
Uniwersytet Przyrodniczy w Poznaniu
Wydział Rolnictwa, Ogrodnictwa i Bioinżynierii

AUTOREFERAT

dr Aneta Sawikowska

1. Imię i Nazwisko: Aneta Sawikowska

2. Posiadane dyplomy, stopnie naukowe lub artystyczne – z podaniem nazwy, miejsca i roku ich uzyskania oraz tytułu rozprawy doktorskiej:

2009 Dyplom doktora nauk matematycznych w zakresie matematyki
Uniwersytet im. Adama Mickiewicza w Poznaniu
Wydział Matematyki i Informatyki
Tytuł rozprawy doktorskiej: „Graphs with minimum eigenvalue
for the number of vertices and edges”

2006 Świadectwo ukończenia kursu kwalifikacyjnego pedagogicznego
dla czynnych zawodowo nauczycieli

2004 Dyplom ukończenia studiów wyższych magisterskich
Politechnika Poznańska
Wydział Budowy Maszyn i Zarządzania
kierunek: Matematyka
specjalność: Matematyczne Metody Informatyki

3. Informacja o dotychczasowym zatrudnieniu w jednostkach naukowych lub artystycznych.

od 09.2020 Starszy specjalista biolog
Pracownia Bioinformatyki
Instytut Chemii Bioorganicznej Polskiej Akademii Nauk

od 10.2016 Adiunkt
Katedra Metod Matematycznych i Statystycznych
Uniwersytet Przyrodniczy w Poznaniu

| | |
|-------------------|--|
| 10.2019 – 06.2020 | Umowy o dzieło Instytut Genetyki Roślin Polskiej Akademii Nauk |
| 04.2017 – 10.2019 | Matematyk Zakład Biometrii i Bioinformatyki Instytut Genetyki Roślin Polskiej Akademii Nauk |
| 12.2016 – 04.2017 | Umowy o dzieło Instytut Genetyki Roślin Polskiej Akademii Nauk |
| 01.2010 – 12.2016 | Adiunkt Zakład Biometrii i Bioinformatyki Instytut Genetyki Roślin Polskiej Akademii Nauk |
| 10.2009 – 01.2010 | Umowa o dzieło Instytut Genetyki Roślin Polskiej Akademii Nauk |
| 10.2004 – 06.2009 | Studia doktoranckie z matematyki Wydział Matematyki i Informatyki Uniwersytet im. Adama Mickiewicza w Poznaniu |

4. Omówienie osiągnięć, o których mowa w art. 219 ust. 1 pkt. 2 ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (Dz. U. z 2020 r. poz. 85 z późn. zm.). Omówienie to winno dotyczyć merytorycznego ujęcia przedmiotowych osiągnięć, jak i w sposób precyzyjny określać indywidualny wkład w ich powstanie, w przypadku, gdy dane osiągnięcie jest dziełem współautorskim, z uwzględnieniem możliwości wskazywania dorobku z okresu całej kariery zawodowej.

Tytuł osiągnięcia naukowego lub artystycznego:

Chemometria w metabolomice roślin

Cykl prac omówionych w autoreferacie:

IB1. Sawikowska A., Piasecka A., Kachlicki P., Krajewski P. (2021). Separation of Chromatographic Co-Eluted Compounds by Clustering and by Functional Data Analysis. *Metabolites* 11(4), 214. IF=4.097; MNiSW=70 pkt.

IB2. Sawikowska A. (2020). Meta-analysis of flavonoids with antiviral potential against coronavirus. *Biometrical Letters* 57: 13-22. MNiSW=20 pkt.

IB3. Piasecka A., Sawikowska A., Kuczyńska A., Ogrodowicz P., Mikołajczak K., Krajewski P., Kachlicki P. (2020). Phenolic metabolites from barley in contribution to phenome in soil moisture deficit. *International Journal of Molecular Sciences* 21(17), 6032. IF=4.556; MNiSW=140 pkt.

IB4. Kruszka D., Sawikowska A., Selvakesavan R.K., Krajewski P., Kachlicki P., Franklin G. (2020). Silver nanoparticles affect phenolic and phytoalexin composition of *Arabidopsis thaliana*. *Science of the Total Environment* 716: 135361. IF=6.551; MNiSW=200 pkt.

IB5. Mieldzioc A., Mokrzycka M., Sawikowska A. (2019). Covariance regularization for metabolomic data on the drought resistance of barley. *Biometrical Letters* 56: 165–181. MNiSW=20 pkt.

IB6. Piasecka A., Sawikowska A., Kuczyńska A., Ogrodowicz P., Mikołajczak K., Krystowiak K., Gudyś K., Guzy-Wróbelska J., Krajewski P., Kachlicki P. (2017). Drought related secondary metabolites of barley (*Hordeum vulgare* L.) leaves and their association with mQTLs. *Plant J*, 89: 898–913. IF= 5.901; MNiSW=45 pkt.

IB7. Swarczewicz B., Sawikowska A., Marczak Ł., Łuczak M., Ciesiołka D., Krystkowiak K., Kuczyńska A., Piślewska-Bednarek M., Krajewski P., Stobiecki M. (2017). Effect of drought stress on metabolite contents in barley recombinant inbred line population revealed by untargeted GC–MS profiling. *Acta Physiol Plant* 39: 158. IF= 1.364; MNiSW=25 pkt.

IB8. Chmielewska K., Rodziewicz P., Swarczewicz B., Sawikowska A., Krajewski P., Marczak Ł., Ciesiołka D., Kuczyńska A., Mikołajczak K., Ogrodowicz P., Krystkowiak K., Surma M., Adamski T., Bednarek P., Stobiecki M. (2016). Analysis of drought-induced proteomic and metabolomic changes in barley (*Hordeum vulgare* L.) leaves and roots unravels some aspects of biochemical mechanisms involved in drought tolerance. *Frontiers in Plant Science* 7: 1108. IF=4.495; MNiSW=40 pkt.

Wprowadzenie

Chemometria jest nauką zajmującą się wykorzystaniem metod komputerowych, matematycznych i statystycznych w analizowaniu danych chemicznych. Natomiast metabolomika zajmuje się badaniem złożonych oddziaływań jakie zachodzą w genomice funkcjonalnej i biologii systemów poprzez identyfikację oraz analizę ilościową niskocząsteczkowych produktów naturalnych zwanych metabolitami. Metabolomika jest nauką interdyscyplinarną, a realizacja jej badań wymaga współpracy z matematykami, informatykami, bioinformatykami i statystykami. Efektem takiej współpracy są wyniki moich badań naukowych.

Metabolomika znajduje zastosowanie w wielu dziedzinach nauki, np. w rolnictwie, medycynie, naukach o żywności i żywieniu, toksykologii, genomice funkcjonalnej oraz nutrigenomice. Obecnie w wielu badaniach naukowych w metabolomice roślin stosuje się nowoczesne, wysokoprzepustowe technologie, które na całym świecie stają się coraz bardziej popularne i szeroko stosowane. Dane z analizy wysokoprzepustowej mają postać wielowymiarową. Zatem w

wyniku takich badań uzyskuje się do analizy bardzo wiele cech i generuje ogromne ilości danych. Gdy liczba prób jest duża (powyżej 30) liczba obserwacji może się wahać od kilkuset tysięcy do kilkudziesięciu milionów. Wiele doświadczeń prowadzonych jest na dużą skalę z wieloma genotypami, pod wpływem różnych warunków środowiskowych, w różnych fazach rozwojowych lub punktach czasowych oraz pod wpływem różnych traktowań, w celu pogłębienia wiedzy na temat metabolizmu roślin. Wszystko to prowadzi do sytuacji, w której prowadzone doświadczenia mają strukturę wieloczynnikową, a liczba próbek do analizy jest bardzo duża. Ponadto liczba powtórzeń biologicznych musi być wystarczająca do prawidłowego oszacowania naturalnej zmienności, co również ma wpływ na liczebność danych. Współczesna metabolomika wykorzystuje coraz bardziej czułe przyrządy wysokoprzepustowe i protokoły, co również wpływa na generowanie dużej liczby danych. Występuje także tendencja do badania całego spektrum obserwowanych cech, np. związków chemicznych, tzw. analiza niecelowana, biorąca pod uwagę wszystkie wykryte związki. W efekcie analiza staje się dużym wyzwaniem.

Ze względu na dużą liczbę danych i ich wielowymiarowość nie da się wysnuć bezpośrednio z surowych danych żadnych wniosków. Wstępne przetwarzanie danych oraz ich analiza statystyczna są do otrzymania tych wniosków niezbędne. Na przestrzeni ostatnich lat dynamicznie rozwijały się technologie, a wraz z nimi rosło zapotrzebowanie na rozwijające się metody obróbki i analizy statystycznej dla wieloczynnikowych doświadczeń wysokoprzepustowych. W rozwoju takich metod udało mi się uczestniczyć. W literaturze, przed pojawieniem się prac stanowiących mój dorobek habilitacyjny, zaproponowano różne metody statystyczne, jednak były to analizy jednoczynnikowe lub w przypadku przetwarzania danych w programach istniały ograniczenia co do liczby danych.

Metody zaprezentowane w omawianych tu publikacjach nie posiadają takich ograniczeń, ponieważ zostały dostosowane do wysokiego poziomu skomplikowania danych moimi własnymi skryptami napisanymi w R i Genstat.

Do przetwarzania big data zastosowałam supercomputing i zrównoleglenie wykonywanych procesów. Dane były przetwarzane własnymi skryptami od wysokoprzepustowych danych surowych poprzez analizę statystyczną do wyników i integracji danych.

Analizując dane metabolomiczne wykonałam konieczne kroki wstępnego przetwarzania danych takie jak: normalizacja, usuwanie linii bazowej i wyrównanie czasu retencji. Następnie przeprowadziłam analizę statystyczną. Zastosowałam analizę wariancji w mieszanych modelach liniowych oraz metody wielowymiarowe. Wykonałam analizę sieci korelacyjnych oraz wprowadziłam różnicowe sieci korelacyjne do badania istotnych różnic między korelacjami w różnych traktowaniach.

Aby w przyszłości możliwe było przeprowadzenie jeszcze dokładniejszej analizy statystycznej dostosowanej do tego typu danych, badałam problem estymacji macierzy kowariancji dla danych metabolomicznych.

Omawiany tutaj cykl prac koncentruje się na analizie dużych zbiorów danych metabolomicznych pochodzących z różnych gatunków roślin i z różnego rodzaju powszechnie używanych urządzeń, tj. chromatografii gazowej sprzężonej ze spektrometrią mas (GC-MS), wysokosprawnej chromatografii cieczowej z detektorem UV (LC-UV) oraz chromatografii cieczowej sprzężonej ze spektrometrią mas (LC-MS). Chromatografia jest szeroko używaną metodą laboratoryjną do rozdzielania związków chemicznych, m.in. metabolitów.

Nakładanie się pików jest częstym problemem rozdzielania mieszaniny związków w danych chromatograficznych. Ze względu na istnienie zjawiska koelucji różnych związków o podobnych właściwościach chromatograficznych rozdzielanie pików staje się trudne. W moich badaniach wprowadziłam dwie nowe metody separacji wspólnie wpływających związków w chromatografii: przy pomocy analizy skupień oraz funkcjonalnej analizy składowych głównych.

Moje badania koncentrowały się również na meta-analizie, czyli procesie połączenia wyników różnych badań dotyczących metabolomu. W odpowiedzi

na sytuację epidemiologiczną, wykonałam meta-analizę danych metabolomicznych o potencjale antywirusowym przeciw koronawirusowi.

Cele badań

Celem prowadzonych przeze mnie badań było wstępne przetwarzanie, analiza statystyczna, separacja związków, meta-analiza i integracja dużych zbiorów danych metabolomicznych oraz analiza ich struktury kowariancyjnej wraz z interpretacją uzyskanych wyników. Szczegółowe cele przedstawiały się następująco:

- wstępne przetwarzanie dużej liczby danych chromatograficznych z różnego rodzaju urządzeń przedstawione na przykładzie metabolitów pierwotnych i wtórnych w jęczmieniu jarym pod wpływem suszy, a w *Arabidopsis thaliana* w badaniach wpływu nanocząsteczek srebra na skład fenolowy i fitoaleksyny; zastosowanie kilkukrokowych metod, tak aby z surowych danych chromatograficznych uzyskać tabelę gotową do przeprowadzenia analizy statystycznej, bez ograniczeń co do liczby danych, co narzucały do tej pory wszystkie programy,
- separacja wspólnie wpływających związków w chromatografii wprowadzając dwie nowe metody, jedna wykorzystująca analizę skupień, druga funkcjonalną analizę składowych głównych, na przykładzie metabolitów wtórnych w jęczmieniu jarym pod wpływem suszy,
- analiza statystyczna doświadczeń wieloczynnikowych na przykładzie metabolitów pierwotnych i wtórnych w jęczmieniu jarym pod wpływem suszy oraz danych z *Arabidopsis thaliana* w badaniach wpływu nanocząsteczek srebra na skład fenolowy i fitoaleksyny,
- analiza sieci korelacyjnych oraz zdefiniowanie różnicowych sieci korelacyjnych w doświadczeniach wieloczynnikowych na przykładzie metabolitów pierwotnych i wtórnych w jęczmieniu jarym pod wpływem suszy,

- estymacja macierzy kowariancji dla danych metabolomicznych z jęczmienia pod wpływem suszy,
- meta-analiza metabolitów z grupy flawonoidów o potencjale antywirusowym przeciw koronawirusowi.

Materiał

Materiał badawczy pochodził:

- a) z badań prowadzonych w ramach projektu POLAPGEN-BD: Narzędzia biotechnologiczne służące do otrzymywania odmian zbóż o zwiększonej odporności na suszę (Załącznik 5, IIF7), w którym byłam zatrudniona,
- b) z badań prowadzonych w ramach projektu HyperNano: Badanie zmian metabolizmu wtórnego u *Hypericum perforatum* pod wpływem nanocząsteczek poprzez zastosowanie zintegrowanego podejścia i technologii "omics" (Załącznik 5, IIF11),
- c) z badań naukowych dostępnych w publicznych bazach danych.

W ramach pierwszego projektu dane uzyskano z badań jęczmienia jarego (*Hordeum vulgare*) wykonanych w warunkach niedoboru wody. Wymienione poniżej zbiory danych były przedmiotem badań w pracach IB1, IB3, IB5, IB6, IB7, IB8 z Załącznika 5.

Dane dotyczące metabolizmu pierwotnego z doświadczenia pilotażowego analizowane w pracy IB5 stanowiły 9 odmian jęczmienia (CamB1, Georgia, Harmal, Lubuski, Maresi, MDingo, Morex, Sebastian, Stratus) pod wpływem suszy i w warunkach kontrolnych, badane w 4 fazach rozwojowych rośliny, w 4 powtórzeniach biologicznych i 2 powtórzeniach technicznych, co stanowiło 422 próby, po odrzuceniu przez doświadczalników pewnych prób ze względów technicznych. Badaną częścią rośliny był liść. Doświadczenie zostało przeprowadzone w warunkach szklarniowych. Przedmiotem badań były metabolity pierwotne wykryte za pomocą chromatografii gazowej sprzężonej ze

spektrometrią mas. Zbiór surowych danych składał się z 51 135 cech dla 422 prób, co stanowi ok 21 mln obserwacji.

Dane dotyczące metabolizmu pierwotnego analizowane w publikacji IB7 stanowiły populację 100 zrekombinowanych linii wsobnych jęczmienia oraz odmiany rodzicielskie Maresi, CamB, badane pod wpływem suszy i w warunkach kontrolnych, w 2 powtórzeniach biologicznych, co stanowiło 408 prób osobno dla liścia i korzenia. Doświadczenie zostało przeprowadzone w warunkach szklarniowych. Cechy stanowiły intensywność absorpcji metabolitów pierwotnych wykrytych za pomocą chromatografu gazowej sprzężonego ze spektrometrem mas. Zbiór surowych danych składał się z dziesiątków milionów obserwacji.

Dane dotyczące metabolizmu wtórnego analizowane w pracy IB1 oraz IB3 uzyskano badając wpływ niedoboru wody na poziom metabolitów wtórnych jęczmienia jarego (*Hordeum vulgare*), wielokrotnie mierzonych w okresie stosowania suszy. Przeprowadzone doświadczenia stanowiły badanie pilotażowe. Dane uzyskano dla 9 odmian jęczmienia, 3 różnych wariantach suszy (I, II, I + II) i w warunkach kontrolnych, w 8 punktach czasowych i 4 powtórzeniach biologicznych. Próbkę liści do analiz metabolomicznych pobrano z roślin poddanych suszy I, II i I + II po 3, 6 i 10 dniach niedoboru wilgoci oraz w warunkach suszy I + II po 1 dobie. Warto zauważyć, że np. w przypadku badania suszy I analizowane są 3 punkty czasowe w trakcie trwania suszy oraz 1 punkt czasowy podczas warunków kontrolnych, natomiast w przypadku łączonej suszy I+II, najpierw zastosowano suszę I, a następnie I+II, co wymagało badania w warunkach kontrolnych po 1 dobie zastosowania traktowania I+II. Pomiary intensywności absorpcji metabolitów przeprowadzono za pomocą ultra-wysokosprawnego chromatografu cieczowego z detektorem PDA. Liczba analizowanych próbek wyniosła 547, przy czym liczba odrębnych punktów czasu retencji w każdym chromatogramie przekroczyła 10000, więc dane stanowiły około 5,5 mln obserwacji.

Dane dotyczące metabolizmu wtórnego analizowane w pracy IB6 stanowiły populację 100 zrekombinowanych linii wsobnych jęczmienia oraz odmiany rodzicielskie Maresi, CamB, badanych pod wpływem suszy i w warunkach kontrolnych, w 2 fazach rozwojowych rośliny (punktach czasowych T1 i T2), w 4 powtórzeniach biologicznych. Liczba analizowanych próbek wyniosła 1632. Badaną częścią rośliny był liść. Doświadczenie zostało przeprowadzone w warunkach szklarniowych. Cechy stanowiły intensywności absorpcji metabolitów wtórnych wykrytych za pomocą ultrasprawnego chromatografu cieczowego z detektorem UV. Zbiór surowych danych składał się z ok 43 milionów obserwacji.

Dane metabolomiczne oraz integrowane z nimi dane proteomiczne będące przedmiotem pracy IB8 stanowiły 2 odmiany jęczmienia (Maresi, CamB), badane pod wpływem suszy i w warunkach kontrolnych, w 2 powtórzeniach biologicznych i 2 powtórzeniach technicznych, osobno dla liścia i korzenia, powtórzone w 2 doświadczeniach. Doświadczenia zostały przeprowadzone w warunkach szklarniowych. Pomiar białek wykonano za pomocą dwukierunkowej elektroforezy dla 64 prób. Metabolity pierwotne zostały wykryte za pomocą chromatografu gazowej sprzężonego ze spektrometrem mas. Analizowano również 14 cech fenotypowych.

