

卒業研究のフロー

- ・データセットの用意と決定

(血液以外に昆虫の型など)

ナミテントウムシのデータセット作成 (1 画像に複数のテントウムシが含まれていること)

画像内のテントウムシがどの型に属するのか Vott にてアノテーションを作成

このときの型をもっと遺伝子的に深堀したい (二斑型 h^C , 斑型 h^A , 紅型 h にわけて、さらに細分化してもよし)

- ・画像解析のモデルの決定

(画像の解析と正解データの csv ファイルの作成)

- ・モデルの勉強

- ・モデルの実装 & 精度の精査

- ・論文の作成

- ・基礎生物学研究所、国立遺伝子学研究所の共著による論文

ナミテントウの翅の斑紋はパニア遺伝子の発現の有無で決まる。

パニア遺伝子が発現すると、黒色色素（メラニン）が定着し、黒くなる

反対に発現しないと、赤色色素（カロテノイド）が定着

パニア遺伝子にはメラニンの合成を促し、カロテノイドの沈着を抑制する機能を持つことが明らかにされている。

主要な斑紋型・・・紅型、斑型、二斑型

パニア遺伝子が斑紋パターンを決定する唯一の遺伝子であり、ゲノム解読を行い、パニア周辺の DNA 配列を取得し、斑紋ごとに配列比較したところ第一イントロンの配列の多様性が斑紋の違いをもたらす可能性が示されました。

論文の最後に記述されていた今後の展望

本研究により、ナミテントウの斑紋パターンを決定する遺伝子として Pannier 遺伝子の特定に成功し、たった一つの Pannier 遺伝子が斑紋パターンを決定することを明らかにしました。今回の成果は、進化の過程で多様性が遺伝子レベルで生じるメカニズムの普遍的な原理の解明に繋がることが期待されます。

↓

あくまでもゲノム解析を行うことで到達できた。斑紋パターンを画像解析にかけることで、ゲノム解析とはまた違ったアプローチで分類に試みる。本来ではゲノムデータを結びつけ

ることで、精度や研究の実益も向上するのだが、ナミテントウのゲノムデータの入手が困難なため、斑紋のパターンを遺伝子型として画像と結び付け研究を進めていく。

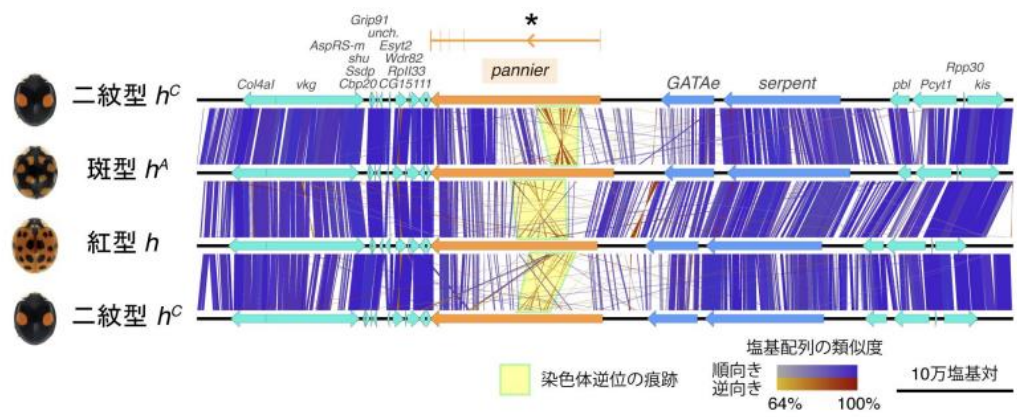


図 3 : 各斑紋型におけるパニア遺伝子周辺の DNA 配列の比較

中間発表

表紙

タイトルは SSD を利用した画像認識による生物の遺伝子多型の同定です（仮題）。

目次

この順番で進んでいきます。

研究背景

第一に画像認識の背景

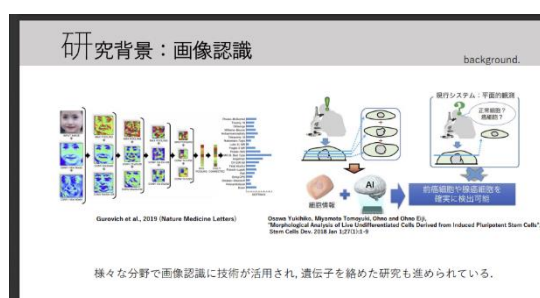
左図は顔の形（パーツ）から遺伝子疾患を予測するものです。その疾患から遺伝子の変異を予測するものです。

