

＜遺伝統計学チームの研究ミッション＞

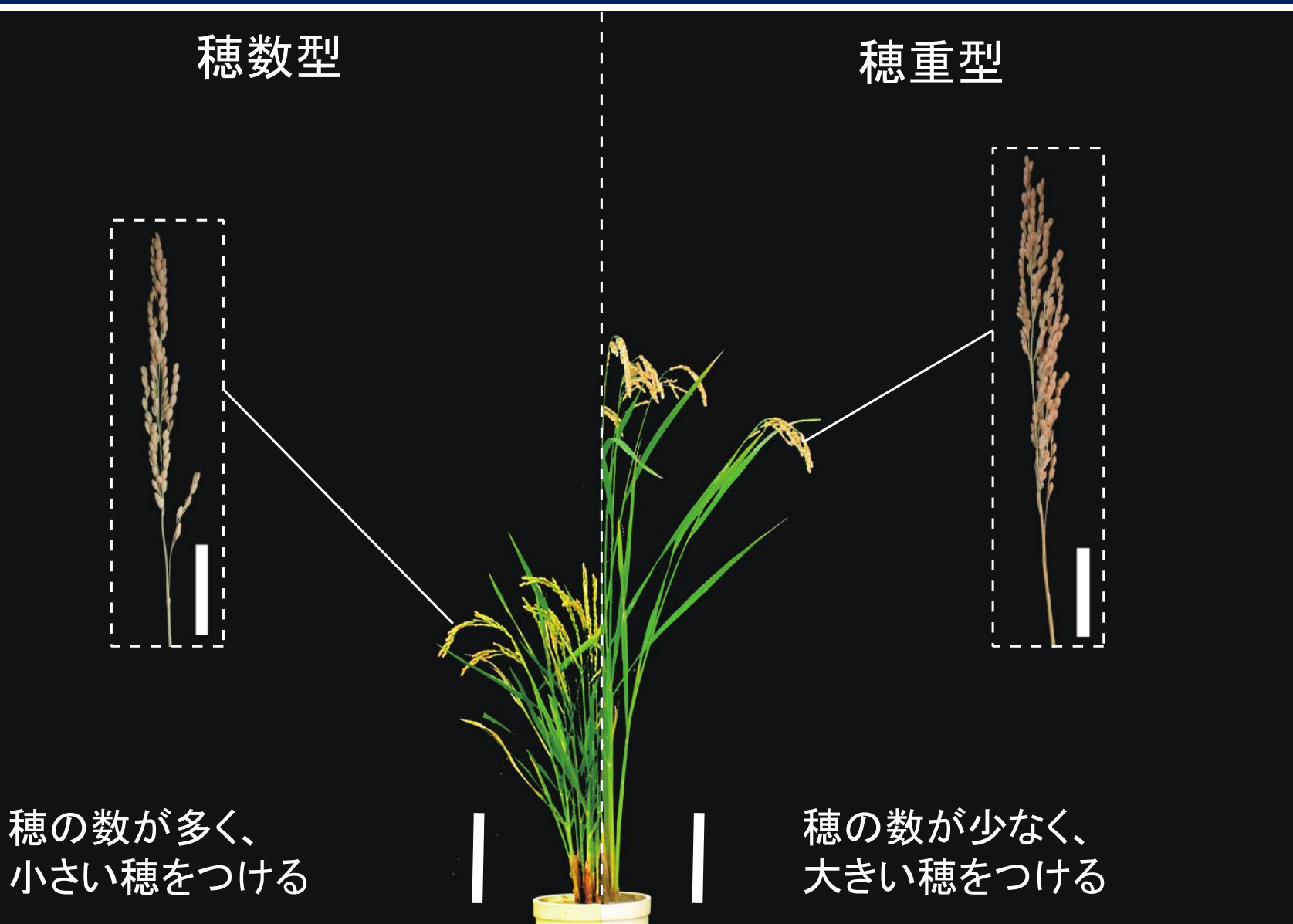
生命科学(医学や農学)のビッグデータを機械学習・人工知能ベースの遺伝統計学で解析を行い、ヒトの疾患や農業形質などの複雑な現象に寄与する要因を探り出す。今回、我々研究チームの取り組んでいる課題のうち、2つのテーマを代表して紹介する。

1. 機械学習を用いたイネの収量に関わる遺伝子の同定(矢野特別研究員)

要約

イネの収量は、イネの丈や穂の数、穂の構造など、複数の要素により決定される。このような複雑な要素を解析し、それに関わる遺伝子を同定するには、新しい解析手法が求められていた。そこで、教師なし機械学習の一つである主成分分析(PCA)とゲノムワイド関連解析(GWAS)を組み合わせることで、収量に関わる遺伝子を同定を試みた。その結果、収量決定の一因となる遺伝子を同定することに成功し、本手法が複雑形質に関わる遺伝子の同定に有効であることが確認された。今後、本研究で、深層学習の適用や、解析対象としたイネだけでなく、他の生物種にも本手法を応用することが期待される。

背景



イネの収量は、草型と呼ばれる形態と密接に関わっている。草型は、一般に穂の数と穂の大きさにより、大きく二つに分けられ、1株当りの穂の数が多く小さい穂をつけるタイプは「穂数型」、穂の数が少なく大きい穂をつけるタイプは「穂重型」と呼ばれる。さらに、イネの穂は、穂軸、一次枝梗(しこう)、二次枝梗、種子などの要素によって構成されており、これらの要素により穂の大きさが決められている。

