こんにちは。私たちは早稲田チームです。

Hello. We are Waseda Team.

初めに、本日はご招待くださりありがとうございます。私たちは、このプロジェクトに関係しているすべての人に感謝の意を表します。

First of all, thank you for inviting us this contest. We thank the all who are concerned with this contest. Thank you.

続けて、私たちの簡単な自己紹介をします。

私たちは、日本の東京にある早稲田大学の学生です。

iGEMの監督をしている木賀大介教授の研究室で合成生物学の研究をしています。

これが私たちの研究室の写真です。

以前木賀教授は深圳にきて、Xiao yi教授らと交流もしています。

私たちは主に、遺伝子回路とタンパク質工学の研究をしています。

そして、数年前、タンパク質工学と機械学習を組み合わせた研究を始めようと考えました。

Then, I will introduce ourselves shortly.

We are the student at Waseda University in Tokyo Japan.

Our professor is Daisuke Kiga, who is one of the directors of Igem.

This is our laboratory picture.

A few months ago, he came to Shenzhen, and interacting with Prof. Xiao.

We are studying gene circuits and protein engineering, and we are trying to study protein engineering with machine learning for some years.

ここから先が、私たちの今回の発表の本題です。

From now, we will introduce our project.

まず、私たちは2つの方法を使って変異体を作成しました。

1つ目は、変異は足し算で考えられると仮定しました。その場合の変異のスコアによるものを用いた方法です。この方法によって、既存の変異を組み合わせることで高い活性を持つＧＦＰを目指しました。

2つ目は、配列の相同性を用いた活性推定方法です。この手法によって、1つ目の方法で作成した変異体を違う側面から評価することができます。

この2つを用いて変異体を生成、評価しました。

First, we use 2 ways to generate mutants.

One is the mutation score, which is assumed additively.

In this approach, we combinate with known mutations and generate assumed high activity GFP mutants which is not in dataset.

The other way is the homology of the mutants. In this method, it can evaluate the mutants that was generated by 1st way by other way.

We use these 2 ways, then generate and evaluate the mutants.

変異体の生成とその活性の評価は次のように行いました。

まず、avGFPの変異体の中からアミノ酸1置換及び2置換のものをピックアップしました。

変異ごとに野生型からどれくらい活性が上昇したかを計算しました。この値をAdditive Scoreと名付けました。

その中から、3種類の変異体を組み合わせて新たな変異体を作りました。

私たちは、変異を組み合わせたとき、その活性の変化は加算であると考えました。

例えば、右下の図のように、Aに変異が入ったもののScoreが0.4, BとCに変異が入ったもののScoreが0.2だったとき、A～Cすべてに変異が入っていると、Scoreは0.6になると考えました。

We generate, and evaluate the mutants as following.

First, We picked up the mutants that has 1 or 2 amino acid replacement.

We calculated the activity for each mutant that are combined three mutants that are picked up the preliminary step. And The difference in activity between the mutant and the wild type is named Additive Score.

For example, if there are 2 mutants on the right bottom figure, on the top sequence has replacement on A, the middle sequence has replacement on B and C. and scores are 0.4 and 0.2 each.

We think that if there is a sequence that has replacement on A B and C, this sequence activity will sum of the Additive Score. So, this activity score can be evaluated as 0.6, we think.

次に、もう1つの方法で変異体の活性評価をしました。

K近傍法を用いて評価しました。

K近傍法のアルゴリズムを説明します。

例えば、右の図の星の活性を評価したいとします。

そのとき、相同性の高い配列の活性値が参考になると考えられます。

そこで、星の配列と相同性が高い配列との距離をハミング距離を用いて計算しました。

マッピングすると、右図のように書けます。

例えば、3つの配列から星の配列の活性を評価しようとすると、相同性の高い3つの配列の平均値を取ればよいので、0.4になるでしょう。

Next, we evaluated mutation activity by the other way.

We used K-nearest neighbor method.

What is K-nearest neighbor method? I will explain this method algorism.

Imagine, for example, when we want to evaluate the activity of the mutant of star, how can I evaluate of that’s activity?

It can be easily assumed that a sequence which has high homology between mutants have similar activity.

In K-nearest neighbor method, calculate average activity which similar sequence data has.

For example, if we want to evaluate an activity that star has. When we decide it was evaluated by 3 mutants, The activity of star will be able to evaluate with the average of A,B and C activity. So it can evaluated as 0.4, we think.

次に、私たちが提出する配列をどのように選んだか説明します。

左図は2通りで評価した変異体の活性値の相関を示しています。

図の通り、これらは相関があることがわかります。

そこで、私たちはそれぞれtop5つずつの配列を提出しました。

Additive Scoreを用いた評価はより挑戦的で、効果的な変異を多数導入します。しかし、変異による活性の喪失をまったく考えていないという問題があります。

一方、K近傍法による評価は保守的です。なぜなら、相同性の高い配列を参考にしているので、周りに活性を低下させるような変異を許しにくいからです。

これは、孫子の教えある”凡勝者、以正合、以奇勝”といえるでしょう。

Next, we explain how we chose the sequences we submitted.

The figure on the left shows the correlation between the activity values of the mutants by evaluated in the two ways.

As you can see in the figure, they are correlated. So, we thought that our approach has convincing to evaluate mutants.

Then, we submitted sequences with top5 each.

Evaluation by the Additive Score is more challenging than that by K-nearest neighbor method. So we assumed that it can introduces a number of effective mutations well. However, it has the problem that it does not consider the complete loss of activity due to fatal mutations at all.

On the other hand, evaluation by the K Nearest Neighbor method is conservative. This is because that it allows us to check whether a homologous sequence causes a fatal mutation.

Our strategy is much with ancient Chinese strategy byスンズー

Translated with DeepL.com (free version)

以上です。

Hello.

First, Thank you for inviting us. We express all of you are concerned with this project the gratitude.

I want to introduce ourselves shortly.

We are