遺伝子疾患と呼ばれる遺伝子の欠損や変異による問題は内臓疾患や精神遅滞などの問題が現れることがあり、また顔などの表面的特徴として現れることがあります、これは子供のころに顕著でも大人になると特徴が薄くなってしまいます。

→遺伝子疾患の病気しか特定できない。

右図は細胞を二次元的から三次元的に捉え、細胞ががん細胞かどうか識別する研究です。焦点を変えることで、対象を立体的に捉え、情報量を増やすことで制度をあげるという研究です。

→細胞ががんでるかどうかしか識別できない。遺伝子にかかわっていない。



背景 2

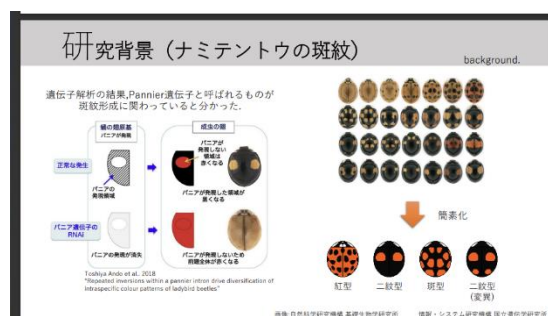
今回解析対象として選んだテントウムシの斑紋について

ナミテントウと呼ばれるテントウムシには様々な斑紋パターンが存在します。

これはパニア遺伝子とよばれるものの発現パターンによって、決定されます。

パニア遺伝子は発現時、黒色色素であるメラニンの合成を促し、赤色色素であるカロテノイドの沈着を抑制します。RNAi（RNA 干渉）によってパニアの発現を抑えたところすべて赤くなった。

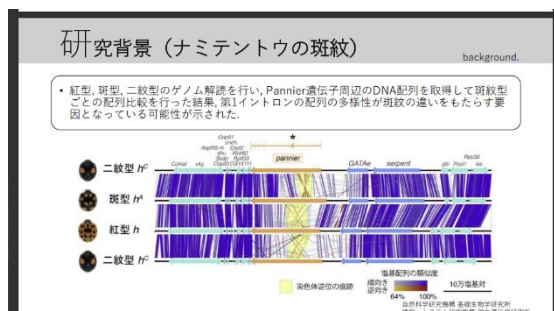
右図の通り、様々な斑紋パターンがあるがこれを簡素化し、本研究では 4 つに分類した。



スライド二枚目（ナミテントウの斑紋形成パターン）

パニア遺伝子周辺の DNA 配列を取得、斑紋パターンごとに比較したところ、第一イントロンの配列に斑紋の違いをもたらす要因となっている可能性が示唆された。

アスタリスクの部分のパニアの第一イントロンの部分にあたるのだが、この部分において染色体の逆位の痕跡が見られた。赤い線は比較した際の塩基配列の類似度を示し、これら赤の部分は逆向きであることを示している。



研究目的

研究目的だが、最初のスライドなどからもわかるように現行の研究は、解析する対象がどの動物であるか、どの動物であるかにとどまっている。

解析する対象から遺伝子の型を特定することができれば、それは多くの分野で役に立つと考え、これに新規性を見出し、研究目的とする。（役に立つのは、例えばその遺伝子の型、遺伝子変異に対する対症療法的なアプローチが可能になったりするのではないかなど）

研究手法

スライド通りに

SSD の説明は長くなるので今回は省略させていただく。

現在の研究結果

テスト画像20枚を用いて、プログラムを実行したところ、プログラム自体は無事稼働した。しかし、結果は脅威の11%。この原因は、単に教師用のデータが圧倒的に少ないことに起因すると考えられる。

現在、正答率をあげるため鋭意活動中。具体的には、教師用データの拡充、1枚当たりの情報量を増やす。

また現在は正答率を出力する形だが、学習後に予測を行った際に、それらを csv ファイルなどの形で出力したいと考えている。