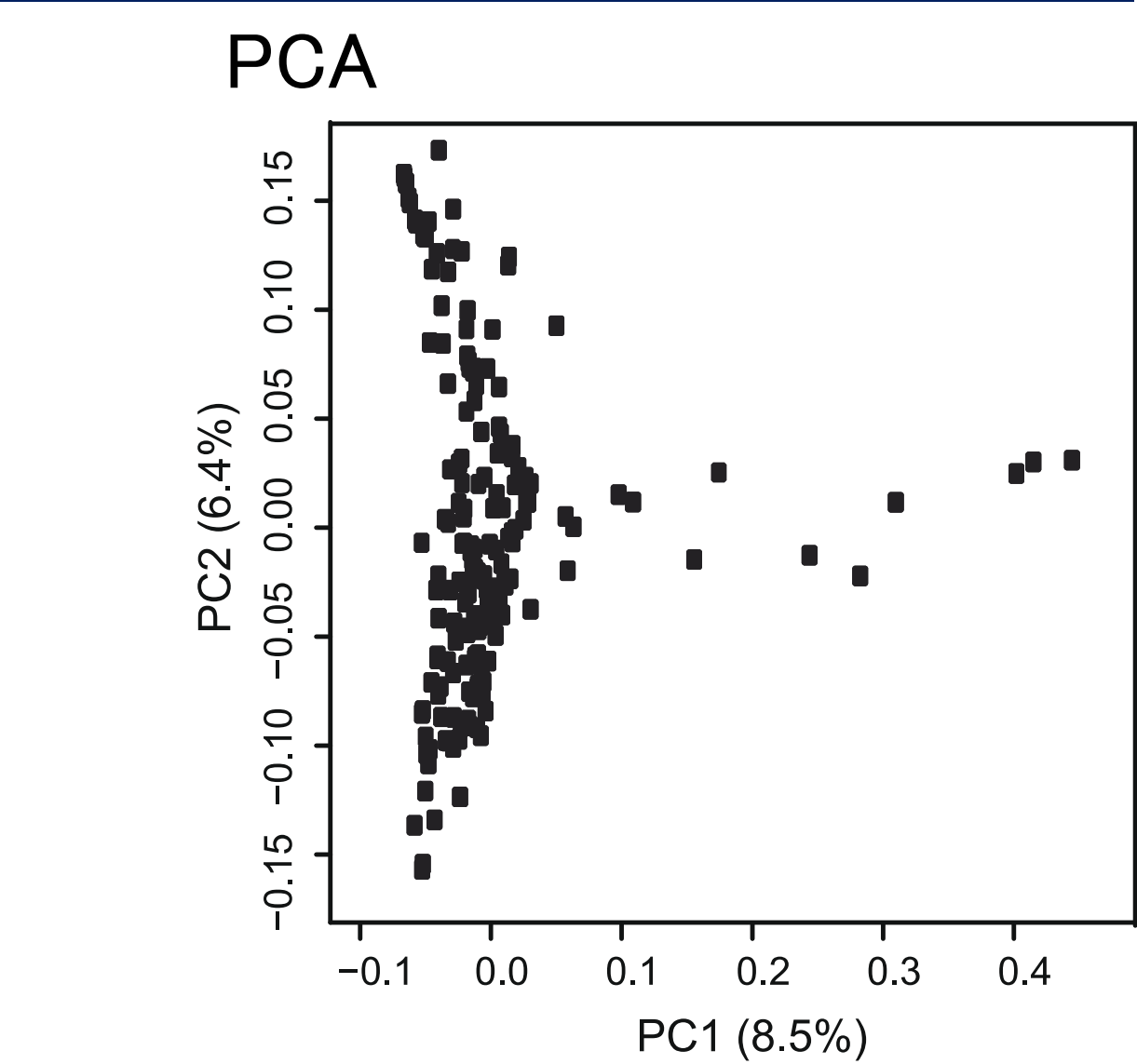
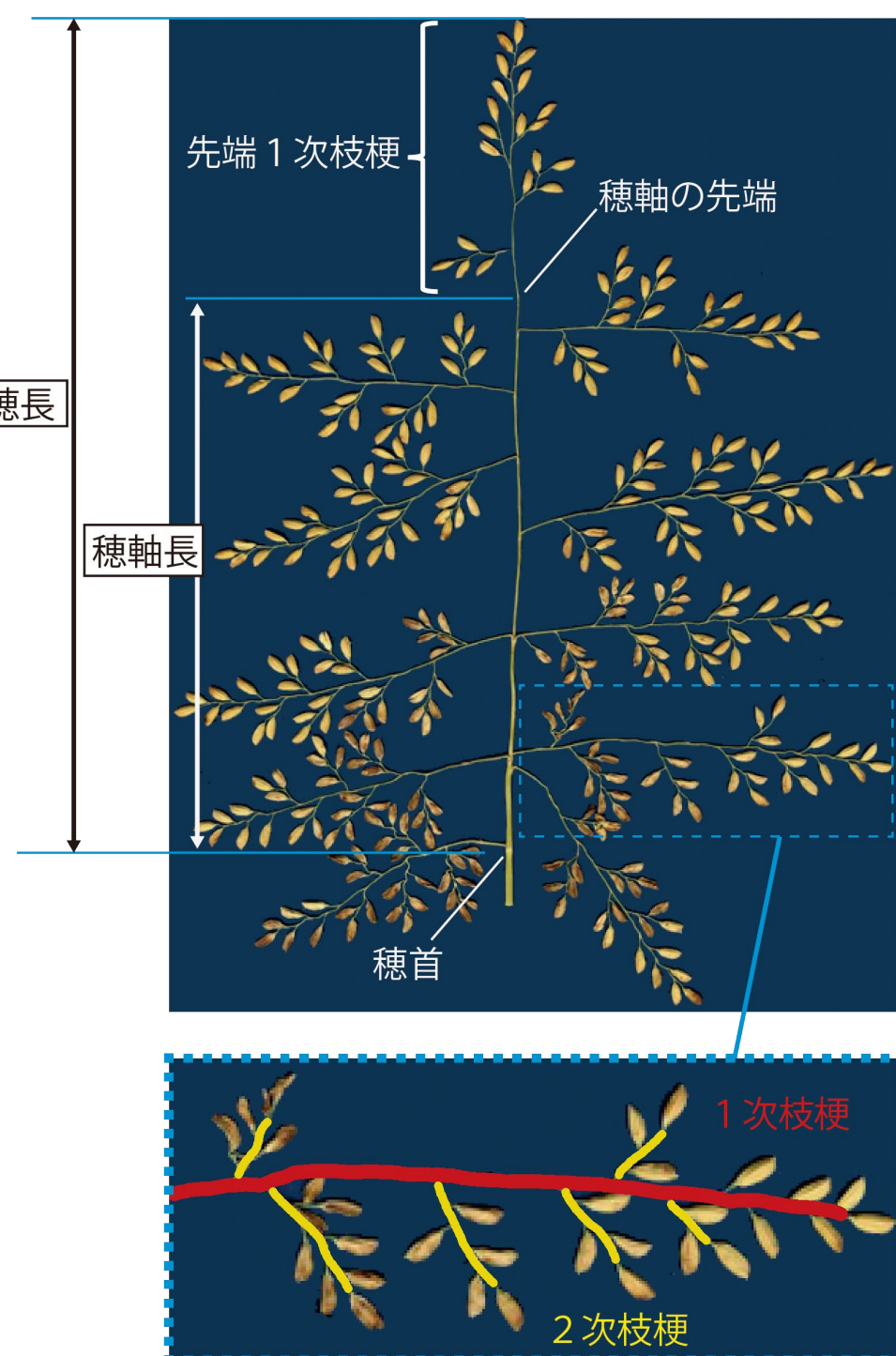
実験材料と方法

