UNIWERSYTET
WARMIŃSKO-MAZURSKI
W OLSZTYNIE

WYDZIAŁ BIOLOGII I BIOTECHNOLOGII

Dziedzina nauk ścisłych i przyrodniczych

Dyscyplina: Nauki biologiczne

Angelika Maria Gomolińska

Rozprawa doktorska

Zmienność genomów organellowych i wybranych genów jądrowych gatunków z rodzaju *Erica*

The variability of the organellar genomes and selected nuclear genes of the genus *Erica*

Praca wykonana w Katedrze Botaniki i Ekologii Ewolucyjnej Uniwersytetu Warmińsko-Mazurskiego w Olsztynie

Promotor: Prof. dr hab. Jakub Sawicki, Katedra Botaniki i Ekologii Ewolucyjnej, Wydział Biologii i Biotechnologii, Uniwersytet Warmińsko-Mazurski w Olsztynie

Olsztyn, 2024

Olsztyn, dn. 20.02.2024 r.

Prof. dr hab. Jakub Sawicki

(~~stopień~~/tytuł naukowy, imię i nazwisko promotora/~~promotora pomocniczego~~*)

Przewodnicząca Rady Naukowej Dyscypliny

Nauki biologiczne

Uniwersytetu Warmińsko-Mazurskiego w Olsztynie

OŚWIADCZENIE

promotora/~~promotora pomocniczego~~* rozprawy doktorskiej

Oświadczam, że rozprawa doktorska **Pani mgr Angeliki Marii Gomolińskiej** pod tytułem **Zmienność genomów organellowych i wybranych genów jądrowych gatunków z rodzaju *Erica*** została przygotowana pod moim kierunkiem i stwierdzam, że spełnia ona warunki do przedstawienia jej w postępowaniu o nadanie stopnia naukowego.

* *niepotrzebne skreślić*

Jakub Sawicki

(podpis promotora)

Olsztyn, dn. 20.02.2024 r.

(miejscowość, data)

Angelika Maria Gomolińska

(imię i nazwisko kandydata)

Nauki biologiczne

(dyscyplina naukowa)

Przewodnicząca Rady Naukowej Dyscypliny

Nauki biologiczne

Uniwersytetu Warmińsko-Mazurskiego w Olsztynie

OŚWIADCZENIE

o samodzielnym napisaniu rozprawy doktorskiej

Świadoma odpowiedzialności prawnej oświadczam, że rozprawa doktorska pod tytułem

„Zmienność genomów organellowych i wybranych genów jądrowych gatunków z rodzaju *Erica*"

została napisana przeze mnie samodzielnie i nie zawiera treści uzyskanych w sposób niezgodny z obowiązującymi przepisami.

Oświadczam również, że przedstawiona rozprawa nie jest i nie była wcześniej podstawądo ubiegania się o nadanie stopnia naukowego doktora.

Ponadto oświadczam, że treść rozprawy doktorskiej przedstawionej przeze mnie do obrony, zawarta na przekazywanym nośniku elektronicznym, jest identyczna z jej wersją drukowaną.

W związku z powyższym oświadczam, że wyrażam zgodę na poddanie rozprawy doktorskiej procedurze antyplagiatowej.

Angelika Maria Gomolińska

(podpis kandydata)

Pragnę wyrazić bezgraniczną wdzięczność

za ogromne pokłady wyrozumiałości, empatii i cierpliwości,

poświęconą energię i czas,

bezcenne zaplecze wsparcia,

unikalne nawigacje merytoryczne,

ponadczasową pomoc w wymiarze naukowym i ogólnoludzkim

Promotorowi mojej pracy doktorskiej – Profesorowi dr hab. Jakubowi Sawickiemu

Składam także najserdeczniejsze podziękowania

wszystkim Pracownikom Wydziału Biologii i Biotechnologii

oraz Uniwersytetu Warmińsko-Mazurskiego,

z którymi zeszły się moje drogi podczas studiów doktoranckich

oraz przygotowania niniejszej dysertacji.