W ramach drugiego projektu dane uzyskano z badań nad *Arabidopsis thaliana* traktowanych nanocząsteczkami srebra. Analiza tych danych została opisana w pracy IB4 i dotyczyła metabolizmu wtórnego, w 3 punktach czasowych dla 12 wariantów traktowania nanocząsteczkami srebra i w warunkach kontrolnych, w 3 powtórzeniach biologicznych, 2 powtórzeniach technicznych, co stanowiło 225 prób, po odrzuceniu przez doświadczalników kilku powtórzeń technicznych ze względu na ich jakość. W tym przypadku analizowane były nasiona. Cechy stanowiły intensywności absorpcji metabolitów wtórnych wykrytych za pomocą ultrasprawnego chromatografu cieczowego z detektorem UV. Zbiór surowych danych składał się z prawie 10 milionów obserwacji.

Analizowane dane dostępne w publicznych bazach danych pochodziły z największych na świecie naukowych baz: Metabolights i Prime, jak również publikacji naukowych. Były to wybrane zbiory danych, w których zidentyfikowano co najmniej jeden z trzech metabolitów: roifolinę, pektolinarynę lub herbacetynę. Dane te zostały uzyskane za pomocą wysokoprzepustowych metod analitycznych m.in. chromatografii cieczowej sprzężonej ze spektrometrią mas. Analizowane zestawy danych obejmują następujące rośliny wraz z liczbą prób i cech:

jęczmień (CamB1, Georgia, Harmal, Lubuski, Maresi, MDingo, Morex, Sebastian, Stratus) – 181 prób, 104 metabolity,

pszenica (Chinese Spring) – 18 prób, 371 metabolitów,

soja (linie wsobne *G. max* i *G. soja*) – 279 prób, 48 metabolitów,

ostrożeń (*Cirsium japonicum*) – 15 prób, 10 metabolitów,

rosiczka (*Drosera peltata*) – 4 prób, 11 metabolitów,

kłosownica dwukłosowa (linie: Bd21, Bd3-1) – 36 prób, 517 metabolitów,

fikus (*Ficus deltoidea*) – 100 prób, 387 metabolitów.

We wszystkich roślinach oprócz soi badane były liście. W przypadku soi badane były nasiona. Dodatkowo nasiona były analizowane w przypadku pszenicy i kłosownicy dwukłosowej.

Wyniki, metody i ich zastosowanie:

1. Wstępne przetwarzanie danych chromatograficznych

Praca z wysokoprzepustowymi danymi metabolomicznymi wymaga odpowiedniego przygotowania i redukcji danych do dalszej analizy. Wybór odpowiedniej metody często zależy od charakteru danych i urządzenia z jakiego są generowane. W prezentowanym cyklu prac wstępne przetwarzanie danych

składało się z kilku zintegrowanych etapów wykonanych w moich własnych skryptach napisanych w R z metodami dostosowanymi do dużej liczby danych otrzymanych w dużych wieloczynnikowych doświadczeniach. W omawianych pracach wykonałam przetwarzanie danych z urządzeń powszechnie stosowanych w metabolomice: GC-MS (IB7, str. 3-4; IB8: str. 4-5; IB5: str. 167), UPLC-UV (IB1: str. 5-6, 13; IB3: str. 17-18, IB4: str. 3,9; IB6: str. 899, 910-911, Supplementary methods: 1-2). W każdym z tych przypadków należy pamiętać o usunięciu niepożądanych informacji jak np. szum, czy też linia bazowa oraz o innych elementach przetwarzania w zależności od instrumentu.

W przypadku UPLC-UV nie istniały żadne programy do przetwarzania dużych danych. Wszystkie zastosowane tu metody zostały zapisane we własnych skryptach w systemie R i uruchomione równolegle na klastrach Poznańskiego Centrum Superkomputerowo-Sieciowego. W przypadku danych otrzymanych z urządzenia UPLC-UV każda próba posiada obserwację (intensywność absorpcji) w pewnym czasie retencji dla całego zakresu fal ultrafioletowych. Zatem dane te mają trójwymiarową naturę. W przeprowadzonych badaniach dokonano wyboru dwóch najbardziej reprezentatywnych długości fal. Analizowane były sygnały (piki), które mogą być interpretowane jako przedziały czasu retencji, w których występuje metabolit lub grupa metabolitów o podobnych własnościach. Wykonałam poszczególne kroki przetwarzania danych, które przedstawiłam na przykładzie odmiany jęczmienia Maresi przy długości fali 330 nm na rysunku Figure 3 w pracy IB1. Surowe dane stanowiące chromatogramy z różnych prób dla różnych wariantów suszy i warunków kontrolnych zostały znormalizowane przez masę próbki. Z porównania wizualnego przy tak dużej liczbie chromatogramów nie da się wysnuć żadnych wniosków. Dlatego zastosowanie wstępnego przetwarzania danych było niezbędne.

Łatwo zauważyć, że występuje tutaj linia bazowa, zatem wykonałam jej usunięcie poprzez różnicowanie danych. Na przykładzie wspomnianego rysunku można zaobserwować, że w poszczególnych fragmentach chromatogramów występują przesunięcia czasowe, dlatego dokonałam

wyrównania czasu retencji przez zastosowanie algorytmu COW (correlation optimised warping). Następnie wyodrębniłam piki. Najpierw wykryłam piki indywidualne dla poszczególnych chromatogramów, a następnie piki wspólne dla wszystkich chromatogramów. Ostatecznie chromatogramy zostały zintegrowane w granicach wspólnych pików. Na ostatnim etapie omawianego rysunku szerokość słupków jest szerokością pików, a wysokość obrazuje wielkość pól pod indywidualnymi chromatogramami w granicach wspólnych pików. Rozważmy poszczególne kroki omawianego przetwarzania danych. Linię bazową można zdefiniować jako złożenie sumy idealnego spectrum i tła z funkcją rozmywającą plus szum. Problemem jest odzyskanie idealnego spectrum. Jednakże wiedza dotycząca tła, funkcji rozmywającej i szumu jest często niekompletna lub całkowicie brakująca. Linia bazowa jest podstawą prawdziwej mierzonej wartości. Problem polega na tym, że zerowe wartości związków nie zawsze są reprezentowane przez 0. Linia bazowa prowadzi do błędnych obliczeń pola w obszarze pików i może zaburzać wszystkie kolejne kroki analizy, zatem musiała zostać usunięta. Aby abstrahować od różnych metod i problemów z oszacowaniem linii bazowej i potrzebnych do tego parametrów, w omawianych publikacjach usunęłam linię bazową poprzez różnicowanie danych, co oznacza, że odjęto od obserwacji następnej obserwację poprzednią.

Kolejnym wykonanym przeze mnie krokiem przetwarzania danych było wyrównanie czasu retencji. Istnieją różne metody rozwiązania tego problemu, najczęściej stosuje się dynamiczną transformację czasową (dynamic time warping), ale ponieważ jest ona wrażliwa na różne intensywności pików i może prowadzić do niezadowolającego wyrównania, w swoich pracach wybrałam do zastosowania COW (correlation optimised warping). Jego zaletą jest wyrównywanie chromatogramów poprzez dopasowanie ich do siebie w odpowiednich fragmentach pod względem kształtu, co nie zmienia znacznie wysokości i pól w ramach pików. Ponadto wyrównanie za pomocą COW nie wymaga wcześniejszego wykrywania piku chromatograficznego, w przeciwieństwie do kilku innych metod. Nadmienię również, że COW nie było

wcześniej zaprogramowane w R. Zatem tak jak we wszystkich innych elementach przetwarzania nie korzystałam z żadnych gotowych pakietów, tylko własnoręcznie napisanych skryptów. Wyjątek stanowił zbiór danych w pracy IB1, gdzie do usunięcia linii bazowej użyłam algorytm „Rolling Ball” z pakietu „baseline” w R ze względu na potrzebę stosowania wymienionych w tej pracy metod na danych, które nie są zróżnicowane jak w pozostałych przypadkach. Stosując algorytm COW chromatogramy podzieliłam na pewną liczbę segmentów, które były rozciągane lub zwężane po czasie. Optymalne wyrównanie, w napisanym przeze mnie programie, odbywało się dla zadanych parametrów m - długość segmentu oraz t - maksymalne dopuszczalne przesunięcie, które było obliczane na podstawie korelacji danego chromatogramu z chromatogramem referencyjnym. Do ustalenia optymalnego wyrównania wykorzystałam macierz, której elementy stanowią maksymalną możliwą skumulowaną funkcję dobroci bazującą na korelacji odpowiednich fragmentach profilu referencyjnego i interpolowanego. W celu wybrania najlepszych parametrów m i t zastosowałam automatyczne COW. Utworzono siatkę parametrów m i t oraz wybrano najlepsze w danej grupie (np. dla odmiany). Do ustalenia siatki parametrów zastosowałam tzw. regułę kciuka oraz przybliżoną średnią szerokość pików w chromatogramie referencyjnym. Do oceny, które parametry są najlepsze wykorzystałam podobieństwo wszystkich wyrównywanych chromatogramów między sobą oraz wpływ wyrównania czasu na zmiany kształtu i pola w obrębie pików. Ponadto wyrównanie czasu retencji w przypadku IB6 odbywało się dla każdej odmiany osobno. Aby wykonać wyrównanie czasu retencji najpierw wybrano chromatogram, do którego czasu retencji były wyrównywane pozostałe chromatogramy. Był to chromatogram najbardziej podobny do wszystkich pozostałych (co zostało oszacowane odpowiednim parametrem). Najpierw wyboru chromatogramu referencyjnego dokonano w obrębie każdej odmiany z osobna, a następnie wśród chromatogramów referencyjnych dla odmian został wybrany chromatogram referencyjny dla danej długości fali (również jako najbardziej podobny do wszystkich w tej grupie). Dla chromatogramów referencyjnych zostało

wykonane wyrównanie czasu retencji względem wspólnego referencyjnego chromatogramu i tak wyrównane chromatogramy referencyjne stały się dalej podstawą do wyrównywania czasu retencji w poszczególnych odmianach.

Następnym niezbędnym tutaj krokiem jest wykrycie pików. Wprowadziłam własną metodę detekcji pików tak, aby można je było analizować statystycznie w całej grupie chromatogramów. Istnieje wiele metod detekcji sygnałów, ale żadna inna znana mi metoda nie wykonywała detekcji wspólnych sygnałów dla grupy chromatogramów, co jest potrzebne do porównywania tych chromatogramów między sobą. Pik w pojedynczym chromatogramie zdefiniowałam jako przedział, w którym wygładzona druga różnica jest ujemna. Wspólne piki dla grupy chromatogramów (w tym przypadku dla wszystkich przy określonej długości fali) zostały wyznaczone jako suma pików dla indywidualnych chromatogramów (suma w sensie dodawania przedziałów). Następnie rozwiązałam problem dekonwolucji pików, czyli nakładania się pików, które mają bliskie sobie czasy retencji. Problem ten rozwiązałam za pomocą własnej metody przy pomocy grupowania hierarchicznego, opisanego szerzej w pracy IB1 oraz następnym rozdziale. Na różnych etapach wprowadziłam eliminację pików nieistotnych, czyli np. zbyt krótkich, aby mogły być prawdziwymi pikami. Niektóre parametry zastosowane podczas przetwarzania danych jak np. m (długość segmentów) i t (przesunięcie) w wyrównywaniu czasu retencji zostały zoptymalizowane automatycznie. Inne parametry, takie jak progi detekcji pików UHPLC-UV, zostały przetestowane z różnymi wartościami oraz przeprowadzono kilkietapowe konsultacje z doświadczalnikami dotyczące liczby i jakości wykrytych metabolitów.

Kolejnym zastosowanym przeze mnie krokiem była integracja chromatogramów w granicach wspólnych pików. W wyniku przetwarzania danych nastąpiła ich duża redukcja. Tak przygotowane dane były danymi wejściowymi gotowymi do przeprowadzenia analizy statystycznej.

Opisane powyżej wstępne przetwarzanie danych różniło się w zależności od własności danych pod względem siatki parametrów m i t , pod względem podziału na grupy, w których wyrównywanie czasu retencji odbywało się

osobno, pod względem separacji pików w zależności od liczby chromatogramów składających się na wspólny pik oraz częstości występowania problemu dekonwolucji. Wspomniany podział na grupy, w których wykonywany był osobno COW, w pracach IB1, IB3, IB6 stanowiły poszczególne odmiany, natomiast w pracy IB4 wszystkie niezbyt liczne chromatogramy stanowiły jedną grupę odmiany *Arabidopsis thaliana*. Ponadto w pracy IB1 zastosowałam inny sposób usunięcia linii bazowej niż w pracach IB3, IB4, IB6, w których zastosowałam różnicowanie danych. W tym przypadku użyłam algorytm „Rolling Ball”, co spowodowane było potrzebą pracy na danych, które nie byłyby zróżnicowane do porównania dwóch metod separacji pików.

Dane z urządzenia chromatograficznego GC-MS, podobnie jak dane z UPLC-UV, mają również trójwymiarową naturę. W tym przypadku wymiarami są czas retencji, stosunek masy do ładunku oraz intensywność (wysokość pików). Dane GC-MS z publikacji IB7 oraz IB8 zostały wyeksportowane do plików NetCDF, na których podstawie zostały wyznaczone sumy z wartości intensywności po skanach oraz zbindowane moimi własnymi skryptami i przy użyciu pakietu „ncdf” w R. Następnie dane przetworzyłam za pomocą pakietu TargetSearch w R. Użyłam bazy danych Golm Metabolome Database jako biblioteki referencyjnej. Czas retencji został przekonwertowany na wskaźnik retencji (RI) metodą wygładzania, a korekcja linii bazowej została wykonana przy użyciu interpolacji liniowej. Specyficzne RI (retention index) dla wszystkich próbek wykryto na podstawie skorelowanych mas selektywnych za pomocą normalizacji mediany i progu korelacji 0,95. Po uśrednieniu skorelowanych mas do dalszej analizy zostały przefiltrowane metabolity o parametrach spełniających założenia wyznaczone przez chemików doświadczalników. Uzyskane dane zostały znormalizowane wg standardu wewnętrznego, jakim był rybitol, a następnie uśrednione po powtórzeniach technicznych. Tak przygotowane dane stanowiły dane wejściowe do analizy statystycznej. W przypadku danych z pracy IB5 były to wyniki otrzymane z urządzenia o mniejszej rozdzielczości i w standardowym formacie. Do celu realizowanego w tej publikacji nie było potrzeby identyfikowania związków, więc dane te nie

wymagały kroków związanych z biblioteką referencyjną. Wykonałam uśrednienie danych po powtórzeniach technicznych, a następnie obliczenie sumy intensywności po masach otrzymując dla każdej próby tzw. TIC (total ion current chromatogram). Każdy TIC reprezentuje intensywność zsumowaną po całym zakresie mas wykrytych w każdym punkcie czasowym retencji. Po tym etapie każda próba biologiczna jest reprezentowana przez jeden chromatogram TIC. Tak przygotowane dane służyły w kolejnych etapach analizy, opisanych w rozdziale 5.

2. Separacja wspólnie wpływających związków w chromatografii za pomocą analizy skupień oraz funkcjonalnej analizy składowych głównych

Nakładanie się pików jest częstym problemem w chromatografii, głównie w przypadku złożonych mieszanin biologicznych jakimi są metabolity. Ze względu na istnienie zjawiska koelucji różnych związków o podobnych właściwościach, separacja pików staje się sporym wyzwaniem. W artykule IB1 opisałam i porównałam dwie metody obliczeniowe separacji takich związków, a także po raz pierwszy zastosowałam je do tego celu i wykonałam to na dużych zbiorach danych chromatograficznych. Ponadto w omawianej publikacji metody te zostały zweryfikowane instrumentalnie. Prowadzą one od surowych obserwacji do danych, które mogą stanowić dane wejściowe do analizy statystycznej. Na początku w obu metodach dane są normalizowane na podstawie masy próbki, usuwana jest linia bazowa, przeprowadzane jest wyrównanie czasu retencji i wykonywane jest wykrywanie pików. W pierwszej metodzie do rozdzielania nakładających się pików zastosowałam analizę skupień, w drugiej metodzie do tego samego celu zastosowałam funkcjonalną analizę składowych głównych (FPCA). Wykonałam symulacje oraz posłużyłam się danymi rzeczywistymi jako przykładami do przedstawienia obu metod i ich porównania. Rzeczywiste dane uzyskano w badaniu zmian metabolomicznych liści jęczmienia (*Hordeum vulgare*) pod wpływem stresu suszy. Uzyskane wyniki

sugerują, że obie metody są odpowiednie do rozdzielania nakładających się pików, ale dodatkową zaletą FPCA jest możliwość oceny zmienności poszczególnych związków występujących w tych samych pikach w różnych chromatogramach. Dlatego metoda ta jest szczególnie przydatna do analiz chromatograficznych mających na celu znalezienie różnic w rozważanych związkach między próbami lub wariantami eksperymentalnymi, w tym w sytuacjach, gdy liczba prób lub wariantów jest duża. Opisanie metody można zastosować do rozdzielenia pików chromatograficznych zarejestrowanych za pomocą detektorów absorbancji światła lub fluorescencji i uzyskanych technikami separacji, takimi jak chromatografia cieczowa lub elektroforeza kapilarna. Omawiane metody po raz pierwszy zostały zastosowane do dekonwolucji pików w danych chromatograficznych.

W przypadku pików podejrzanych o to, że są mieszaniną większej liczby pików, pierwsza metoda bazująca na analizie skupień dzieli zestawy profili chromatograficznych w określonych czasach retencji, na grupy za pomocą grupowania hierarchicznego oraz bootstrapu. W tym celu wykorzystałam pakiet `pvclust` w R. Następnie wykonałam detekcję pików w każdej z grup. Oczywiście w różnych uzyskanych grupach chromatogramów, mogą występować piki o nakładających się czasach retencji jak i piki będące częścią innych pików w innych grupach. Zatem kolejnym zaproponowanym przeze mnie krokiem jest przypisanie pików między grupami za pomocą wprowadzonego „procentu nakładania się” (`percentage of overlap`), podanego wzorem w publikacji IB1. Opisaną w tej pracy procedurą piki o procencie nakładania się powyżej 80% są przypisywane do innych pików.

Natomiast w przypadku metody bazującej na FPCA kluczowe jest zaobserwowanie, że FPCA jest to ciągła alternatywa dla klasycznej, wielowymiarowej analizy składowych głównych (PCA). Tutaj do obliczeń zastosowałam pakiet `fda` w R. Metoda polega na zastosowaniu FPCA do poszczególnych pików uzyskanych w ostatnim etapie przetwarzania danych, którym jest wykrywanie wspólnych pików dla wszystkich chromatogramów. Stężenia związków, analizowanych za pomocą chromatografii, reprezentowane

są przez wartości kilku pierwszych wyników (scores) FPCA, a ich liczba jest określona przez skumulowany procent wyjaśnianej zmienności. Liczba ta jest równa najmniejszej liczbie pierwszych wyników FPCA, dla których skumulowany procent wyjaśnianej zmienności jest większy niż ustalony próg 80%. Następnie dane są przygotowane jako dane wejściowe do analizy statystycznej, np. analizy wariancji z czynnikami znajdującymi się w doświadczeniu.

