

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2024年12月5日(05.12.2024)



(10) 国際公開番号

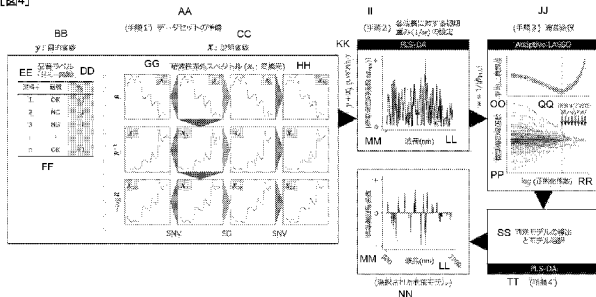
WO 2024/248014 A1

- (51) 国際特許分類:
G01N 21/3563 (2014.01)
- (21) 国際出願番号: PCT/JP2024/019621
- (22) 国際出願日: 2024年5月29日(29.05.2024)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願 2023-089036 2023年5月30日(30.05.2023) JP
- (71) 出願人: 国立大学法人九州大学(KYUSHU UNIVERSITY, NATIONAL UNIVERSITY CORPORATION) [JP/JP]; 〒8190395 福岡県福岡市西区元岡744 Fukuoka (JP).
- (72) 発明者: 松田 修(MATSUDA, Osamu); 〒8190395 福岡県福岡市西区元岡744 国立大学法人九州大学内 Fukuoka (JP).
- (74) 代理人: 鎌田 光宜, 外(KAMADA, Mitsunori et al.); 〒5410044 大阪府大阪市中央区伏見町四丁目1番1号 明治安田生命大阪御堂筋ビル 高島国際特許事務所 Osaka (JP).
- (81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU,

(54) Title: METHOD FOR SORTING SEEDS

(54) 発明の名称: 種子の選別方法

【図4】



- AA (Procedure 1) Preparation of data set
BB Objective variable
CC Explanatory variable
DD Quality label (dummy variable)
EE Quality label
FF Sample, Quality
GG Short-wavelength IR spectrum
HH Conversion source
II (Procedure 2) Setting of initial weighting with respect to each wavelength
JJ (Procedure 3) Wavelength selection
KK Any
LL Wavelength
MM, PP Standardized partial regression coefficient
NN (Selected discrimination model)
OO Mean squared error
QQ Wavelength combination for selection
RR Regularization factor
SS Discrimination model derivation and model selection
TT (Procedure 4)

(57) **Abstract:** The present invention provides a method for sorting plant seeds having a desired trait, the method including the following first to seventh steps: (First step) A step for extracting some seeds from a seed group. (Second step) A step for constructing a data set with respect to the extracted seeds, the step for constructing a data set including the following (a) to (c): (a) beaming light onto the extracted seeds and acquiring spectral data and/or spectral data derived therefrom; (b) subjecting the seeds for which spectral data has been acquired to quality inspection with respect to the desired trait, and determining the advance quality rate of the extracted seeds, the advance quality rate being defined by the following formula: [Advance quality rate (%)] = [number of seeds having desired trait from among seeds subjected to quality evaluation] / [number of seeds subjected to quality evaluation] × 100; (c) associating the spectral data for each seed and the result of the quality inspection. (Third step) A step for applying multivariate discriminant analysis to the data set constructed in the second step and thereby deriving candidates for a good/poor discrimination model for calculating a discrimination score for the desired trait in each seed. (Fourth step) A step for selecting a discrimination model from the candidates for a good/poor discrimination model derived in the third step, the step including selecting as a good/poor discrimination model a candidate that satisfies the criterion that the conformity rate exceeds the advance quality rate under

LY, MA, MD, MG, MK, MN, MU, MW, MX, MY,
MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL,
PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK,
SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA,
UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

- (84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類:

一 国際調査報告(条約第21条(3))

a standard sorting condition, where the recovery rate and the conformity rate are defined by the following formulae, and a sorting condition that the advance quality rate and the recovery rate have the same value is defined as the "standard sorting condition": [Recovery rate (%)] = [Number of seeds selected on the basis of discrimination score from among seeds subjected to quality evaluation] / [number of seeds subjected to quality evaluation] × 100 [Conformity rate (%)] = [Number of seeds with desired trait from among seeds subjected to quality evaluation and selected on the basis of discrimination score] / [number of seeds selected on the basis of discrimination score from among seeds subjected to quality evaluation] × 100 (Fifth step) A step for beaming light onto the remaining seeds not extracted from the seed group in the first step and/or other seeds obtained under essentially the same conditions as those for said seeds and acquiring spectral data, and applying the discrimination model selected in the fourth step to the spectral data to determine the discrimination score for each seed. (Sixth step) A step for determining a threshold for the discrimination score on the basis of the discrimination scores determined in the fifth step. (Seventh step) A step for comparing the discrimination scores of the respective seeds determined in the fifth step with the threshold for the discrimination score determined in the sixth step, and thereby recovering seeds predicted to have the desired trait.

(57) 要約: 本発明は、所望の形質を有する植物種子の選別方法であって、以下の第1工程～第7工程を含む、方法を提供する: (第1工程) 種子集団から一部の種子を取り出す工程、(第2工程) 取り出された種子に対するデータセットを構築する工程であって、該データセットを構築する工程は以下(a)～(c)を含み: (a) 取り出された種子に対して光を照射してスペクトルデータおよび/またはその派生スペクトルデータを取得すること、(b) スペクトルデータを取得した種子を所望の形質に対する品質検査に供し、該取り出された種子の事前良品率を決定すること、ここで、事前良品率は次式で定義される: [事前良品率(%)] = [品質評価に供された種子のうち、所望の形質を有していた種子数] / [品質評価に供された種子数] × 100 (c) 各種子のスペクトルデータと品質検査の結果とを対応させること、(第3工程) 第2工程で構築されたデータセットに対して多変量判別分析を適用することにより、各種子における所望の形質に対する判別スコアを算出するための良否判別モデルの候補を導出する工程、(第4工程) 回収率および適合率を次式で定義し: [回収率(%)] = [品質評価に供された種子のうち、判別スコアに基づいて選択される種子数] / [品質評価に供された種子数] × 100 [適合率(%)] = [品質評価に供された種子であって、判別スコアに基づいて選択される種子のうち、所望の形質を有している種子数] / [品質評価に供された種子のうち、判別スコアに基づいて選択される種子数] × 100 かつ、事前良品率と回収率を同値とする選別条件を「標準選別条件」と定義するとき、該標準選別条件下で、適合率が事前良品率を上回るとの基準を満たす候補を良否判別モデルとして選択することを含む、第3工程で導出された良否判別モデルの候補から判別モデルを選択する工程、(第5工程) 第1工程で種子集団から取り出されなかった残りの種子、および/または該種子と実質的に同一の条件下で得られた別の種子に対して光を照射してスペクトルデータを取得し、該スペクトルデータに第4工程で選択された判別モデルを適用することで各種子の判別スコアを決定する工程、(第6工程) 第5工程で決定された判別スコアに基づいて、判別スコアの閾値を決定する工程、(第7工程) 第5工程で決定された各種子の判別スコアと第6工程で決定された判別スコアの閾値とを比較することで、所望の形質を有すると予測される種子を回収する工程。

明 細 書

発明の名称： 種子の選別方法

技術分野

[0001] 本発明は、種子の選別方法に関し、詳細には、種子のスペクトルデータを多変量統計分析で解析することによる、所望の形質を有する植物種子の選別方法に関する。

背景技術

[0002] 世界人口は増加を続け、2022年には80億人、2030年には85億人、2050年には97億人に達すると予測されている。従って、人類社会を存続させていくためには、食料の安定的生産および供給の重要性がますます高まっている。かかる問題を解決するには、多収品種を開発し、世界中に普及させることが有効な手段の一つと考えられている。しかし、どのような品種で作物を生産するにしても、良質な苗がなければその品種の優れた形質は発揮され得ない。古来日本には「苗半作（なえはんさく）」という言葉があり、この言葉は「成長の半分は苗で決まる」との意味を有する。さらに踏み込めば、良質な種子なしには、良質な苗を作ることはできない。したがって、高品質の種子をどのように生産し調製するか、その品質を劣化させずにどのように保存するか、そしてその品質の現状をどのように評価するかは、人類が協力して取り組むべき緊急の技術的課題である。

[0003] 種子の品質管理の重要性は、国内外の公的機関によって強く認識されている。国際種子検査協会（ISTA）は、主要な作物種のほぼすべての種子についての品質検査方法に関するガイドラインを提供している。また、日本政府は、市場に流通する農作物生産用の種子の品質基準を定めた法律「種苗法」を施行している。このような公的監視が行われる理由の1つは、種子の品質（とりわけ発芽性や遺伝的純度）は、通常、種子の外観やその他の簡易な基準によっては識別することが非常に難しいことにある。低品質の種子が誤って拡散してしまうと、地域的な食糧不足につながるのみならず、農家の貧

困にもつながる可能性がある。

[0004] 公的な品質管理は確かに種子がもたらす作物の不作を防ぐのに役立つものであるが、何事にも必ず2つの側面が存在する。例えば、公的な品質管理の観点から種子が満たさなければならない良品率が90%と設定されている場合、当該90%から僅か1%低い良品率（即ち、89%の良品率）を有する種子ロットは、その商品価値を失ってしまう。なお、実状としては、種子供給業者らによる熾烈な品質競争により、そのような法的基準でさえも無関係或いは競争力のないものとなっている。その結果として、作物生産に貢献する可能性があるにもかかわらず、商業的価値がない種子が大量に生み出され続けており、この状況は無慈悲な「シードロス」と称し得る。

[0005] 主要な作物種であっても、種子の良品率は、収穫後の選別されていない状態では通常90%には及ばない。不良種子の発生は母植物での種子の成熟不全が一因であるが、収穫後の種子の調製プロセス中の損傷、劣化、または汚染などもまたその一因となり得る。多段階の選別を経て初めて法定基準内の高い良品率を達成することができるが、かかる選別において使用される大規模な装置としては、スパイラル比重選別機、トロンメル回転式選別機、振動選別機、磁力選別機、重量選別機、色彩選別機などが挙げられる。しかし、選別後の最終的な良品率が目標を下回ってしまった場合、播種、栽培、収穫および選別を経てようやく調製された種子はただのゴミの山と化す。現状、この段階から「最後の1パーセント」であっても良品率を改善させる方法は存在しない。

[0006] 上述される種子の選別装置は、種子の品質が、その外表面または外部から検出可能な特徴に反映されるとの前提に基づくものである。種子の外表面または外部から検出可能な特徴としては、形状、表面の質感や粘着性、比重および色などが含まれる。上述の装置は大量選別に適しており、収穫された種子のほとんどを高品質の状態に調製し得る。対照的に、これらの装置は動作条件の微調整が難しいとの欠点もある。さらに、種子の生化学的特性に着目すると、外部から隠れた胚や胚乳の変化は、露出した種皮における変化より

も、種子の機能に遥かに大きな影響を及ぼし得る。たとえば、胚乳脂質の酸化や特定の害虫による胚の摂食は、種子の外観や重量に直ちに変化を引き起こすわけではないが、発芽能を大きく毀損させる。「シードロス」は、このような不良種子によってもたらされるものと推定される。

[0007] 近赤外（NIR）分光法は、主に有機化合物で構成される物質の化学組成の違いや変化を非破壊的に評価する有効な手法である。換言すれば、NIR分光法は、食品、医薬品、農産物との親和性が高く、それらの品質管理への応用が広がりつつある。主にリモートセンシングの分野において発展してきたイメージング分光法も、不等質であることを特徴とするこれらの商品に関する品質管理において注目を集める技術である。これらの新技術が融合した近赤外イメージング分光法は「シードロス」をなくす鍵を提供すると大いに期待される。

[0008] 実際、生存、成長力、遺伝的純度、虫害や菌害の有無などの観点から種子の品質を評価する際に、NIRイメージング分光法を応用した事例も報告されている（非特許文献1～4）。しかし、未だに「シードロス」が続いているということは、この技術がまだ実用レベルに達していない、あるいは社会実装が十分に進んでいないことを如実に物語っているように思われる。

先行技術文献

非特許文献

[0009] 非特許文献1: Rahman A. et al., Assessment of seed quality using non-destructive measurement techniques: A Review., Seed Science Research, Volume 26, Issue 4, December 2016, pp.285-305.

非特許文献2: Feng L. et al., Hyperspectral imaging for seed quality and safety inspection: A Review, Plant Methods, 15, Article number: 91 (2019)

非特許文献3: Venkatesan S. et al., Role of near-infrared spectroscopy in seed quality evaluation: A review. Agric Rev. 2020;41.

非特許文献4: Long W. et al., A Review of Artificial Intelligence Meth

ods for Seed Quality Inspection based on Spectral Imaging and Analysis, J. Phys.: Conf. Ser., 1769, 012013.

発明の概要

発明が解決しようとする課題

[0010] 上述の先行技術文献において報告されている手法は、いずれも一定の条件下において非破壊的な種子の選別を可能とするものであり有用性を有するものである。しかしながら、かかる先行技術はいずれも特定の植物種子の、特定の形質への適用に制限されていることから、より多くの品種および形質に対応できる汎用性の高い新たな種子選別手法の構築には依然として強いニーズがある。

課題を解決するための手段

[0011] 本発明者は、上記課題に対して鋭意検討した結果、光学的分析手法等を用いて構築した各植物種子に対するデータセットを多変量判別分析に適用して種子の判別モデル候補を多数生成させ、当該モデル候補から予測精度の高い判別モデルを選択する際に、「標準選別条件」との概念を導入することで予測精度の高いモデルの選択が可能となること等を見出し、かかる知見に基づいてさらに研究を進めることによって本発明を完成するに至った。

すなわち、本発明は以下の通りである。

[0012] [1]

所望の形質を有する植物種子の選別方法であって、以下の第1工程～第7工程を含む、方法：

(第1工程) 種子集団から一部の種子を取り出す工程、

(第2工程) 取り出された種子に対するデータセットを構築する工程であって、該データセットを構築する工程は以下(a)～(c)を含み：

(a) 取り出された種子に対して光を照射してスペクトルデータおよび／またはその派生スペクトルデータを取得すること、

(b) スペクトルデータを取得した種子を所望の形質に対する品質検査に供し、該取り出された種子の事前良品率を決定すること、

ここで、事前良品率は次式で定義される：

$$[\text{事前良品率}(\%)] = [\text{品質評価に供された種子のうち、所望の形質を有していた種子数}] / [\text{品質評価に供された種子数}] \times 100$$

(c) 各種子のスペクトルデータと品質検査の結果とを対応させること、
(第3工程) 第2工程で構築されたデータセットに対して多変量判別分析を適用することにより、各種子における所望の形質に対する判別スコアを算出するための良否判別モデルの候補を導出する工程、

(第4工程) 回収率および適合率を次式で定義し：

$$[\text{回収率}(\%)] = [\text{品質評価に供された種子のうち、判別スコアに基づいて選択される種子数}] / [\text{品質評価に供された種子数}] \times 100$$
$$[\text{適合率}(\%)] = [\text{品質評価に供された種子であって、判別スコアに基づいて選択される種子のうち、所望の形質を有している種子数}] / [\text{品質評価に供された種子のうち、判別スコアに基づいて選択される種子数}] \times 100$$

かつ、事前良品率と回収率を同値とする選別条件を「標準選別条件」と定義するとき、

該標準選別条件下で、適合率が事前良品率を上回るとの基準を満たす候補を良否判別モデルとして選択することを含む、第3工程で導出された良否判別モデルの候補から判別モデルを選択する工程、

(第5工程) 第1工程で種子集団から取り出されなかった残りの種子、および／または該種子と実質的に同一の条件下で得られた別の種子に対して光を照射してスペクトルデータを取得し、該スペクトルデータに第4工程で選択された判別モデルを適用することで各種子の判別スコアを決定する工程、

(第6工程) 第5工程で決定された判別スコアに基づいて、判別スコアの閾値を決定する工程、

(第7工程) 第5工程で決定された各種子の判別スコアと第6工程で決定された判別スコアの閾値とを比較することで、所望の形質を有すると予測される種子を回収する工程。

[2]

2以上の所望の形質を有する植物種子の選別方法である、〔1〕記載の選別方法。

〔3〕

形質が、発芽能、発芽勢、遺伝子型、ストレス耐性、休眠性、病害抵抗性、虫害抵抗性、QTL特性、食味性、出穂期、および形態学的な特徴からなる群から選択される形質の少なくとも1つである、〔1〕または〔2〕記載の選別方法。

〔4〕

2以上の所望の形質の予測が、以下の(i)～(iii)のいずれかの方法によって行われることを特徴とする、〔1〕～〔3〕のいずれか記載の選別方法：

(i) 形質ごとの良否に着目して形質ごとの判別モデルにより個別に種子をスコア化し、すべてのスコアが閾値以上のものを良種子と予測する方法、

(ii) 形質ごとの良否は顧みず、所望の形質のすべてが良好である場合を良好とし、それ以外の場合は不良とすることで単一の判別モデルを選択し、当該単一の判別モデルにより種子をスコア化し、スコアが閾値以上のものを良種子と予測する方法、

(iii) 形質ごとの良否に着目して形質ごとの判別モデルを選択した後、当該形質ごとの判別モデルを用いて得られたスコアを統合して統合スコアとし、該統合スコアの順位に基づき、各種子の良否を予測する方法。

〔5〕

判別モデルの精度を評価する工程をさらに含む、〔1〕～〔4〕のいずれか記載の選別方法。

〔6〕

植物が園芸作物である、〔1〕～〔5〕のいずれか記載の選別方法。

発明の効果

[0013] 本発明によれば、所望の形質を有する植物種子と、所望の形質を有さない植物種子の集団から、非破壊的、高効率、且つ高精度に所望の形質を有する

植物種子を選別することができる。

図面の簡単な説明

[0014] [図1]品質区分別の野菜種子の外観を示す図である。枠下の数値は、表2の判別モデルにより算出した各種子のスコアを表す。

[図2]品質区分別の野菜種子の近赤外反射スペクトルを示す図である。A～Iでは区分ごとの平均反射スペクトル、Jでは個別種子の反射スペクトルを示す。

[図3]良否判別モデルの導出手順を示すフロー図である。

[図4]良否判別モデルの導出手順を示す概略図である。

[図5]判別スコアの分布域と判別精度の関係を示す図である。Aは高精度、Bは低精度な判別モデルを使用した仮想的な状況を示す。事前良品率は80%とする。黒及び灰色の三角印は、それぞれ事前良品率及び標準閾値（表3脚注2参照）を示す。略号：eCDF（empirical Cumulative Distribution Function）、経験的累積分布関数。（a）全種子における判別スコアの密度分布と横軸スコアを上回る種子比率のeCDF。（b）良否別種子における判別スコアの密度分布。（c）良否別種子における判別スコアの分布域（箱ひげ図及び散布図）。（d）スコア下限閾の設定値と判別精度の関係。（e）種子回収率と判別精度の関係。