植物材料
日本で育成されたイネ169品種

ゲノム解析
Illumina HiSeq 2000 system
平均デプス: 5.8
総リード数: 3,689,176,148

381,007 SNPs・58,886 INDELS

対象形質



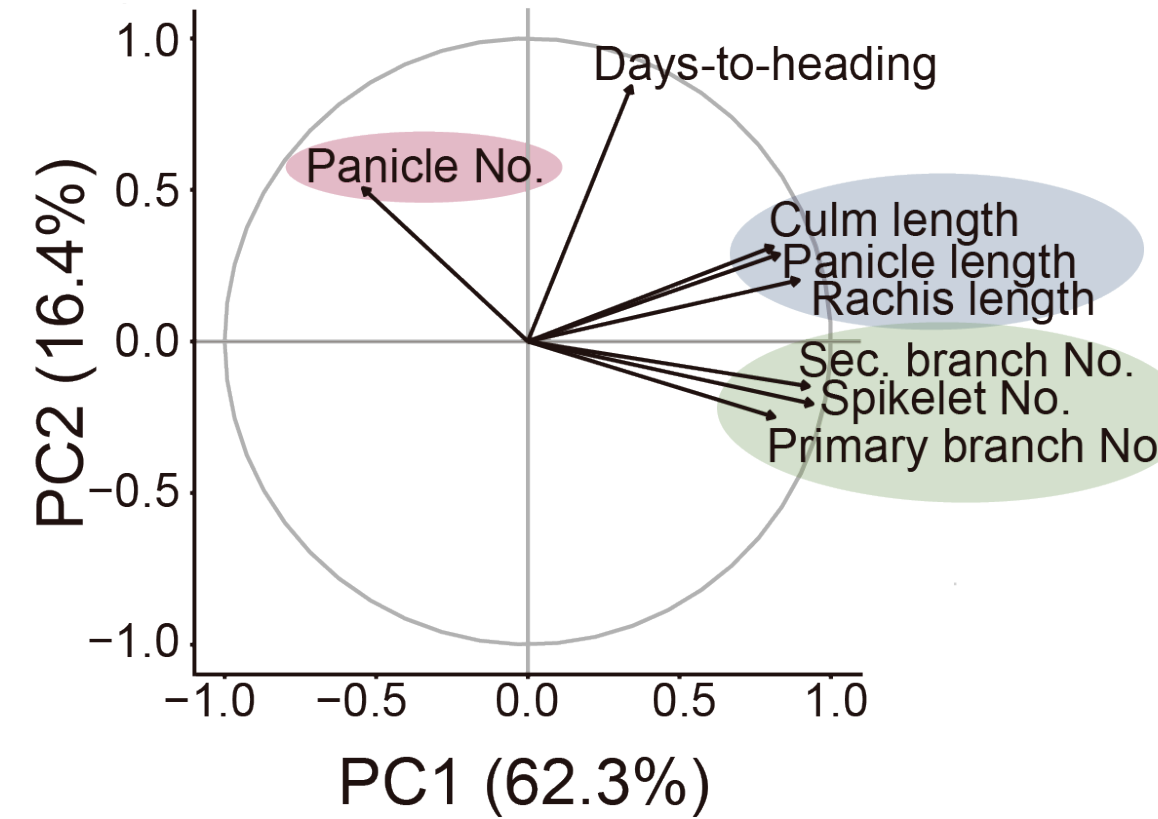
草型の指標となる8形質(花が咲くまでの日数、イネの丈、穂の数、穂の大きさなど)を計測した。