Każde pomniejsze doświadczenie traktuję jako uniwersalną lekcję życia.

Spis treści

Spis skrótów użytych w pracy 6

STRESZCZENIE 8

ABSTRACT 9

1. **Wstęp** 10

1.1. Genetyka konserwatorska 15

1.2. Charakterystyka genomów chloroplastowych roślin 18

1.2.1. Genom chloroplastowy 18

1.3. Charakterystyka regionu organizacji jąderka (NOR) 24

1.4. Rodzina *Ericaceae* 27

1.5. Charakterystyka rodzaju *Erica* 33

2. **Cele pracy** 39

3. **Materiał i metody** 42

3.1. Materiał użyty do badań 42

3.2. Metody badawcze 55

3.2.1. Izolacja DNA 56

3.2.2. Kontrola jakości i ilości wyizolowanego DNA 58

3.2.3. Sekwencjonowanie chloroplastowego DNA z zastosowaniem metody nowej generacji NGS 59

3.2.4. Procedura i etapy przygotowania bibliotek *Erica versicolor* 62

3.2.5. Procedura i etapy przygotowania bibliotek pozostałych gatunków z rodzaju *Erica*…… 67

3.2.6. Analiza bioinformatyczna sekwencji genomowych otrzymanych metodą NGS…… .....……..75

3.2.7. Otrzymanie i analiza sekwencji jądrowych NOR 82

4. **Wyniki**…. 84

4.1. Wyniki sekwencjonowania bibliotek i organizacji otrzymanych genomów chloroplastowych 84

4.1.1. Genom chloroplastowy *Erica versicolor* – wyniki sekwencjonowania i organizacja 86

4.1.2. Genom chloroplastowy *Erica arborea* – wyniki sekwencjonowania i organizacja. 91

4.2. Charakterystyka otrzymanych genomów 96

4.2.1. Analiza sekwencji powtórzeniowych.....96

4.2.2. Analiza regionów odwróconego powtórzenia (IR) 101

4.2.3. Analiza porównawcza genomów *Erica*.....113

4.2.4. Porównanie genomów chloroplastowych z rodzaju *Erica* względem plastomów fotosyntetyzujących przedstawicieli rodziny *Ericaceae* 132

4.3. Otrzymanie sekwencji NOR.....145

4.3.1. Zmienność rDNA w rodzaju *Erica* 148

4.3.2. Struktura jednostek rDNA i porównania międzygatunkowe 152

4.4. Analiza filogenetyczna.....155

4.4.1. Filogenetyczna informatywność.....155

4.4.2. Wzorce zróżnicowania sekwencji powtórzonych w obrębie regionu IGS – zdetektowane sekwencje motywów 160

4.5. Filogenetyka rodzaju *Erica* 166

4.5.1. Informatywność filogenetyczna poszczególnych grup genów chloroplastowych w rodzinie *Ericaceae* 186

5. **Dyskusja**…..... 217

5.1. Genom chloroplastowy *Erica* a inne genomy chloroplastowe 217

5.1.1. Analiza porównawcza struktury i zawartości genów genomów chloroplastowych gatunków fotosyntetyzujących z rodziny *Ericaceae* 219

5.1.2. Analiza porównawcza regionów powtórzeniowych oraz struktury i zawartości genów w plastomach okrytonasiennych 228

5.2. Analiza NOR..... 241

5.3. Potencjał superbarkordingowy plastomów oraz regionów NOR 247

6. **Wnioski** …….251

Spis literatury..... 255

Spis rycin……..289

Spis tabel…… 299

Spis fotografii 302

Suplement…….304

Spis skrótów użytych w pracy

ATP adenozyno-5′-trifosforan (potocznie adenozynotrójfosforan)

