghosts against one PacMan on the same board is measured, and all 3 agents different parse trees. Individuals are randomly paired for evaluations, and an individual's fitness is determined by the average of the (negated) scores of the evaluations it participated in, after which it is adjusted for parsimony pressure.

6.1 GP Parameters

See Table 8 for Red5 GP Parameters.

Parameter Name	Value
μ	200
λ	50
Mutation Rate	0.40
Depth Limit	7
Constant Terminal Range	[-5,20]
Bloat Control	Parsimony Pressure with Coefficient 1
Parent Selection Method	K-Tournament With Replacement ($k = 9$)
Survival Selection	K-Tournament Without Replacement ($k = 13$)
Recombination Method	Subtree Crossover
Mutation Method	Subtree Mutation
Tree Size Metric	Node Count

Table 8: Red5 GP Parameter Values

6.2 Results

See Figure 6 for an Evals vs Population Mean and Population Max fitness. The plot also shows non-penalized fitness (without parsimony pressure). Based on the shape of this plot, we can see that the at least some of the Ghosts quickly found a near perfect strategy against a randomly moving PacMan, and regularly achieve fitness near 100, which corresponds to a -100 score in the game.

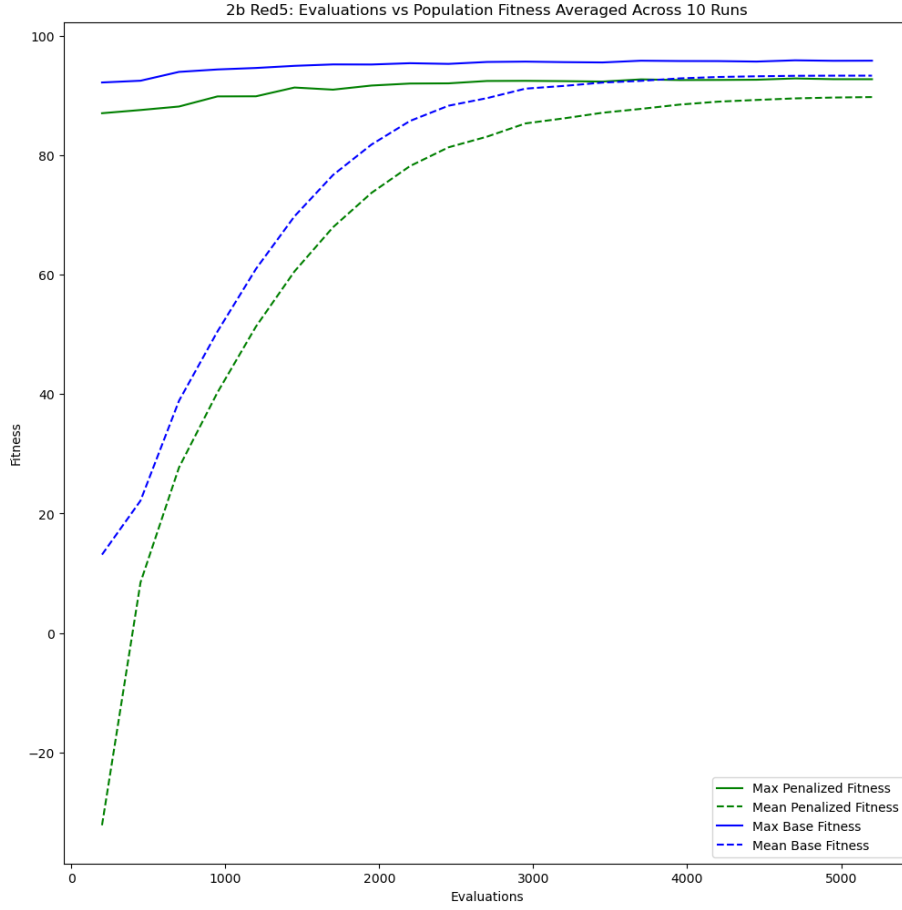


Figure 6: Red5 Evals vs Base Fitness and Penalized Fitness

6.3 Solutions

The best run from Red5 GP yielded a best single-game score of -97.79 during run 0. The parse trees (Genotype) for the 3 Ghosts are shown below in the form of an in-order traversals that show the underlying formulas found.

1. (P **RAND** 5.231)
2. ($G - M$)
3. ($M - M$)

The log of this game is saved in the ‘data/2b/red5’ directory. It is immediately obvious that each Ghost is controlled by a different controller. The top-right and bottom-left ghosts have terrible performance, and they simply alternate between going up or down. These are probably the (P **RAND** 5.231)

and $(M - M)$ controllers, as those two formulas would rarely, if ever, be motivated to move toward a PacMan. The remaining controller, and the one that won the game for the other two useless ghosts, strongly resembles the best solution from Red 4, which was $G - 2M$. This ghost moves directly toward the PacMan and ended the game at a nearly identical pace to the similiarly performing one from Red 4.