Do testowania obu metod zastosowałam scenariusz symulacji odpowiadający sytuacji, w której porównuje się dwa warianty eksperymentalne (np. grupy pacjentów, genotypów roślin, traktowań), z których każdy jest reprezentowany przez szereg chromatogramów odpowiadających poszczególnym próbkom (replikacjom). Wykonałam symulacje danych dla dwóch sytuacji: po pierwsze, gdy obserwuje się pojedynczy pik reprezentujący jeden związek, a po drugie, gdy obserwuje się pik reprezentujący mieszaninę dwóch związków, tj. pik podwójny. Dla pojedynczego pikę założyłam, że średnie stężenie związku różni się między dwoma wariantami doświadczalnymi; dla pikę podwójnego założyłam, że średnie stężenia różnią się tylko dla jednego związku z mieszaniny. Symulowany zestaw danych dla każdego wariantu reprezentowany jest przez 50 chromatogramów opartych na 100 punktach czasu retencji. Użyłam 6 funkcji B-sklejanych rzędu 3 do wygenerowania poszczególnych chromatogramów.

Wykonana przeze mnie analiza danych symulowanych wykazała, że obie metody dały oczekiwaną liczbę rozdzielonych pików i obie były odpowiednie do rozwiązania problemu wspólnie wpływających związków. Przeprowadziłam test statystyczny na istotność różnic między wariantami i porównałam liczbę odrzuceń hipotezy zerowej o równości wariantów za pomocą dwóch metod w serii 1000 symulacji, aby odkryć, która metoda jest lepsza do wykrywania różnic między wariantami. W przypadku pojedynczego pikę do przeprowadzenia testu równości średnich zastosowałam test t, w przypadku podwójnego pikę zastosowałam test T^2 Hotellinga, aby przetestować dwuwymiarową hipotezę zerową. W przypadku metody wykorzystującej analizę skupień, obliczenia

przeprowadziłam na wartościach scałkowanych pików, w przypadku metody stosującej FPCA na pierwszym wyniku lub dwóch pierwszych wynikach (scores). W przypadku pojedynczego pików obie metody dały bardzo podobne wyniki. W przypadku podwójnego pików FPCA dała nieco lepsze wyniki, więc metoda ta lepiej zachowała różnice między wariantami.

Przykłady dotyczące danych rzeczywistych odnosiły się do najczęstszych problemów związanych z konwolucją pików. Pierwszy rozważany przykład wybrałam tak, aby pokazać problem wypływania różnych związków przy bardzo podobnych czasach retencji, ale obecnych w różnych chromatogramach, które są analizowane razem. Drugi przykład ilustruje problem nakładających się pików występujących wspólnie na każdym pojedynczym chromatogramie. W obu przypadkach walidacji metod dokonano na podstawie analizy instrumentalnej, która wskazała na istnienie dwóch różnych metabolitów. W pierwszym przykładzie wykazały to również obie metody, w drugim przykładzie wykazała to jedynie metoda stosująca FPCA, podczas gdy metoda wykorzystująca analizę skupień wykazała tylko jedną grupę, zamiast dwóch. Zatem metodą tą nie jest możliwe uzyskanie grup wskazujących na dwa różne pików, w przypadku, gdy wszystkie chromatogramy mają podobny kształt.

W metodzie stosującej analizę skupień stężenia związków we wszystkich próbkach obliczono przez całkowanie chromatogramów w granicach pików otrzymanych po rozdzieleniu. W drugiej metodzie stężenia były reprezentowane przez wartości kilku pierwszych wyników FPCA, przy czym ich liczba była określona przez skumulowany procent wyjaśnionej zmienności większy niż próg 80%.

Metoda rozdzielania pików poprzez grupowanie jest podejściem, które pozwala na ustalenie grup chromatogramów posiadających różne wzorce sygnałów. W przypadku rzeczywistych danych z różnymi pikami w tym samym czasie w różnych profilach, chromatogramy były dobrze pogrupowane, oddzielając w ten sposób pików. Jednak w przypadku nakładających się pików w każdym profilu o podobnym kształcie, grupowanie nie było w stanie stworzyć grup odpowiadających różnym składowym. Gdyby grupowanie mogło

wskazywać na pewne grupy, powstałyby one z powodu różnic w intensywności lub z powodu niewielkich przesunięć czasu retencji, co nie jest interesujące i znaczące. FPCA dało równie zadowalające wyniki w obu rozważanych sytuacjach. Obie metody dekonwolucji pozwoliły na rozróżnienie dwóch pików w przypadku pierwszego przykładu danych rzeczywistych, i rzeczywiście, analiza danych surowych wykazała, że piki miały różne maksima absorpcji. Istnienie dwóch różnych metabolitów zostało również udowodnione przez spektrometrię mas w przypadku drugiego przykładu danych rzeczywistych, ale tutaj prawidłowej odpowiedzi udzielił tylko FPCA. Metoda FPCA daje wiele zmiennych, a liczba użytych zmiennych, czyli funkcjonalnych składowych głównych, może być ograniczona przez próg wyjaśnianej wariancji. W tym konkretnym przypadku każdy pik można interpretować jako złożenie dwóch pików, więc w przypadku tej metody wyniki można zawsze przedstawić w dwuwymiarowym układzie współrzędnych. Taka prezentacja wyników nie zawsze jest możliwa dla metody wykorzystującej grupowanie. Ponadto FPCA nie oddziela pików jawnie, ale wykrywa te o największej zmienności, zapewnia optymalną, wielowymiarową reprezentację pików. Żadna z wcześniej wprowadzonych metod przedstawionych w literaturze nie koncentruje się na zachowaniu różnic między wariantami eksperymentalnymi, co jest kluczowe dla analizy statystycznej. Zastosowanie metody dekonwolucji z wykorzystaniem FPCA do danych chromatograficznych daje tę dodatkową korzyść. Rozważając porównanie między próbami, uwypukla się, w tej metodzie, piki o różnej powierzchni, co jest zgodne z nurtem niecelowanej metabolomiki porównawczej.

3. Analiza statystyczna dużych doświadczeń wieloczynnikowych dla danych metabolomicznych

W większości badań naukowych mamy do czynienia z traktowaniem i kontrolą i istnieje wiele programów do analizy tego typu danych. Jednak w sytuacjach, w których występuje więcej czynników nie odzwierciedlają one

interakcji między czynnikami. Odpowiednie podejście do analizy danych z wieloma czynnikami i dużą liczbą danych na początku mojej drogi naukowej stanowiło duże wyzwanie. We wszystkich analizowanych doświadczeniach z cyklu publikacji mamy do czynienia z większą liczbą czynników. W pracy IB4 pochodzącej z projektu HyperNano (Załącznik 5, IIF11) czynnikami były warianty traktowania oraz punkty czasowe, natomiast w pracach IB3, IB6, IB7, IB8 pochodzących z projektu POLAPGEN-BD (Załącznik 5, IIF7) występowały odmiany, traktowania i punkty czasowe.

Analizę danych metabolicznych dotyczących *Arabidopsis thaliana* traktowanej nanocząsteczkami srebra opisaną w pracy IB4 (str. 3, 9-12, Supplementary Data 1, Supplementary Data 2) przeprowadziłam w celu zbadania ich zmian ilościowych w 3 punktach czasowych (odpowiednio T1, T2, T3), w 13 wariantach traktowania (wraz z warunkami kontrolnymi). Dane przekształciłam przez $\log_2(10^6x)$, w celu sprowadzenia ich rozkładu do rozkładu normalnego. Przeprowadziłam dwuczynnikową analizę wariancji z wariantem traktowania i punktem czasowym jako czynnikami. Istotne efekty traktowania, punktu czasowego i ich interakcji wybrałam przy p-wartości $P < 0,01$. Aby wykryć statystycznie istotne związki o obniżonej lub podwyższonej zawartości w stosunku do kontroli, przetestowałam różnice między średnimi wartościami dla każdego wariantu traktowania i punktu czasowego, stosując test Fishera (FPLSD, na poziomie 1%) i wizualizowałam je dla cech o istotnej statystycznie interakcji. Różnice między średnimi wartościami traktowania również poddałam testowi Fishera (FPLSD, na poziomie 1%) i wizualizowałam dla cech o istotnym wpływie traktowania, z wyłączeniem cech o istotnej interakcji. W ten sam sposób przetestowałam również różnice między średnimi wartościami punktów czasowych oraz zwizualizowałam cechy z istotnym efektem punktu czasowego, z wyłączeniem cech o istotnej interakcji. Analizę statystyczną przeprowadziłam we własnych skryptach napisanych w Genstat. Analiza składowych głównych próbek pogrupowanych według punktów czasowych wykazała, że trzeci punkt czasowy był znacznie różny od poprzednich. Wszystkie metabolity pogrupowałam w odniesieniu do reakcji w czasie. Na podstawie analizy

wariancji metabolity zostały sklasyfikowane pod względem znaczenia: (i) średnich różnic między wariantami, (ii) średniego efektu punktu czasowego oraz (iii) interakcji między wariantami traktowania i punktami czasowymi. Niektóre wyróżniające się w analizie związki zostały opisane bardziej szczegółowo na podstawie wyników, np. kamaleksyna. Na podstawie analizy wykazano, że nanocząsteczki srebra mogą wywoływać istotne zmiany w metabolizmie wtórnym roślin *Arabidopsis thaliana*. Stwierdzono, że nanocząstki wpływają na 47 metabolitów wtórnych w zależności od ich wielkości, stężenia i punktu czasowego. Wykazano również, że nanocząsteczki srebra mogą działać jako elicytory fitoaleksyn, takich jak kamaleksyna i pokrewne związki w *Arabidopsis thaliana*. Zdolność nanocząstek do indukowania produkcji fitoaleksyn i innych metabolitów wtórnych w roślinach pokazała, że zjawisko to można wykorzystać do zwiększenia odporności roślin na patogeny, o ile zostaną zastosowane w zoptymalizowany sposób.

Celem pracy IB3 (str 3-6, 9-16, 18-19, Supplementary Table S1 A, S1 B, S2, S3, S4) była korelacja między dynamiką zmian profili związków fenolowych a zjawiskiem zewnętrznym dla ośmiu odmian jęczmienia pod wpływem suszy w różnych punktach czasowych. Zatem analizie statystycznej poddałam nie tylko zbiór danych metabolomicznych, ale również fenotypowych oraz analizowałam ich korelację. Do wyników wstępnego przetwarzania danych chromatograficznych przekształconych przez $\log_2(10^9x)$ (w celu sprowadzenia ich rozkładu do rozkładu normalnego) oraz do danych fenotypowych zastosowałam analizę wariancji opartą na algorytmie REML (restricted maximum likelihood), do każdej cechy niezależnie. Istotność zmienności sprawdziłam testami opartymi na aproksymacji F statystyki Walda. W przypadku cech fenotypowych mierzonych na dojrzałych roślinach model ANOVA obejmował (stałe) efekty odmiany (V), traktowania (D) i interakcji VD. Istotne efekty wybrałam przy p-wartości $P < 0,01$. W przypadku metabolitów model ANOVA obejmował (stałe) efekty odmiany (V), czas obserwacji (T), traktowanie (D) i efekty wszystkich możliwych interakcji. W celu korekty porównań wielokrotnych do wyznaczenia istotnych efektów, niezależnie w każdej grupie

V, T, D, VT, VD, TD, VTD, zastosowałam wskaźnik błędu FWER (family-wise error rate) przy użyciu poprawki Bonferroniego dla liczby metabolitów. Interesujące do interpretacji były wyniki dotyczące efektów D, VD, TD i VTD, ponieważ reprezentują one efekty niedoboru wody w poziomie absorpcji metabolitów oraz modyfikacje tych efektów w zależności od odmiany, czasu obserwacji i interakcji między nimi. Dlatego te właśnie wyniki są omawiane w publikacji IB3. Podane zostały liczby metabolitów o istotnych różnicach w akumulacji dla różnych wariantów suszy. Dane wskazują na większą mobilizację metabolitów fenolowych podczas deficytu wody na wczesnym etapie rozwoju niż w fazie liścia flagowego w wariacie suszy I + II, co zostało przetestowane testem niezależności χ^2 z wykorzystaniem permutacji. Efekty traktowania obliczyłam jako różnice między stężeniem w niedoborze wody a stężeniem w kontroli. Analogicznie obliczyłam średnie efekty czynników z odpowiednich wartości średnich. Metabolity, dla których efekty niedoboru wody były istotnie różne w czasie, czyli należące do grupy TD, pogrupowałam na 4 klasy zgodnie z profilem czasowym efektów przy użyciu kryterium maksymalnej międzygrupowej sumy kwadratów. Obliczone zostały odległości Mahalanobisa między grupami. Taki podział ze względu na profile czasowe efektów interakcji TxD (uśrednionych po odmianach) dla każdego wariantu suszy osobno zaowocował szeregiem wniosków dla poszczególnych metabolitów. Do wizualizacji podobieństw między odmianami a cechami fenotypowymi wykorzystano biploty i przeprowadzono wielowymiarową analizę dla cech, dla których zaobserwowano efekty suszy specyficzne dla odmiany. Policzono również współczynniki korelacji Pearsona między istotnymi efektami specyficznymi dla odmiany dla wariantu suszy I, dla cech fenotypowych i metabolomicznych przy dwóch poziomach istotności: $P < 0,05$ oraz $P < 0,01$. Najbardziej znaczący wzrost zawartości związków w niedoborze wody zaobserwowano dla glikokoniugatów chryzoeriolu i apigeniny acylowanych metoksylovanymi kwasami hydroksycynamonowymi, które wzmacniały skuteczność ochrony przed promieniowaniem UV. Ponadto inne dobre przeciwutleniacze, takie jak pochodne luteoliny i hordatyny, były również

indukowane niedoborem wody. Zróżnicowanie strukturalne metabolitów pod względem ich zawartości zmieniło się w odpowiedzi na niedobór wody, co wskazuje na ich wielotorową aktywność w warunkach stresu. Rośliny narażone na deficyt wody w stadium siewek mobilizowały dwa razy więcej metabolitów niż rośliny narażone na ten stres w stadium liścia flagowego. W długotrwałej aklimatyzacji uczestniczyły specyficzne metabolity, takie jak kwasy metoksyhydroksycynamonowe. Szereg różnego rodzaju dodatkowych wniosków można znaleźć w pracy IB3.

W pracy IB6 (str 899-902, 906-908, 911, Supporting Information: Table S1, Table S3), w której w różnych częściach współuczestniczyłam, analizę statystyczną przeprowadzono dla 98 cech (spośród 135), dla których liczba obserwacji niezerowych we wszystkich warunkach była wystarczająca. Celem analizy było zbadanie zmian ilościowych metabolitów wtórnych w dwóch punktach czasowych (T1, T2) podczas stosowania suszy w jęczmieniu jarym. Obserwacje równe zeru (poniżej poziomu wykrywalności) zostały zastąpione przez połowę minimalnej niezerowej obserwacji dla każdego metabolitu. Następnie w celu sprowadzenia rozkładu danych do rozkładu normalnego obserwacje przekształcono przez $\log_2(10^9x)$. Analiza statystyczna została przeprowadzona osobno dla T1 i T2 za pomocą własnych skryptów napisanych w Genstat oraz systemie R. Analizę wariancji zastosowano osobno dla T1 i T2, aby przetestować istotność różnicy między warunkami kontrolnymi a suszą w odniesieniu do poziomu metabolitu (efekt stały, $P < 0,05$, z poprawką Bonferroniego dla liczby metabolitów), a także różnice między liniami RIL i interakcją RIL z traktowaniem (efekty losowe, komponenty wariacyjne były istotne, jeśli były większe niż trzy razy błąd standardowy). Wpływ suszy mierzono różnicą średnich wartości między suszą a warunkami kontrolnymi, co wskazywało na dodatnią (wzrost) lub negatywną (spadek) regulację akumulacji metabolitów podczas suszy (istotność poszczególnych efektów suszy dla linii została oceniona przez test t przy $P < 0,05$). Analizę wariancji oddzielnie dla kontroli i suszy zastosowano do oszacowania odziedziczalności cech metabolicznych. Analizę współrzędnych głównych (ze współczynnikami

podobieństwa opartymi na odległościach euklidesowych) przeprowadzono dla wszystkich cech i przeprowadzono grupowanie hierarchiczne dla linii RIL i cech (za pomocą funkcji heatmap2 w R) dla podzbioru cech wykazujących interakcję linii i traktowania. Tak przeprowadzona analiza współrzędnych głównych wykazała, że dyspersja linii była największa w T2 w warunkach kontrolnych (10 dni), a różnice między liniami RIL w warunkach kontrolnych i suszy były większe w T2 niż w T1 (5 dni). Analiza statystyczna przeprowadzona dla omawianego tu doświadczenia pozwoliła wysunąć szereg wniosków. Udało się zidentyfikować metabolity o akumulacji, która była głównie zróżnicowana w dwóch reżimach nawadniania. Znalezienie metabolitów z efektem interakcji linii i traktowania było szczególnie ważne, ponieważ są to związki, które w reakcji na suszę zmieniają się różnie w różnych liniach, a zatem są interesujące z punktu widzenia genetyki roślin. Analiza statystyczna wykazała wyższy poziom metabolitów w odmianie Maresi niż w CamB, a podczas stresu suszy zaobserwowano odmienne przeprogramowanie metaboliczne u tych dwóch odmian rodzicielskich. Pozwoliło to wysnuć wniosek, że obie odmiany rodzicielskie stosują różne systemy przeciwutleniaczy podczas stresu suszy. Liczba metabolitów o wyżej zawartości w suszy spadła z T1 do T2 w Maresi i wzrosła w CamB. Może to oznaczać lepszą ochronę metabolomiczną liści w CamB niż w Maresi. W przypadku populacji RIL na podstawie analizy wnioskujemy, że rośliny reagowały na warunki suszy wzrostem w czasie liczby metabolitów o wyżej zawartości w suszy. Co więcej, zróżnicowanie reakcji linii na suszę również wzrosło w czasie. Analiza pozwoliła również wyodrębnić pewne grupy chemiczne metabolitów i wysnuć wnioski o ich reakcji na suszę. Interesujące wnioski wysnuto również dla poszczególnych związków i ich reakcji na stres suszy w poszczególnych punktach czasowych. Wszystko to pokazało, że zastosowanie stresu drastycznie wpłynęło na dynamikę metabolitów, szczególnie w T2.