さらに、穂の大きさに関わる形質として、以下の5形質を測定した。

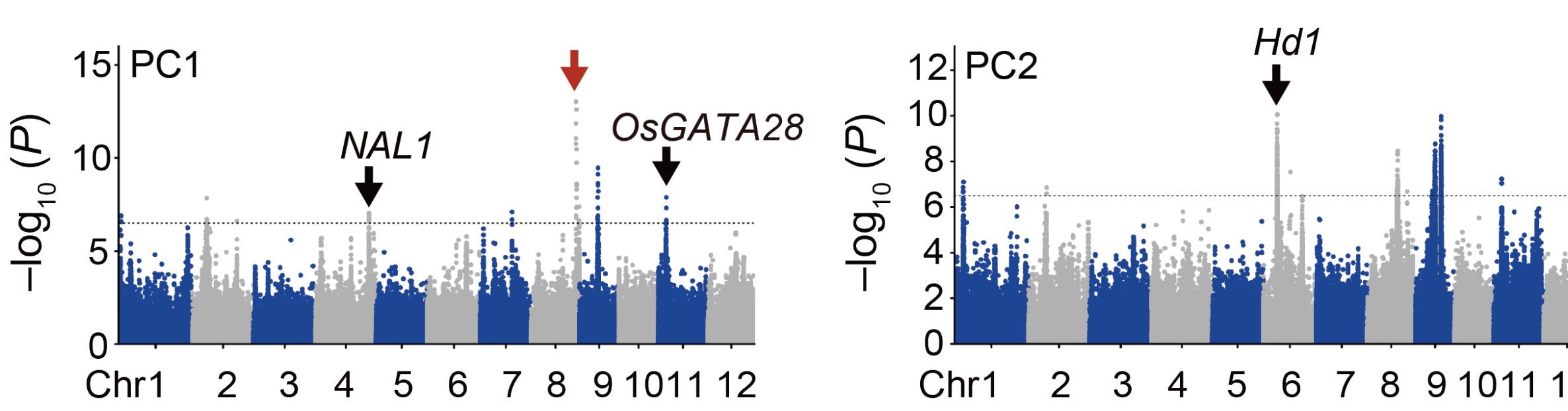
- ・穂の長さ
- ・穂軸の長さ
- ・一穂当たりの1次枝梗の数
- ・一穂当たりの2次枝梗の数
- ・一穂当たりの種子の数

結果 I: 候補遺伝子の同定

草型の指標となる8形質を用いて主成分分析を行い、これらの形質の特徴量を抽出した。



PC1に対して、イネの丈と穂の大きさに関わる6つの形質は正の相関(0.80-0.92)を示した。一方、穂の数は負の相関(-0.54)を示した。この結果から、PC1が草型の指標になりうることを示唆している。そこで、PC1の主成分値を用いてGWASを行った。