[図6]良否判別モデルにおける選択波長と変数重要度を示す図である。略号：SPRC（Standardized Partial Regression Coefficient）、標準偏回帰係数；VIP（Variable Importance in Projection）、射影における変数重要度。

[図7]カリフラワー種子の良否判別におけるスコア下限閾と精度指標の関係を示す図である。A．内的妥当性検証、B．外的妥当性検証。発芽形質、交配形質をそれぞれモデル $m-Ca47g$ 、 $m-Ca47x$ により逐次的に予測する際、両モデルの判別スコアに対する下限閾の設定値と、良種子（正常発芽能を持つF1雑種）の判別における精度指標との関係を示す。灰色の三角印は標準閾値（表3脚注2参照）を示す。

[図8]良否判別モデルの内的妥当性検証を示す図である。A. カボチャ 品種 5、B. エンドウ 品種 19、C. レタス 品種 803、D. ネギ 品種 22、E. トマト 品種 94、G. ネギ 品種 51、H. トマト 品種 221、F・I～K. カリフラワー 品種 47。Fは発芽形質のみ、Iは交配形質のみの判別の結果と精度を示す。Jは良種子（正常発芽能を持つF1雑種）かそれ以外かの判別の結果と精度を示す。KはF及びIで使用した2つの判別モデルに基づく両形質の予測を反映した、統合スコアによる良否判別の結果と精度を示す。記号、略号及び（a）～（e）のグラフの概要は、図5における説明と共通する。

[図9]良否判別モデルの外的妥当性検証を示す図である。A. カボチャ 品種 5、B. エンドウ 品種 19、C. レタス 品種 803、D. ネギ 品種 22、E. トマト 品種 94、F～I. カリフラワー 品種 47。A～Fは図8A～F、G～Iは図8I～Kと対応する。記号、略号及び（a）～（e）のグラフの概要は図5における説明と共通する。

[図10]カリフラワー種子の良否判別におけるPR／rPR曲線を示す図である。A. PR曲線、B. rPR曲線。発芽形質、交配形質をそれぞれモデル $m-Ca47g$ 、 $m-Ca47x$ により逐次的に判別する際、各形質の判別における再現率と、良種子（正常発芽能を持つF1雑種）の判別における適合率（A）及び相対適合率（B）との関係を示す。適合率及び相対適合率は曲面、全体的な判別精度は、その曲面下体積として表れるが、図11及び図12の表現に合わせ、これらも「曲線」及び「AUC（曲線下面積）」と称する。略号：（r）PR（(relative) Precision-Recall）、（相対）適合率と再現率の関係；AUC（Area Under Curve）。

[図11]良否判別モデルとデータセットの各組み合わせにおけるPR曲線を示す図である。略号：PR（Precision-Recall）、適合率と再現率の関係。

[図12]良否判別モデルとデータセットの各組み合わせにおけるrPR曲線を示す図である。相対適合率（Relative Precision）を次式により定義する。[相対適合率]＝（適合率－初期適合率）／（1－初期適合率）。尚、本明細書

において「適合率」は「事後良品率」と、「初期適合率」は「事前良品率」とそれぞれ同義である。再現率と相対適合率の関係を「rPR曲線」、その曲線下面積を「rPR-AUC」と称することとする。

[図13]良否判別スコアの画像表示を示す図である。A. カボチャ 品種5、B. エンドウ 品種19。

[図14]良否判別スコアにより区分した種子の発芽及び初期成長特性を示す図である。A・B. エンドウ 品種19、C. レタス 品種803、D. ネギ 品種22、E. カリフラワー 品種47。BはAの緑葉に対応する画素のみを抽出した画像である。A～D及びEの上段に示す順位は発芽形質、Eの下段に示す順位は交配形質の判別モデルによる予測順位である。

発明を実施するための形態

[0015] 以下、本発明について具体的に説明する。

[0016] 本発明は、所望の形質を有する植物種子の選別方法であって、以下の第1工程～第7工程を含む、方法（以下、「本発明の方法」と称することがある）を提供する：

（第1工程）種子集団から一部の種子を取り出す工程、

（第2工程）取り出された種子に対するデータセットを構築する工程であって、該データセットを構築する工程は以下（a）～（c）を含み：

（a）取り出された種子に対して光を照射してスペクトルデータおよび／またはその派生スペクトルデータを取得すること、

（b）スペクトルデータを取得した種子を所望の形質に対する品質検査に供し、該取り出された種子の事前良品率を決定すること、

ここで、事前良品率は次式で定義される：

$$[\text{事前良品率}(\%)] = [\text{品質評価に供された種子のうち、所望の形質を有していた種子数}] / [\text{品質評価に供された種子数}] \times 100$$

（c）各種子のスペクトルデータと品質検査の結果とを対応させること、

（第3工程）第2工程で構築されたデータセットに対して多変量判別分析を適用することにより、各種子における所望の形質に対する判別スコアを算出

するための良否判別モデルの候補を導出する工程、

(第4工程) 回収率および適合率を次式で定義し：

〔回収率(%)〕＝〔品質評価に供された種子のうち、判別スコアに基づいて選択される種子数〕／〔品質評価に供された種子数〕×100

〔適合率(%)〕＝〔品質評価に供された種子であって、判別スコアに基づいて選択される種子のうち、所望の形質を有している種子数〕／〔品質評価に供された種子のうち、判別スコアに基づいて選択される種子数〕×100
かつ、事前良品率と回収率を同値とする選別条件を「標準選別条件」と定義するとき、

該標準選別条件下で、適合率が事前良品率を上回るとの基準を満たす候補を良否判別モデルとして選択することを含む、第3工程で導出された良否判別モデルの候補から判別モデルを選択する工程、

(第5工程) 第1工程で種子集団から取り出されなかった残りの種子、および／または該種子と実質的に同一の条件下で得られた別の種子に対して光を照射してスペクトルデータを取得し、該スペクトルデータに第4工程で選択された判別モデルを適用することで各種子の判別スコアを決定する工程、

(第6工程) 第5工程で決定された判別スコアに基づいて、判別スコアの閾値を決定する工程、

(第7工程) 第5工程で決定された各種子の判別スコアと第6工程で決定された判別スコアの閾値とを比較することで、所望の形質を有すると予測される種子を回収する工程。

[0017] 本発明の方法において、「植物種子」とは、あらゆる植物の種子が意図され、特に限定されない。植物としては、種子植物であれば特に限定されず、例えば、被子植物及び裸子植物のいずれであってもよい。また、植物が被子植物である場合、植物は双子葉類及び単子葉類のいずれであってもよい。また、植物が双子葉類であるとき、植物は、合弁花類及び離弁花類のいずれであってもよい。一態様において、本発明の方法は、高付加価値を有する植物の種子の選別に用いられ得る。かかる高付加価値を有する植物としては、例

例えば、園芸作物や、建築用木材として用いられ得る植物が挙げられるが、これらに限定されない。園芸作物には、野菜、果樹および花卉が含まれる。尚、野菜としては、例えば、カボチャ、エンドウ、レタス、ネギ、トマト、カリフラワー、ニガウリ、オクラ、タマネギ、ミョウガ、大豆、フキ、アスパラ、ニラ、ソラマメ、セロリ、ニンジン、水菜、小松菜、春菊、大根、ブロッコリー、ハウレンソウ、ハクサイ、ルッコラ、レンコン、かぶ、アボカド、キュウリ、パプリカ、ニンニク、トウモロコシ、ズッキーニ、パセリ、パクチー、ナス、およびピーマン等が挙げられるがこれらに限定されない。また、果樹としては、ウメ、イチジク、アケビ、アセロラ、アボカド、オリーブ、オレンジ、カキ、カリン、グアバ、クランベリー、クルミ、グレープフルーツ、サクランボ、およびザクロ等が挙げられるがこれらに限定されない。また、花卉としては、アサガオ、ケイトウ、コスモス、ジニア、セイヨウオダマキ、センニチコウ、ペチュニア、ニチニチソウ、ハボタン、ヒマワリ、ハウセンカ、マツバボタン、ハナスベリヒユ、フウセンカズラ、マリーゴールド、カスミソウ、キンギョソウ、キンセンカ、スイートピー、ストック、ダイアンサス、デージー、ニゲラ、ネメシア、ネモフィラ、ハナビシソウ、バーベナ、パンジー、ビオラ、ヒナゲシ、ヤグルマギク、ルピナス、およびワスレナグサ等が挙げられるがこれらに限定されない。また、建築用木材として用いられ得る植物としては、スギ、ヒノキ、ヒバ、クリ、ケヤキ、サクラ、ブナ、ウォルナット、ファルカタ、およびアカマツ等が挙げられるがこれらに限定されない。

[0018] 本発明の方法において、「所望の形質」とは、種子そのものが有する形質および種子を生育させた際に当該種子から発芽した植物が発現する形質の両方を包含する概念とする。本発明の方法において選別の対象となる「形質」としては、例えば、発芽能、発芽勢、遺伝子型、ストレス耐性、休眠性、病害抵抗性、虫害抵抗性、QTL特性、食味性、出穂期、および、葉の大きさなどの形態学的な特徴等が挙げられるがこれらに限定されない。

[0019] (第1工程)

本発明の第1工程においては、種子集団から一部の種子を取り出すことを特徴とする。種子集団を構成する種子数は2以上であればよく、特に上限はない。例えば、種子集団は、通常、2粒～10,000,000粒、好ましくは1,000粒～10,000,000粒、より好ましくは10,000粒～10,000,000粒の種子からなる種子集団であってよいがこれらに限定されない。或いは、種子集団は、重量で示されるものであってもよい。例えば、種子集団は、1g～10,000kg、好ましくは1kg～10,000kg、より好ましくは10kg～10,000kgの重量の種子集団であってよいがこれらに限定されない。

[0020] また、種子集団の「一部」としては、選別の対象となる「形質」によって異なり得るが、通常、50粒～50,000粒、好ましくは100～30,000粒、より好ましくは200粒～3,000粒程度であるが、これらに限定されない。

[0021] (第2工程)

本発明の第2工程は、取り出された種子に対するデータセットを構築する工程である。なお、データセットの構築は、少なくとも以下の(a)～(c)を含む。

[0022] (a) 取り出された種子に対して光を照射してスペクトルデータおよび／またはその派生スペクトルデータを取得すること。

(b) スペクトルデータを取得した種子を所望の形質に対する品質検査に供し、該取り出された種子の事前良品率を決定すること。尚、(b)における事前良品率は次式で定義される：

$$[\text{事前良品率}(\%)] = [\text{品質評価に供された種子のうち、所望の形質を有していた種子数}] / [\text{品質評価に供された種子数}] \times 100$$

(c) 各種子のスペクトルデータと品質検査の結果とを対応させること。

[0023] 尚、本明細書において、「事前良品率」を「初期適合率」と表記することがある。

[0024] データセットの構築の(a)における、種子に対して光を照射してスペク

トルデータ（反射、吸収、および／または透過スペクトルデータ）および／またはその派生スペクトルデータを取得する方法は、光学的分析の技術分野において一般的に用いられる自体公知の方法を用いることができる。また、種子に照射する光は、スペクトルデータを得られるものであれば特に限定されない。種子に照射する光としては、例えば、マイクロ波、テラヘルツ波、赤外光、可視光、紫外光、X線、およびγ線であり得るが、これらに限定されない。本発明の好ましい一態様において、光は可視光または赤外光であり得る。尚、スペクトルデータの取得は、例えば、本願の実施例で用いた方法や条件のほか、特許第6782408号等で教示される方法を例示することができるがこれらに限定されない。

[0025] また、得られたスペクトルデータの派生スペクトルデータを生成する方法も光学的分析の技術分野における自体公知の方法を用いればよい。例えば、得られたスペクトルデータ（R）に対して、逆数（ $1/R$ ）変換、対数変換、標準正規化、SG（Savitzky-Golay）フィルタによる平滑化および平滑化微分、又はこれらを任意に組み合わせた変換を行うことで、派生スペクトルデータを生成することができる。尚、スペクトルデータおよび派生スペクトルデータは、そのいずれか1つのみを用いてもよいし、2つ以上を用いてもよい。好ましい一態様において、用いられるスペクトルデータはスペクトルデータおよびそれより生成した派生スペクトルデータのうち、最も精度の高い予測モデルを導き得る1つのスペクトルデータであり得る。

[0026] 尚、スペクトルデータを取得するための機器についても、自体公知の機器を用いればよい。例えば、生体組織の分光特性は、光ファイバー式の分光器による点計測が可能なほか、リモートセンサの一種であるハイパースペクトルカメラを用いることにより、座標情報（画像）と併せて計測（測定）することができる（面計測）。計測に際しては、対象とする波長域において高い検出感度を示す機器を適宜選択すればよい。本発明の方法においては、例えば、CCD、CMOS、CQD、InGaAs、HgCdTe（MCT）、Type II 超格子（T2SL）等の光検出器を備えた機器を用いることが

できるが、これらに限定されない。また、スペクトルデータを取得する際に、必要に応じて、反射率補正画像を生成し、種子認識モデルを適用することもできる（図3のST1～3を参照）。

[0027] データセットの構築の（b）における品質検査もまた、所望の形質に対する事前良品率が決定できる限り、特に限定されない。例えば、所望の形質が「発芽能」である場合、スペクトルデータを取得した種子を、例えばISTAが定める国際種子検査規定に準拠した発芽試験やこれに準ずる発芽試験に供することで容易に確認することができる。そして、発芽試験に供した種子数をA（個）、実際に発芽した種子数をB（個）としたとき、[事前良品率（％）（即ち、発芽率）]は、[品質評価に供された種子のうち、発芽した種子数（B個）] / [品質評価に供された種子数（A個）] × 100との定義式から、 $100B/A$ （％）と決定することができる。

[0028] データセットの構築の（c）における、「各種子のスペクトルデータと品質検査の結果とを対応させること」を、具体的に形質が「発芽能」の場合を用いて説明するに、種子A、種子Bおよび種子Cに対して光を照射して得られた各種子のスペクトルデータ（反射、吸収、若しくは透過スペクトルデータまたはその派生スペクトルデータ）を、それぞれ、SA、SB、SCとし、また、当該種子A、種子Bおよび種子Cを品質検査に供することにより種子Aが正常発芽、種子Bが異常発芽し、種子Cが不発芽であるとの検査結果を得た場合に、（c）においては、SAは正常発芽、SBは異常発芽、SCは不発芽といったように、スペクトルデータと品質結果とを対応させることが意図される。なお、この際に、良否の判別を容易にする目的で、ダミー変数を充ててもよい。例えば、「発芽能」の場合、正常発芽には1（良好）、異常発芽および不発芽には-1（不良）といった変数を充てることができるが、これに限定されない。

[0029] （第3工程）

本発明の第3工程は、第2工程で構築されたデータセットに対して多変量判別分析を適用することにより、各種子における所望の形質に対する判別ス

コアを算出するための良否判別モデルの候補を導出する工程である。

[0030] 本工程における多変量判別分析は、自体公知の方法を用いればよく、本発明の所望の効果を得られる限り、いずれの方法を用いてもよい。本発明の方法において、多変量統計分析技法としては、部分的最小二乗法（Partial Least Squares；PLS）または主成分分析（Principal component analysis；PCA）を用いることができるがこれらに限定されない。多変量判別分析は、好ましくは部分的最小二乗判別分析（Partial Least Square-Discriminant Analysis；PLS-DA）であり得る。

[0031] 一例として、PLS-DAを用いて多変量判別分析を行う場合を概説する。

[0032] PLS-DAは、設定された説明変数および目的変数を用いて次式1のような式をモデリングするものである。

[0033] （式1）

$$y = a + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4 + \dots$$

[0034] ここで、 x_1 、 x_2 、 x_3 および x_4 は説明変数であり、 y は目的変数であり、 a は切片（定数）であり、 b_1 、 b_2 、 b_3 および b_4 は偏回帰係数（定数）である。

[0035] 第2工程で取得されたスペクトルデータを説明変数とし、品質検査により得られた結果を目的変数とし、各種子における所望の形質に対する判別スコアを算出するための良否判別モデルの候補を導出する。良否判別モデルの候補の導出方法もまた、自体公知の方法を用いればよいが、以下に、本発明の実施例において用いた方法を一例として説明する。

[0036] 種子に光を照射して取得した反射、吸収、若しくは透過スペクトルデータ（R）から、自身（X1）を含めて、上述した変換方法およびその任意の組み合わせから、派生スペクトルデータXsを生成させる。判別分析では、その全派生スペクトルの、説明変数としての使用を試みるが、導出された判別モデルの予測性能が同等の場合は、なるべく単純な派生パターンを採択すれば

よい。

[0037] (良否判別モデルの候補の導出手順)

データセットの最終準備を次のように行う。

反射、吸収、または透過スペクトルを取得し、さらにその派生スペクトルを生成させ、説明変数データを準備する。必要に応じて、目的変数および説明変数のいずれか又は両方に対して標準化を行う。多変量判別分析により判別モデルの導出を行う。例えば、線形スパースモデリングによる判別モデルの導出は次のように行うことができる。目的変数と説明変数データのすべての組み合わせについて、最初にPLS回帰やRidge回帰により、各説明変数の重みの初期値を決定する。続いてAdaptive LASSO回帰により、正則化係数 λ と偏回帰係数、及び偏回帰係数が0とならない説明変数の個数の関係を表す解パスを算定する。回帰における残差平方和を最小化する説明変数の個数を p とすると、説明変数を p 個以下に削減する場合に適した説明変数の組み合わせの候補は、解パスより決定できる。最終的な判別モデルはPLS-DAにより導出する。Adaptive LASSO回帰における解パスより決定した、 $2 \sim p$ 個の説明変数を使用して判別モデルを導出する。導出された判別モデルを良否判別モデルの候補とする。尚、この段階では多数の候補判別モデルが提示される。良種子に対して、より高い相対スコアを与える判別モデルの選択は、次の工程4において、実施される。

[0038] (第4工程)