Uzyskane dane dotyczące metabolitów pierwotnych w jęczmieniu jarym analizowanych pod wpływem suszy (IB7, str. 3-10) po wstępnym przetworzeniu znormalizowałam do standardu wewnętrznego, którym był rybitol i uśredniłam po powtórzeniach technicznych. Analizę statystyczną, w której

współuczestniczyłam, przeprowadzono w Genstat, tylko w przypadku metabolitów, w których co najmniej 20% obserwacji było powyżej poziomu wykrywalności. W przypadku cech o wartości poniżej poziomu wykrywalności przyjęte zostały wartości stanowiące połowę niezerowej wartości minimalnej. Następnie obserwacje zostały przekształcone przez $\log_2(10^6x)$, w celu sprowadzenia rozkładu danych do rozkładu normalnego. Analizę wariacji wykonano dla każdego metabolitu w modelu ze stałymi efektami linii i suszy; istotność efektów oceniono na podstawie wskaźnika błędu FWER (family-wise error rate) $< 0,05$ przy użyciu poprawki Bonferroniego. Efekty suszy zostały oszacowane za pomocą różnicy (kontrola - susza). Dla form rodzicielskich wykonano klasyfikację ze względu na istotność i znak efektu suszy, w podziale na liście i korzenie, co pozwoliło porównać liczbę metabolitów między wymienionymi wariantami. Podobnie jak w przypadku metabolitów wtórnych została wykonana analiza współrzędnych głównych (ze współczynnikami podobieństwa opartymi na odległościach euklidesowych) oraz przeprowadzono grupowanie hierarchiczne oparte na odległości euklidesowej, metodzie pełnego wiązania oraz algorytmie (funkcje heatmap.2 oraz pvclust w R dla 1000 symulacji). Analiza statystyczna pozwoliła wysnuć wnioski, że reakcja linii jęczmienia na suszę była raczej konserwatywna; dla większości genotypów uległ zmianie skład metabolomu, niezależnie w liściach i korzeniach. Na podstawie analizy wariacji sklasyfikowano metabolity co do istotności różnicy między liniami, efektu suszy (rozumianego jako różnica między poziomem metabolitu w suszy a poziomem metabolitu w kontroli), oraz interakcji linia x susza. Ujawnione zmiany w akumulacji niektórych metabolitów, np. proliny i innych aminokwasów, węglowodanów lub kwasów karboksylowych były podstawową strategią roślin do uzyskania tolerancji na suszę. Na podstawie wyników wyciągnięto pewne ogólne wnioski. Zmiany metabolitów biorących udział w odpowiedzi na suszę w jęczmieniu były podobne jakościowo między badanymi liniami, natomiast różniły się ilościowo. Zidentyfikowano grupy związków nagromadzonych podczas suszy. Wyciągnięto wnioski dotyczące specyfiki narządu (liść, korzeń) w odpowiedzi na suszę na poziomie

metabolomu we wszystkich grupach metabolitów. Ponadto znaleziono metabolity, które różnicują badane genotypy pod wpływem suszy i te związki można uznać za potencjalne biomarkery związane z tolerancją na suszę w jęczmieniu.

W publikacji IB8 z danymi metabolomicznymi zestawione zostały dodatkowo dane proteomiczne, stanowiąc również przedmiot moich analiz. Analizując dane proteomiczne dotyczące jęczmienia badanego pod wpływem suszy z pracy IB8 (str. 3, 5), istotność statystyczną względnej zmiany akumulacji plam białkowych określiłam za pomocą testu t-Studenta. Natomiast dane metabolomiczne z tej samej publikacji (IB8: str. 4-5) poddałam analizie wariacji w Genstat w celu zidentyfikowania istotnych różnic między średnimi dla badanych odmian, między warunkami kontrolnymi a suszą oraz w celu zidentyfikowania istotnej interakcji odmiana x traktowanie suszą. Analiza pozwoliła wysnuć wnioski dotyczące mechanizmów w metabolomie i proteomie, które mogą przyczynić się do zwiększonej tolerancji na suszę obserwowanych odmian. Zaobserwowano kilka wyraźnych zmian w akumulacji białek i metabolitów, które były podobne w obu badanych odmianach.

4. Analiza sieci korelacyjnych i różnicowych sieci korelacyjnych

Kolejnym ważnym elementem analizy zawartym w omawianych pracach jest analiza sieci korelacyjnych i różnicowych sieci korelacyjnych. Obecnie, gdy wysokoprzepustowe urządzenia generują ogromną ilość danych, uzyskuje się wiele cech do analizy, a ich integracja w dużych interdyscyplinarnych projektach naukowych daje dodatkowe wnioski. Do takiej integracji wykorzystuje się sieci korelacyjne bazujące na korelacjach lub ich przekształceniach. Taka analiza dostarcza szerszy ogląd analizowanego dużego zbioru danych, a także daje możliwość wnioskowania o grupach cech, jak i o wyraźnych zależnościach między indywidualnymi cechami. Taka analiza daje doświadczalnikom, biologom, chemikom wyraźny obraz korelacji wielu cech dzięki wizualizacji. Sieci korelacyjne są popularne w wielu najnowszych badaniach w nauce.

Natomiast różnicowe sieci korelacyjne po raz pierwszy wprowadziłam w publikacji IB6 (Figure S2) dla metabolitów wtórnych, w celu badania istotnych różnic między korelacjami w różnych traktowaniach, w tym przypadku między warunkami kontrolnymi a suszą. Można je stosować do porównania zmian w korelacjach cech między dwoma warunkami, punktami czasowymi, traktowaniami itd.

W moim cyklu prac zaprezentowałam metodę analizy dużych zbiorów danych za pomocą sieci korelacyjnych oraz różnicowych sieci korelacyjnych w celu porównania korelacji między metabolitami pierwotnymi (IB7, str. 4, 9, 11-12) i metabolitami wtórными (IB6, str. 902-904, 908-909, 911, Supporting information: Figure S2, Table S2, Methods S1). Analizę sieci korelacyjnych przeprowadziłam za pomocą pakietu WGCNA w systemie R. Macierz korelacji Pearsona przekształciłam w macierz przyległości wierzchołków i podniosłam ją do potęgi wyznaczonej kryterium topologicznym oraz przekształciłam do macierzy TOM (topological overlap matrix). Elementy macierzy TOM wykorzystałam do wizualizacji sieci. Moduły, czyli grupy silnie skorelowanych metabolitów wykryłam poprzez grupowanie hierarchiczne oraz algorytm *dynamic tree cut*. Wyznaczyłam także *huby*, czyli cechy o największej liczbie silnych korelacji z innymi cechami. W pracy IB6 aby porównać sieci korelacyjne dla metabolitów wtórnych z różnych traktowań i punktów czasowych przeanalizowałam parametry takie jak: *connectivity* i *clustering coefficients* dla całych sieci oraz dla każdego modułu. W IB7 *clustering coefficients* wykorzystałam do porównania struktury korelacji dla metabolitów pierwotnych między warunkami kontrolnymi a warunkami pod wpływem suszy, zarówno dla liścia jak i korzenia. W celu znalezienia korelacji metabolitów lub białek, które istotnie różnią się między suszą a kontrolą lub między punktami czasowymi wprowadziłam po raz pierwszy różnicowe sieci korelacyjne. Skonstruowałam je w pracach IB6, IB7 na podstawie testu opartego na transformacji Fishera Z z poprawką Bonferroniego. Wizualizację sieci wykonałam w programie Cytoscape.

W przypadku metabolitów pierwotnych (IB7) skonstruowałam 4 sieci korelacyjne metabolitów zaobserwowanych w warunkach kontrolnych i suszy, osobno w liściu i korzeniu. Skonstruowałam także 2 różnicowe sieci korelacyjne dla liścia i korzenia, w których każda krawędź wskazuje dla pary metabolitów istotną zmianę w korelacjach między warunkami kontrolnymi a suszą. W przypadku metabolitów wtórnych (IB6) skonstruowałam 4 sieci korelacyjne metabolitów zaobserwowanych po 5 i 10 dniach (co odpowiadało punktom czasowym T1 i T2), osobno w warunkach kontrolnych i suszy, a także 2 różnicowe sieci korelacyjne dla metabolitów wtórnych zaobserwowanych po 5 i 10 dniach, w celu badania istotnych różnic w korelacjach związków między warunkami kontrolnymi a suszą. Moduły wyznaczyłam dla wszystkich 4 wariantów sieci korelacyjnych, osobno w warunkach kontrolnych i suszy, w 2 punktach czasowych. Ponadto w publikacji IB6 skonstruowano sieć metabolitów wtórnych z markerami SSR i SNP na podstawie analizy loci cech ilościowych (mQTL).

Opisane powyżej metody pozwoliły na sformułowanie wielu wniosków dotyczących korelacji metabolitów w jęczmieniu jarym. W przypadku metabolitów pierwotnych (IB7) struktura korelacji była silniejsza w liściach niż w korzeniach, co zostało wywnioskowane na podstawie *clustering coefficient*. Większość dużych korelacji, zobrazowanych w sieci krawędziami, była dodatnia, z wyjątkiem trzech ujemnych korelacji w suszy w liściu i jednej w kontroli w korzeniu. Siła korelacji zmieniła się istotnie między warunkami kontrolnymi a suszą w korzeniu. Jednakże, zarówno w liściu, jak i korzeniu istniały podsieci składające się głównie z aminokwasów, które zostały zachowane między warunkami kontrolnymi a suszą. Metabolity pierwotne reprezentowane przez kwadraty, czyli wierzchołki z największą liczbą połączeń, należą do grup: aminokwasów, lipidów lub grupy nieznanymi metabolitów (lipidy tylko w warunkach suszy). Sieci różnicowe dla metabolitów pierwotnych w liściach jęczmienia wykazały, że ogólnie korelacje między tymi metabolitami zmniejszyły się w suszy w porównaniu z warunkami kontrolnymi. Zostały

wskazane nieliczne pary metabolitów, których korelacje istotnie wzrosły z warunków kontrolnych do suszy.

Korelacje metabolitów wtórnych (IB6) były silniejsze w warunkach kontrolnych niż podczas suszy przez cały okres traktowania suszą. Metabolity wtórne, które wykazywały szczególne zmiany w korelacji z innymi, charakteryzowały się także różnymi reakcjami linii RIL w różnych warunkach suszy. Analiza poszczególnych modułów dla każdej sieci korelacyjnej, osobno w warunkach kontrolnych i suszy, po 5 i 10 dniach, pozwoliła wysnuć wnioski o poszczególnych grupach metabolitów wtórnych. Na przykład moduł „cyan” dotyczący sieci w suszy po 10 dniach składał się z dwóch metabolitów: „isoorientin 2''-O-glucoside” oraz „isovitexin 7-O-glucoside”: oba metabolity pozostały razem w jednym module we wszystkich sieciach. Konstrukcja różnicowej sieci korelacyjnej pozwoliła zaobserwować znaczny wzrost liczby istotnych zmian korelacji w czasie przechodząc od warunków kontrolnych do suszy. Otrzymano również szereg wniosków dotyczących sieci metabolitów wtórnych połączonych z markerami molekularnymi SSR i SNP, skonstruowanej za pomocą analizy mQTL, po 10 dniach trwania suszy. Między innymi metabolity wtórne, należące do tego samego modułu, wykazały tendencję do łączenia się ze wspólnymi markerami; np. skorelowane metabolity o nr 13, 45, 61, 100 i 104 (IB6) zostały połączone z SNP 3101-111. Niektóre metabolity zostały połączone z izolowanymi markerami, ale znaleziono również grupy związków ze wspólnymi grupami QTL. Niektóre połączenia między takimi grupami były statystycznie istotne tylko w warunkach kontrolnych.

5. Estymacja macierzy kowariancji dla danych chromatograficznych

Jednym z elementów analizy danych chromatograficznych była analiza struktury kowariancyjnej za pomocą regularyzacji oraz estymacji macierzy kowariancji dla danych metabolomicznych z jęczmienia pod wpływem suszy. Celem takiej estymacji jest lepsze poznanie struktury danych, aby móc przeprowadzać w przyszłości jeszcze dokładniejszą analizę statystyczną

dostosowaną do tego typu danych. Rozważałam w tym przypadku olbrzymi zbiór danych, w których liczba mierzonych cech przewyższa liczbę niezależnych prób, a zatem trudno jest estymować macierz kowariancji. Jednym z rozwiązań tego problemu jest nałożenie na macierz kowariancji pewnej struktury zależności, a ponieważ nie wiadomo, w jaki sposób cechy mogą być ze sobą powiązane, starano się „dopasować” strukturę ze zbioru znanych, rozważanych w literaturze struktur kowariancyjnych.

W pracy IB5 koncentrowałam się na wybraniu odpowiedniej struktury kowariancyjnej, co nazywa się regularyzacją. Rozważałam strukturę całkowicie symetryczną, trójdziagonalną i pięciodziagonalną Toeplitza oraz autoregresję pierwszego rzędu. Problem charakteryzacji struktury kowariancji zbadaliśmy w pracy IB5 za pomocą dwóch metod. Pierwsza oparta jest na normie Frobeniusa, a druga na entropijnej funkcji straty. Dane analizowane w tej pracy pochodziły z urządzenia chromatograficznego GC-MS, które jak już wspomniano powyżej ma trójwymiarową naturę. Takie surowe dane po wstępnym przetworzeniu i przygotowaniu do dalszej analizy zostały zlogarytmowane. Analizowane dane dotyczyły metabolitów pierwotnych.

Pierwszą rozważaną strukturą jest struktura całkowicie symetryczna, czyli struktura, w której wariancje cech są jednakowe i kowariancje cech również są jednakowe. Zatem tym przypadku obserwacje są jednakowo skorelowane, co oznacza, że współczynnik korelacji nie zależy do odległości między cechami. Następną strukturą jest macierz Toeplitza, która ma jednorodne wariancje i niejednorodne korelacje między elementami. Skoncentrowano się na strukturze Toeplitza trójdziagonalnej i pięciodziagonalnej. Natomiast autoregresja pierwszego rzędu charakteryzuje się tym, że wariancje obserwacji są jednorodne, a wielkość korelacji między dwiema obserwacjami zależy od odległości (czasu retencji w danych GC-MS) między nimi.

Niech A będzie zbiorem struktur: $A = \{\Psi_{CS}, \Psi_{T_1}, \Psi_{T_2}, \Psi_{AR}\}$. Spośród wymienionych macierzy celem było wybranie tej, która najlepiej odzwierciedla potencjalną strukturę kowariancyjną. Takie działanie to regularyzacja lub, innymi słowy, dobranie najbliższej struktury $\Psi \in A$ w sensie wybranej funkcji.

Zatem interesująca jest w tym przypadku minimalizacja odległości między macierzą kowariancji, a strukturą, $\min_{\Psi \in A} f(\Omega, \Psi)$. Nie jest znana prawdziwa macierz kowariancji Ω , więc do badania odległości wykorzystałam jej estymator największej wiarygodności:

$S = \frac{1}{n} X_{1n} X^T$, gdzie X jest macierzą obserwacji oraz $Q_{1n} = I_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T$, gdzie I_n jest macierzą jednostkową stopnia n , a $\mathbf{1}_n$ jest n -wymiarowym wektorem jedynek.

Funkcjami, które były minimalizowane były odpowiednio norma Frobeniusa i entropijna funkcja straty, zdefiniowane następująco:

$$f_F(S, \Psi) = \|S - \Psi\|_F$$

oraz

$$f_E(S, \Psi) = \text{tr}(S^{-1} \Psi) - \ln |S^{-1} \Psi| - m.$$

Minimum odległości w zależności od metody zostały zdefiniowane następująco

$$\min_{\Psi \in A} f_F(S, \Psi),$$

$$\min_{\Psi \in A} f_E(S, \Psi).$$

i dla rozważanych struktur kowariancyjnych minima te są znane z literatury. Do wyznaczenia minimum w sensie normy Frobeniusa użyto algorytmu przedstawionego w pracy IIA3. Macierz próbkowa S jest nieosobliwa. Aby sobie poradzić z tym problemem podzieliłam macierz na 3 podzbiory i dla każdego z nich obliczyłam macierz S . Wykonano symulacje, aby zweryfikować, czy odległości w sensie normy Frobeniusa i entropijnej funkcji straty właściwie wyznaczają prawdziwą macierz kowariancji z danych symulowanych z zadaną strukturą kowariancji. Zauważono, że obie funkcje rozpoznają strukturę właściwie. Odległości w sensie normy Frobeniusa i entropijnej funkcji straty trudno porównać, więc zostały znormalizowane. Obliczono odległości i

znormalizowane odległości dla rozważanych struktur kowariancyjnych, dla 3 podzbiorów. Najlepiej dopasowaną strukturą w sensie normy Frobeniusa jest struktura kompletnie symetryczna lub autoregresja pierwszego rzędu, natomiast w sensie entropijnej funkcji straty struktura kompletnie symetryczna. Porównano parametry struktury kowariancyjnych dla trzech podzbiorów trzema metodami. Otrzymano różne estymatory z powodu użycia różnych metod, w których funkcje prowadziły do różnych rozwiązań. Jednakże zaobserwowano, że estymatory są porównywalne dla struktury kompletnie symetrycznej, trójdzielnej i pięciokątnej Toeplitza są porównywalne. Norma Frobeniusa jako mniej obciążona od entropijnej funkcji straty została w pracy IB5 wskazana jako bardziej trafna do regularyzacji oraz estymacji. Zatem najlepiej dopasowaną strukturą okazała się struktura wyznaczona w sensie normy Frobeniusa, tj. struktura kompletnie symetryczna lub autoregresja pierwszego rzędu. Jednakże warto również pamiętać, że entropijna funkcja straty dała niewielkie wskazanie na korzyść struktury kompletnie symetrycznej. Uzyskane w tej pracy wyniki były badaniami wstępnymi nad regularyzacją macierzy kowariancji przy pomocy znanych z literatury metod. Prowadzę dalsze badania nad strukturą blokową macierzy kowariancji.

6. Meta-analiza danych metabolomicznych

Jednym z aspektów analizy danych metabolomicznych prezentowanego cyklu była meta-analiza metabolitów z grupy flawonoidów o potencjale antywirusowym przeciw koronawirusowi, w tym szczep SARS-CoV-2, który zapoczątkował epidemię w Chinach. W pracy IB2 przeanalizowałam wszystkie publicznie dostępne dane z eksperymentów naukowych, które zawierają fitozwiązki znane z blokowania rozwoju koronawirusa.

Wszystkie koronawirusy produkują białka, które wywierają negatywny wpływ na nasz układ odpornościowy. Z literatury wiemy, że szczególnie trzy roślinne produkty naturalne: roifolina, pektolinaryna i herbacetyna z grupy fawonoidów, blokują aktywność tego białka. Łączą się z nim i tym samym

deaktywują białko, unieszkodliwiając wirusa. W publikacji IB2 skoncentrowano się na najlepszym źródle metabolitów o potencjale przeciwwirusowym przeciwko koronawirusowi, ponieważ problem ten nie był wcześniej rozważany w literaturze i może być pomocny w walce z SARS-CoV-2.