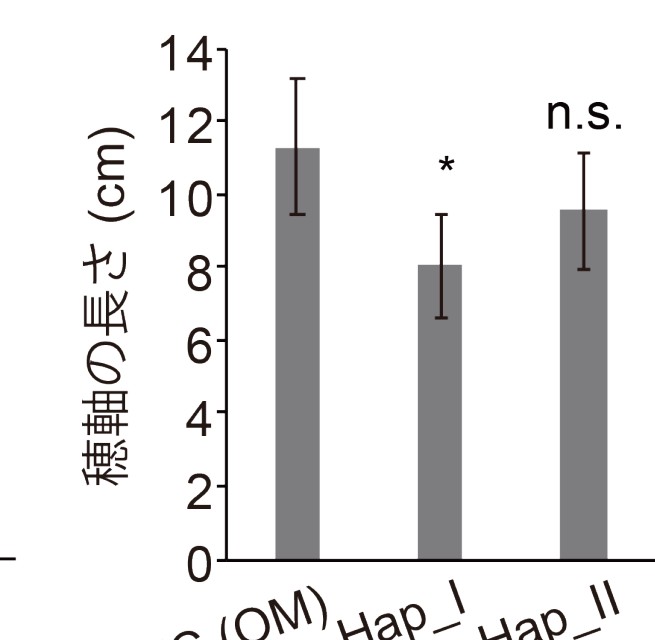
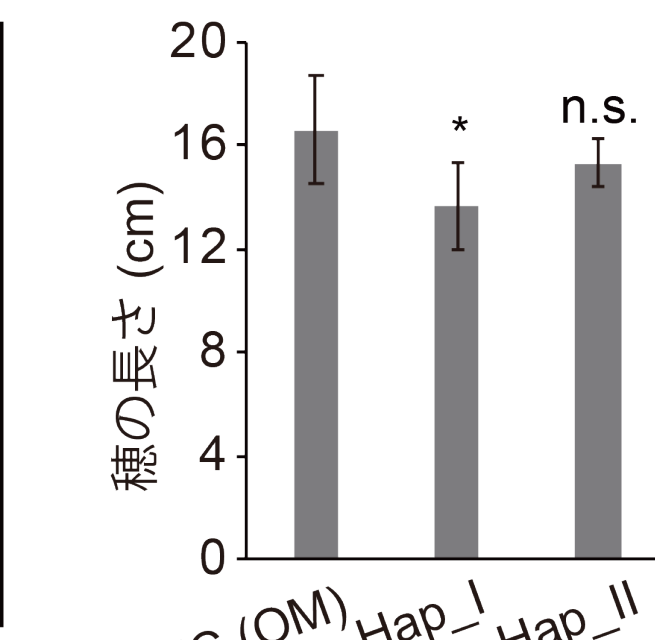
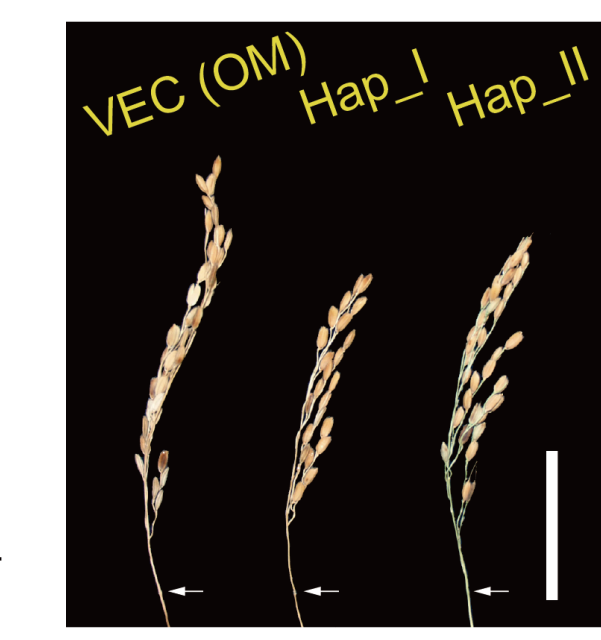
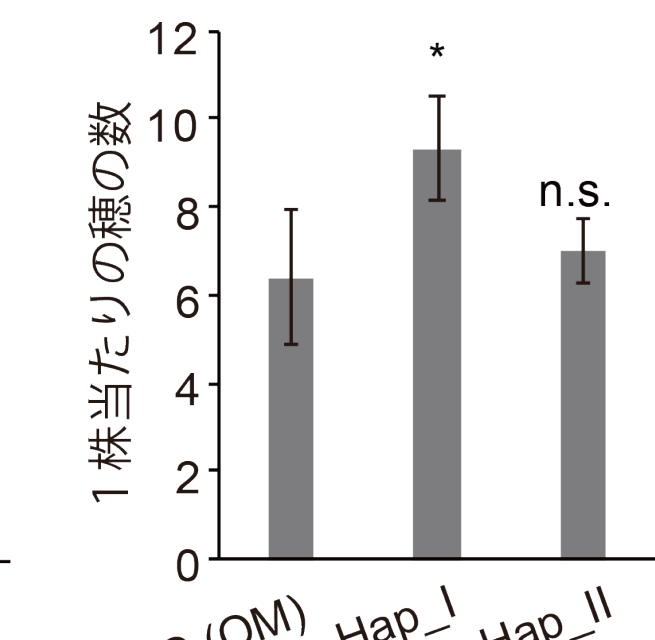
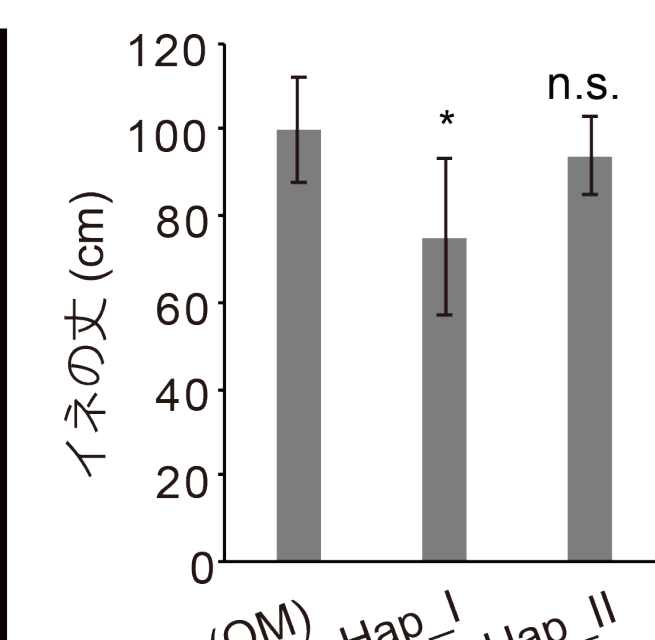
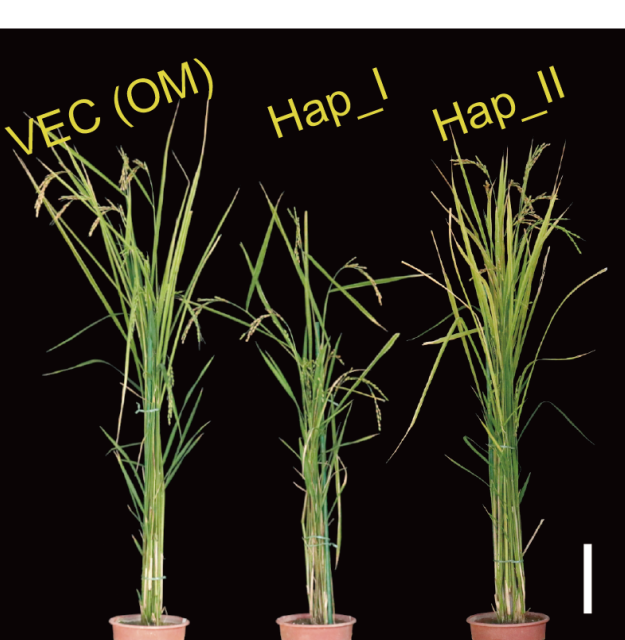
第8番染色体に最も高いピークが検出されたため、このピークに着目し以降の解析を行った。