本発明の第4工程は、第3工程で導出された良否判別モデルの候補から判別モデルを選択する工程である。

[0039] 第4工程における、「回収率」、「適合率」、「標準選別条件」は、以下のように定義される。

[0040] [回収率(%)] = [品質評価に供された種子のうち、判別スコアに基づいて選択される種子数] / [品質評価に供された種子数] × 100

[0041] [適合率(%)] = [品質評価に供された種子であって、判別スコアに基づいて選択される種子のうち、所望の形質を有している種子数] / [品質評価

に供された種子のうち、判別スコアに基づいて選択される種子数] × 100

[0042] [標準選別条件] : [事前良品率] と [回収率] を同値とする選別条件

[0043] 回収率、適合率および標準選別条件を上述のとおり定義した時、適合率が事前良品率を上回るとの基準を満たす候補を良否判別モデルとして選択することで、第3工程で得られた多数のモデル候補から、所望の形質に対して良否を判別し得る判別モデルを選択することができる。なお、適合率が事前良品率を大きく上回るほど、優れた判別モデルであると考えられる。また、適合率を必要に応じて簡便に調整できることを可能とする目的で、回収率の増減と適合率の増減とに高い順位相関（即ち、回収率を減少させると適合率が確実に増加し、回収率を増加させると適合率が確実に減少する関係）を有する判別モデルを選択することが好ましい。

[0044] 尚、本明細書において、「適合率」を「事後良品率」と表記することがある。

[0045] (第5工程)

本発明の第5工程は、第1工程で種子集団から取り出されなかった残りの種子、および／または該種子と実質的に同一の条件下で得られた別の種子に対して光を照射してスペクトルデータを取得し、該スペクトルデータに第4工程で選択された判別モデルを適用することで各種子の判別スコアを決定する工程である。換言すれば、第5工程は、種子集団のうち、非訓練種子に対してのスペクトルデータを取得し、これに第4工程で得られた判別モデルを適用することで、各種子の判別スコアを決定する工程である。スペクトルデータの取得方法や条件などは、訓練種子において使用したものと同一ものを用いればよく、また、所望の効果を得られる限り、部分的に異なってもよい。

[0046] (第6工程)

本発明の第6工程は、第5工程で決定された判別スコアに基づいて、判別スコアの閾値を決定する工程である。閾値は、種子集団における事前良品率が「k %」であったとすると、上位から「k %」に位置するスコアが標準的

な選別を行うための閾値（標準閾値）と設定することができる。尚、上述の第4工程において、回収率の増減と事後良品率の増減とに高い順位相関を有する判別モデルを選択した場合は、標準閾値を基準として当該閾値を適宜増減すれば、適合率を調整することもできる。

[0047] （第7工程）

本発明の第7工程は、第5工程で決定された各種子の判別スコアと第6工程で決定された判別スコアの閾値とを比較することで、所望の形質を有すると予測される種子を回収する工程である。種子の回収は、所望の形質を有すると予測される種子を選別して取り出すことで達成してもよいし、或いは、所望の形質を有しないと予測される種子を選別して取り出し、取り出されなかった種子を、所望の形質を有する種子として回収してもよい。

[0048] 本発明の一態様において、本発明の方法は、2以上の所望の形質を有する植物種子の選別方法であり得る。

[0049] 例えば、2以上の形質としては、先に列挙した形質（即ち、発芽能、発芽勢、遺伝子型、ストレス耐性、休眠性、病害抵抗性、虫害抵抗性、QTL特性、食味性、出穂期、および、葉の大きさなどの形態学的な特徴等）のいずれか2つ以上の組み合わせが挙げられるが、これらに限定されない。

[0050] 本発明の方法が、2以上の所望の形質の選別方法である場合、2以上の所望の形質の予測は、以下の（i）～（i i i）のいずれかの方法によって行われることを特徴とする。

[0051] （i）形質ごとの良否に着目して形質ごとの判別モデルにより個別に種子をスコア化し、すべてのスコアが閾値以上のものを良種子と予測する方法（「複数モデルによる逐次予測」）。

[0052] （i i）形質ごとの良否は顧みず、所望の形質のすべてが良好である場合を良好とし、それ以外の場合は不良とすることで単一の判別モデルを選択し、当該単一の判別モデルにより種子をスコア化し、スコアが閾値以上のものを良種子と予測する方法（「単一モデルによる直接予測」）。

[0053] （i i i）形質ごとの良否に着目して形質ごとの判別モデルを選択した後に

、当該形質ごとの判別モデルを用いて得られたスコアを統合して統合スコアとし、該統合スコアの順位に基づき、各種子の良否を予測する方法（「統合モデルによる直接予測」）。

[0054] いずれの方法を用いても、2以上の所望の形質を有する種子の選別が達成できる。

[0055] また、本発明の一態様において、本発明の方法は、判別モデルの精度を評価する工程をさらに含んでもよい。尚、精度評価とは、選択した判別モデルの「汎化的な性能」を評価するものである。良否判別モデルの精度評価は、具体的には以下のような手法により実施することができる。

[0056] 上述した工程の実施により得られた判別モデルについて、その導出時に使用した内部データ（即ち、データセットを構築した種子集団）に対して適用した際の内的妥当性検証および当該モデルの導出とは独立に取得した外部データに対して適用した際の外的妥当性検証を実施する。

[0057] スコア下限閾の設定値に依存しない、良否判別モデルの全般的な精度の評価には、ROC (Receiver Operating Characteristic) 曲線や、PR (Precision-Recall) 曲線が用いられ得る。とりわけ、誤判別のうち、良品の棄却 (False Negative、偽陰性) よりも不良品の採択 (False Positive、偽陽性) を深刻視する品質検査においては、偽陽性に対して鋭敏な、PR曲線が用いられることが多い。

[0058] ROC曲線、PR曲線のいずれにおいても、AUC (Area Under Curve, 曲線下面積) が最大値である1に近い値を取る判別モデルほど優れているとされる。しかし、ROC-AUCでは、最小値が必ず0.5であるのに対し、PR-AUCでは、初期適合率、すなわち事前良品率が最小値となり、適用先のデータセットにより変動する。

[0059] PR-AUCは、特定のデータセットに対し、数ある判別モデルのいずれが優れた性能を示すかを検証する目的には適しているが、単一の判別モデルが異なるデータセットに対し、同等の性能を発揮するかを検証する目的には適さない。これは、事前良品率に影響を受けてしまうためである。

[0060] P R 曲線における前記の欠点を補完するため、次式により表される相対適合率 (Relative Precision) を定義する。

[0061] [相対適合率] = (適合率 - 初期適合率) / (1 - 初期適合率)

[0062] 相対適合率と再現率の関係を rPR (relative Precision-Recall) 曲線、その曲線下面積を $rPR-AUC$ と称する。 $rPR-AUC$ は、適用先のデータセットに依存せず、必ず閉区間 $[0, 1]$ の値を取る。

[0063] 良否判別において n 個のモデルを使用する場合、適合率や相対適合率と再現率との関係は、 $n + 1$ 次元空間を占める超曲面の形状を成すこととなる。ここでは便宜上、使用する判別モデルの数によらず、適合率及び相対適合率と再現率との関係は PR 曲線及び rPR 曲線、それらが成す超曲面より下部の比率は $PR-AUC$ 及び $rPR-AUC$ と称する。

[0064] 導出したすべての判別モデルについて、それぞれが対象とする作物種に由来するデータセットに適用した際の PR 曲線及び rPR 曲線を作成する。また、各曲線から $PR-AUC$ 及び $rPR-AUC$ を算出する。作成された曲線の形状や AUC 値から、判別モデルの精度評価が達成される。

[0065] 以下の実施例において本発明を更に具体的に説明するが、本発明はこれらの例によってなんら限定されるものではない。

実施例

[0066] [出荷要件を満たす品質を備えた野菜種子の判別]

(材料) 供試資料として、トキタ種苗株式会社の提供による、2015～2022年産の野菜種子（カボチャ、エンドウ、レタス、ネギ、トマト、カリフラワー）を用いた（表1）。

[0067]

[表1]

作物種	科・属	品種番号	採種地	採種年	ロット記号	データセット	派生判別モデル	判別対象	品質内容				
									良	不良	計	事前良品率 (初期適合率)	
カボチャ (<i>C. moschata</i> s. <i>moschata</i>)	ウリ科 カボチャ属	5	中国	2020	A	dc-Sq5A	m-Sq5	発芽	194	17	29	240	80.8%
									128	25	31	184	69.6%
エンドウ (<i>P. sativum</i>)	マメ科 エンドウ属	19	日本 (愛媛)	2022	A	dc-Pe19A	m-Pe19	発芽	259	47	93	399	64.9%
									156	28	129	313	49.8%
レタス (<i>L. sativum</i>)	キク科 マキノノグサ属	803	中国	2020	A	dc-Le803A	m-Le803	発芽	345	58	8	411	83.9%
									515	63	19	597	86.3%
ネギ (<i>A. fistulosum</i>)	ヒガンバナ科 ネギ属	22	フランス	2021	A	dc-Bo22A	m-Bo22	発芽	363	58	58	479	75.8%
									337	77	96	510	66.1%
トウモロコシ (<i>Z. mays</i>)	マメ科 トウモロコシ属	51	トルコ	2018	A	dc-Bo51A	m-Bo51	発芽	380	71	19	480	81.3%
									254	16	66	336	75.0%
トウモロコシ (<i>Z. mays</i>)	マメ科 トウモロコシ属	94	タイ	2015	B	dc-Im94B	m-Im94	発芽	272	29	35	336	81.0%
									457	50	125	332	47.3%
カボチャ (<i>C. moschata</i> s. <i>moschata</i>)	ウリ科 カボチャ属	47	中国	2017	A	dc-Ca47A	m-Ca47	発芽・交配	415	46	11	472	87.9%
									351	91	42	442	79.4%
カボチャ (<i>C. moschata</i> s. <i>moschata</i>)	ウリ科 カボチャ属	47	中国	2017	A	dc-Ca47A	m-Ca47	発芽・交配	329	143	472	472	69.7%
									494	52	27	573	86.2%
カボチャ (<i>C. moschata</i> s. <i>moschata</i>)	ウリ科 カボチャ属	47	中国	2017	A	dc-Ca47A	m-Ca47	発芽・交配	418	120	38	576	77.7%
									382	191	573	573	66.7%
カボチャ (<i>C. moschata</i> s. <i>moschata</i>)	ウリ科 カボチャ属	47	中国	2020	B	dc-Ca47B	m-Ca47	発芽・交配	413	94	39	546	75.6%
									404	75	479	479	84.3%
カボチャ (<i>C. moschata</i> s. <i>moschata</i>)	ウリ科 カボチャ属	47	中国	2020	B	dc-Ca47B	m-Ca47	発芽・交配	342	204	546	546	62.6%
									32428	13107	65535	65535	80.0%
カボチャ (<i>C. moschata</i> s. <i>moschata</i>)	ウリ科 カボチャ属	47	中国	2020	B	dc-Ca47B	m-Ca47	発芽・交配	32428	13107	65535	65535	80.0%
									32428	13107	65535	65535	80.0%

[0068] 種子に対する近赤外光の照射は、24V 250Wのアルミミラー付ハロゲンランプ（河北ライティングソリューションズ社製JTR24V250W1

OH/5-AL、GX5.3口金)2灯への直流電圧印加により行い、近赤外ハイパースペクトル画像は、ラインスキャン型ハイパースペクトルカメラ(住友電気社製CV-N801HS)を用いて撮影した。近赤外レンズは、焦点距離が30mmの像側テレセントリックレンズ(住友電気社製)であり、撮影時の作動距離は28cm、空間分解能は90ppi、波長サンプリング間隔は6nmであった。また、使用したカメラの波長感度域は980~2,350nmであった。各波長における反射率は、コントラストターゲット(米Labsphere社製SRT-MS-050)を構成する、反射率が99%の標準反射板及び暗電流(反射率0%)の撮影画像をもとに、拡散反射率相当に校正した。種子の反射スペクトルは、各種子が占める領域に含まれる画素ごとに記録された反射スペクトルの平均を使用した。

[0069] 種子の外観は、白色LED光源(Lemac社製IDBA-HMS150WHV-S)の照射下において、8kカラーラインスキャンカメラ(e2v社製EV71C4CCL8005-BA0)を使用して撮影した。レンズは光学倍率が0.4倍の物体側テレセントリックレンズ(オプトアート社製FT04-150CL)であり、空間分解能は2,032ppiであった。カメラのホワイトバランス調整には、コントラストターゲット(前記SRT-MS-050)を構成する反射率が25%(種子表面が暗色の場合)又は50%(種子表面が明色の場合)の標準反射板を使用した。

[0070] (方法)

1. データセットの構築

(1) 種子に対し、ハロゲンランプより光を均一に照射し、近赤外波長域の分光情報を含む近赤外ハイパースペクトル画像を撮影した。種子の一部については、8kラインカメラにより、高解像度の外観画像も併せて撮影した(図1)。

[0071] (2) 近赤外ハイパースペクトル画像から、各種子が占める領域の近赤外反射スペクトルを算出した(図2)。

[0072] (3) 前記の種子は、ISTA(International Seed Testing Association

）が定める国際種子検査規定に準拠した発芽試験に供し、それぞれの発芽形質を正常発芽、異常発芽、不発芽の3群に区分した。

[0073] (4) カリフラワー 品種47 (以下、品種番号は省略する) は、F1雑種として利用されるが、不完全な自家不和合性により、自殖種子が形成される場合がある。それぞれの種子の交配形質がF1、自殖のいずれであったかについては、芽生えを用いたアイソザイム型解析により決定した。

[0074] (5) 良否それぞれの種子にダミー変数を充て、これを後述の判別分析における目的変数とした。ダミー変数の値は、発芽形質の判別では、正常発芽種子を1 (良好)、それ以外を-1 (不良、以下同様) とした。カリフラワーにおける交配形質の判別では、F1雑種を1、それ以外を-1とした。また、発芽及び交配形質が揃って良種子の要件を満たしているかの判別では、正常発芽能を持つF1雑種を1、それ以外を-1とした。

[0075] II. 野菜種子の良否判別モデルの導出

(6) 種子の近赤外反射スペクトルから、それが所望の形質を備えているか (すべての作物種において正常発芽すること、カリフラワーではさらにF1雑種であること) を予測するための判別分析を行った。

[0076] (7) 判別分析の手順は、特開2023-125301に記載の手法に準拠した。本発明においては、近赤外反射スペクトルの平滑化微分や、異なる微分次数のスペクトルの列結合は省略した (図3)。

[0077] (8) 説明変数となる近赤外反射スペクトル (R) には、逆数 ($1/R$) 及び対数 (疑似吸光度、 $-\log R$) 変換のほか、標準正規化 (SNV, Standard Normal Variate; 有効波長範囲は980~2,200 nm)、SG (Savitzky-Golay) フィルタによる平滑化 (5点, 3次多項式近似) を任意の組み合わせで施し、11通りの派生スペクトルを生成した。原スペクトルを含む計12通りのスペクトルを、説明変数として使用した (図4)。

[0078] (9) 種子の良否判別モデルは、Adaptive LASSO (Least Absolute Shrinkage and Selection Operator) 及びPLS-DA (Partial Least Squares-Discriminant Analysis) を組み合わせた多変量線形スパースモ

デリングの手法により導出した。この段階では多数の候補モデルが提示されるが、良種子に対し、より高い相対スコアを与えるものを選び出す必要がある。以下にモデル選択の方法を具体的に示す。

[0079] III. 良否判別モデルの選択

(10) 前記において提示された多数の候補の中から、望ましい判別モデルを選択する手法を考案するため、以下の思考実験を行った。

[0080] 1) 良否それぞれの種子に対し、明確に離れたスコアを与える高精度な判別モデルと、幅広く重なるスコアを与える低精度な判別モデルを仮定する。ここでは、事前良品率を80%とし、良否それぞれの種子に対し、前者が平均1及び-1、標準偏差が共に0.25の正規分布に従うスコアを与え、後者が平均0.05及び-0.05、標準偏差が共に0.25の正規分布に従うスコアを与えるものとする(図5)。

[0081] 2) スコアの降順に種子を回収する選種作業を行うものとする。高精度な判別モデルを使用した場合には、回収率が事前良品率と一致する80%に至るまで、適合率(事後良品率)は100%近傍を維持する(図5A(e))。一方、低精度な判別モデルを使用した場合には、適合率は速やかに事前良品率に収束する(図5B(e))。いずれの場合も、回収率が事前良品率と一致した時点で、適合率と再現率(良種子の回収率)は同値となる。

[0082] 3) 判別モデルを利用した良種子の選別では、全ての種子の計測を終えた後に、スコアが上位を占める種子を回収するのではなく、実用上は予め定めた下限閾を上回るスコアを示す種子を順次回収することとなる。スコアとその順位の関係は容易に求められるため、適合率や再現率を含む判別精度にかかわる各指標と回収率との関係(図5(e))は、判別スコアとの関係(図5(d))に描き替えることができる。

[0083] 4) ここで、事前良品率と回収率を同値とする選別条件を「標準選別条件」と定義し、且つ、その際に回収の対象となる種子の判別スコアの下限閾を「標準閾値」と定義する。

[0084] 前記より、標準選別条件を用いた選別を行ったときに、適合率(即ち、事

後良品率) 及び、それと同値となる再現率が、事前良品率を大きく上回ることが、優れた判別モデルの要件であることが分かる。

[0085] (11) 加えて、良種子の選別では、適合率を必要に応じて調整できることが望ましい。すなわち、回収率を下げる(スコアの下限閾を上げる)に従い、適合率を高確度に上げられる関係が成り立っていることがより好ましい。

[0086] (12) 以下の表2-1、表2-2及び表2-3に示した各作物種子に対する良否判別モデルは、前記2点を満たすことを要件として選択したものである。また、同表の各モデルにおいて、必要な説明変数として採択された波長帯と、標準偏回帰係数及び変数重要度(VIP, Variable Importance in Projection)との関係を図6に示した。

[0087]

[表2-1]