cpDNA (chloroplast DNA) - chloroplastowe DNA

CDS (coding DNA sequence) – sekwencje regionów kodujących DNA

CFR (The Cape Floristic Region) – Region Przylądkowy w Afryce Południowej

DNA (deoxyribonucleic acid) – kwas deoksyrybonukleinowy

EMR/ER (ericoid micorrhiza) – mikoryza erikoidalna

IGS (intergenic spacer) – międzygenowy dystansowy DNA

InDel (Insertion Deletion Polymorphism) – polimorfizm insercyjno-delecyjny

ITS (internal transcibed spacer) – wewnętrznie transkrybowany region

IRA/IRB (inverted spacer region A, B) – region odwróconych powtórzeń A i B

LSC (large single copy region) – duży region jednokopijny

NADH dinukleotyd nikotynoamidoadeninowy

NGS (Next Generation Sequencing) – sekwencjonowanie nowej generacjji

NOR (nucleus organizer region) – organizator jąderka

NS (nonsynonymous substitution) – mutacja niesynonimiczna

NTS (nontranscribed spacer) – nietranskrybowane dystansowe DNA

MSA (multiple sequence alignment) – dopasowanie wielu sekwencji

ML (maximum likelihood) – maksymalne podobieństwo

MP (maximum parsimony) – maksymalna parsymonia

ORF (open reading frame) – otwarta bramka odczytu; część sekwencji nukleotydów między kodonem START a kodonem STOP, włącznie z odcinkami niekodującymi

PCR (polymerase chain reaction) – reakcja łańcuchowa polimerazy

PS I/ PS II (photosystem system I and II/ plastocyanin–ferredoxin oxidoreductase) – integralny kompleks białek błonowych dwóch fotosystemów

RAPD (Random Amplification of Polymorphic DNA) – losowa amplifikacja polimorficznego DNA

RNA (ribonucleid acid) – kwas rybonukleinowy

rRNA (ribosomal RNA) – rybosomalny RNA

tRNA (transfer RNA) – transportujący RNA

SANBI (South African National Biodiversity Institute) – Południowoafrykański Narodowy Instytut Bioróżnorodności

SNP (Single Nucleotide Polymorphism) – polimorfizm pojedycznego nukleotydu

SSC (small single copy region) – mały region jednokopijny

SSR (Simple Sequence Repeats) – amplifikowane sekwencje mikrosatelitarne

TIS (transcription initiation site) – miejsce inicjacji transkrypcji

TR (tandem repeats) – powtórzenia tandemowe

TTS (transcription termination site) – miejsce terminacji transkrypcji

UPGMA (Unweighted Pair-Group Method using arithmetic Averages)

STRESZCZENIE

Dokumentacja genomiczna rodzaju *Erica*, pomimo prężnego rozkwitu badań filogenetycznych wciąż pozostawała niekompletna. Intrygujący charakter tego rodzaju, którego 80% przedstawicieli stanowią gatunki endemiczne oraz brak opracowanych kompletnych genomów chloroplastowych oraz sekwencji NOR rodzaju *Erica*, stanowiły główną motywację do przeprowadzenia wnikliwych badań.

Niniejsza rozprawa doktorska stanowi istotny etap badań genetycznych wybranych przedstawicieli z rodzaju *Erica*, rozszerzający zasoby genomowe tych roślin w opaciu o otrzymane kompletne genomy organellowe wraz z ich analizą porównawczą, operony plastydowe wraz z ich analizą komparatywną, filogenetyczną i analizą zmienności, oraz sekwencje jądrowe NOR. Rekonstrukcja filogenetyczna rodziny *Ericaceae* w oparciu o kompletne genomy oraz skonstruowana na podstawie regionów odwróconych powtórzeń uwzględniła nie tylko wybranych przedstawicieli z rodzaju *Erica*, ale także gatunek *Chamaedaphne calyculata*, dla którego także po raz pierwszy opracowano szczegółową budowę IR.

Przeprowadzone badania dostarczyły niezwykle frapujących informacji z zakresu analiz porównawczych, przynosząc nowe dane ekspandujące dotychczasową wiedzę ze zmienności genetycznej i zasobów genomowych rodzaju *Erica*, wzbogacając także dokumentację całej rodziny *Ericaceae*.