LOC_Os08g44510 (OsSPY)
Chr 8 (+ strand)

	+26	+2498
Hap_I	AGC (S)	CGC (R)
Hap_II	ACC (T)	CTC (L)
Hap_III	ACC (T)	CGC (R)

候補領域の中には、植物ホルモン「ジベレリン」のシグナル制御に関わるN-acetyl glucosamine transferase をコードする遺伝子(OsSPY)が含まれていた。ジベレリンは、植物の丈に関わることが知られているホルモンであることから、このホルモンに関わるOsSPYが有力な候補遺伝子として考えられた。

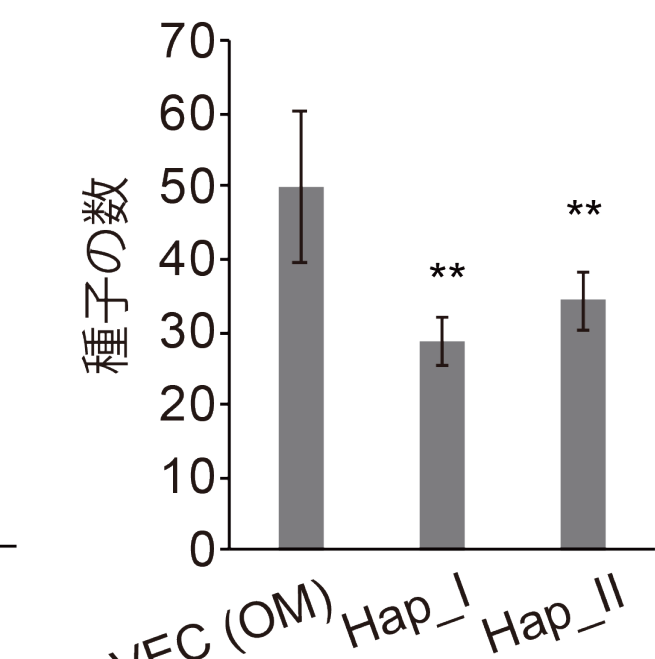
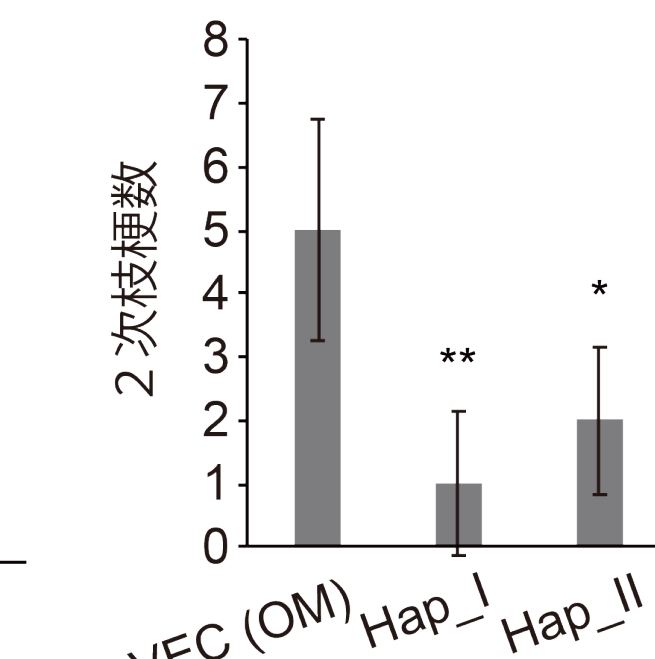
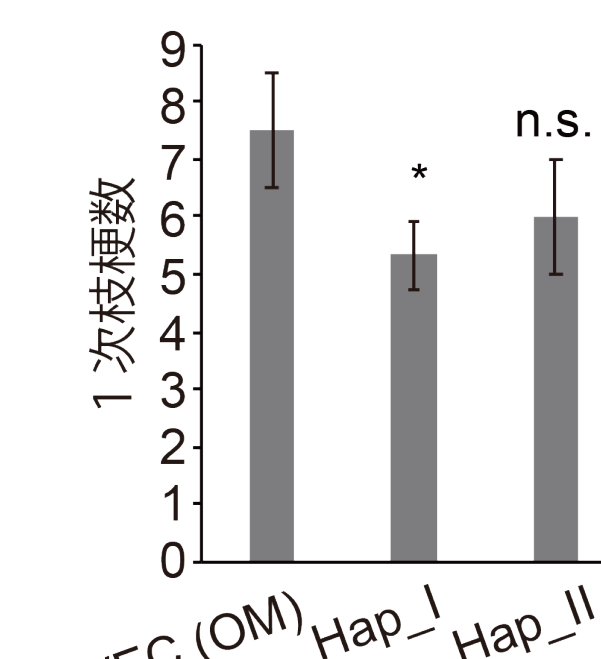
結果 II: 収量形質に対する効果