判別モデル		m-Sq5		m-Jc19		m-Ic803		m-Bo22	
対象作物種		カボチャ		エンドウ		レタス		ネギ	
品種番号		5		19		803		22	
判別対象		発芽		発芽		発芽		発芽	
スペクトル x		R^{-1}		$-\log R$		R^{-1}		R	
前処理	平滑化	+		-		-		-	
	SNV変換	-		-		-		-	
判別式		切片 a		-0.1650		1.6803		0.7284	
[スコア y]	波長帯	λ (nm)	PRC b	λ (nm)	PRC b	λ (nm)	PRC b	λ (nm)	PRC b
	$y = a + \sum_{i=1}^n b_i x_i$								
	1	985.2	-17.285	979.1	13.745	979.1	-1.9022	979.1	-3.3260
	2	1040.5	-12.926	1114.1	-35.776	991.4	1.7534	985.2	-21.909
	3	1108.0	73.895	1401.9	34.653	1009.8	-3.1212	991.4	-26.860
	4	1132.5	71.535	1408.0	86.219	1052.8	-3.2869	997.5	-7.6271
	5	1267.3	-97.841	1524.1	30.978	1058.9	-1.6954	1034.4	68.075
	6	1346.8	-105.98	1664.4	-215.16	1095.7	-0.0826	1273.4	-25.176
	7	1420.2	59.458	1701.0	162.91	1261.1	2.7192	1432.4	-30.897
	8	1426.3	61.298	1859.4	-22.545	1291.8	3.7851	1579.0	20.989
	9	1511.9	-27.860	1871.6	-27.074	1334.6	2.2499	1609.5	12.900
	10	1640.0	-108.35	1895.9	-85.158	1408.0	2.2381	1621.7	33.270
	11	1682.7	58.364	1938.5	40.669	1432.4	1.8627	1646.1	35.181
	12	1688.8	63.131	1950.6	11.127	1456.9	1.8168	1682.7	61.441
	13	1719.3	-43.763	1956.7	-29.433	1542.4	-1.2534	1719.3	0.1395
	14	1725.4	-57.638	1968.9	-48.051	1572.9	-2.4464	1731.5	0.0355
	15	1786.3	6.3219	1981.0	-24.417	1627.8	-0.6800	1737.6	4.1772
	16	1810.7	39.016	2011.4	5.1748	1682.7	-1.4694	1743.7	28.052
	17	1841.1	24.196	2017.5	19.607	1737.6	-1.7589	1749.8	44.771
	18	1865.5	17.660	2029.7	50.729	1743.7	-1.7105	1762.0	-13.721
	19	1871.6	37.848	2035.7	46.753	1877.6	-3.4651	1774.1	-20.582
	20	1920.2	-24.555	2041.8	-1.4392	1968.9	1.2489	1798.5	-23.735
	21	1950.6	16.076	2060.0	8.2482	2017.5	-0.2014	1804.6	-24.590
	22	1968.9	22.535	2066.1	15.190	2041.8	-2.0604	1822.9	-31.388
	23	2023.6	-44.804	2084.3	-25.130	2066.1	0.6819	1829.0	-45.526
	24	2047.9	-47.988	2090.4	-46.696	2096.5	-1.3287	1835.0	-36.668
	25	2090.4	-11.892	2096.5	39.109	2114.7	0.5566	1956.7	19.528
	26	2102.5	40.137	2175.4	8.2341	2126.8	0.6221	1999.3	36.552
	27	2126.8	5.5655	2187.5	-13.881	2132.9	0.6588	2054.0	3.9937
	28	2151.1	-22.296			2145.0	0.4663	2066.1	-69.468
	29	2187.5	20.376			2163.2	0.2316	2096.5	-38.050
	30					2181.4	0.6626	2157.2	12.418
	31					2193.6	0.1541	2169.3	9.1979
	32					2199.6	0.4806	2181.4	31.197
	33							2187.5	-1.4424
	34							2193.6	17.216
	35							2199.6	-10.033

[0088]

[表2-2]

判別モデル		m-Tm94		m-Ca47g		m-Ca47x		m-Ca47	
対象作物種		トマト		カリフラワー		カリフラワー		カリフラワー	
品種番号		94		47		47		47	
判別対象		発芽		発芽		交配		発芽・交配	
スペクトル x		$-\log R$		$-\log R$		$-\log R$		R^{-1}	
前処理	平滑化	+		+		+		+	
	SNV変換	-		+		+		-	
判別式		切片 a		-32.816		-5.7311		9.7297	
[スコア y] $= a + \sum_k b_k x_k$	波長帯	λ (nm)	PRC b	λ (nm)	PRC b	λ (nm)	PRC b	λ (nm)	PRC b
	$i=1$	979.1	32.439	1108.0	-2.9088	1083.4	-4.0739	1015.9	9.2786
	2	991.4	-61.744	1114.1	-4.8038	1089.6	-4.6297	1077.3	-12.817
	3	997.5	35.689	1138.6	-0.8299	1138.6	-4.7749	1114.1	-13.515
	4	1028.2	-11.282	1163.1	-10.570	1144.7	-3.3836	1138.6	-8.1682
	5	1077.3	-61.183	1212.2	-9.1599	1163.1	-7.2185	1163.1	-30.373
	6	1163.1	77.499	1273.4	1.4640	1187.6	-4.9512	1206.0	25.235
	7	1285.6	-14.395	1383.5	9.1878	1206.0	2.9174	1218.3	5.4793
	8	1328.5	83.516	1389.6	1.7531	1218.3	4.4311	1322.4	17.676
	9	1377.4	-100.36	1505.8	5.2224	1230.5	8.8339	1328.5	10.189
	10	1444.7	40.109	1511.9	5.7933	1255.0	8.7222	1395.8	-14.470
	11	1456.9	51.721	1518.0	6.0745	1389.6	-10.480	1432.4	-16.855
	12	1481.3	42.881	1542.4	4.5440	1395.8	-12.589	1499.6	7.7883
	13	1591.2	-55.641	1615.6	-12.465	1444.7	-4.3872	1511.9	18.149
	14	1597.3	-79.390	1640.0	-13.301	1658.3	3.7392	1615.6	-31.707
	15	1615.6	-91.241	1646.1	-13.122	1694.9	-18.320	1664.4	19.798
	16	1652.2	51.800	1688.8	-7.0869	1841.1	4.6177	1688.8	-20.324
	17	1658.3	62.580	1707.1	-2.1428	1908.1	9.7230	1694.9	-12.651
	18	1682.7	87.257	1835.0	6.3120	1944.6	-0.4903	1707.1	-22.506
	19	1719.3	-6.817	1859.4	6.1748	1987.1	-0.7094	1737.6	7.4684
	20	1841.1	-81.024	1908.1	-2.4662	1999.3	-11.946	1822.9	7.7184
	21	1853.3	-54.645	1968.9	-1.8750	2023.6	-4.3224	1841.1	4.4263
	22	1859.4	22.189	1975.0	-1.5667	2047.9	0.1108	1908.1	5.4179
	23	1920.2	-26.301	2078.3	2.2857	2090.4	-4.6157	1950.6	-1.8920
	24	1975.0	12.137	2096.5	2.9516	2168.6	-2.5500	1968.9	-4.4392
	25	1999.3	31.550	2102.5	1.7054	2163.2	2.4813	1987.1	-3.1778
	26	2017.5	-10.549	2120.8	1.5381	2187.5	1.9852	1999.3	-4.4901
	27	2023.6	18.521	2157.2	1.7330			2005.4	-3.0414
	28	2029.7	41.529	2175.4	-2.8884			2090.4	-1.6392
	29	2096.5	-44.815					2108.6	-1.6385
	30	2114.7	27.423					2151.1	-2.8028
	31	2120.8	-24.472					2175.4	-0.6425
	32	2145.0	0.0523						
	33	2187.5	-2.9051						
	34								
	35								

[表2-3]

判別モデル		m-BoS1		m-Tm221	
対象作物種		ネギ		トマト	
品種番号		51		221	
判別対象		発芽		発芽	
スペクトル x		$-\log R$		R	
前処理	平滑化	+		+	
	SNV 変換	+		+	
判別式		切片 a		-2.2358	
[スコア y] $= a + \sum_i b_i x_i$	波長帯	λ (nm)	PRC b	λ (nm)	PRC b
	$i=1$	991.4	-0.3200	979.1	-27.475
	2	997.5	-1.5337	991.4	-7.2701
	3	1009.8	-1.2544	1046.6	60.754
	4	1034.4	2.7833	1095.7	59.207
	5	1175.4	-6.0212	1187.6	-36.608
	6	1401.9	12.811	1193.8	-47.203
	7	1481.3	16.716	1199.9	-43.470
	8	1487.4	-8.7768	1218.3	-60.465
	9	1518.0	-12.355	1279.5	60.087
	10	1688.8	-7.8318	1414.1	-20.665
	11	1694.9	-7.2889	1524.1	-86.194
	12	1774.1	6.6653	1560.7	95.957
	13	1786.3	4.1531	1566.8	75.772
	14	1859.4	9.2668	1646.1	-61.118
	15	1889.8	2.7611	1682.7	-93.961
	16	2181.4	-0.0746	1707.1	53.825
	17			1731.5	66.247
	18			1847.2	81.463
	19			1908.1	150.57
	20			1956.7	-28.889
	21			1968.9	-3.8018
	22			1975.0	-30.586
	23			1999.3	-157.58
	24			2017.5	-92.367
	25			2041.8	-41.259
	26			2090.4	78.075
	27			2096.5	10.621
	28			2102.5	43.024
	29			2175.4	-0.8358
	30				
	31				
	32				
	33				
	34				
	35				

[0090] I V. 複数の判別スコアの統合処理

複数の形質が評価される実施形態をこれ以降説明する。

[0091] (13) カリフラワーにおける良種子を(1) 正常な発芽能を有し、且つ(2) F1雑種と定義したとき、カリフラワーの良種子を判別するには、以下2通りの方法を採用し得る。

[0092] 方法1) 発芽及び交配形質における良否を、2つのモデル($m-Ca47g$ 及び $m-Ca47x$)により個別にスコア化し、両スコアが共に上位のものを良種子と予測する。方法2) 個々の形質における良否は顧みず、良種子であるかどうかのみを予測する($m-Ca47$)。

[0093] (14) 方法1)において、希望の判別精度を得るためには、2つのモデルに対してスコアの下限閾を設定する必要がある。 n 個の判別モデルを使用する場合、スコア下限閾の設定値と判別精度にかかわる各指標との関係は、 $n+1$ 次元の超曲面の形状を成すこととなる。本例のように $n=2$ の場合は、各指標は曲面をなすが、ヒートマップとして平面的に描くこともできる(図7)。しかし、本方法は、方法2)に比べ、運用の簡便性は著しく劣る。

[0094] (15) 複数の判別モデルのスコアを統合し、設定すべき下限閾を1つで済むようにするため、以下の数理的処理を案出した。

[0095] 1) 形質1~ k の良否を判別するモデルを $M1 \sim Mk$ 、これらに説明変数 X を代入して得られる判別スコアを $y1(X) \sim yk(X)$ とする。

[0096] 2) 特定のロットを構成する十分数 n (数百程度)の種子を計測し、それぞれに対応する説明変数 $x(x1 \sim xn)$ から得られるスコア $y1(x) \sim yk(x)$ を求める。形質ごとの事前良品率が $p1 \sim pk$ と想定される場合、 $y(y1 \sim yk)$ の上位から $n \times p(p1 \sim pk)$ 番目のスコアを $yp(yp1 \sim ypk)$ とする。また、 $y(y1 \sim yk)$ の標準偏差を $s(s1 \sim sk)$ とする。

[0097] 3) スコア $y(y1 \sim yk)$ を次式によりスコア $z(z1 \sim zk)$ に変換する。

$$z = (y - yp) / s$$

[0098] 4) スコア z ($z_1 \sim z_k$) の最小値を、 $M_1 \sim M_k$ を使用した判別における統合スコアとする。

[0099] 5) 統合スコアの順位に基づき、各種子の良否を判別する。統合スコアを用いた判別においても、標準条件及び標準閾値を決定できる。

[0100] この処理を $m-Ca_47g$ 及び $m-Ca_47x$ に適用し、モデル $m-Ca_47gx$ を生成した (表 3-1 及び表 3-2)。

[0101]

[表3-1]

判別 モデル	判別 対象	データ セット	妥当性検証		判別精度指標								
			種類	内部データ との差異	標準条件:					全値			
					回収率	事後良品率 (適合率)	再現率	正精度	特異度	標準偏差	PR- AUC	F1- AUC	
仮定モデル (スコアのみのみ)													
m-Sq5	発生	de-Sq5A	内部	無	80.8%	93.0%	100.0%	100.0%	100.0%	100.0%	-0.1080	1.0000	1.0000
			外部	相対	80.8%	82.4%	82.4%	71.9%	79.6%	0.1807	0.8529	0.3646	
観測予測													
m-Sq5	発生	de-Sq5A	内部	無	80.8%	93.0%	100.0%	100.0%	100.0%	100.0%	0.2713	0.9731	0.8398
			外部	相対	80.8%	82.4%	82.4%	75.5%	59.8%	0.4962	0.9079	0.6973	
m-Pe19	発生	de-Pe19A	内部	無	64.9%	83.0%	83.0%	77.9%	68.6%	0.4880	0.8904	0.6877	
			外部	相対	64.9%	85.6%	85.6%	85.6%	85.7%	-1.0275	0.8828	0.7663	
m-Lx803	発生	de-Lx803A	内部	無	83.9%	89.6%	89.6%	82.5%	45.5%	0.3492	0.9496	0.6864	
			外部	相対	83.9%	88.4%	88.4%	80.1%	27.4%	0.4333	0.9200	0.4176	
m-Bx22	発生	de-Bx22A	内部	無	75.8%	86.1%	86.1%	78.9%	56.5%	0.2395	0.9071	0.6165	
			外部	相対	75.8%	83.8%	83.8%	78.6%	68.5%	0.6538	0.8884	0.6711	
m-Bx51	発生	de-Bx51A	内部	無	81.3%	82.8%	82.8%	72.1%	25.6%	0.3881	0.8581	0.6884	
			外部	相対	81.3%	86.2%	86.2%	75.1%	41.7%	0.3566	0.8991	0.4635	
m-Bx21	発生	de-Bx21A	内部	無	75.8%	79.9%	79.9%	69.5%	37.1%	0.6701	0.8115	0.2718	
			外部	相対	75.8%	92.9%	92.9%	89.3%	75.0%	0.2340	0.9697	0.8758	
m-Tm04	発生	de-Tm04B	内部	無	81.0%	90.8%	90.8%	85.1%	60.9%	0.4120	0.9295	0.6298	
			外部	相対	81.0%	77.8%	77.8%	78.6%	79.7%	0.5277	0.8297	0.6770	
m-Tm21	発生	de-Tm21A	内部	無	47.3%	86.0%	86.0%	86.7%	87.4%	0.0250	0.9316	0.8703	
			外部	相対	47.3%	85.8%	85.8%	78.6%	56.1%	0.4824	0.9007	0.6011	

[0102]

[表3-2]

判別モデル	判別対象	妥当性検証		判別精度指標							全般		
		データセット	検別	内部データとの差異	基準良品率 (初期適合率)			標準条件				PR-AUC	FPR-AUC
					回復率	事後良品率 (適合率)	正精度	検出率	特異度	標準閾値			
部分予測													
m-Ca47g	基準	de-Ca47Ag	内部	なし	87.9%	95.2%	95.2%	95.2%	64.9%	0.3840	0.9807	0.8399	
		de-Ca47Ag	外部	同ロット	86.2%	93.4%	93.4%	93.4%	58.9%	0.0404	0.9654	0.7487	
		de-Ca47Bg	外部	異ロット	75.6%	92.5%	92.5%	92.5%	76.7%	-0.0424	0.9579	0.8270	
		de-Ca47Ax	内部	なし	79.4%	90.5%	90.5%	90.5%	63.2%	0.2110	0.9645	0.8274	
m-Ca47x	交配	de-Ca47AX	外部	同ロット	77.7%	86.7%	86.7%	79.4%	53.8%	0.2283	0.9342	0.7081	
		de-Ca47Bx	外部	異ロット	84.3%	90.8%	90.8%	90.8%	50.7%	-0.0913	0.9785	0.8656	
特徴モデルによる逐次予測													
m-Ca47g + m-Ca47x	基準 + 交配	de-Ca47A	内部	なし	69.7%	86.3%	86.3%	81.0%	68.5%	0.3840 ~ 0.2110	0.9496	0.8472	
		de-Ca47A	外部	同ロット	66.7%	81.4%	81.4%	76.2%	62.6%	0.0404 ~ 0.2283	0.9331	0.8093	
		de-Ca47B	外部	異ロット	63.6%	82.9%	82.9%	78.7%	71.3%	-0.0424 ~ -0.0913	0.9223	0.6465	
		統一モデルによる逐次予測											
m-Ca47g + m-Ca47x	基準 + 交配	de-Ca47A	内部	なし	69.7%	86.8%	86.8%	81.0%	69.6%	0.1651	0.9302	0.7995	
		de-Ca47A	外部	同ロット	66.7%	81.7%	81.7%	75.8%	63.4%	0.2537	0.8883	0.6648	
		de-Ca47B	外部	異ロット	62.6%	80.6%	80.6%	75.0%	67.4%	0.1549	0.8609	0.6254	
統合モデルによる逐次予測													
m-Ca47g + m-Ca47x	基準 + 交配	de-Ca47A	内部	なし	69.7%	86.5%	86.5%	81.3%	68.9%	0.0015	0.9315	0.7739	
		de-Ca47A	外部	同ロット	66.7%	83.0%	83.0%	77.3%	66.0%	0.0321	0.9081	0.7244	
		de-Ca47B	外部	異ロット	62.6%	82.9%	82.9%	78.0%	71.3%	0.0043	0.9114	0.7628	

[0103] V. 良否判別モデルの精度評価

(16) 選択した全ての判別モデルについて、全種子又は良否別種子の判別スコアの分布域及び、スコア下限閾の設定値又は種子回収率と判別精度にか

かわる各指標との関係を、図 8 及び図 9 に示した。図の構成は図 5 と共通である。

[0104] (17) 図 8 及び図 9 はそれぞれ、各判別モデルを、その導出時に使用した内部データに対して適用した際の内的妥当性検証及び、モデル導出とは独立に取得した外部データに対して適用した際の外的妥当性検証の結果に相当する。図 9 はすなわち、モデルの汎化性能の評価結果である。

[0105] (18) スコア下限閾の設定値に依存しない、良否判別モデルの全般的な精度の評価には、ROC (Receiver Operating Characteristic) 曲線や、PR (Precision-Recall) 曲線が用いられることが多い。とりわけ、誤判別のうち、良品の棄却 (False Negative, 偽陰性) よりも不良品の採択 (False Positive, 偽陽性) を深刻視する品質検査においては、偽陽性に対して鋭敏な、PR 曲線が用いられることが多い。

[0106] (19) ROC 曲線、PR 曲線のいずれにおいても、AUC (Area Under Curve, 曲線下面積) が最大値である 1 に近い値を取る判別モデルほど優れているとされる。しかし、ROC-AUC では、最小値が必ず 0.5 であるのに対し、PR-AUC では、初期適合率、すなわち事前良品率が最小値となり、適用先のデータセットにより変動する。

[0107] (20) PR-AUC は、特定のデータセットに対し、数ある判別モデルのいずれが優れた性能を示すかを検証する目的には適しているが、単一の判別モデルが異なるデータセットに対し、同等の性能を発揮するかを検証する目的には適さない。これは、事前良品率に影響を受けてしまうためである。