Przeprowadziłam bioinformatyczne meta-analizy największych na świecie dostępnych publicznie baz danych naukowych: Metabolights i Prime, jak również publikacji naukowych, w których zidentyfikowano opisane substancje. Bazy te zawierają dane metabolomiczne pochodzące z wysokoprzepustowych metod analitycznych m.in. chromatografii cieczowej sprzężonej ze spektrometrią mas. Bazy danych przeszukiwano słowami kluczowymi: rhoifolin, pectolinarin, herbacetin. Wyszukałam zestaw eksperymentów dla każdej z szukanych fraz. Przeanalizowałam rośliny znane z dużej aktywności fitochemicznej: jęczmień, pszenica, kłosownica dwukłosowa, a także ostrożeń, rosziczka, soja oraz figus. Wszystkie zestawy danych dotyczyły różnych rodzajów traktowania roślin i różnych warunków doświadczalnych. Dlatego do analizy wybrałam tylko próbki z warunków kontrolnych.

W przypadku rosziczki analizowaną cechą był procent zawartości herbacetyny w łącznej zawartości wszystkich znanych z literatury metabolitów fenyłowych w każdej próbce. Obliczyłam średnią wartość cechy dla wszystkich próbek. W przypadku figusa, ostrożenia, jęczmienia, pszenicy, kłosownicy dwukłosowej i soi analizowana cecha stanowi procent rozpatrywanej zawartości metabolitu w łącznej zawartości wszystkich znanych z literatury metabolitów w każdej próbce. W przypadku jęczmienia przeprowadziłam dwuczynnikową analizę wariancji dla czynników: odmiana i punkt czasowy (etap wzrostu roślin). Tylko dla czynnika – odmiana, różnica między średnimi była istotna statystycznie, przy p-wartości równej 0,0106. Wykonałam dodatkowo jednoczynnikową analizę wariancji dla czynnika - odmiana. Średnia zawartość roifoliny była największa dla odmian: Stratus i Morex. Różnice między średnimi dla odmian badałam pod kątem istotności stosując test HSD Tukeya, Tukey's honest significant difference test (na poziomie 5%). Wykonałam mapę ciepła ze średnimi wartościami procentowymi roifoliny dla każdej odmiany jęczmienia i w

sześciu punktach czasowych. Dendrogramy zobrazowały podział na grupy odmian i punktów czasowych. Stratus wyraźnie się wyróżnił. Zawartość procentowa roifoliny w dwóch liniach kłosownicy: Bd21 i Bd3-1 oraz pszenicy, w liściach i korzeniach, obliczona wśród wszystkich znanych literaturowo metabolitów odpowiednio dla liści i korzeni Bd21, Bd31 oraz pszenicy zaprezentowałam przy pomocy boxplotów. Analizę statystyczną oraz wszystkie wizualizacje wykonałam w R.

Na podstawie meta-analizy wywnioskowałam, że wszystkie badane w publikacji IB2 rośliny należy wziąć pod uwagę w badaniach nad opracowaniem leków przeciwko koronawirusom, w tym SARS-CoV-2, ze względu na wysoką zawartość roifoliny, w szczególności w liściach figusa i jęczmienia, ale także antywirusowe metabolity przeciwwirusowe w rosicdze i ostrożeńiu. Pomocne mogą także okazać się dalsze badania dotyczące pszenicy, kłosownicy oraz soi. Planowane są dalsze badania w tym kierunku.

Podsumowanie

W zaprezentowanym cyklu prac opracowałam różne chemometryczne metody analizy danych metabolomicznych występujące w różnych stadiach analizy wykonywanej pod różnymi aspektami, począwszy od przetwarzania wielowymiarowych surowych danych, separacji związków, po analizę statystyczną, analizę sieci korelacyjnych, różnicowych sieci korelacyjnych, estymację macierzy kowariancji oraz meta-analizę.

Opracowane metody chemometryczne w badaniach metabolomu roślin da się zastosować do wszelkich danych pochodzących z wymienionych instrumentów, a w przypadku analizy danych o podobnym układzie doświadczalnym oraz w przypadku sieci korelacyjnych i różnicowych sieci korelacyjnych do wszelkich danych omicznych. Poszczególne elementy opisywanej analizy statystycznej w zależności od potrzeb i warunków, można zastosować do różnych danych biologicznych oraz chemicznych, w różnych układach doświadczalnych. Opracowaną meta-analizę związków chemicznych

można przeprowadzić w odniesieniu do wielu innych chorób i zastosowań badanych związków.

Ponadto moje zainteresowania naukowe obejmują szeroko rozumiane zagadnienia związane z przetwarzaniem i analizą danych wysokoprzepustowych pochodzących z roślin oraz algorytmy i metody algebry liniowej, które mogą być zastosowane do tego typu danych.

W swojej pracy naukowej zajmuję się analizą danych chromatograficznych dotyczących roślin i badaniem metabolitów w różnych układach doświadczalnych i pochodzących z różnego rodzaju urządzeń (np. LC-MS, GC-MS, UPLC-UV). Ponadto moje badania koncentrują się na analizie statystycznej danych metabolomicznych, fenotypowych oraz dotyczących lipidów w roślinach oraz na analizie poziomu ekspresji genów. Analizowałam dane RNA-seq i CHIP-seq dotyczące *Arabidopsis thaliana* w celu zbadania podstaw molekularnych plastyczności kontroli czasu kwitnienia w reakcji na czynniki środowiskowe (temperaturę i długość dnia). Wykonałam również analizę sieci korelacyjnych oraz konstruuje sieci korelacyjne różnicowe w celu znalezienia korelacji cech, które istotnie różnią się w dwóch badanych traktowaniach, a także wizualizację wymienionych sieci. Przeprowadziłam analizę dużych zbiorów danych dotyczących metabolitów pierwotnych i wtórnych pochodzących z jęczmienia jarego pod wpływem suszy, metabolitów wtórnych w liściach traw pod wpływem suszy oraz zasolenia, w roślinach z rodzaju *Kalanchoe*, ziarnach różnych odmian pszenicy jarej poddanej infekcji patogena *Fusarium*, w różnych gatunkach roślin z rodzaju *Passiflora*, w roślinach *Brachypodium* infekowanych *Fusarium*, a także prowadziłam analizy badań na epilobium i szałwi. Wykonałam detekcję metabolitów o zróżnicowanej akumulacji w gatunkach z rodziny Brassicaceae, w *Arabidopsis thaliana* traktowanej różnymi grzybami podczas niedoboru fosforu oraz w roślinach Poaceae traktowanych *Fusarium*. Wykonałam analizę danych RNA-seq i CHIP-seq dotyczącą *Arabidopsis* oraz analizę statystyczną danych pochodzących z jęczmienia jarego, grochu, traw i innych roślin. W swoich

badaniach wykorzystałam metody matematyczne (np. funkcjonalną analizę danych) i statystyczne (analiza wariancji w mieszanych modelach liniowych, metody wielowymiarowe). Prowadziłam i kontynuuję także badania na temat estymacji macierzy kowariancji dla danych rzeczywistych.

5. Informacja o wykazywaniu się istotną aktywnością naukową albo artystyczną realizowaną w więcej niż jednej uczelni, instytucji naukowej lub instytucji kultury, w szczególności zagranicznej.

Prowadzona przeze mnie działalność naukowa była realizowana w sześciu jednostkach naukowych: Uniwersytet im. Adama Mickiewicza w Poznaniu, Instytut Genetyki Roślin Polskiej Akademii Nauk w Poznaniu, Uniwersytet Przyrodniczy w Poznaniu, Instytut Chemii Bioorganicznej Polskiej Akademii Nauk w Poznaniu (obecnie jestem zatrudniona w dwóch ostatnich), The College of William and Mary w Williamsburg oraz The James Hutton Institute w Dundee (w wyniku staży zagranicznych) i obejmowała poza wymienionymi zagadnieniami 8 zasadniczych kierunków:

Temat 1. Wyznaczanie metabolitów o zróżnicowanej akumulacji w gatunkach z rodziny Brassicaceae - analiza wysokoprzepustowych danych z najnowszych urządzeń do chromatografii cieczowej LC-MS

Załącznik nr 5, praca IIA1.

Obecnie współpracę z pracownikami Instytutu Chemii Bioorganicznej Polskiej Akademii Nauk, gdzie znajdują się najnowocześniejsze nie tylko w Polsce, ale i w Europie urządzenia do chromatografii cieczowej sprzężonej ze spektrometrią mas (LC-MS). W związku z tym w ramach tej współpracy mogę pracować na danych z najnowszych technologii. W pracy IIA1, w wyniku

międzynarodowej współpracy m.in. z Max Planck Institute for Plant Breeding Research w Kolonii, znajdują się wyniki moich analiz danych metabolomicznych dla *Arabidopsis thaliana*, *Capsella rubella*, *Cardamine hirsuta*, *Eutrema salsugineum* traktowanych oligopeptydem pochodzącym z bakterii (flg22). Zatem w tym przypadku analizowałam 4 gatunki, traktowane flg22 i w warunkach kontrolnych, w 5 powtórzeniach biologicznych, w 3 doświadczeniach, w jonizacji dodatniej i ujemnej, co stanowiło 9616 metabolitów w 120 próbach. Problem z tego typu danymi polega na tym, że wielu doświadczalników, biologów, chemików korzysta z programów np. MZmine2, MetaboAnalyst lub XCMS, które nie nadają się do doświadczeń wieloczynnikowych, są w stanie przeanalizować jedynie dane typu traktowanie – kontrola. Zastosowania w tym przypadku takich programów są błędne i nie przedstawiają prawdziwej sytuacji, w której na wyniki mają wpływ interakcje czynników. Natomiast ja zastosowałam wieloczynnikowe metody analiz statystycznych we własnych skryptach w programie Genstat, a do wizualizacji własne skrypty w R. Obserwacje poniżej progu detekcji zastąpiłam połową minimalnej niezerowej obserwacji dla danego metabolitu. Następnie na zlogarytmowanych danych przeprowadziłam dwuczynnikową analizę wariancji z doświadczeniem jako blokiem (efekt losowy) i dwoma stałymi czynnikami: gatunek i traktowanie flg22. Analizę wykonałam razem dla dodatniej i ujemnej jonizacji. Celem analizy było wyznaczenie metabolitów o zróżnicowanej akumulacji (Differentially Accumulated Metabolites, DAMs), które zdefiniowałam osobno dla każdego gatunku w przypadku spełnienia wszystkich trzech warunków: 1) efekt traktowania lub efekt interakcji traktowania i gatunku był istotny przy q-wartości < 0.05 (fdr – false discovery rate), 2) różnica między traktowaniem a warunkami kontrolnymi była istotna, co zostało przetestowane testami indywidualnymi dla każdego gatunku przy p-wartości < 0.05 , 3) $|FC| > 1.5$, gdzie FC (fold change) jest ilorazem akumulacji metabolitu w traktowaniu flg22 i warunkach kontrolnych. Wykonałam również analizę składowych głównych, mapę ciepłą, venn diagram oraz wykres słupkowy.

Publikacja IIA1 pokazuje jak ewoluowała odpowiedź immunologiczna gatunków z rodziny Brassicaceae.

Dwie prace powstałe we współpracy z ICHB PAN dotyczące wyznaczania metabolitów o zróżnicowanej akumulacji z danych LC-MS w innych roślinach zostały wysłane do recenzji, trzecia jest obecnie w przygotowaniu. W jednej z prac zajmowałam się wyznaczaniem metabolitów o zróżnicowanej akumulacji w *Arabidopsis thaliana* traktowanej różnymi grzybami podczas niedoboru fosforu i przy pełnym fosforze, w dwóch doświadczeniach, 4 powtórzeniach biologicznych, jonizacji dodatniej i ujemnej, co stanowiło zbiór 6415 metabolitów w 142 próbach. Przeprowadziłam analizę wyznaczającą DAMy ze względu na porównanie grzybów do kontroli, a także ze względu na traktowanie niedoborem fosforu w porównaniu do pełnego fosforu, w obu przypadkach osobno w liściach i korzeniach. Wyniki wskazują na możliwość użycia w rolnictwie grzybów zamiast nawożenia fosforem. W kolejnej pracy wyznaczałam metabolity wykazujące istotne efekty w *Brachypodium* w trzech różnych organach (liść, korzeń, kłos) dla dwóch linii, w 4 powtórzeniach biologicznych, w dwóch doświadczeniach, jonizacji dodatniej i ujemnej, co stanowiło 22307 metabolitów w 48 próbach. W pracy, do której jeszcze analizuję dane wyznaczam DAMsy w roślinach Poaceae traktowanych *Fusarium* w dwóch punktach czasowych, 4 powtórzeniach biologicznych, 2 doświadczeniach, jonizacji dodatniej i ujemnej co stanowi 16781 metabolitów w 128 próbach. Metody zastosowane w tym przypadku wymagały szczególnych sformułowań ze względu na liczbę czynników, rozbijając problem znalezienia DAMsów na pewne grupy efektów i ich interakcji wraz z innymi warunkami np. dla fold change, który jest ilorazem akumulacji danego metabolitu podczas traktowania i w warunkach kontrolnych.

Temat 2. Analiza sekwencjonowania nowej generacji (NGS) dla doświadczeń na roślinie modelowej *Arabidopsis thaliana*

Załącznik nr 5, prace z p. II A): 3, 9.

W wyniku współpracy z Max Planck Institute for Developmental Biology, Umeå University, University Tübingen, Universidad de Málaga, University of Basel, Beijing Forestry University w ramach mojego zatrudnienia do projektu FlowPlast (Załącznik nr 5, IIF5) analizowałam dane RNA-seq i CHIP-seq dotyczące *Arabidopsis thaliana* w celu zbadania podstaw molekularnych plastyczności kontroli czasu kwitnienia w reakcji na czynniki środowiskowe (temperaturę i długość dnia). Moje wyniki, ujęte w pracy IIA1, pozwoliły wysnuć wnioski na temat zachodzących modyfikacji chromatyny H3K4me3 i H3K27me3 i równoczesnych zmian ekspresji genów w merystemie w czasie w odpowiedzi na kwitnienie indukowane fotoperiodem. Stwierdzono, że obecność H3K4me3 jest dobrym predyktorem ekspresji genów oraz że H3K27me3 może towarzyszyć zarówno aktywnym, jak i nieaktywnym genom. Opisano występowanie nakładających się modyfikacji H3K4me3 i H3K27me3 i zaproponowano wyjaśnienie tego faktu poprzez specyficzne funkcje komórek merystemu wierzchołkowego. Następnie kontynuowałam badania poprzez analizę danych opisujących położenie modyfikacji chromatyny oraz zmiany ekspresji genów we floemie w odpowiedzi na zmiany fotoperiodu, a wyniki zostały opublikowane w pracy IIA7. Moje analizy pozwoliły zaobserwować dodatnią korelację między zmianami w ekspresji i poziom H3K4me3 genów zaangażowanych w podstawowe funkcje we floemie, w tym regulację metabolizmu, rytm okołodobowy, rozwój i modyfikacje epigenetyczne. Zaobserwowaliśmy również, że zmiany w H3K27me3 przyczyniają się w niewielkim stopniu do zmian w ekspresji genów. Zidentyfikowaliśmy wcześniej niescharakteryzowane białko MORN-MOTIF REPEAT PROTEIN REGULATING FLOWING1 (MRF1), które było silnie podwyższone we floemie w odpowiedzi na indukcyjny fotoperiod. Mutacja mrf1 opóźniła kwitnienie, podczas gdy nadekspresja MRF1 miała odwrotny efekt, co wskazuje, że MRF1 działa jako promotor kwitnienia.

Analizę danych RNA-seq wykonałam zaczynając od mapowania odczytów NGS do referencyjnego transkryptomu *Arabidopsis thaliana*, z pewnymi regionami rybosomalnymi RNA zamaskowanymi, przy użyciu TopHat

2.0.13. Liczbę odczytów NGS obejmujących transkrypcje obliczyłam za pomocą funkcji `summarizeOverlaps` w R. Ekspresowane geny zostały zdefiniowane za pomocą fragmentów na kilobazę transkryptu, na milion odwzorowanych odczytów (FPKM), jako te o $FPKM > 1$ dla co najmniej jednego punktu czasowego. Liczba odczytów posłużyła do analizy różnicowej ekspresji genów w `Deseq2` wykonanej przy $FDR < 0.05$. Regularyzowane logarytmy liczby odczytów, obliczone za pomocą `Deseq2`, zastosowałam do analizy zależności między poziomem ekspresji genów a modyfikacjami histonów.

Analizę danych CHIP-seq dotyczącą modyfikacji histonów przeprowadziłam zaczynając od mapowania odczytów NGS do genomu referencyjnego *Arabidopsis thaliana* za pomocą `Bowtie2` 2.2.4. Istotnie wzbogacone regiony w próbkach H3K4me3 i H3K27me3 w stosunku do kontroli zidentyfikowałam za pomocą `MACS2` 2.1.0. Wykonałam filtrowanie regionów ze względu na różnice między biologicznymi powtórzeniami, dla każdego punktu czasowego oddzielnie. Regiony będące sumą regionów z dwóch replikacji (dla każdego przeciwciała i punktu czasowego) stanowiły ocenę modyfikacji histonów. Analizę różnicową modyfikacji dla tych ocen wykonałam przy użyciu `DiffBind68` (przy $FDR < 0.01$), porównując punkty czasowe w parach. Oceny modyfikacji zestawiałam z adnotowanymi genami, aby uzyskać listy genów z szczególnym pokryciem przez H3K4me3 i H3K27me3. Współczynniki korelacji rang Spearmana obliczyłam i przetestowałam pod kątem istotności w `Genstat` 18.

Następnie wykonano analizę Gen Ontology (GO) używając program `AmiGO 2`, aby ocenić nadreprezentację i niedoreprezentację w procesach biologicznych o istotnych różnicach w próbkach H3K4me3 i funkcjach molekularnych w przeciwstawnych stanach modyfikacji H3K4me3-H3K27me3. W celu korekty porównań wielokrotnych P-wartości zostały skorygowane przy użyciu poprawki Bonferroniego i jako próg istotności zastosowano $P < 0.05$. Aby zmniejszyć złożoność, zbędne podrzędne terminy oparte na hierarchii GO zostały usunięte.

Temat 3. Analiza danych fenotypowych pochodzących z jęczmienia i pszenicy. Załącznik nr 5, prace z p. II A): 4, 10, 11, 13, 15.

Cztery pierwsze z wymienionych prac powstały w ramach mojej pracy w projekcie POLAPGEN-BD (Załącznik 5, IIF7) i dotyczyły jęczmienia, kolejna praca dotyczyła badań pszenicy.