OsSPYに着目し、分子遺伝学的実験により検証を行った。

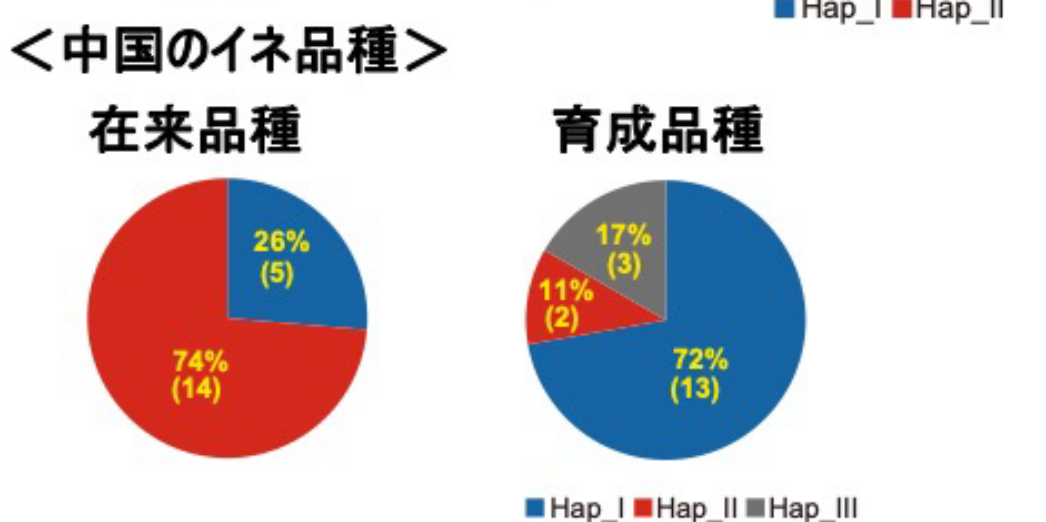
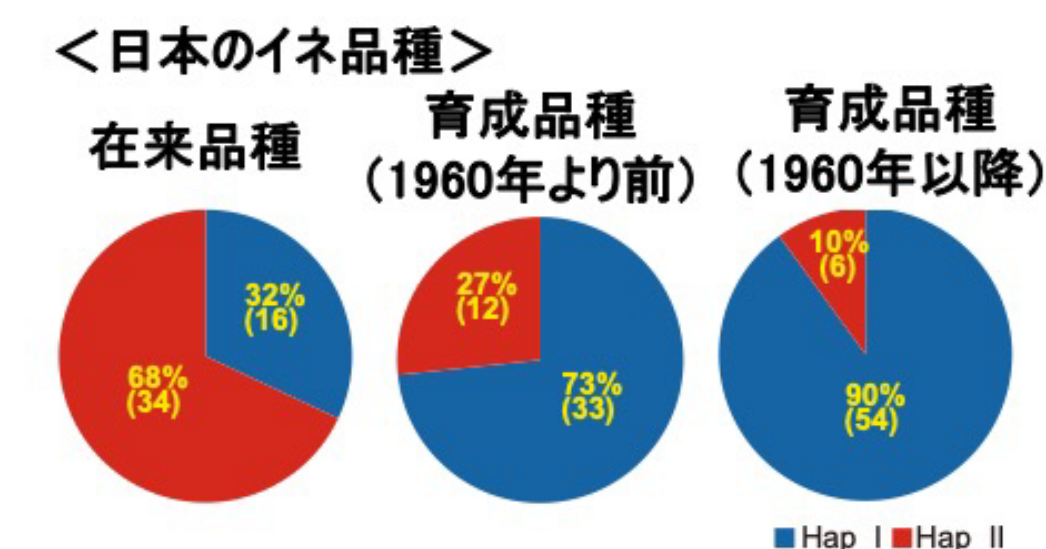
結果、イネの丈だけでなく、穂の数や穂の構造の制御に多面的に関わることがわかった。