[0108] (21) PR 曲線における前記の欠点を補完するため、次式により表される相対適合率 (Relative Precision) を定義する。

[0109] [相対適合率] = (適合率 - 初期適合率) / (1 - 初期適合率)

[0110] (22) 相対適合率と再現率の関係を rPR (relative Precision-Recall) 曲線、その曲線下面積を rPR-AUC と称する。rPR-AUC は、適用先のデータセットに依存せず、必ず閉区間 [0, 1] の値を取る。

[0111] (23) 良否判別において n 個のモデルを使用する場合、適合率や相対適合

率と再現率との関係も、 $n + 1$ 次元の超曲面の形状を成すこととなる。ここでは便宜上、使用する判別モデルの数によらず、適合率及び相対適合率と再現率との関係はPR曲線及びrPR曲線、それらが成す超曲面より下部の比率はPR-AUC及びrPR-AUCと称する。

[0112] (24) 導出したすべての判別モデルについて、それぞれが対象とする作物種に由来するデータセットに適用した際のPR曲線(図10A及び図11)及びrPR曲線(図10B及び図12)を示した。また、各曲線から算出されたPR-AUC及びrPR-AUCを表3-1及び表3-2に併記した。

[0113] VI. 判別モデルに基づく良種子の選別

(25) 種子を撮影した近赤外ハイパースペクトル画像に対し、表2に示した判別モデルによるスコアを、画素又は種子が占有する領域ごとに計算し、疑似色又は輝度値に変換して可視化するソフトウェアを作成した(図13)。

[0114] (26) 前記のソフトウェアを使用し、発芽又は交配形質にかかわる判別スコアが、上位、下位を占める種子をそれぞれ回収した。各区分について、種子の発芽及び初期成長を評価すると共に(図14)、カリフラワーについてはアイソザイム型を解析し、事後良品率(適合率)を求めた(表4)。

[0115]

[表4]

作物種	品種 番号	データ セット	事前製品率 (初期選別率)	判別 モデル	スコア順位	品質内容						事後製品率 (割合率)
						発芽		交配		計		
						正発	不良	正発	不良			
大豆	19	ds-1p19A	49.8%	m-1p19	F1位 20%	76	4	4	-	84	90.4%	
					F1位 20-30%	59	6	6	-	59	71	83.0%
					F1位 30-40%	10	9	41	-	19	60	16.7%
米	22	ds-1p22A	66.1%	m-1p22	F1位 20%	11	18	80	-	11	16.9%	
					F1位 30%	134	16	3	-	134	147	91.2%
					F1位 40%	79	33	79	-	78	211	33.4%
小麦	803	ds-1p803A	86.7%	m-1p803	F1位 20%	173	7	3	-	173	181	95.6%
					F1位 30%	118	26	13	-	118	159	74.2%
					F1位 40% F1位 40%	107	4	0	109	0	106	111
ライム草	47	ds-1p47B	62.6%	m-1p47A	F1位 40% F1位 20%	72	3	0	38	34	74	51.4%
					F1位 20% F1位 60%	16	36	28	49	7	13	82

[0116] (結果・考察)

(1) いずれの作物種においても、発芽試験後に明らかとなった種子の良否を、発芽前の外観及び近赤外反射スペクトルの概形から、主観的に検知する

ことは不可能であった（図 1 及び図 2）。

[0117] （2）導出した判別モデルを、その対象品種に適用した場合、スコアの分布域は例外なく、良種子から不良種子に向けて、高い側から低い側の順序となることが確認された（図 8 並びに図 9（b）及び（c））。

[0118] （3）発芽形質を予測するモデルにおいては、異常発芽種子が示すスコアの分布域は、正常発芽及び不発芽種子が示すスコアの間中に位置しており（図 8 A～H 並びに図 9 A～F（b）及び（c））、判別スコアは、概ね種子に備わる成長力を反映しているものと推察された。

[0119] （4）判別スコアの分布域は、良種子と不良種子の間で重複が生じており、いずれの作物種においても、誤判別を避けることはできなかった。また、トマトの品種 9 4、2 2 1の間では、判別モデルの互換性は確認されたが、同品種に適用した場合と比べて、判別精度が劣ることが示唆された（表 3 並びに図 1 2 H 及び I）。ネギにおいては、品種 2 2 と 5 1 の間で、判別モデルの共用は実用的でないと判断された（表 3 並びに図 1 2 F 及び G）。

[0120] （5）判別スコアの分布域は、同品種かつ同ロットの種子を使用した場合でも、試験の機会が異なれば同一とはならなかった（図 8、図 9（a）～（c）；トマト 品種 9 4 は、図 8 と図 9 で使用ロットが異なる）。導出した判別モデルは、少なくとも同品種に対し、繰り返し適用できるものであったが、良種子を選別する際のスコアの下限閾は、機会ごとに最適化することが望まれた。

[0121] （6）無作為に抽出した種子から得られた近赤外スペクトルに判別モデルを適用し、スコアを算出した時点では、個々の種子の良否は不明である。また、スコアの密度分布は特徴のないひと山型を示すのみであり（図 9 C、EーI（a）；A、B、Dは無作為抽出ではない）、この形状から希望に適った選別を行うためのスコアの下限閾を推定することは困難である。一方、推定される事前良品率と同率の種子を回収するための標準閾値は決定可能であり、これを標準的な選別条件（標準条件）と定めることは合理的である。よって、使用する種子の事前良品率は、可能な限り正確に把握しておくことが望

まれる。

- [0122] (7) 前記の標準条件よりも事後良品率（適合率）を上げたい場合には、判別スコアの下限閾を上げ、再現率（良種子の回収率）を上げたい場合には、スコアの下限閾を下げれば良い。
- [0123] (8) 前記において、判別スコアの下限閾をどの程度変更すれば、適合率や再現率にどの程度の影響が及ぶのかを推定することも容易ではない。スコアから標準閾値を減じ、さらにスコアの標準偏差で除算する変換は、変数の標準化（正規化）と同じく、スコアのばらつきを一律にできる利点をもたらす。このことは、単一のモデルにより良種子を選別する場合においても有効である。
- [0124] (9) 前記のスコア変換により、複数の判別モデルを使用する良種子選別においても、調整すべき閾値を一つに統合することができる。
- [0125] (10) カリフラワーでは、良種子である正常発芽能を備えた F1 雑種を予測するに際し、良否を直接的に判別する単一モデルを使用するよりも、発芽及び交配形質を個別に予測するモデルを組み合わせ使用した場合に、より高い判別精度が認められた（表3並びに図12L及びM）。
- [0126] (11) 発芽形質を予測するモデルを適用し、判別スコアが上位、下位を占める種子について、発芽及び初期成長を比較したところ、スコアが上位の種子は、発芽率が高いのみならず、旺盛な発芽後成長を示す傾向が認められた（図14）。導出した良否判別モデルの性能がやや劣っていたレタスにおいても傾向は同じであった。このことから、発芽形質に対する判別スコアは、種子に備わる成長力を反映しているものと推察された。
- [0127] (12) 標準選別条件における適合率（事後良品率と同義であり、再現率と一致する）が事前良品率を十分に上回ることを確認することにより、多数の候補判別モデルから、好ましいモデルを効率的に絞り込むことができた（表3及び図3）。さらに、相対適合率を定義し、これと再現率との関係を表す rPR 曲線の曲線下面積（ $rPR-AUC$ ）を算出することにより、モデルとデータセットの組み合わせにかかわらず、判別精度を客観的に比較する

ことが可能となった（表 3 及び図 1 2）。

[0128] （1 3）本発明技術では、種子の品質評価試験の結果を受けてから概ね 1 日以内には、その良否判別の可否が明らかとなり、可能な場合は具体的な判別モデルを得ることができる。よって、不良種子の発生が当該品種の遺伝的性質等に起因しない偶発的なものであったとしても、出荷基準に足る良品率を回復させるための選種作業を、問題の発覚から、程なく開始することができる。

[0129] （1 4）前記を可能とするためには、品質評価試験の結果と紐づけできる、種子の近赤外ハイパースペクトル画像を予め取得しておく必要がある。

[0130] （1 5）先述の通り、判別モデルの導出には、多大な時間や労力を要するわけではない。特定の作物品種に適用したモデルを、別の機会に再利用することを考慮して、精度向上のための更新を続けることも可能であるが、不良種子の発生事由は機会ごとに異なる場合が多い。様々な作物品種に対応した判別モデルのライブラリを充実させるだけでなく、問題の発生機会ごとに最適な良否判別モデルを導出し、解決に向けて即応できる種子の品質管理体制を整えることが重要である。

[0131] （1 6）近赤外ハイパースペクトル画像は一度に取得できる一方で、そこから得られるデータは、種子の発芽特性や遺伝子型といった、複数の形質の判別に使用することができる。同画像は、粒径に関する情報を含む上、粒重の推定にも有効である。本発明は、機能の異なる複数の選別機を導入、設置することなく、低コストかつ省スペースに、高度な種子選別を実施する手段を与えるものである。

産業上の利用可能性

[0132] 本発明によれば、所望の形質を有する植物種子と、所望の形質を有さない植物種子の集団から、非破壊的、高効率、且つ高精度に所望の形質を有する植物種子を選別することができる。従って、本発明は、例えば、種苗産業において極めて有用である。

[0133] 本出願は、日本で出願された特願 2 0 2 3 - 0 8 9 0 3 6（出願日：2 0

23年5月30日)を基礎としており、その内容は本明細書に全て包含されるものである。

請求の範囲

[請求項1] 所望の形質を有する植物種子の選別方法であって、以下の第1工程～第7工程を含む、方法：

(第1工程) 種子集団から一部の種子を取り出す工程、

(第2工程) 取り出された種子に対するデータセットを構築する工程であって、該データセットを構築する工程は以下(a)～(c)を含み：

(a) 取り出された種子に対して光を照射してスペクトルデータおよび／またはその派生スペクトルデータを取得すること、

(b) スペクトルデータを取得した種子を所望の形質に対する品質検査に供し、該取り出された種子の事前良品率を決定すること、

ここで、事前良品率は次式で定義される：

〔事前良品率(%)〕＝〔品質評価に供された種子のうち、所望の形質を有していた種子数〕／〔品質評価に供された種子数〕×100

(c) 各種子のスペクトルデータと品質検査の結果とを対応させること、

(第3工程) 第2工程で構築されたデータセットに対して多変量判別分析を適用することにより、各種子における所望の形質に対する判別スコアを算出するための良否判別モデルの候補を導出する工程、

(第4工程) 回収率および適合率を次式で定義し：

〔回収率(%)〕＝〔品質評価に供された種子のうち、判別スコアに基づいて選択される種子数〕／〔品質評価に供された種子数〕×100

〔適合率(%)〕＝〔品質評価に供された種子であって、判別スコアに基づいて選択される種子のうち、所望の形質を有している種子数〕／〔品質評価に供された種子のうち、判別スコアに基づいて選択される種子数〕×100

かつ、事前良品率と回収率を同値とする選別条件を「標準選別条件」

と定義するとき、

該標準選別条件下で、適合率が事前良品率を上回るとの基準を満たす候補を良否判別モデルとして選択することを含む、第3工程で導出された良否判別モデルの候補から判別モデルを選択する工程、

(第5工程) 第1工程で種子集団から取り出されなかった残りの種子、および／または該種子と実質的に同一の条件下で得られた別の種子に対して光を照射してスペクトルデータを取得し、該スペクトルデータに第4工程で選択された判別モデルを適用することで各種子の判別スコアを決定する工程、

(第6工程) 第5工程で決定された判別スコアに基づいて、判別スコアの閾値を決定する工程、

(第7工程) 第5工程で決定された各種子の判別スコアと第6工程で決定された判別スコアの閾値とを比較することで、所望の形質を有すると予測される種子を回収する工程。

[請求項2] 2以上の所望の形質を有する植物種子の選別方法である、請求項1記載の選別方法。

[請求項3] 形質が、発芽能、発芽勢、遺伝子型、ストレス耐性、休眠性、病害抵抗性、虫害抵抗性、QTL特性、食味性、出穂期、および形態学的な特徴からなる群から選択される形質の少なくとも1つである、請求項1または2記載の選別方法。

[請求項4] 2以上の所望の形質の予測が、以下の(i)～(iii)のいずれかの方法によって行われることを特徴とする、請求項1または2記載の選別方法：

(i) 形質ごとの良否に着目して形質ごとの判別モデルにより個別に種子をスコア化し、すべてのスコアが閾値以上のものを良種子と予測する方法、

(ii) 形質ごとの良否は顧みず、所望の形質のすべてが良好である場合を良好とし、それ以外の場合は不良とすることで単一の判別モデ

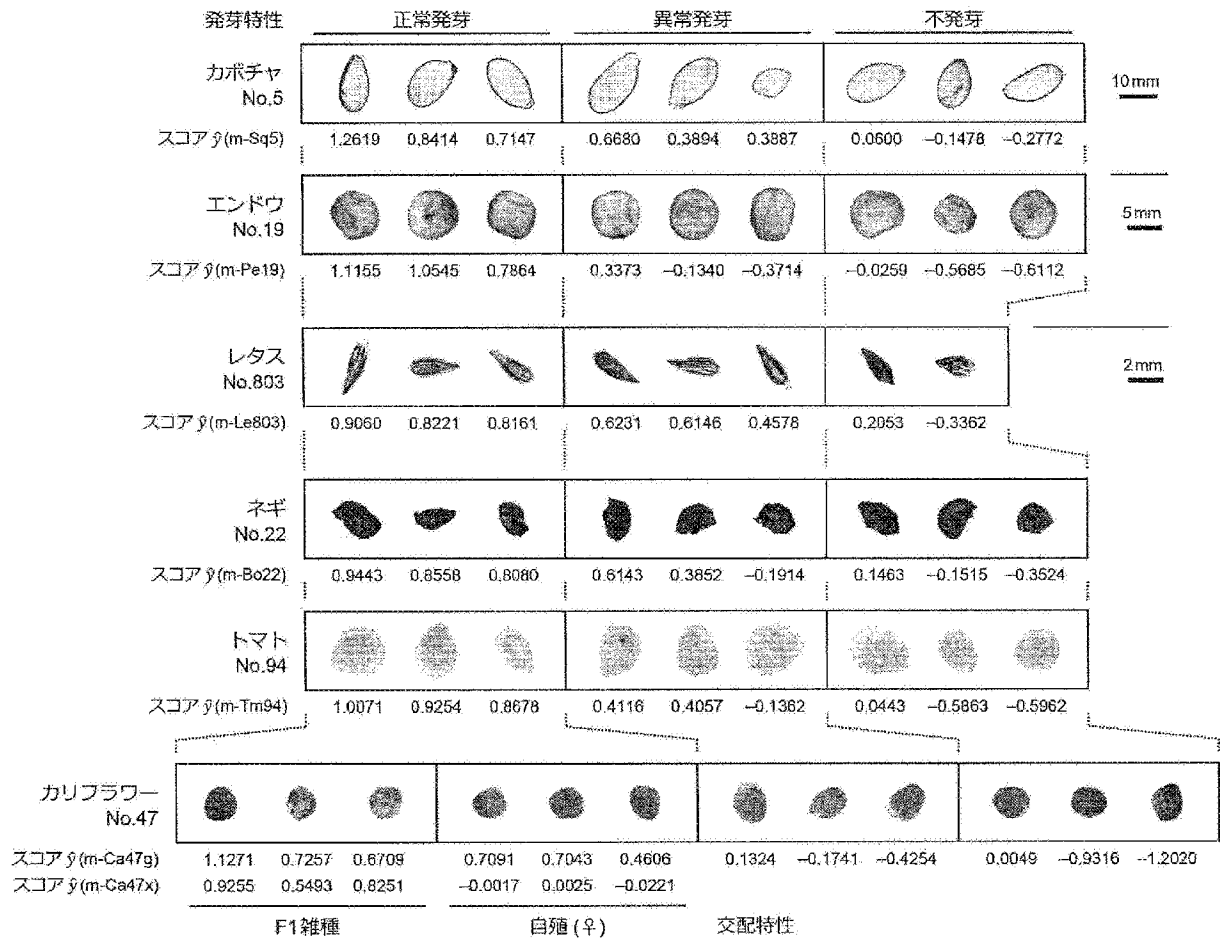
ルを選択し、当該単一の判別モデルにより種子をスコア化し、スコアが閾値以上のものを良種子と予測する方法、

(i i i) 形質ごとの良否に着目して形質ごとの判別モデルを選択した後に、当該形質ごとの判別モデルを用いて得られたスコアを統合して統合スコアとし、該統合スコアの順位に基づき、各種子の良否を予測する方法。

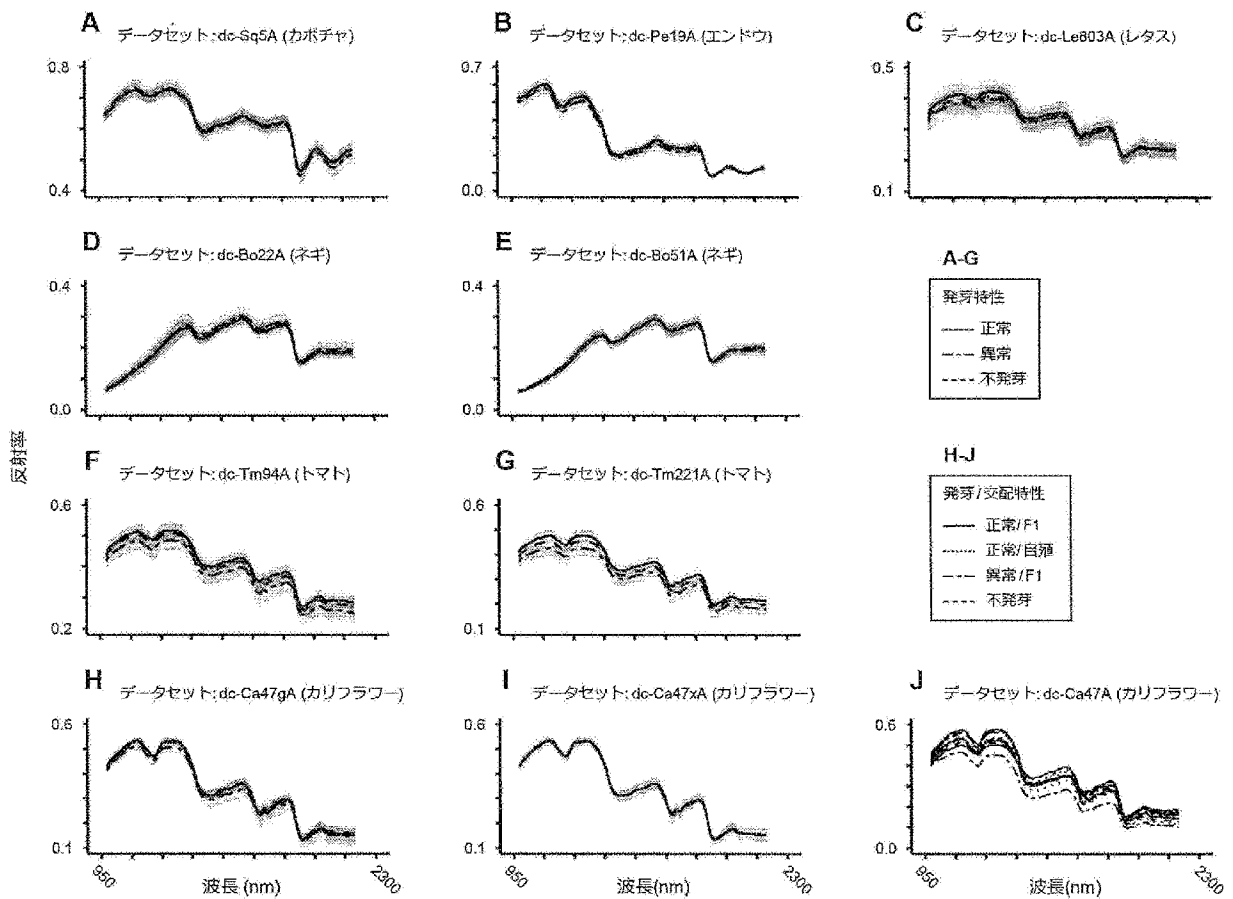
[請求項5] 判別モデルの精度を評価する工程をさらに含む、請求項 1 または 2 記載の選別方法。