W pracy IIA13 analizowałam statystycznie dane w celu zbadania potencjału plonowania rekombinowanych linii wsobnych jęczmienia (RIL) wyprowadzonych z trzech krzyżówek europejskich oraz syryjskich odmian jęczmienia jarego: Maresi × CamB, Lubuski × CamB oraz Georgie × Harmal oraz w celu identyfikacji loci cech ilościowych (ang. quantitative trait loci – QTL) w tych trzech populacjach. Linie RIL pochodziły z doświadczeń polowych prowadzonych przez okres trzech lat (2011-2013) i były genotypowane za pomocą markerów SNP (ang. single nucleotide polymorphism) oraz markerów SSR (ang. simple sequence repeats). Skonstruowano mapę genetyczną dla każdej populacji, a następnie jedną wspólną mapę. Biologiczną interpretację zidentyfikowanych QTL uzyskano przez odniesienie do Ensembl Plants. Analiza QTL wyróżniła 12 regionów genomów badanych linii. Większość QTL zidentyfikowano na chromosomie 2H. Odmiany rodzicielskie syryjskie wniosły allele obniżające wartości cech w większości QTL dla masy ziarna, liczby ziaren, długości kłosów i czasu do kłoszenia oraz liczne allele zwiększające długość łodygi. Wyróżniono linie o akceptowalnym potencjale plonowania ziarna łączące pożądane cechy lub allele od ich rodziców. Brałam udział w analizie statystycznej wykonanej we własnych skryptach w programie Genstat. Postawiono hipotezę, że linie, które były stabilne w naszych eksperymentach przeprowadzonych w różnych latach, w różnych warunkach pogodowych są bardziej odporne na suszę.

Badania te kontynuowaliśmy w pracach IIA10 i IIA11. W pierwszej z nich celem była identyfikacja loci cech ilościowych (QTL) powiązanych z wysokością roślin i potencjałem plonowania linii wsobnych (RIL) jęczmienia jarego

hodowanych w różnych warunkach wilgotności gleby. Materiał roślinny stanowiła krzyżówka Maresi × CamB badana w trzyletnich doświadczeniach szklarniowych. Wykryto łącznie 103 QTL wpływające na badane cechy, a 36 z nich wykazało stabilny wpływ środowiska. Stwierdzono, że dziesięć QTL ma istotne znaczenie wyłącznie w przypadku niedoboru wody. Dziewięć QTL wpływających na długość głównej łodygi zostało wykrytych na chromosomach 2H-6H. W przypadku czterech wykrytych QTL, allele pochodzące z Maresi miały negatywny wpływ na tę cechę, najbardziej istotne okazało się QLSt-3H.1-1 w grupie sprzężeń 3H.1. Ścisłe powiązanie między zidentyfikowanymi QTL wokół locus *sdw1/denso*, z pozytywnymi allelami wniesionymi przez Maresi, wskazuje, że półkarłowata odmiana Maresi może służyć jako dawca korzystnych cech dających poprawę plonowania, także w warunkach niedoboru wody. Analizy molekularne ujawniły, że również CamB przyczyniło się do zwiększenia potencjału plonowania. Obserwacje cech fenotypowych poddano analizie wariancji w modelu mieszanym ze stałymi efektami roku, traktowania suszą oraz interakcji roku i traktowania. Algorytm resztowej największej wiarygodności (REML) został wykorzystany do estymacji komponentów wariancyjnych dla efektów losowych. Statystyka F została obliczona w celu oceny istotności efektów stałych. Dziedziczność obliczono z odpowiednich komponentów wariancyjnych. Korelacje Pearsona zostały wykorzystane do oceny związków między długością głównej łodygi oraz inne cech. Wszystkie analizy zostały wykonane przy użyciu programu Genstat. W publikacji IIA11 celem była identyfikacja loci cech ilościowych (QTL) powiązanych z wczesnością roślin i potencjałem plonowania linii wsobnych (RIL) jęczmienia jarego hodowanych w różnych warunkach wilgotności gleby. Materiał roślinny stanowiła krzyżówka Lubuski × CamB badania w trzyletnich doświadczeniach szklarniowych. Na wszystkich chromosomach jęczmienia wykryto 60 QTL. Największa liczba QTL została znaleziona na chromosomie 2H. Główny QTL został zlokalizowany na chromosomie 2H, zidentyfikowany i powiązany z pewnymi cechami plonu. Niniejsze badanie wykazały, że allel wczesności z CamB, wprowadzony do genomu odmiany europejskiej może poprawić wydajność plonowania

jęczmienia w warunkach suszy. Analizę statystyczną przeprowadzono w podobny sposób jak w przypadku pierwszej publikacji. Dodatkowo wykonano biploty.

Praca IIA4 dotyczy 60 odmian jęczmienia badanych przez 3 lata zarówno w warunkach szklarniowych jak i polowych. Jest to praca integrująca wyniki badań fenotypowych oraz genotypowych dla jęczmienia jarego pod wpływem suszy. Mierzono wiele cech związanych z plonowaniem jak i własnościami fizykochemicznymi. Doświadczenia polowe były przeprowadzone w 14 lokalizacjach, w których cechy mierzono przez 3 lata miesięcznie wraz z temperaturą powietrza oraz opadami. Wyodrębniono 5 odmian o wysokiej wydajności plonu oraz 5 odmian stosunkowo stabilnych w obu warunkach: szklarniowych i polowych. Cechy związane z wydajnością plonu zaobserwowane w różnych lokalizacjach były powiązane ze zmiennymi środowiskowymi istotnymi dla dostępności wody. Suma opadów w kwietniu i maju była ujemnie skorelowana z masą 1000 ziaren i pozytywnie z wysokością rośliny. Pozytywne korelacje stwierdzono między wysokością roślin a temperaturami w czerwcu i lipcu. Pięć markerów wykrytych wcześniej jako powiązane z loci cech ilościowych w populacjach mapujących zostały zidentyfikowane jako mające spójny efekt wśród odmian o różnym rodowodach. Względne skutki suszy obliczono jako różnice między wartością w czasie suszy a wartością w warunkach kontrolnych pomnożoną przez 100 i podzieloną przez wartość w warunkach kontrolnych. Analiza wariancji dla danych szklarniowych została przeprowadzona na surowych obserwacjach, ze stałym modelem uwzględniającym efekty suszy (D), odmian (V) i interakcji $D \times V$. Analiza wariancji na danych polowych została przeprowadzona na średnich dla odmian (a) z modelem efektów stałych obejmującym efekty lat (Y), lokalizacji (L), odmian (V) i efekty $Y \times L$, $Y \times V$, $L \times V$ oraz (b) z modelem efektów stałych z uwzględnieniem efektów kompleksów glebowych (S), odmian (V) i interakcji $S \times V$. Wykonano również m.in. biploty, grupowanie hierarchiczne, wyliczono korelacje Pearsona i zastosowano do nich test t.

W pracy IIA15 analizowałam statystycznie dane dotyczące plonowania linii RIL pszenicy jarej traktowanej grzybem *Fusarium*. Doświadczenia obejmowały krzyżówkę odmian Zebra i Saar w liczbie 198 linii i odmian rodzicielskich, w których zaszczepiono patogena *Fusarium culmorum*, obserwowaną w ciągu trzech lat. Poziomy odporności oszacowano oceniając objawy choroby na ziarniakach. Celem niniejszych badań była ocena zmienności populacji linii RIL pszenicy, wyprowadzonej z form rodzicielskich o umiarkowanej odporności, nie wnoszących genów głównych odporności na *Fusarium*, pod kątem odporności na fuzariozę kłosów oraz cech związanych z plonem. Pomimo podobnej reakcji rodziców na infekcję *F. culmorum*, stwierdzono istotne różnice między liniami we wszystkich analizowanych cechach. Wykazano również, iż genotypy scharakteryzowane jako ościste odznaczały się niższą wartością stopnia porażenia ziarna, oraz wyższą wartością masy 1000 ziaren (zarówno w roślinach kontrolnych, jak i porażonych). W badaniach wykorzystano również markery mikrosatelitarne związane z QTL-ami odporności na *Fusarium*. Analizy molekularne wykazały m.in. iż w badanej populacji RIL markery Xgwm389 i Xgwm533 związane były zarówno ze stopniem porażenia ziarna, jak i masą 1000 ziaren. Do dwóch cech: stosunek do wagi i liczba nasion została zastosowana dwuczynnikowa analiza wariancji, w której lata i linie były źródłem zmienności (efekty stałe). W przypadku masy ziarna z kłosa oraz masy 1000 ziaren obserwowanych na poletkach infekowanych i kontrolnych przeprowadzono trójczynnikową analizę wariancji, a jako źródła zmienności przyjęto linie, lata i traktowanie. Współczynniki odziedziczalności obliczono za pomocą estymowanych komponentów wariacyjnych wykorzystując algorytm resztowej największej wiarygodności (REML). Istotność związków między cechami fenotypowymi a markerami SSR oceniano za pomocą testu F.

Temat 4. Analiza danych lipidomicznych w trawach.

Załącznik nr 5, praca IIA14.

W pracy dotyczącej przebudowy składu glicerolipidów liści komórkowych w warunkach suszy i ponownego nawodnienia w trawach pastewnych z kompleksu *Lolium-Festuca* zajmowałam się analizą statystyczną danych lipidomicznych. Dwa blisko spokrewnione genotypy introgresji *Lolium multiflorum* (życica włoska) i *Festuca arundinacea* (kostrzewa wysoka) zostały tutaj użyte jako model dla innych gatunków traw do opisanie przegrupowań lipidów podczas suszy i ponownego nawodnienia. Genotypy różniły się poziomem zdolności fotosyntezy podczas suszy oraz zdolnością do regeneracji błon po ustaniu stresu. Dane pochodziły z ultrasprawnej chromatografii cieczowej sprzężonej ze spektrometrią mas. Analizowałam 120 lipidów dla 2 genotypów, w 5 punktach czasowych i 4 powtórzeniach biologicznych. Wyniki moich badań jednoznacznie wykazały, że deficyt wody miał istotny wpływ na metabolizm lipidów w badanych trawach pastewnych. Wykazałam, że strukturalne i metaboliczne gatunki lipidów zmieniły swoją liczebność podczas okresów suszy i ponownego nawadniania, a także zaobserwowałam, że pewne istotne różnice zależą od genotypu. Genotyp introgresji charakteryzujący się zdolnością do regeneracji błon po ponownym nawodnieniu wykazał wyższy poziom akumulacji większości chloroplastów i licznych pozachloroplastowych gatunków lipidów błonowych na początku suszy. Ponadto ten genotyp wykazał również znaczne zmniejszenie akumulacji większości lipidów chloroplastowych po ponownym nawodnieniu, w porównaniu z innym genotypem introgresji bez zdolności do regeneracji błon. W pracy IIA14 omówiono potencjalny wpływ zaobserwowanych zmian lipidomicznych na stabilność błony komórkowej i zdolność fotosyntezy. Wyższa tolerancja traw na suszę może być związana z wcześniejszą reakcją lipidomową na sygnał stresowy oraz z regeneracją błon po ustaniu stresu. Dane po wstępnym przetwarzaniu komercyjnym oprogramowaniem, dołączonym do urządzenia chromatograficznego, zostały normalizowane oraz zlogarytmowane. Wykonałam analizę wariancji w układzie powtarzanych pomiarów z czynnikiem powtarzanych pomiarów – czasem. Istotne efekty interakcji genotyp, czas i genotyp × czas wybrałam przy $P < 0,01$. Obliczyłam różnice między średnimi wartościami w kolejnych punktach

czasowych a początkowymi wartościami kontrolnymi oraz wykonałam dla nich testy pod kątem istotności za testu LSD Fishera (ang. Fisher's least significant difference of means) na poziomie 5%. Wykonałam wizualizacje za pomocą map ciepła. Wskaźnik podwójnego wiązania DBI (ang. Double Bond Index) obliczono na podstawie zlogarytmowanych intensywności lipidów.

Temat 5. Analiza danych proteomicznych pochodzących z jęczmienia jarego.

Załącznik nr 5, praca IIA3.

Wykonana na szeroką skalę analiza proteomiczna tkanki liścia i korzenia wyodrębniła białka, które reagują na suszę w sposób specyficzny dla genotypu jęczmienia. Uzyskane dane poddano analizie zależności między markerami a profilami białkowymi i identyfikacji loci cech ilościowych związanych z białkami (pQTL). Taka analiza jest wartościowa w kontekście wskazania procesów biochemicznych leżących u podstaw odpowiedzi jęczmienia na stres suszy oraz może pomóc w poszukiwaniu genów kandydatów zaangażowanych w te procesy. Dane proteomiczne będące przedmiotem pracy IIA3 analizowałam w ramach mojego zatrudnienia do projektu IIF7. Analizowane przeze mnie dane stanowiły populację 100 zrekombinowanych linii wsobnych jęczmienia oraz odmiany rodzicielskie Maresi, CamB, pod wpływem suszy i w warunkach kontrolnych, w 2 powtórzeniach biologicznych i 2 powtórzeniach technicznych, osobno dla liścia i korzenia. Doświadczenie zostało przeprowadzone w warunkach szklarniowych. Pomiar białek wykonano za pomocą dwukierunkowej elektroforezy dla 1632 prób.

Po wstępnym przetwarzaniu danych, przekształciłam je przez $\ln(10^3x)$. Analizy statystyczne przeprowadziłam we własnych skryptach w Genstat i R. Zastosowałam analizę wariancji (ANOVA) w celu wyznaczenia istotnych różnic między średnimi linii RIL i genotypami rodzicielskimi, różnic między akumulacją białek w warunkach suszy i warunkach kontrolnych (efekt suszy) oraz w celu identyfikacji białek o różnym poziomie reakcji na suszę wśród badanych linii RIL

(interakcja genotyp \times środowisko ($G \times E$)). Efekty istotne wybrałam przy $P < 0,05$ z zastosowaniem poprawki Bonferroni ($P < 0,05/z$, gdzie z jest liczbą obserwowanych białek). Test chi-kwadrat zastosowałam do zidentyfikowania znaczących odchyłeń od jednorodnych rozkładów w grupach funkcyjnych białek. Grupowanie hierarchiczne przeprowadziłam przy użyciu pakietu *pvcust* w R bazując na efektach suszy dla białek z istotną interakcją $G \times E$ przy minimum 60 obserwacjach i oparłam na odległości euklidesowej oraz metodzie pełnego wiązania, a brakujące dane zastąpiłam wartościami średnimi. Dzięki grupowaniu znaleziono homogeniczne grupy białek. Analizę współrzędnych głównych przeprowadziłam dla wszystkich białek oraz dla ich grup funkcyjnych. Statystyczna ocena danych proteomicznych na dużą skalę ułatwiła rozróżnienie między białkami, które w podobnym stopniu zareagowały na suszę wśród badanych genotypów, a białkami, dla których zmiana akumulacji silnie wahała się wśród genotypów jęczmienia. Zostały zidentyfikowane konkretne izoformy białkowe wykazujące silne interakcje $G \times E$, które mogą stanowić potencjalne cele dalszych badań nad odpornością na suszę.

Wykonałam również analizę sieci korelacyjnych osobno dla warunków kontrolnych i suszy, za pomocą pakietu *WGCNA* w systemie R. Macierz korelacji Pearsona przekształciłam do macierzy TOM, a jej elementy wykorzystałam do wizualizacji sieci. Moduły wykryłam poprzez grupowanie hierarchiczne oraz algorytm *dynamic tree cut*, wyznaczyłam *huby*, czyli cechy o największej liczbie silnych korelacji z innymi cechami. Współczynniki korelacji obliczyłam dla wszystkich par białek oraz zastosowałam test Manna – Whitneya w celu przetestowania istotności różnic w tych współczynnikach. Wizualizację sieci przeprowadziłam w programie *Cytoscape*.

Na strukturę korelacji białek zaangażowanych w mechanizmy obronne susza wpłynęła bardziej niż na przykład na strukturę korelacji białek związanych z metabolizmem węgla w liściach lub metabolizmem azotu w korzeniach. Moduły białek, czyli zbiory najbardziej skorelowanych związków, składały się z białek o różnych funkcjach. Zidentyfikowana struktura korelacji między aktywatorami Rubisco A i B została zachowana w obu warunkach

środowiskowych, co może wskazywać na wspólne mechanizmy regulacyjne podczas tego stresu. Grupy najbardziej skorelowanych białek (modułów) zdefiniowane w sieciach korelacyjnych składały się z białek o różnych funkcjach. Zidentyfikowane zostały białka o największej liczbie korelacji w liściu oraz korzeniu. Kilka podzbiorów białek zachowało swoją strukturę korelacji między warunkami kontrolnymi a suszą. Zostały wyciągnięte wnioski na temat procentowego składu białek z określonych grup funkcyjnych w pewnych modułach. Wartości te zostały porównane między liściem a korzeniem oraz w różnych warunkach środowiskowych. W liściu 73% i 77% korelacji było dodatnich odpowiednio w warunkach kontrolnych i suszy, natomiast w korzeniach 59% i 58% korelacji było dodatnich odpowiednio w warunkach kontroli i suszy.

Temat 6. Analiza metabolitów wtórnych i pierwotnych w różnych roślinach – analiza danych z urządzeń UPLC-UV, GC-MS oraz mniejszych doświadczeń LC-MS.

Załącznik nr 5, prace z p. II A): 7, 8, 12, 16.

W ramach swoich badań przeprowadziłam również analizy metabolitów wtórnych w roślinach z rodzaju *Kalanchoe*, w różnych gatunkach roślin leczniczych z rodzaju *Passiflora*, a także prowadziłam analizę chemometryczną na ziołach takich jak epilobium (*Epilobium augustifolium*) i szalwia (*Salvia miltiorrhiza* i *Salvia przewalskii*) oraz jęczmieniu jarym (*Hordeum Vulgare*), a także metabolitów pierwotnych w trawach. Dane dotyczące dwóch ostatnich roślin analizowałam w ramach projektów IIF7 i IIF8, w których byłam wykonawcą.

W wymienionych analizach stosowałam takie metody jak analiza wariancji, test Tuckey'a dla prób niezależnych, grupowanie hierarchiczne, wykonywałam diagramy Venna, dendrogramy przy użyciu metody bootstrap, a także wykonywałam wstępnie przetwarzanie danych.

Wyniki dotyczące traw, pochodzące z projektu IIF8, zostały wysłane do recenzji. W tym przypadku analizowałam 2 gatunki traw: kostrzewę trzcinową i życicę trwałą, traktowanych zasoleniem, chłodem, suszą, w 4 punktach czasowych, 4 powtórzeniach biologicznych i 2 powtórzeniach technicznych. Obliczenia ze względu na złożoność wstępnego przetwarzania danych wymagały użycia superkomputerów w Poznańskim Centrum Superkomputerowo-Sieciowym. Analizę statystyczną wykonałam we własnych skryptach w programie Genstat, natomiast wizualizacje w skryptach w R.

Temat 7. Estymacja nieznanymi struktur kowariancyjnych w modelach wielowymiarowych.

Załącznik nr 5, praca z p. IIA6.