Słowa kluczowe: genom chloroplastowy, NOR, sekwencje jądrowe, *Erica*, *Ericaceae*

ABSTRACT

The genomic documentation of the genus *Erica*, despite the dynamic development of phylogenetic research, still remained incomplete. The intriguing nature of this genus, of which 80% species are endemic, and the lack of complete chloroplast genomes and NOR sequences of the *Erica* genus were the main motivation to conduct in-depth research.

This doctoral dissertation is an important stage of genetic research on selected representatives of the genus *Erica*, expanding the genomic resources of these plants based on the obtained complete organellar genomes along with their comparative analysis, plastid operons along with their comparative, phylogenetic and variability analysis, and NOR nuclear sequences. The phylogenetic reconstruction of the *Ericaceae* family based on complete genomes and constructed on the basis of inverted repeat regions included not only selected representatives of the *Erica* genus, but also the species *Chamaedaphne calyculata*, for which a detailed IR structure was also developed for the first time.

The conducted research provided extremely interesting information in the field of comparative analyses, bringing new data expanding the existing knowledge of the genetic variability and genomic resources of the *Erica* genus, also enriching the documentation of the entire *Ericaceae* family.

Key words: chloroplast genome, NOR, nuclear sequences, *Erica*, *Ericaceae*

1. Wstęp

Objęty badaniami w niniejszej pracy rodzaj *Erica* (wrzosiec) należy do klasy *Magnoliopsida* (dwuliścienne), podklasy *Dilleniidae* (ukęślowe), nadrzędu *Ericanae*, rzędu *Ericales* (wrzosowce), rodziny *Ericaceae* (wrzosowate), podrodziny *Ericoideae* (wrzosowe) oraz sekcji *Ericeae*. Spośród 145 rodzajów wchodzących do rodziny wrzosowatych (*Ericaceae*), rodzaj *Erica* stanowi, zaraz po rodzaju *Rhododenron* (różanecznik/azalia/rododendron), rodzaj najobfitszy w liczbę gatunków. Ewolucja tego rodzaju, pomimo prężnego rozwoju badań filogenetycznych na przestrzeni ostatnich lat, wciąż pozostaje słabo poznana i ubogo opisana. Rodzaj ten jest szczególnie interesujący w ujęciu badawczym, między innymi ze względu na to, iż tworzą go głównie, w niemal 80%, gatunki endemiczne. Istnieje także spora luka w dokumentacji genomicznej przedstawicieli *Erica*, co uniemożliwia odtworzenie historii ewolucyjnej, wytypowanie gatunku ancestralnego oraz wyznaczenia dokładnych relacji pomiędzy poszczególnymi gatunkami (Gillespie i Kron, 2010; Van der Niet i in. 2020; Fu Nan i in. 2022).

Rodzaj *Erica* taksonomicznie jest udokumentowany bardzo dobrze, jednak wiedza na temat relacji gatunkowych, szczególnie ta opierająca się na badaniach molekularnych jest niepełna, a także nieugruntowana została kwestia pochodzenia tego rodzaju (Kron i in. 2005). Skomplikowany system taksonomiczny, bardzo duża liczba gatunków, znacząco rozległy, zaburzony zasięg występowania, to czynniki, które przyczyniają się do faktu, iż niewielu badaczy podejmuje się badań nad gatunkami tego rodzaju i relacji między nimi. Dotychczas prowadzone badania mające na celu uściślenie i dookreślenie dokładnego systemu klasyfikacyjnego rodzaju *Erica*, szczególnie na poziomie podrodzaju i sekcji, z próbami ustalenia także na poziomie gatunkowym, były wykonywane w oparciu o cechy morfologiczne. Szereg analiz palinologicznych – poprzez kompleksową syntezę porównawczą wszystkich diagnostycznych, morfologicznych cech pyłku. Badania morfologiczne stanowiły wielokrotnie podstawę do podjęcia prób rewizji usystematyzowania rodzaju *Erica*, jednak póki co wszystkie te próby były nie do końca skuteczne i niezadawalające pod względem całego spektrum czynników, które powinny być brane pod uwagę (Wrońska-Pilarek i in. 2018).