[請求項6] 植物が園芸作物である、請求項 1 または 2 記載の選別方法。

[図1]

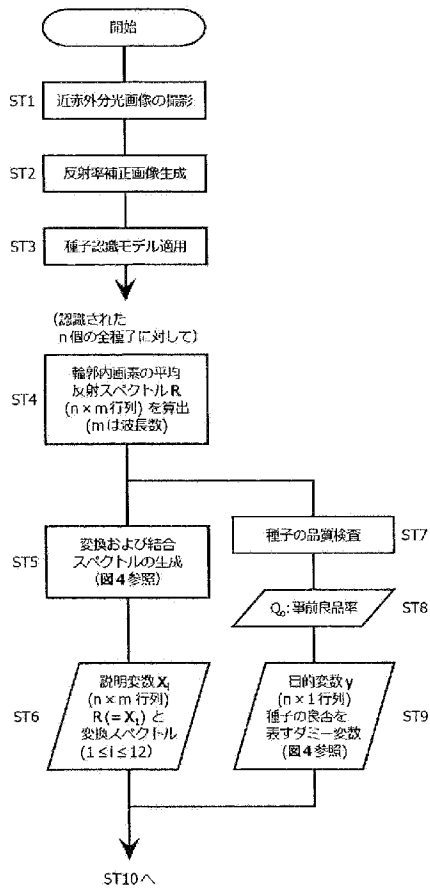


[図2]

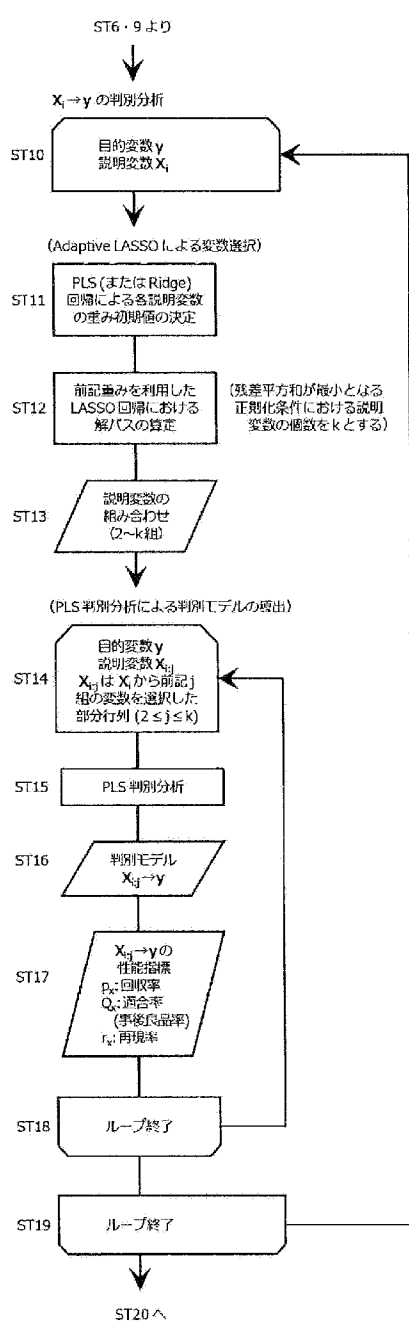


[図3]

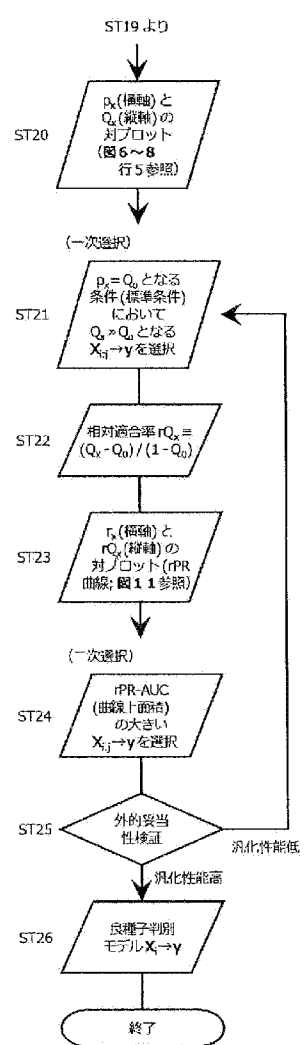
1) データセットの構築フェーズ



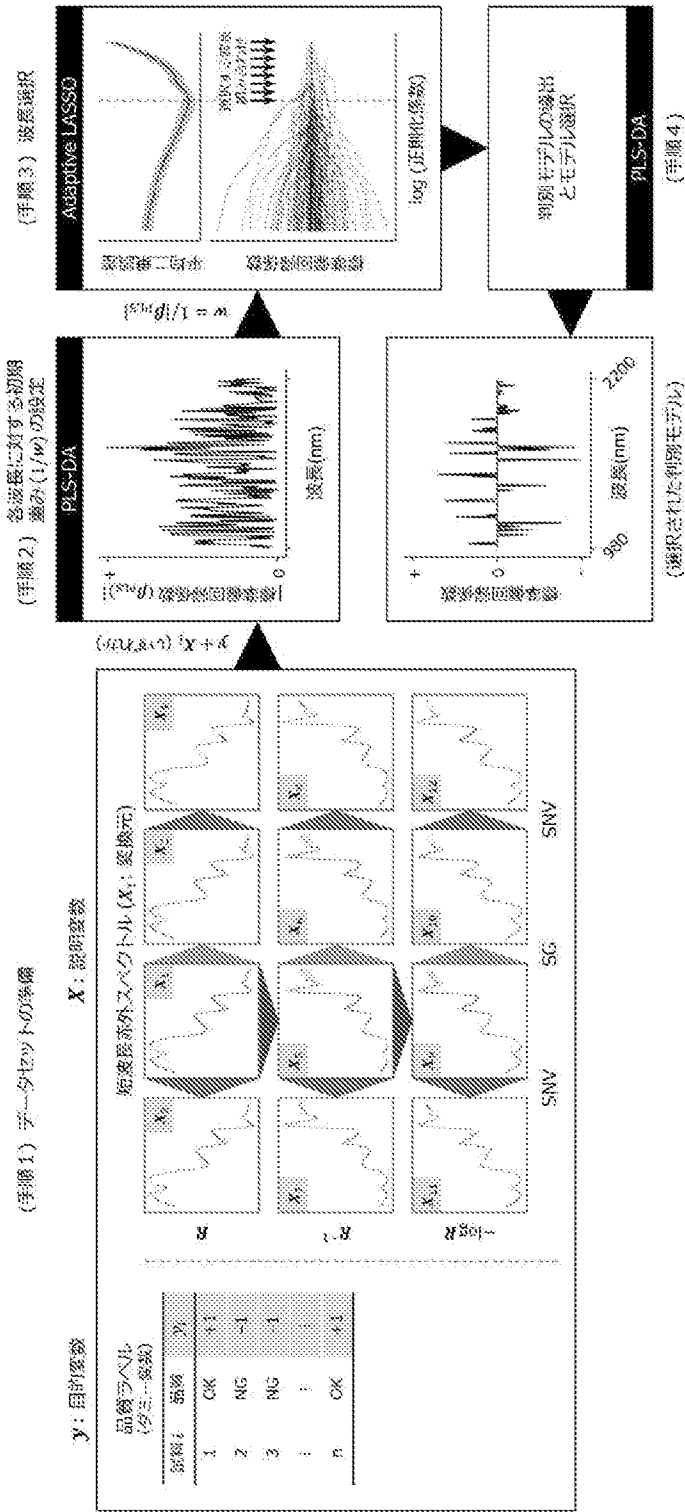
2) 判別モデルの導出フェーズ



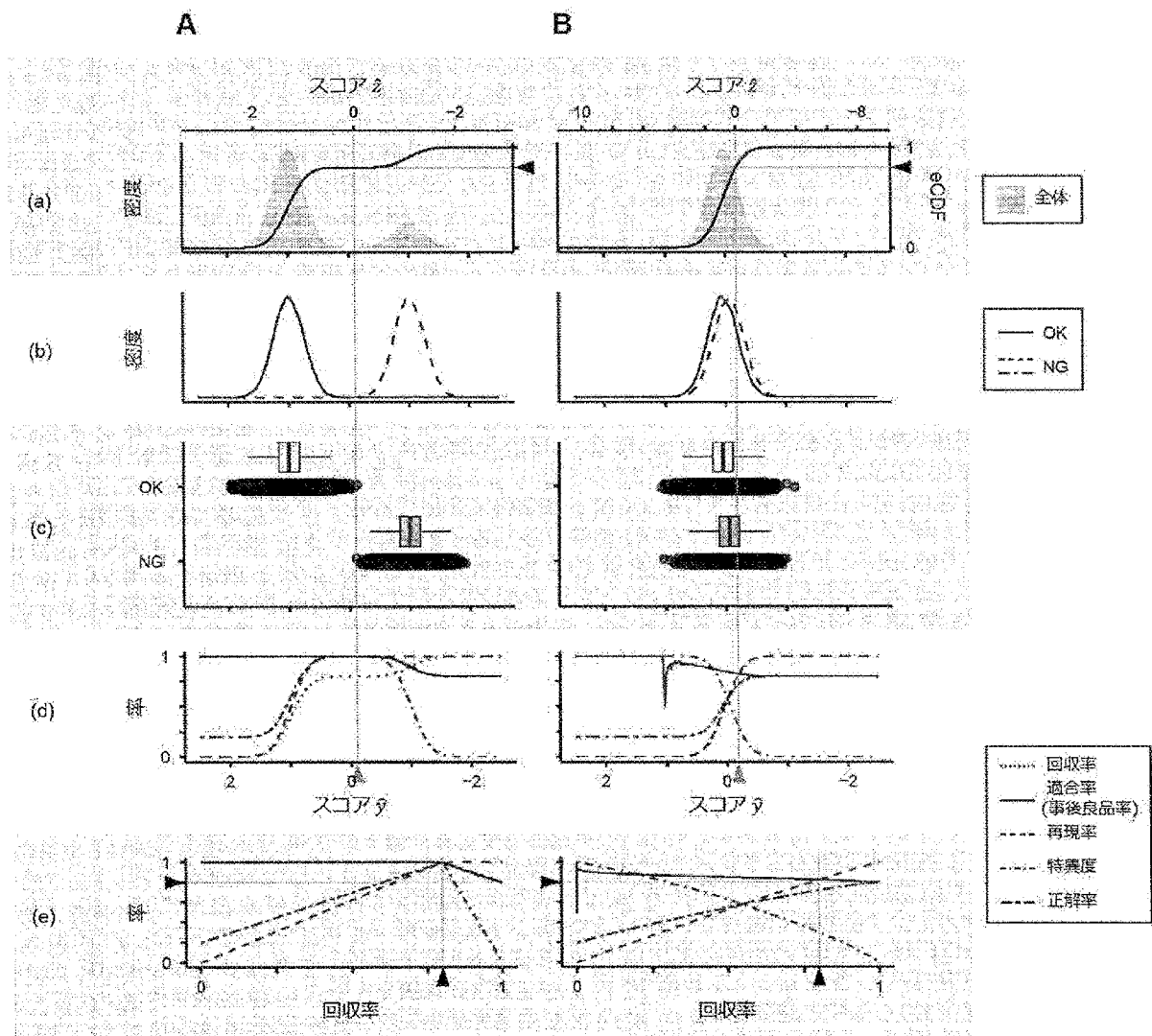
3) 判別モデルの選択フェーズ



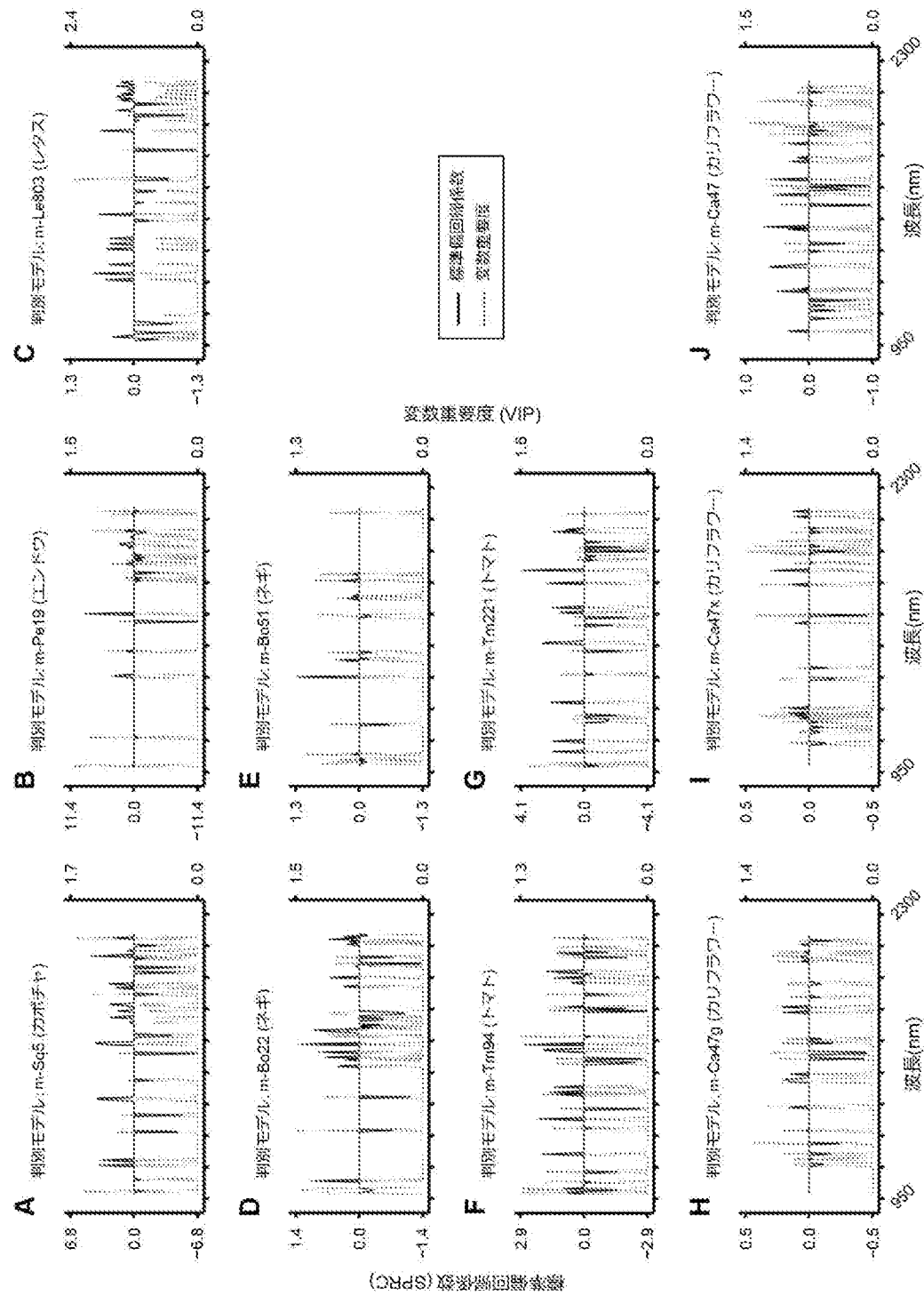
[図4]