W wyniku pracy IIA6 nad aproksymacją macierzy dodatnio określonych pokazaliśmy, że rzutowanie na przestrzeń liniową macierzy Toeplitza nie zawsze zachowuje nieujemną określoność oraz podaliśmy metodologię i algorytm numeryczny rzutowania na stożek wypukły nieujemnie określonych macierzy Toeplitza. Ponadto wskazaliśmy pewną nieścisłość z literatury, gdzie mylnie przyjęto stożek jako stożek macierzy Toeplitza dowolnego wymiaru, a jest on jedynie stożkiem asymptotycznym. W omawianej pracy przeprowadziłam badania symulacyjne w celu porównania właściwości statystycznych estymatorów uzyskanych przez rzutowanie na stożek o zadanym wymiarze macierzy oraz na stożek asymptotyczny. Przedstawiony problem ma zastosowanie w estymacji nieznanymi struktur kowariancyjnych w modelach wielowymiarowych i kontynuując te badania zastosowałam aproksymację macierzy dodatnio określonych za pomocą macierzy Toeplitza, do estymacji macierzy kowariancji dla danych chromatograficznych w omawianej pracy IB5 z prezentowanego cyklu.

Wyniki tych obu prac wykorzystuję do kontynuacji badań dotyczących wyboru odpowiedniej struktury kowariancji dla danych uzyskanych w badaniu zmian metabolomicznych liści jęczmienia (*Hordeum vulgare*) poddanych

stresowi suszy, otrzymanych dzięki chromatografii gazowej sprzężonej ze spektrometrem mas (GC-MS). W pracy IB5 rozważyliśmy trzy wybrane nieosobliwe podzbiory danych dotyczących jęczmienia. Obecnie analizuję cały zbiór danych wykonując obliczenia na macierzy osobliwej.

Temat 8. Narzędzia algebry liniowej i teorii grafów w zastosowaniach.

Załącznik nr 5, prace z p. II A): 17, 18, 19.

W początkach mojej pracy naukowej koncentrowałam się na badaniach dotyczących własności macierzy informacji związanych z układami optymalnymi i prowadziłam je w ramach międzyuczelnianego projektu badawczego IIF18. Wyniki mojej pracy zostały zawarte w publikacji IIA19, której celem było scharakteryzowanie układów E-optymalnych ze względu na estymację efektów obiektowych w pewnych klasach układów. Wyniki te zostały uzyskane w oparciu o własności algebraiczne macierzy informacji. Podane zostały także wskazówki dotyczące konstrukcji układów E-optymalnych. Badania te są pomocne w planowaniu optymalnych eksperymentów i w związku z tym leżą w kręgu zainteresowań nauk rolniczych, biologicznych oraz medycznych.

Kolejnym tematem moich badań była minimalizacja najmniejszej wartości własnej macierzy przyległości wierzchołków grafów prostych, nieskierowanych o zadanej liczbie wierzchołków krawędzi, co zostało opublikowane po uzyskaniu stopnia doktora w pracy IIA18. Praca była kontynuacją współpracy z prof. Charlesem Johnsonem i odbytych dwukrotnie staży na The College of William and Mary w Williamsburgu. Problem sprowadza się do maksymalizacji wartości szczególnej w zbiorze macierzy zero-jedynkowych o zadanej liczbie jedynek. Problem grafów minimalizujących, został postawiony na gruncie teorii grafów, ale w całości rozwiązywany na gruncie algebry liniowej, ponieważ problem ten sprowadza się do minimalizacji najmniejszej wartości własnej w pewnej klasie macierzy zero-jedynkowych. W pracy IIA18 zidentyfikowałam strukturę grafu minimalizującego, która jest związana z dwudzielnością grafu, dlatego dalej moje badania skoncentrowałam

na minimalizujących grafach dwudzielnych. Postawiłam hipotezę mówiącą o tym, że w przypadku grafów dwudzielnych grafami minimalizującymi są grafy pełne dwudzielne lub grafy tzw. „nearly complete” lub grafy składające się z jednego z tych grafów oraz wierzchołków izolowanych. Hipotezę tę udowodniłam dla pewnych szczególnych przypadków. Rozważałam dwa przypadki wymienionego problemu, tj. gdy liczba krawędzi jest na tyle mała, że konstrukcja grafu dwudzielnego jest możliwa oraz gdy liczba krawędzi jest na tyle duża, że graf dwudzielny nie może istnieć. Opisany powyżej problem znajduje zastosowanie w programowaniu dodatnio półokreślonym, które ma szereg praktycznych zastosowań w tworzeniu portfela inwestycyjnego, sterowaniu optymalnym i in., a także jest interesujący sam w sobie w teorii grafów.

Następnie koncentrowałam się na analizie nowych algorytmów obliczania wyznaczników macierzy wielodiagonalnych oraz wielodiagonalnych z narożnikami. W pracy IIA17 sformułowałam te algorytmy oraz porównałam do innych znanych algorytmów w przypadku macierzy pięciodiagonalnych i pięciodiagonalnych z narożnikami. Przedstawione zostały również zastosowania zaprezentowanych algorytmów do teorii eksperymentów w celu scharakteryzowania układów D-optymalnych.

Ponadto jako wykonawca projektu IIF4 brałam udział w analizie danych dotyczących sekwencjonowania DNA oraz RNA w celu badania cechy wczesności kwitnienia u łubinu białego i łubinu żółtego, co zostało opisane w IIA5. Natomiast w projekcie IIF7, w którym byłam zatrudniona, poza analizą szeregu zróżnicowanych danych dotyczących jęczmienia jarego, co zaowocowało wieloma publikacjami, byłam również zaangażowana w integrację danych, co zaowocowało wspomnianą już publikacją IIA4 oraz rozdziałem w monografii E1, w której opisałam proces kolekcji danych w bazie, ich kurowania, standaryzacji oraz usuwania danych odstających.

6. Informacja o osiągnięciach dydaktycznych, organizacyjnych oraz popularyzujących naukę lub sztukę.

A) Osiągnięcia dydaktyczne

W ramach pracy na różnych uczelniach w Poznaniu prowadziłam zajęcia z przedmiotów takich jak: statystyka, matematyka, informatyka, dydaktyka informatyki II, elementy metod numerycznych, metody numeryczne I, numeryczna algebra liniowa.

W trakcie pracy na Uniwersytecie Przyrodniczym prowadziłam ćwiczenia z analizy matematycznej, statystyki matematycznej, matematyki oraz technologii informacyjnych. Opracowałam również i wdrożyłam nowy przedmiot dla studentów programu Erasmus: STATISTICAL INFERENCE IN GENETICS, Kod: AGRO 6.1, ECTS: 5. Przygotowałam wiele ciekawych ćwiczeń oraz skryptów w programie R dla studentów. Materiały dydaktyczne umieszczałam na Wirtualnym Dziekanacie oraz mailowo.

B) Opieka naukowa nad studentami

opieka naukowa nad studentką (stażystką), 08-09.2014, Instytut Genetyki Roślin, Polska Akademia Nauk

C) Udział w komitetach organizacyjnych międzynarodowych i krajowych konferencji naukowych

International Conference on Trends and Perspectives in Linear Statistical Inference LinStat'2018, 2018, Będlewo, członek komitetu organizacyjnego

Mat Triad 2017, 2017, Będlewo, członek komitetu organizacyjnego

III Ogólnopolska Konferencja IGR PAN: Genetics and genomics in improving plants - from model plant to new variety, 2017, Poznań, członek komitetu organizacyjnego

Mat Triad 2009, 2009, Będlewo, członek komitetu organizacyjnego

D) Dodatkowe informacje organizacyjne

2018-2019 Prowadzenie seminarium Katedry Metod Matematycznych i Statystycznych Uniwersytetu Przyrodniczego w Poznaniu

7. Oprócz kwestii wymienionych w pkt. 1-6, wnioskodawca może podać inne informacje, ważne z jego punktu widzenia, dotyczące jego kariery zawodowej.

A) Międzynarodowe i krajowe nagrody za działalność naukową

- Nagroda za wyróżniające osiągnięcia, 2021, Rektor Uniwersytetu Przyrodniczego w Poznaniu, za wyróżniające wyniki pracy naukowej w ostatnich czterech latach
- Nagroda zespołowa II stopnia, 2020, Rektor Uniwersytetu Przyrodniczego w Poznaniu, za osiągnięcia naukowe udokumentowane publikacjami
- Nagroda RID, 2019, Wielkopolska Regionalna Inicjatywa Doskonałości w obrębie nauk o żywieniu Uniwersytetu Przyrodniczego w Poznaniu, za najlepsze artykuły opublikowane w roku 2019
- Dodatek specjalny, 2019, Rektor Uniwersytetu Przyrodniczego w Poznaniu, za wyróżniające osiągnięcia naukowe potwierdzone publikacjami w latach 2017-2018
- Nagroda zespołowa III stopnia, 2018, Rektor Uniwersytetu Przyrodniczego w Poznaniu, za osiągnięcia naukowe udokumentowane publikacjami
- Award for the best talk for young scientists, 2009, Komitet Naukowy MAT-TRIAD 2009, za najlepszy wykład dla młodych naukowców

B) Współpraca międzynarodowa wynikająca z publikacji i grantów:

- Department of Molecular Biology, Max Planck Institute for Developmental Biology, Tübingen, Germany
- Max Planck Institute for Plant Breeding Research, Cologne, Germany
- Max Planck Institute of Molecular Plant Physiology, Potsdam, Germany
- Evolutionary Dynamics and Biophysics Group, Max Planck Institute for Developmental Biology, Tübingen, Germany
- Zentrum für Molekularbiologie der Pflanzen, Eberhard-Karls-Universität Tübingen, Germany
- Leeds University, Leeds, United Kingdom
- Wageningen University, Wageningen, Netherlands
- Department of Mathematics, The College of William and Mary, Williamsburg, USA
- Department of General Genetics, Center for Plant Molecular Biology, University Tübingen, Germany
- Department of Plant Physiology, Umeå Plant Science Centre, Umeå University, Umeå, Sweden
- Beijing Advanced Innovation Centre for Tree Breeding by Molecular Design, Beijing Forestry University, Beijing, China
- Department of Pharmaceutical Science, Health and Biological Science Institute, Federal Rural University of Rio de Janeiro, Seropédica, Brazil
- Facultad de Ciencias, Departamento de Biología Molecular y Bioquímica, Instituto de Hortofruticultura Subtropical y Mediterránea, Universidad de Málaga–Consejo Superior de Investigaciones Científicas, Universidad de Málaga, Spain
- Biozentrum, University of Basel, Basel, Switzerland

C) Współpraca krajowa wynikająca z publikacji i grantów:

- Department of Biometry and Bioinformatics, Institute of Plant Genetics, Polish Academy of Sciences, Poznań

-
- Department of Environmental Stress Biology, Institute of Plant Genetics, Polish Academy of Sciences, Poznan
 - Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan
 - Department of Mathematical and Statistical Methods, Poznan University of Life Sciences, Poznan
 - Faculty of Mathematics and Computer Science, Adam Mickiewicz University, Poznan
 - Poznan University of Technology, Poznan
 - Department of Genetics, Faculty of Biology and Environmental Protection, University of Silesia, Katowice
 - Department of Botany and Nature Protection, Faculty of Biology and Environmental Protection, University of Silesia, Katowice
 - Department of Plant Physiology, University of Agriculture in Krakow, Krakow
 - Institute of Plant Physiology, Polish Academy of Sciences, Krakow
 - Institute of Agricultural and Forest Environment, Polish Academy of Sciences, Poznan
 - Institute of Agrophysics, Polish Academy of Sciences, Lublin
 - Institute of Soil Science and Plant Cultivation – National Research Institute, Pulawy
 - Poznan Plant Breeders Ltd., Tulce
 - Danko Plant Breeders Ltd., Choryn
 - The State Vocational School of Higher Education, Gorzow Wlkp.
 - Research Centre for Cultivar Testing, Słupia Wielka
 - Department of Botany, Breeding and Agricultural Technology for Medicinal Plants, Institute of Natural Fibres and Medicinal Plants, Poznan
 - Department of Pharmaceutical Botany and Plant Biotechnology, Poznan University of Medical Sciences, Poznan

- Department of Pharmacology, Poznan University of Medical Sciences, Poznan
- Department of Pharmacology and Phytochemistry, Institute of Natural Fibers and Medicinal Plants, Poznan
- Department of Pathogen Genetics and Plant Resistance, Metabolomics Team, Institute of Plant Genetics of the Polish Academy of Science, Poznan
- Department of Clinical Chemistry and Molecular Diagnostics, Poznan University of Medical Sciences, Poznan
- Division of Perinatology and Women's Diseases, Poznan University of Medical Sciences, Poznan
- Laboratory of Molecular Biology, Poznan University of Medical Sciences, Poznan
- Department of Nursing, University of Medical Sciences, Poznan
- Department of Molecular Phytopathology, Faculty of Agriculture and Biotechnology, University of Technology and Life Sciences, Bydgoszcz
- Department of Plant Pathology, Plant Breeding and Acclimatization Institute NRI, Radzików

D) Materiały konferencyjne

Piasecka A., Sawikowska A., Witaszak N., Waśkiewicz A., Kaczmarek J. (2019). Resistance-related diversity in cereals response to pathogen infection, Plant Genomes, Systems Biology & Engineering, 4-7.12.2019, New York, USA, Book of abstracts: 79.

Sawikowska A., Krajewski P. (2019). Sieci w badaniach metabolitów i białek w jęczmieniu w warunkach suszy, Genetyka i genomika w doskonaleniu roślin uprawnych, 5-7.11.2019, Poznan, Poland, Książka streszczeń: 65.

Sawikowska A. (2019). Networks in research on omic changes in barley under drought, Multiomics to Mechanisms - Challenges in Data Integration, 11-13.09.2019, Heidelberg, Germany, Abstract of papers: 177.

Piasecka A., Frerigmann H., Sawikowska A., Krajewski P., Hacquard S., Schulze-Lefert P., Bednarek P. (2019). Metabolomic approaches reveal the impact of growth promoting fungal endophytes on Arabidopsis metabolome, 9th Conference of the Polish Society of Experimental Plant Biology, 9-12 September 2019, Toruń, Poland, Abstract book: 100.

Sawikowska A., Krajewski P., Piasecka A., Kachlicki P. (2019). Analiza dużych zbiorów danych w metabolomice niecelowanej. VII Konferencja Chemometria i Metrologia w Analityce, 6 – 8.03.2019, Poznan, Poland, Książka streszczeń: 25.

Błaszczak L., Basińska-Barczak A., Kosmala A., Marczak Ł., Perlikowski D., Sawikowska A., Witaszak N. (2018). Reaction of wheat (*triticum aestivum l.*) to *trichoderma* spp. root colonization. The Integrative Plant Biology Conference, 7-9.11.2018, Poznań, Poland, Abstract Book: 35.

Ćwiek-Kupczyńska H., Mokrzycka M., Sawikowska A., Krajewski P. (2018). Integrating experimental data for integrative plant biology. The Integrative Plant Biology Conference, 7-9.11.2018, Poznań, Poland, Abstract Book: 11.

Mieldzioc A., Mokrzycka M., Sawikowska A. (2018). Regularization of the covariance matrix for metabolomic data. The 48th International Biometrical Colloquium in Honour of the 90th Birthday of Professor Tadeusz Caliński and VI Polish - Portuguese Workshop on Biometry, 9-13.09.2018, Szamotuły, Poland, Abstracts: 30.

Mieldzioc A., Mokrzycka M., Sawikowska A. (2018). Estimation of block partitioned covariance matrix using Frobenius norm and its application to real data example. International Conference on Trends and Perspectives in Linear Statistical Inference LinStat'2018, 20-24.08.2018, Będlewo, Poland, Book of Abstracts: 135.

Mieldzioc A., Mokrzycka M., Sawikowska A. (2018). Estimation of covariance matrices structured by CS, AR(1) or Toeplitz banded matrices using Frobenius norm and entropy loss function. International Conference on Trends and Perspectives in Linear Statistical Inference LinStat'2018, 20-24.08.2018, Będlewo, Poland, Book of Abstracts: 116.

Krajewski P., Sawikowska A. (2017). Genome-metry: statistical methods in genome-level data analysis. 47th International Biometrical Colloquium, 10-14.09.2017, Kiry, Poland, Book of abstracts: 21.

Sawikowska A., Krajewski P., Piasecka A., Kachlicki P. (2017). A data analysis protocol for monitoring metabolomic changes in barley under drought stress. 47th International Biometrical Colloquium, 10-14.09.2017, Kiry, Poland, Book of abstracts: 33.

Sawikowska A., Krajewski P. (2016). Large chromatographic data sets analysis on the example of metabolomic data. XLII Conference on Mathematical Statistics, 27.11– 2.12.2016, Bedlewo, Poland, Book of abstracts: 43.

Kuczyńska A., Krystkowiak K., Mikołajczak K., Ogrodowicz P., Surma M., Adamski T., Kempa M., Sawikowska A., Krajewski P. (2016). Introgression of the LTP2 gene through marker assisted backcross breeding in barley (*Hordeum vulgare* L.) with LTP2 gene analysis expression, 20th Eucarpia, 29.08 – 1.09.2016, Zurych, Switzerland, Book of abstracts: 216.

Sawikowska A. (2015). Metody obliczeniowe dla dużych danych chromatograficznych na przykładzie badań metabolomu jęczmienia jarego, Wielkopolskie Centrum Zaawansowanych Technologii, 1-2.12.2015, Poznan, Poland, Program i streszczenia: 50.

Sawikowska A., Piasecka A., Kachlicki P., Krajewski P. (2015). A data analysis protocol for monitoring metabolomic changes in genetic experiments: a study of barley (*Hordeum vulgare*) leaves under drought stress. XVI-th meeting of the Eucarpia Section Biometrics in Plant Breeding, 9-11.09.2015, Wageningen, Netherlands, Book of abstracts: 90.

Perlikowski D., Kierszniowska S., Sawikowska A., Krajewski P., Rapacz M., Kosmala A. (2015). Remodeling of the membrane lipids composition under drought and rehydration in *Lolium multiflorum*/*Festuca arundinacea* introgression forms. 10th International Conference Plant Functioning Under Environmental Stress, 16-19.09.2015, Kraków, Poland, Book of abstracts: 44.

Piasecka A., Sawikowska A., Krajewski P., Kachlicki P. (2014). Phenolic metabolites expression in leaves of barley inbred lines - comparison of greenhouse and field experiment. Genetics and genomics in improving plants - from model plant to new variety, 5-7.11.2014, Poznan, Poland, Program i streszczenia: 121.

Mikołajczak K., Kuczyńska A., Surma M., Krajewski P., Adamski T., Ogradowicz P., Krystkowiak K., Sawikowska A., Szarejko I., Guzy-Wróbelska J., Gudyś K. (2014). Mapping of QTLs for the plant height and yield forming traits in RIL population of spring barley (*Hordeum vulgare* L.) under various environments. Genetics and genomics in improving plants - from model plant to new variety, 5-7.11.2014, Poznan, Poland, Program i streszczenia: 108.