つまり、この遺伝子が草型の穂重型と穂数型を決めている一因であることが明らかになった。

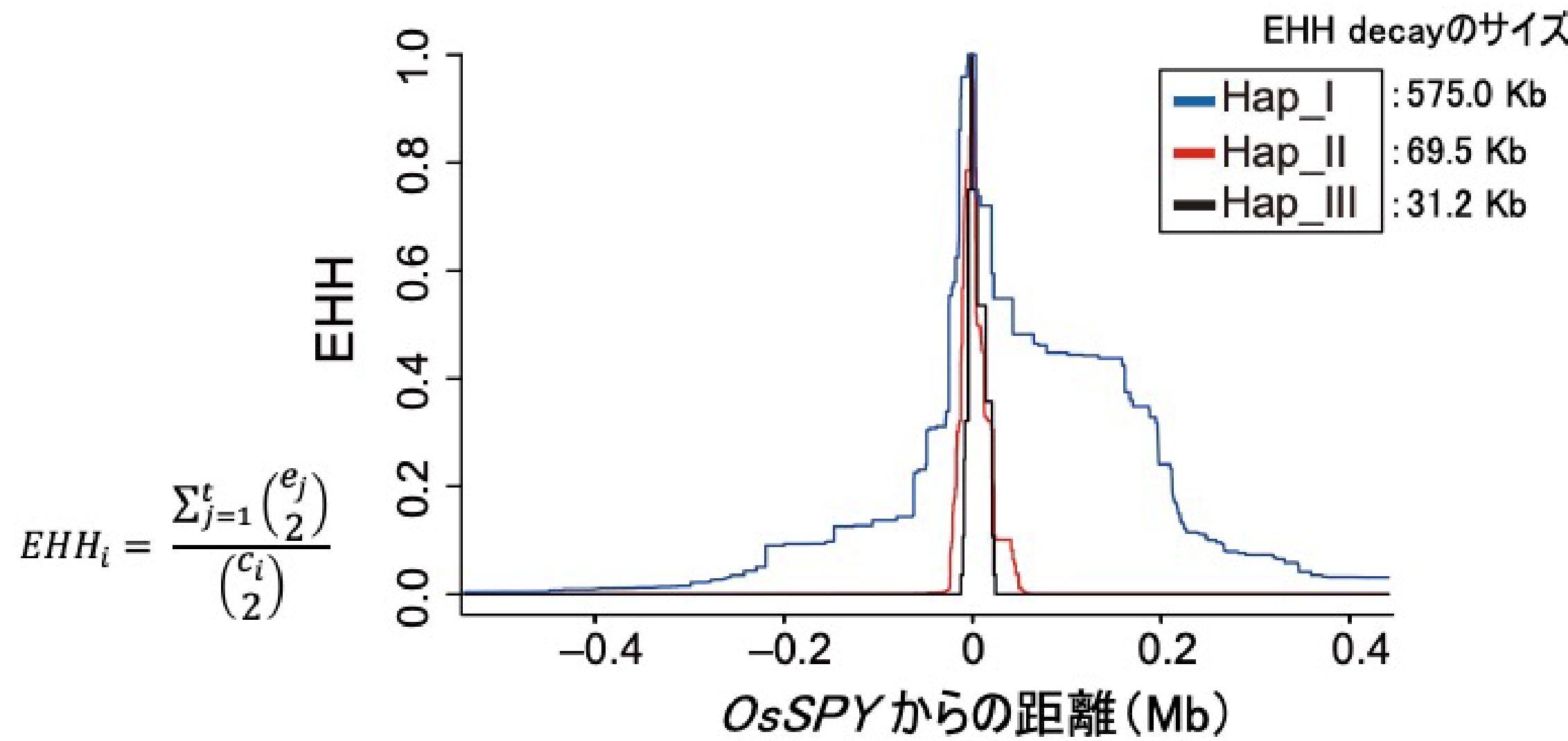


結果 III: 育種過程におけるハプロタイプ頻度の変遷

イネは、古くから各地に存在する「在来品種」と比較的最近の在来種をもとに育成された「近代品種」に分類される。草型との関連で、在来品種では穂重型、近代品種では穂数型が多い傾向があることが知られている。そこで、在来品種と近代品種の違いにOsSPYが関わっているか調べた。



ハプロタイプ割合の比較から、OsSPYは、他の領域に比べ在来品種から近代品種の変移の過程で選抜された可能性が高いことがわかった。また、イネの原産国である中国の品種においても同様の結果が確認された。



在来品種と近代品種の遺伝距離(遺伝的分化の尺度)を調査した。結果、在来品種から近代品種へ変移する過程でOsSPYの特定のハプロタイプが選抜された可能性があることが明らかとなった。この結果は、育種過程でOsSPYが重要な選抜対象だったことを示唆している。

2. バイオバンクにおけるエラー検出法の開発(櫻井特別研究員)

手法の概要

・質問票に対する外れ値検出法

→ 質問票の設計パターンから外れた個体を検出

例) 喫煙する人に尋ねる質問項目について、データベースに格納されるデータの欠損パターンは、下表のような集計でb,cが0でなければ異常パターンである。

	喫煙する	喫煙しない
非欠損	A	b
欠損	c	D

文献 Sakurai R, Ueki M, Tamiya G, et al. Outlier detection for questionnaire data in biobanks. *International journal of epidemiology*, 2019, 48.4: 1305-1315.

拡張

東北メディカルメガバンク機構では三世代の家族に対しデータが収集され、長期間フォローアップしている。

親子間相関、夫婦間相関のパターン

および
経時データにおける時系列のパターン
に対して手法を拡張する