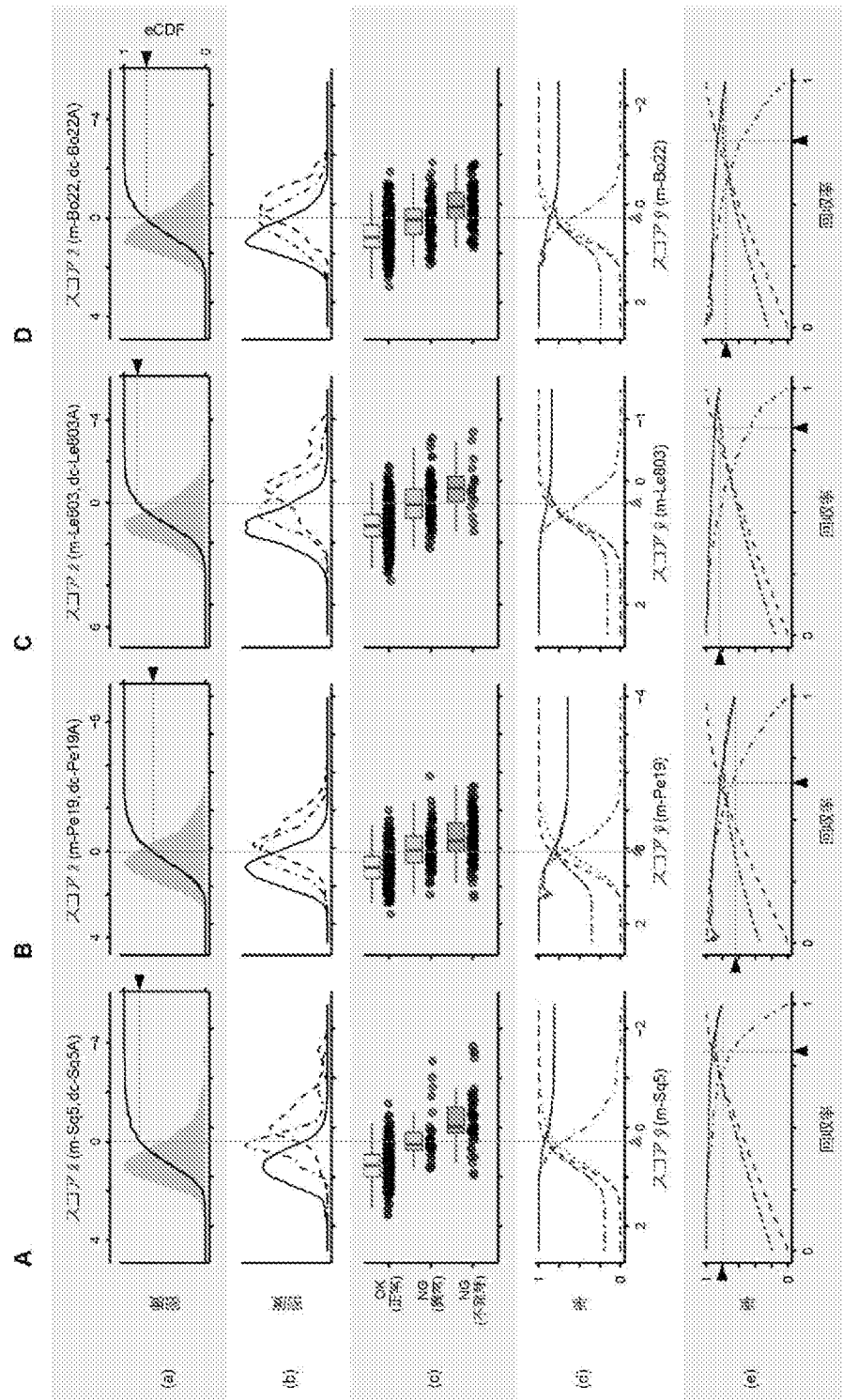
[図5]



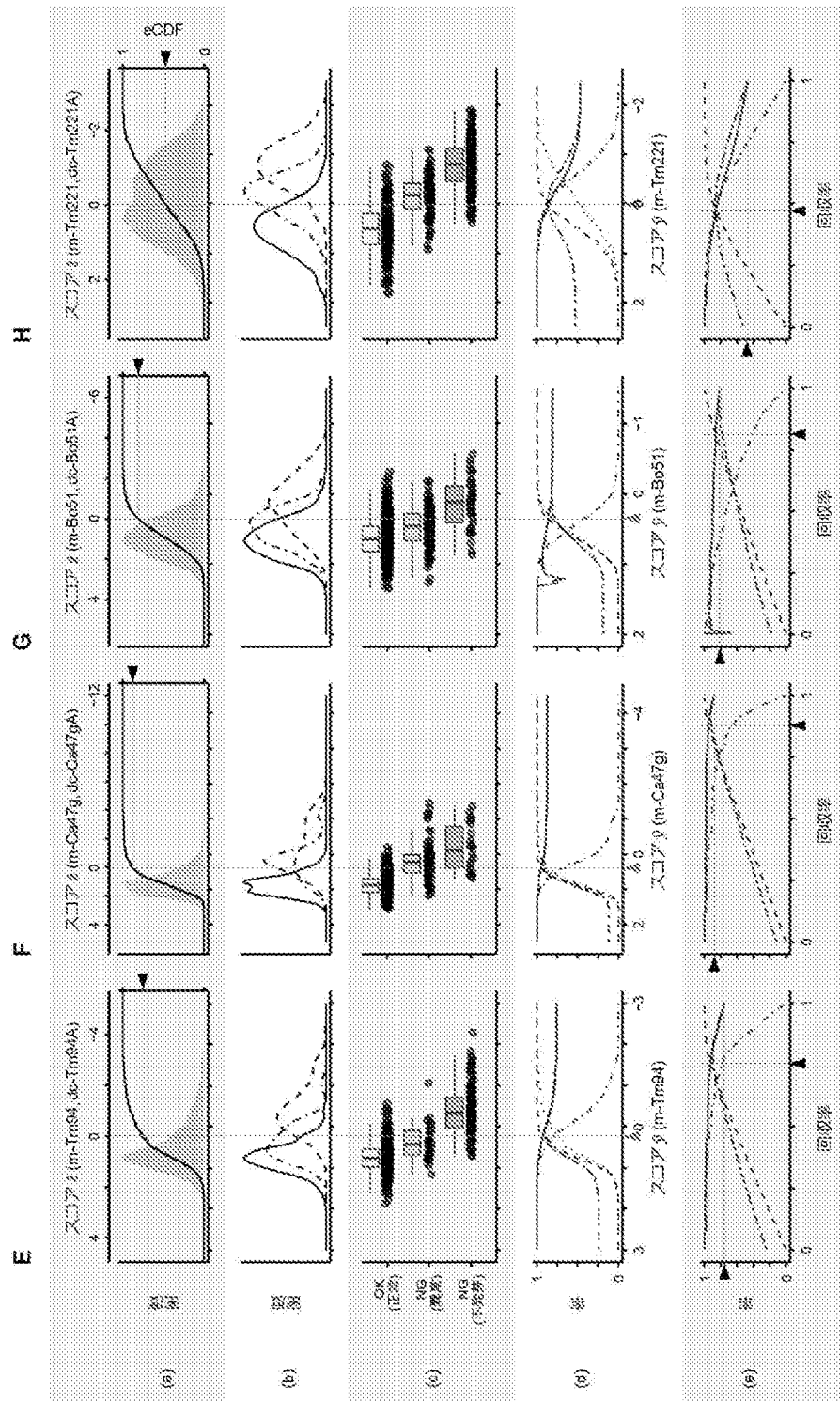
[図6]



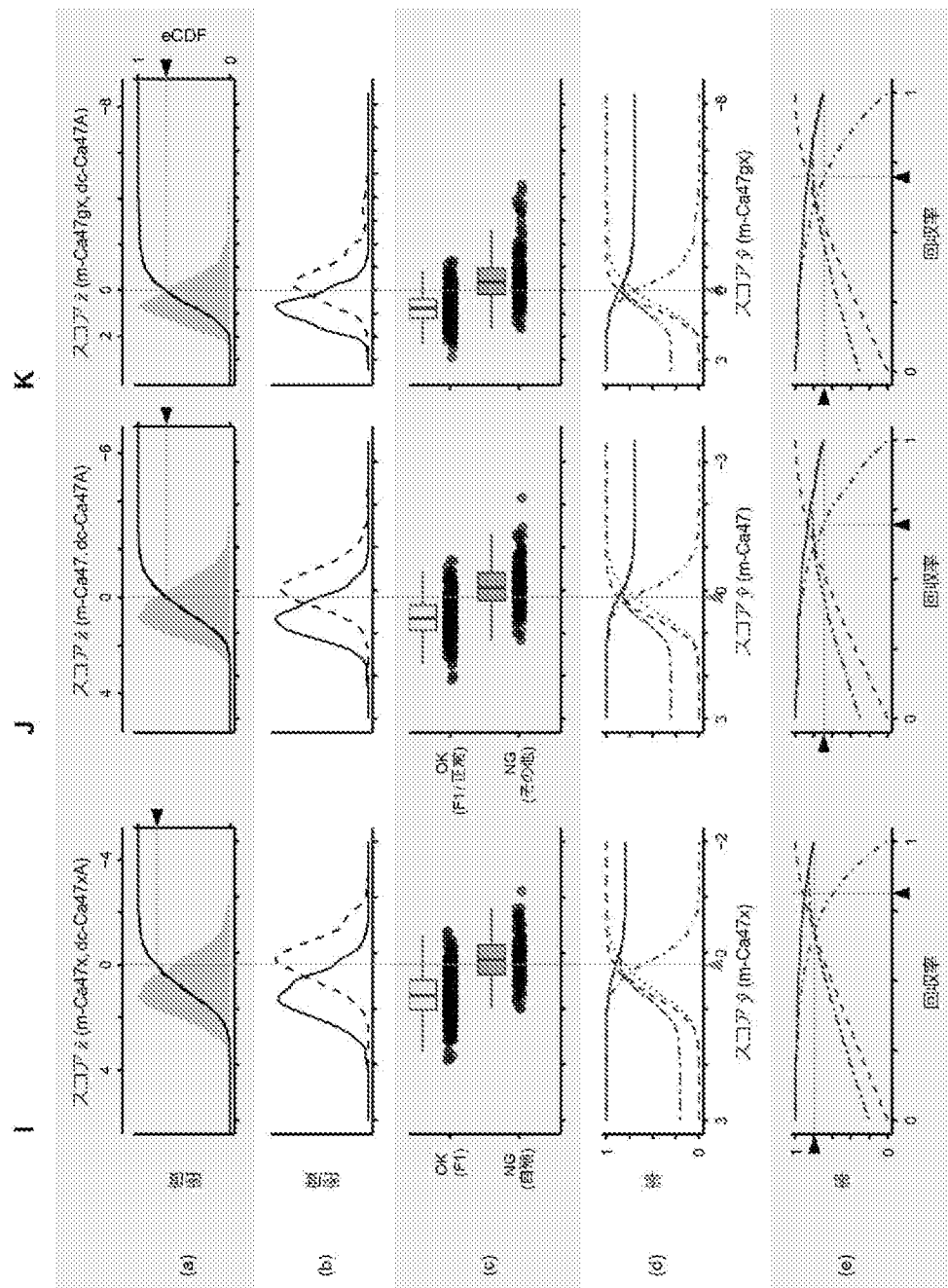
[図8-1]



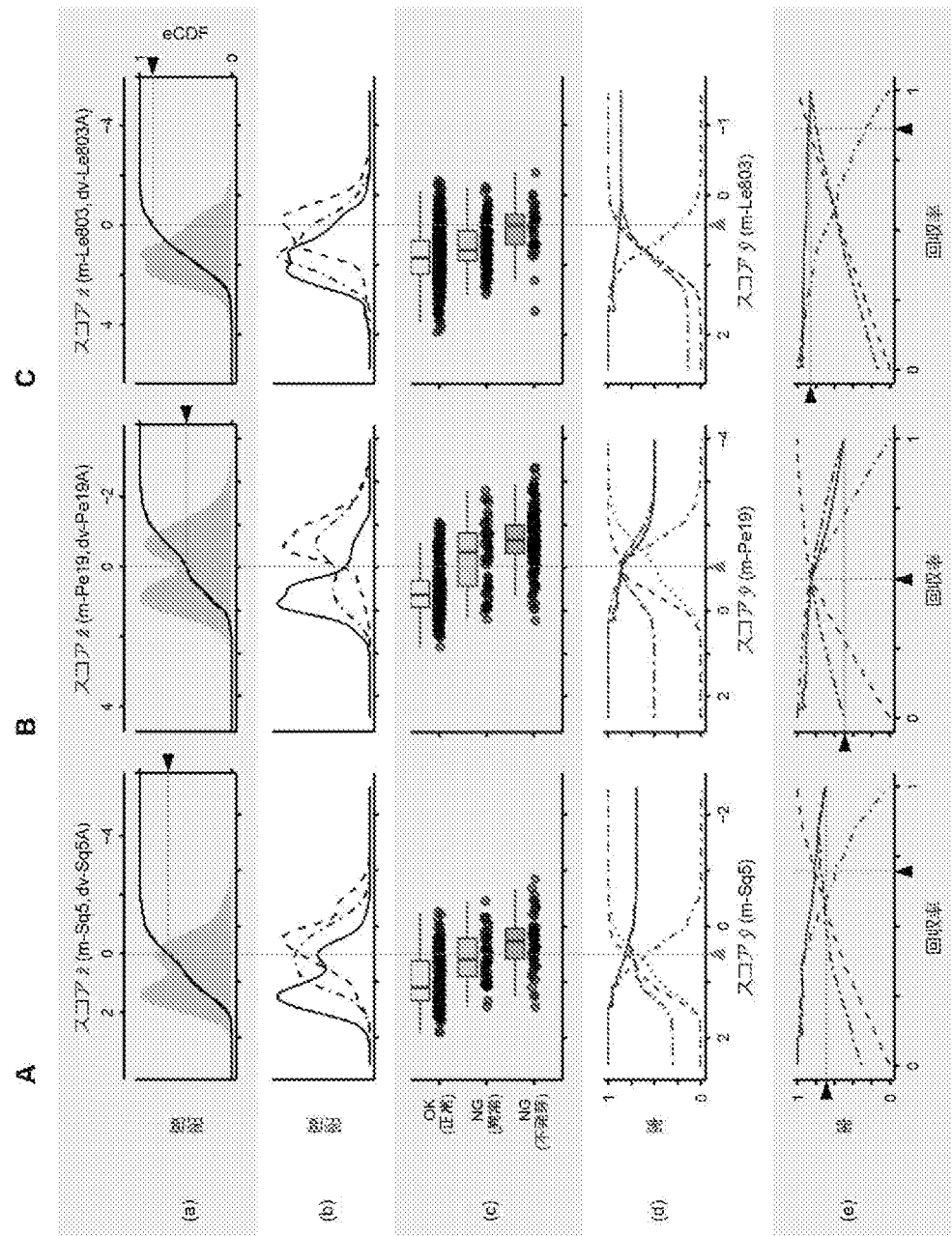
[図8-2]



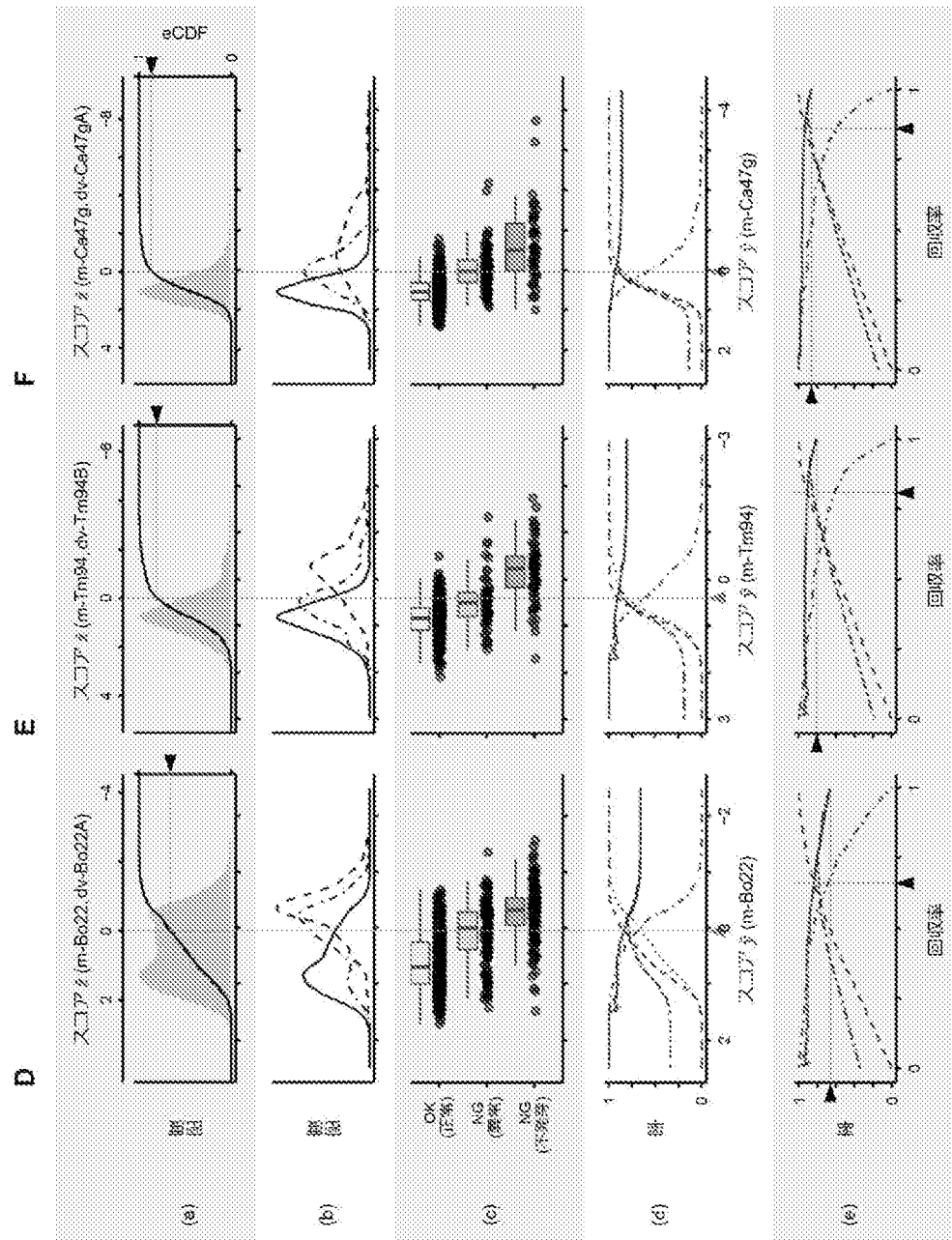
[図8-3]



[図9-1]



[図9-2]