Ogradowicz P., Kuczyńska A., Krystkowiak K., Adamski T., Krajewski P., Surma M., Mikołajczak K., Sawikowska A., Gudyś K., Guzy-Wróbelska J., Szarejko I. (2014). Identification of QTL associated with yield and earliness in barley RIL lines derived from hybrids between European and Syrian varieties differentiated in tolerance to water deficiency. Genetics and genomics in improving plants - from model plant to new variety, 5-7.11.2014, Poznan, Poland, Program i streszczenia: 114.

Piasecka A., Sawikowska A., Sekuła A., Irzykowski W., Jędrzycka M., Kachlicki P. (2014). Comparative metabolomics of the species of ornamental and medicinal plant *Kalanchoe*. Genetics and genomics in improving plants - from model plant to new variety, 5-7 November 2014, Poznań. Program i streszczenia: 104.

Czyżniejewski M., Sawikowska A., Swarczewicz B., Kachlicki P. (2014). Flavonoids and other metabolites of fodder grasses involved in tolerance to cold stress. Genetics and genomics in improving plants - from model plant to new variety, 5-7.11.2014, Poznań. Program i streszczenia: 57.

Swarczewicz B., Rodziewicz P., Sawikowska A., Krajewski P., Stobiecki M. (2014). Analysis of proline and P5CS in leaves of barley mapping population Maresi x Cam/B1/Cl subjected to drought by GC-MS and 2DE MALDI-

TOF/TOF. 4th Conference of Polish Mass Spectrometry Society, 26-29.05.2014, Trzebnica, Poland, Program and Abstracts: 90.

Czyżniejewski M., Sawikowska A., Swarcewicz B., Kachlicki P. (2014). Metabolites of fodder grasses involved in tolerance to cold stress. 4th Conference of Polish Mass Spectrometry Society, 26-29.05.2014, Trzebnica, Poland, Program and Abstracts: 55.

Sawikowska A., Cwiek H., Frohmberg W., Kaczmarek Z., Krajewski P., Konsorcjum POLAPGEN (2014). Data collection, processing and integration for systems biology research on drought tolerance. PAG, 10-14.01.2014, San Diego, USA, <https://pag.confex.com/pag/xxii/webprogram/Paper12952.html>.

Sawikowska A., Cwiek H., Frohmberg W., Kaczmarek Z., Krajewski P. i Konsorcjum POLAPGEN. (2013). Gromadzenie standaryzacja i adnotacja danych w projekcie POLAPGEN-BD. Polski Kongres Genetyki, 10-13.09.2013, Poznan, Poland, Streszczenia: 194-195.

Kachlicki P., Piasecka A., Czyżniejewski M., Sawikowska A., Krajewski P., Kuczyńska A., Zwierzykowski Z. (2012). Profiling of phenolic compounds in monocotyledonous plants using HPLC-MS. 47th Congress of the Polish Biochemical Society, Polish-German Biochemical Societies Joint Meeting, 11 – 14.09.2012, Poznan, Poland. Acta Biochim. Polon. 59 suppl. 3: 24.

Piasecka A., Sawikowska A., Kuczyńska A., Ogrodowicz P., Mikołajczak K., Kachlicki P., Krajewski P. (2012). Analysis of metabolome changes in barley leaves in response to drought conditions. Biotechnology and Plant Breeding: perspectives towards food security and sustainability, 10-12.09.2012, Radzikow, Poland. Abstracts: 72.

Krajewski P., Sawikowska A., Kaczmarek Z., Cwiek H., Frohmberg W., Consortium POLAPGEN (2012). Data collection and processing in POLAPGEN-BD, a project on biotechnology for breeding cereals with increased resistance to drought. XV-th Meeting of the EUCARPIA Section - Biometrics in Plant Breeding, 5-7.09.2012, Hohenheim, Germany. Abstracts: 54.

Swarcewicz B., Piasecka A., Sawikowska A., Kuczyńska A., Ogrodowicz P., Mikołajczak K., Krystkowiak K., Stobiecki M., Kachlicki P., Krajewski P. (2012). Monitoring changes in barley (*Hordeum vulgare*) leaf metabolome under drought stress and their description with computational methods. 29th LC/MS Montreux Symposium, 7-9 November 2012, Montreux, Switzerland. Abstracts: 44.

Sawikowska A., Madrigal P., Piasecka A., Kuczyńska A., Krystkowiak K., Ogrodowicz P., Mikołajczak K., Kachlicki P., Krajewski P. (2012). Comparison of classical and functional principal components applied to chromatographic data. XV-th Meeting of the EUCARPIA Section - Biometrics in Plant Breeding, 5-7 September 2012, Hohenheim, Germany. Abstracts: 77.

Sawikowska A., Piasecka A., Swarcewicz B., Kuczyńska A., Ogrodowicz P., Mikołajczak K., Stobiecki M., Kachlicki P., Krajewski P. (2012). Computational methods for metabolite profiling in POLAPGEN-BD, a biotechnology project for breeding cereals with increased resistance to drought. EUCARPIA - 19th General Congress - Plant Breeding for Future Generations, 21-24 May 2012, Budapest, Hungary. Abstracts: 377.

Swarcewicz B., Piasecka A., Sawikowska A., Krajewski P., Stobiecki M., Kachlicki P. (2012). Monitoring changes in barley (*Hordeum vulgare*) leaf metabolome under drought stress. Joint Conference of Polish Mass Spectrometry Society and German Mass Spectrometry Society, 4-7 March 2012, Poznań, Poland. Abstracts: 256.

Kuczyńska A., Kaczmarek M., Piasecka A., Sawikowska A., de Mezer M. (2011) Narzędzia biotechnologiczne i materiały służące do otrzymywania odmian zbóż odpornych na stresy środowiskowe; 28-29.11.2011, Konferencja „Misja chemo-, bio- i nanotechnologii w Wielkopolskim Centrum Zaawansowanych Technologii – Materiały i Biomateriały”, Poznan, Poland. Abstracts.

Piasecka A., Sawikowska A., Krajewski P., Kachlicki P. (2011). Flavonoids and other phenolic compounds in response of barley (*Hordeum vulgare* L.) plants to

drought. 5th Conference of the Polish Society of Plant Experimental Biology, 6-9.09.2011 Wrocław, Poland, Book of Abstracts: 68.

Piasecka A., Sawikowska A., Zwierzykowski Z., Kachlicki P. (2010). Chemotaxonomic relationships within the Festuca-Lolium complex on the basis of phenolic secondary metabolites. 1st EUCARPIA Festulolium Working Group Workshop, 7-8.10.2010, Poznań, Poland, Abstract book: 43.

E) Plakaty

Sawikowska A., Krajewski P. (2019). Sieci w badaniach metabolitów i białek w jęczmieniu w warunkach suszy, Genetyka i genomika w doskonaleniu roślin uprawnych, Poznan, 5 – 7 November 2019.

Sawikowska A. (2019). Networks in research on omic changes in barley under drought, Multiomics to Mechanisms - Challenges in Data Integration, Heidelberg, Germany, 11 – 13 September 2019.

Piasecka A., Frerigmann H., Sawikowska A., Krajewski P., Hacquard S., Schulze-Lefert P., Bednarek P. (2019). Metabolomic approaches reveal the impact of growth promoting fungal endophytes on Arabidopsis metabolome, 9th Conference of the Polish Society of Experimental Plant Biology, Toruń, Poland, 9 – 12 September 2019.

Ćwiek-Kupczyńska H., Mokrzycka M., Sawikowska A., Krajewski P. (2018). Integrating experimental data for integrative plant biology. The Integrative Plant Biology Conference, Poznań, Poland, 7-9 November 2018.

Błaszczak L., Basińska-Barczak A., Kosmala A., Marczak Ł., Perlikowski D., Sawikowska A., Witaszak N. (2018). Reaction of wheat (*triticum aestivum* L.) to *trichoderma* spp. root colonization. The Integrative Plant Biology Conference, Poznań, Poland, 7-9 November 2018.

Sawikowska A. (2018). Computational peak deconvolution in chromatographic data by clustering and by functional data analysis. 17-th European Conference on Computational Biology, Ateny, 8-12 September 2018.

Sawikowska A., Piasecka A., Kachlicki P., Krajewski P. (2015). A data analysis protocol for monitoring metabolomic changes in genetic experiments: a study of barley (*Hordeum vulgare*) leaves under drought stress. XVI-th meeting of the Eucarpia Section Biometrics in Plant Breeding, 9-11 September 2015, Wageningen, Netherlands.

Perlikowski D., Kierszniowska S., Sawikowska A., Krajewski P., Rapacz M., Kosmala A. (2015). Remodeling of the membrane lipids composition under drought and rehydration in *Lolium multiflorum*/*Festuca arundinacea* introgression forms. 10th International Conference Plant Functioning Under Environmental Stress, 16-19 September 2015, Kraków.

Piasecka A., Sawikowska A., Sekuła A., Irzykowski W., Jędrzycka M., Kachlicki P. (2014). Comparative metabolomics of the species of ornamental and medicinal plant *Kalanchoe*. Genetics and genomics in improving plants - from model plant to new variety, 5-7 November 2014, Poznań.

Czyżniejewski M., Sawikowska A., Swarczewicz B., Kachlicki P. (2014). Flavonoids and other metabolites of fodder grasses involved in tolerance to cold stress. Genetics and genomics in improving plants - from model plant to new variety, 5-7 November 2014, Poznań.

Sawikowska A., Ćwiek H., Frohmberg W., Kaczmarek Z., Krajewski P. (2014), Development of data storage and processing infrastructure for plant genetic research. International Biometric Conference, 6-11 July 2014, Florence.

Swarczewicz B., Rodziejewicz P., Sawikowska A., Krajewski P., Stobiecki M. (2014). Analysis of proline and P5CS in leaves of barley mapping population Maresi x Cam/B1/Cl subjected to drought by GC-MS and 2DE MALDI-TOF/TOF. 4th Conference of Polish Mass Spectrometry Society, 26-29 May 2014, Trzebnica.

Czyżniejewski M., Sawikowska A., Swarczewicz B., Kachlicki P. (2014). Metabolites of fodder grasses involved in tolerance to cold stress. 4th Conference of Polish Mass Spectrometry Society, 26-29 May 2014, Trzebnica.

Piasecka A., Ożarowski M., Sawikowska A., Kachlicki P., Thiem B. (2014). Chemotaxonomic analysis of secondary metabolites in leaves and callus culture of three passiflora species. International Symposium on Chromatography of Natural Products, 26-29 May 2014, Lublin.

Sawikowska A., Ćwiek H., Frohberg W., Kaczmarek Z., Krajewski P., Konsorcjum POLAPGEN (2014). Data collection, processing and integration for systems biology research on drought tolerance. PAG, 10-14 January 2014, San Diego, USA.

Sawikowska A., Ćwiek H., Frohberg W., Kaczmarek Z., Krajewski P., Konsorcjum POLAPGEN (2013). Gromadzenie, standaryzacja i adnotacja danych w projekcie POLAPGEN-BD. IV Polski Kongres Genetyki, 10-13 September 2013, Poznań.

Sawikowska A., Madrigal P., Piasecka A., Kuczyńska A., Krystkowiak K., Ogrodowicz P., Mikołajczak K., Kachlicki P., Krajewski P. (2012). Comparison of classical and functional principal components applied to chromatographic data. XV-th Meeting of the EUCARPIA Section - Biometrics in Plant Breeding, 5-7 September 2012, Hohenheim, Germany.

Swarcewicz B., Piasecka A., Sawikowska A., Kuczyńska A., Ogrodowicz P., Mikołajczak K., Krystkowiak K., Stobiecki M., Kachlicki P., Krajewski P. (2012). Monitoring changes in barley (*Hordeum vulgare*) leaf metabolome under drought stress and their description with computational methods. 29th LC/MS Montreux Symposium, 7-9 November 2012, Montreux, Switzerland.

Sawikowska A., Madrigal P., Piasecka A., Kuczyńska A., Krystkowiak K., Ogrodowicz P., Mikołajczak K., Kachlicki P., Krajewski P. (2012). Comparison of classical and functional principal components applied to chromatographic data. XV-th Meeting of the EUCARPIA Section - Biometrics in Plant Breeding, 5-7 September 2012, Hohenheim, Germany.

Sawikowska A., Piasecka A., Swarcewicz B., Kuczyńska A., Ogrodowicz P., Mikołajczak K., Stobiecki M., Kachlicki P., Krajewski P. (2012). Computational methods for metabolite profiling in POLAPGEN-BD, a biotechnology project for

breeding cereals with increased resistance to drought. EUCARPIA - 19th General Congress - Plant Breeding for Future Generations, 21-24 May 2012, Budapest, Hungary.

Swarcewicz B., Piasecka A., Sawikowska A., Krajewski P., Stobiecki M., Kachlicki P. (2012). Monitoring changes in barley (*Hordeum vulgare*) leaf metabolome under drought stress. Joint Conference of Polish Mass Spectrometry Society and German Mass Spectrometry Society, 4-7 March 2012, Poznań.

Piasecka A., Sawikowska A., Krajewski P., Kachlicki P. (2011). Flavonoids and other phenolic compounds in response of barley (*Hordeum vulgare* L.) plants to drought. 5th Conference of the Polish Society of Plant Experimental Biology, Wrocław, 6-9 September 2011.

Piasecka A., Sawikowska A., Zwierzykowski Z., Kachlicki P. (2010). Chemotaxonomic relationships within the *Festuca-Lolium* complex on the basis of phenolic secondary metabolites. 1st EUCARPIA *Festulolium* Working Group Workshop, 7-8 October 2010, Poznań.

F) Udział w warsztatach i konferencjach:

| | | |
|------------------|-----------------------|--|
| 02.11-06.11.2020 | Online Event/Będlewo | XX Multivariate and Mixed Linear Models |
| 16-17.04.2020 | Oline Event/Amsterdam | 10th World Congress on Chromatography |
| 16.02-22.02.2020 | Będlewo | Multivariate and Mixed Linear Models |
| 4-7.12.2019 | New York, USA | Plant Genomes, Systems Biology & Engineering |
| 10-16.11.2019 | Będlewo | Multivariate and Mixed Linear Models |
| 5-7.11.2019 | Poznań | Genetyka i genomika w doskonaleniu roślin uprawnych |
| 11-13.09.2019 | Heidelberg/Germany | Multimics to Mechanisms - Challenges in Data Integration |

| | | |
|-----------------|-------------------|--|
| 04.05.2019 | Będlewo | Big Data Analytics & Applications |
| 6-8.03.2019 | Poznań | VII Konferencja Chemometria i Metrologia w Analityce |
| 11-17.11.2018 | Będlewo | Research Group Meeting on Multivariate and Mixed Linear Models |
| 7-9.11.2018 | Poznań | The Integrative Plant Biology Conference |
| 8-12.09.2018 | Athens/Greece | 17-th European Conference on Computational Biology |
| 20-24.08.2018 | Będlewo | International Conference on Trends and Perspectives in Linear Statistical Inference LinStat'2018 |
| 19-23.03.2018 | Będlewo | Research Group Meeting on Multivariate and Mixed Linear Models |
| 5-11.11.2017 | Będlewo | Research Group Meeting on Multivariate and Mixed Linear Models |
| 25-29.09.2017 | Będlewo | Mat Triad 2017 |
| 10-14.09.2017 | Kiry | 47th International Biometrical Colloquium |
| 12-18.03.2017 | Będlewo | Research Group Meeting on Multivariate and Mixed Linear Models |
| 27.11-2.12.2016 | Będlewo | XLII Conference on Mathematical Statistics |
| 4-7.09.2016 | Hague/Netherlands | ECCB 2016 - 15th European Conference on Computational Biology |
| 4.09.2016 | Hague/Netherlands | ECCB 2016 Workshop 14: Computational challenges of third Generation DNA Sequencing data analysis |

| | | |
|---------------|------------------------|---|
| 3.09.2016 | Hague/Netherlands | ECCB 2016 Tutorial 4: Genome-wide association analysis and post-analytic interrogation with R |
| 3.09.2016 | Hague/Netherlands | ECCB 2016 Tutorial 3: Protein Interactomes: build and explore protein protein interaction networks |
| 1-2.12.2015 | Poznań | Materiały i biomateriały. Wielkopolskie Centrum Zaawansowanych Technologii |
| 9-11.09.2015 | Wageningen/Netherlands | XVI-th meeting of the Eucarpia Section Biometrics in Plant Breeding |
| 5-7.11.2014 | Poznań | III Ogólnopolska Konferencja IGR PAN: Genetics and genomics in improving plants - from model plant to new variety |
| 6-7.07.2014 | Florence/Italy | International Biometric Conference |
| 10-13.09.2013 | Poznań | IV Polski Kongres Genetyki |
| 25-26.03.2013 | Kraków | IV Warsztaty Naukowe projektu POLAPGEN-BD |
| 26.10.2012 | Poznań | workshops: Podstawy programowania równoległego |
| 05-07.09.2012 | Hohenheim/Niemcy | XV-th Meeting of the EUCARPIA Section - Biometrics in Plant Breeding |
| 10-11.05.2012 | Puławy | III Warsztaty Naukowe projektu POLAPGEN-BD |
| 4-7.03.2012 | Poznań | Joint Conference of Polish Mass Spectrometry Society and German Mass Spectrometry Society |
| 28-29.11.2011 | Poznań | Misja chemo-, bio- i nanotechnologi Wielkopolskim Centrum Zaawansowanych Technologii – Materiały i Biomateriały |

| | | |
|---------------|-----------------|--|
| 18.11.2011 | Poznań | workshops: Korzystanie z zasobów obliczeniowych dla początkujących |
| 04.11.2011 | Poznań | Warsztaty Usługi Powszechnej Archiwizacji |
| 12-16.07.2011 | Tomar/Portugal | Mat Triad 2011 |
| 11-12.01.2011 | Poznań | II Warsztaty Naukowe POLAPGEN-BD |
| 05-06.01.2010 | Poznań | I Warsztaty Naukowe POLAPGEN-BD |
| 12-16.04.2010 | Poznań | Warsztaty Komunikacji Naukowej |
| 04.12.2009 | Poznań | LXXVII sesja naukowa pt. Biometria: teoria i praktyka |
| 23-27.03.2009 | Będlewo | Mat Triad 2009 |
| 01-06.06.2008 | Zeuthen/Germany | Householder Symposium XVII |
| 22-24.03.2007 | Będlewo | Mat Triad 2007 |
| 02-05.02.2006 | Będlewo | German-Polish Workshop for Young Researchers in Applied and Numerical Linear Algebra |
| 03-05.03.2005 | Będlewo | Mat Triad 2005 |

G) Informacje dodatkowe:

| | |
|---------------|---|
| 2018-2019 | Prowadzenie seminarium Katedry Metod Matematycznych i Statystycznych Uniwersytetu Przyrodniczego w Poznaniu |
| 12-16.04.2010 | Uczestnictwo w Warsztatach Komunikacji Naukowej |
| 06.2007 | Zdobycie certyfikatu Londyn Tests of English (Level 3) |