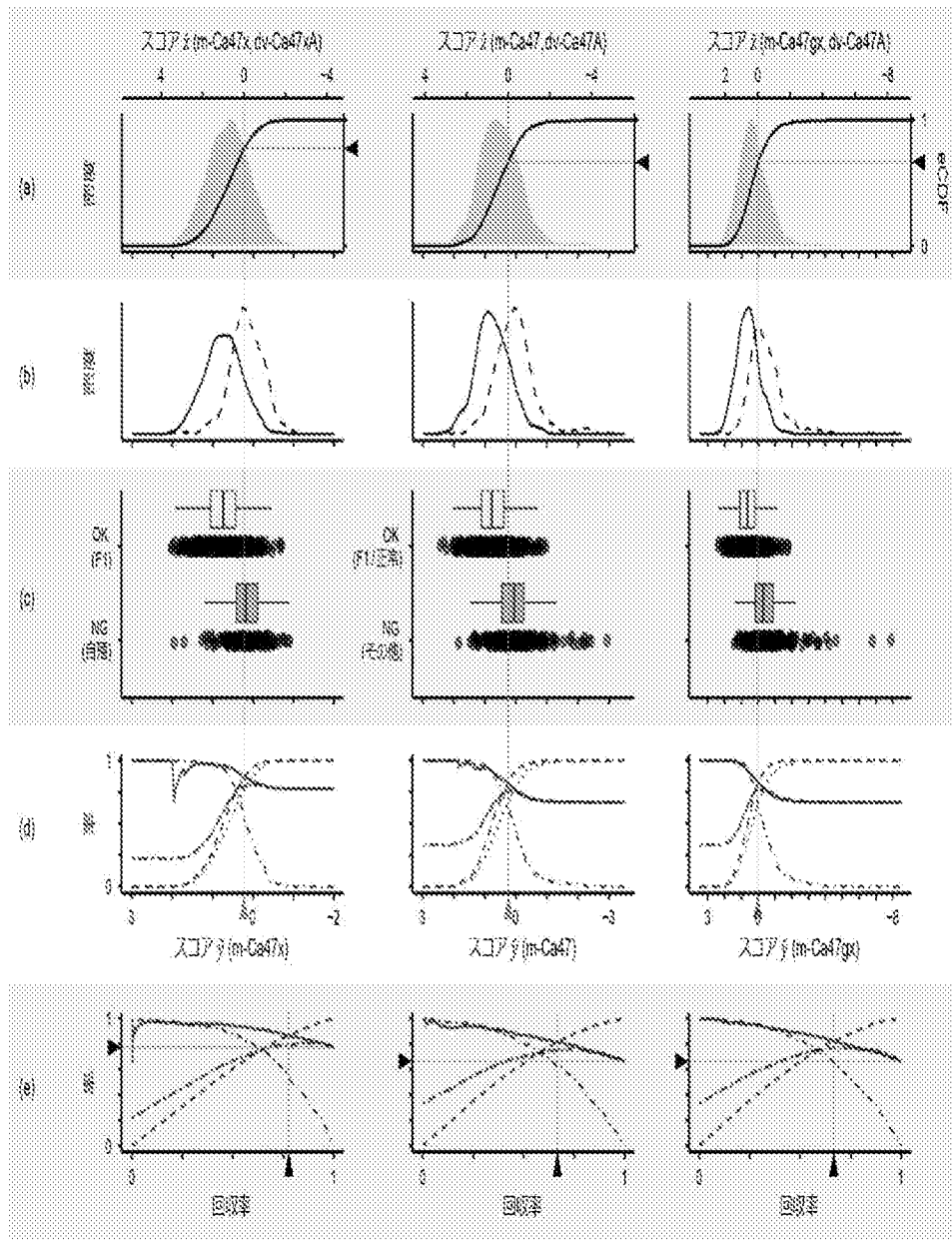


[図9-3]

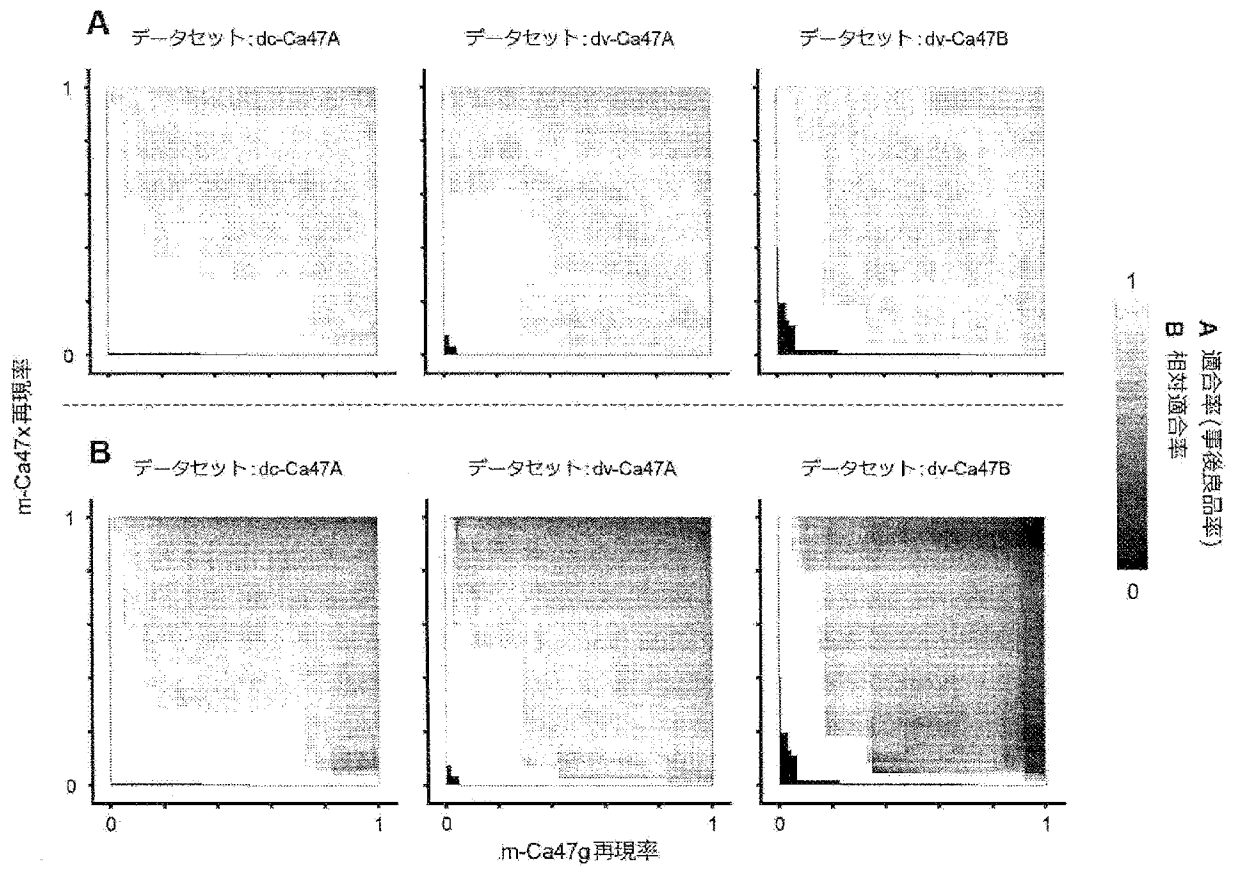
G

H

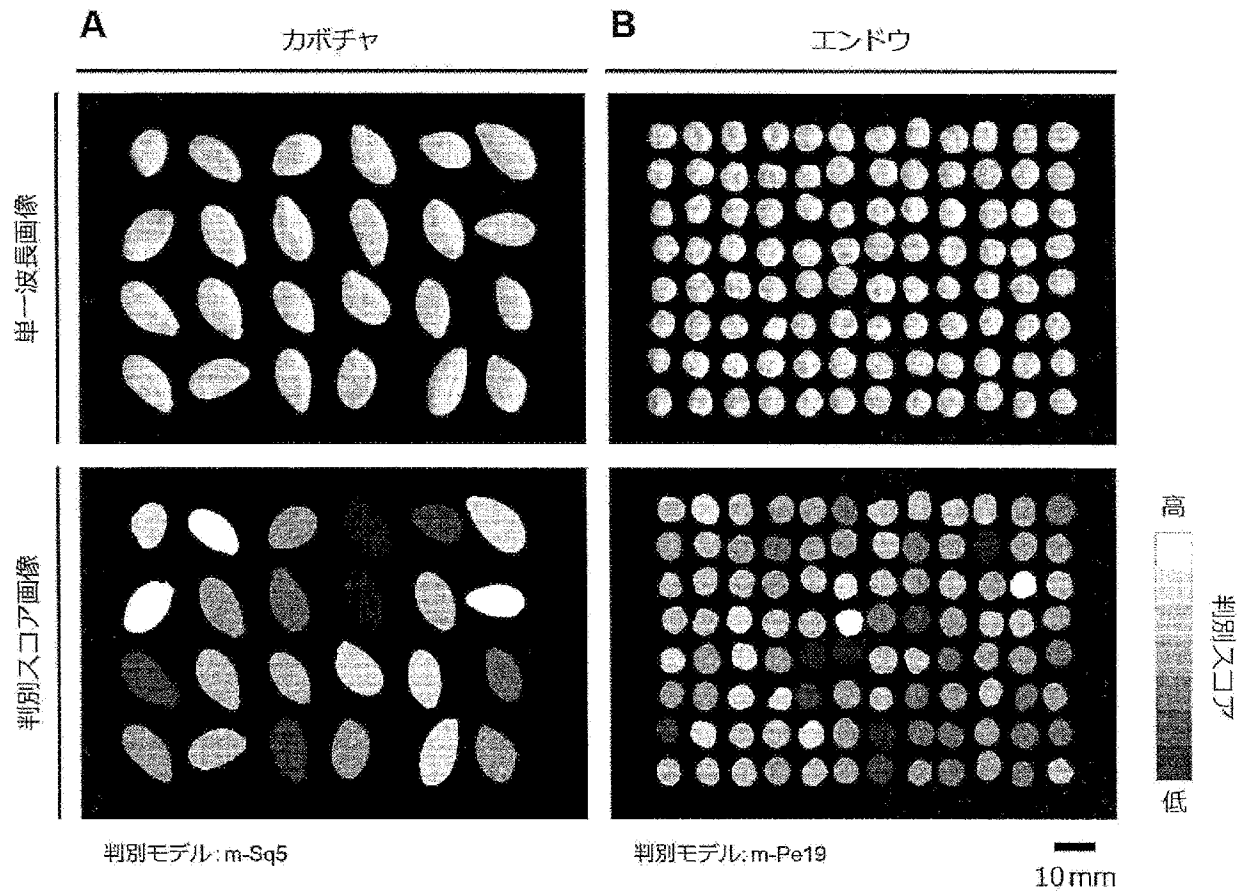
I



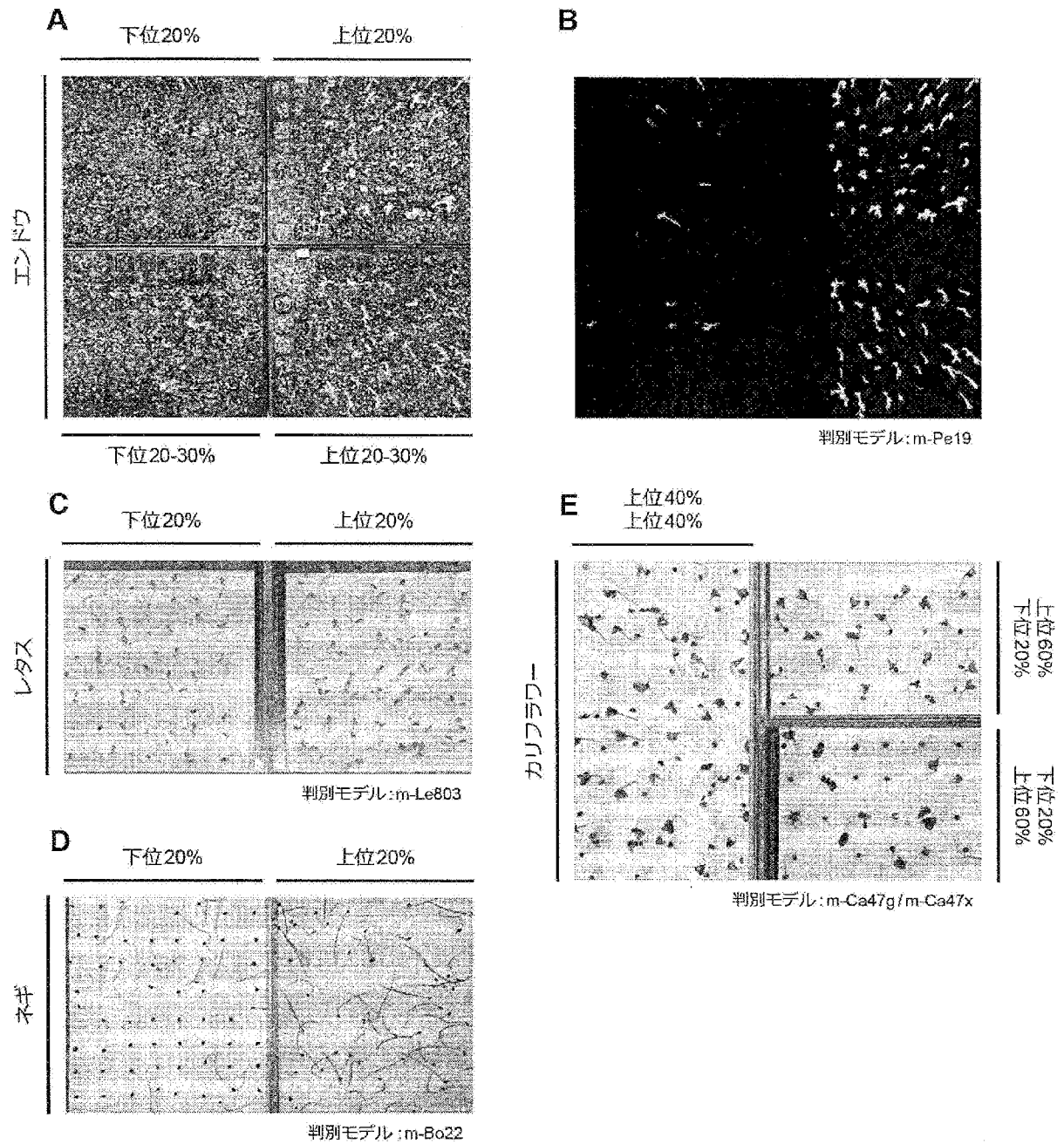
[図10]



[図13]



[図14]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2024/019621

A. CLASSIFICATION OF SUBJECT MATTER G01N 21/3563 (2014.01)i FI: G01N21/3563 According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) G01N21/00-21/958; A01G1/00-33/02 Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Published examined utility model applications of Japan 1922-1996 Published unexamined utility model applications of Japan 1971-2024 Registered utility model specifications of Japan 1996-2024 Published registered utility model applications of Japan 1994-2024 Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) JSTPlus/JSTChina/JST7580 (JDreamIII)		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2016/084452 A1 (SUMITOMO FORESTRY CO., LTD.) 02 June 2016 (2016-06-02) entire text, all drawings	1-6
A	KR 10-2112088 B1 (THE INDUSTRY & ACADEMIC COOPERATION IN CHUNGNAM NATIONAL UNIVERSITY (IAC)) 18 May 2020 (2020-05-18) entire text, all drawings	1-6
A	WO 2022/175309 A1 (KWS SAAT SE & CO. KGAA) 25 August 2022 (2022-08-25) entire text, all drawings	1-6
A	WO 2005/111583 A1 (THE NEW INDUSTRY RESEARCH ORGANIZATION) 24 November 2005 (2005-11-24) entire text, all drawings	1-6
A	TIGABU, Mululem. Multivariate Discriminant Analysis of Single Seed Near Infrared Spectra for Sorting Dead-Filled and Viable Seeds of Three Pine Species: Does One Model Fit All Species. Forests. 2019, vol. 10, no. 469, pp. 1-13, DOI:10.3390/f10060469 entire text, all drawings	1-6
☐ Further documents are listed in the continuation of Box C. ☒ See patent family annex.		
<p>* Special categories of cited documents:</p> <p>“A” document defining the general state of the art which is not considered to be of particular relevance</p> <p>“D” document cited by the applicant in the international application</p> <p>“E” earlier application or patent but published on or after the international filing date</p> <p>“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>“O” document referring to an oral disclosure, use, exhibition or other means</p> <p>“P” document published prior to the international filing date but later than the priority date claimed</p> <p>“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</p> <p>“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art</p> <p>“&” document member of the same patent family</p>		
Date of the actual completion of the international search 06 August 2024		Date of mailing of the international search report 13 August 2024
Name and mailing address of the ISA/JP Japan Patent Office (ISA/JP) 3-4-3 Kasumigaseki, Chiyoda-ku, Tokyo 100-8915 Japan		Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/JP2024/019621

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)		Publication date (day/month/year)
WO	2016/084452	A1	02 June 2016	EP	3225089	A1
KR	10-2112088	B1	18 May 2020	(Family: none)		
WO	2022/175309	A1	25 August 2022	EP	4047350	A1
				CN	117157516	A
WO	2005/111583	A1	24 November 2005	(Family: none)		

A. 発明の属する分野の分類（国際特許分類（IPC））

G01N 21/3563(2014.01)i
FI: G01N21/3563

B. 調査を行った分野

調査を行った最小限資料（国際特許分類（IPC））
G01N21/00-21/958; A01G1/00-33/02

最小限資料以外の資料で調査を行った分野に含まれるもの
日本国実用新案公報 1922-1996年
日本国公開実用新案公報 1971-2024年
日本国実用新案登録公報 1996-2024年
日本国登録実用新案公報 1994-2024年

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）
JSTPlus/JSTChina/JST7580 (JDreamII)

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	WO 2016/084452 A1 (住友林業株式会社) 02.06.2016 (2016 - 06 - 02) 全文、全文	1-6
A	KR 10-2112088 B1 (THE INDUSTRY & ACADEMIC COOPERATION IN CHUNGNAM NATIONAL UNIVERSITY (IAC)) 18.05.2020 (2020 - 05 - 18) 全文、全文	1-6
A	WO 2022/175309 A1 (KWS SAAT SE & CO. KGAA) 25.08.2022 (2022 - 08 - 25) 全文、全文	1-6
A	WO 2005/111583 A1 (財団法人新産業創造研究機構) 24.11.2005 (2005 - 11 - 24) 全文、全文	1-6
A	Muluaem Tigabu, Multivariate Discriminant Analysis of Single Seed Near Infrared Spectra for Sorting Dead-Filled and Viable Seeds of Three Pine Species: Does One Model Fit All Species, Forests, 2019, Vol.10 No.469, pp.1-13, DOI:10.3390/f10060469 全文、全文	1-6

☐ C欄の続きにも文献が列挙されている。

☒ パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー
“A” 特に関連のある文献ではなく、一般的技术水準を示すもの
“D” 国際出願で出願人が先行技術文献として記載した文献
“E” 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
“L” 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）
“O” 口頭による開示、使用、展示等に言及する文献
“P” 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献

“T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの
“X” 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
“Y” 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
“&” 同一パテントファミリー文献

国際調査を完了した日
06.08.2024

国際調査報告の発送日
13.08.2024

名称及びあて先
日本国特許庁(ISA/JP)
〒100-8915
日本国
東京都千代田区霞が関三丁目4番3号

権限のある職員（特許庁審査官）

田中 洋介 2W 3009

電話番号 03-3581-1101 内線 3258

様式 PCT/ISA/210（第2ページ）（2022年7月）

国際調査報告
パテントファミリーに関する情報

国際出願番号

PCT/JP2024/019621

引用文献			公表日	パテントファミリー文献			公表日
WO	2016/084452	A1	02.06.2016	EP	3225089	A1	
KR	10-2112088	B1	18.05.2020	(ファミリーなし)			
WO	2022/175309	A1	25.08.2022	EP	4047350	A1	
				CN	117157516	A	
WO	2005/111583	A1	24.11.2005	(ファミリーなし)			