W większości były to badania fragmentaryczne, dotyczące całej rodziny wrzosowatych (*Ericaceae*), czy rzędu *Ericales* (wrzosowce, dwurożne), stanowiła wycinek badań regionalnych flory, a niewiele z nich skupiało się tylko i wyłącznie na rodzaju *Erica*. Na ich podstawie ustalono cechy diagnostyczne ziaren pyłku, w tym jednostki dyspersyjne. Innymi badaniami mającymi rozwinąć wiedzę na temat systematyki wrzośców były badania opiewające o cechy fenotypowe – jak kolor kwiatów oraz typ zapylaczy, badania oparte o modele biogeograficzne i podobieństwa nisz klimatycznych. (Kron i in. 2005; Sarwar Gollam 2011; Wrońska-Pilarek i in. 2018).

Systematyczne stanowisko rodzaju *Erica* zostało uwzględnione w różnych badaniach związków filogenetycznych w obrębie rodziny *Ericaceae*. Wskaźniki dywersyfikacji, wyznaczanie potencjalnych regionów wspólnego przodka czy wzorce ewolucji rodzaju *Erica* były prowadzone na podstawie filogenezy jądrowych i plastydowych markerów sekwencji DNA dla reprezentantów stanowiących niewiele ponad połowę znanych gatunków. Dotychczas publikowane badania z zakresu filogenezy molekularnej nad rodzajem *Erica* opierały się, m.in. o markery *atp*B-*rbc*L, *mat*K (kodowane w chloroplastach) i dotyczyły badań plemion w podrodzinie *Ericoidae*. W badaniach tych wykorzystano metodę maksymalnej parsymonii, największego prawdopodobieństwa i analizę Bayesa celem poprawy stanu wiedzy o relacjach filogenetycznych w obrębie podrodziny *Ericoidae* i rozwikłania braków wiedzy ograniczających potencjał dalszych analiz kladystycznych, które wynikają z niewyjaśnionych relacji w najgłębszych węzłach (Gillespie i Kron, 2010).

Pierwsze badanie filogenetyczne obejmujące stricte rodzaj *Erica*, bazowało na 3 regionach DNA, które we wcześniej prowadzonych analizach wykazały użyteczność u rodziny *Ericaceae*; 2 sekwencjach kodowanych przez chloroplasty *atp*B-*rbc*L i częściowym *mat*K oraz 1 sekwencji z genomu jądrowego (wewnętrzny region transkrybowany jądrowego DNA rybosomalnego, ITS). Inspiracją dla tych badań była wąska dystrybucja rodzaju, bardzo wyjątkowa dla całej rodziny wrzosowatych, która podsyciła spekulacje na temat właściwego pochodzenia rodzaju wrzośców, nieustalonego w innym typie badań. Celem było ustalenie na podstawie danych molekularnych z trzech wspomnianych regionów DNA, gdzie leży centrum pochodzenia rodzaju *Erica*.

Analiza porównawcza wykazała na monofiletyczność grupy taksonów afrykańskich, pochodzących z taksonów europejskich. Ustalono, że szeroki zakres geograficzny bikontynentalnego gatunku *Erica arborea*, wskazuje, że wspólny przodek także był szeroko rozpowszechniony na obu kontynentach i prawdopodobnie rozszerzył zasięg na Afrykę po połączeniu Europy i Afryki w połowie miocenu (McGuire i Kron, 2005).

Istotna analiza filogenetyczna, uwzględniająca rodzaj *Erica* z wykorzystaniem wyłącznie sekwencji ITS z włączeniem badań morfologicznych obejmowała zakresem aż 379 gatunków, 45% gatunków z zakresu geograficznego rodzaju, wraz z innymi podgrupami *Ericoidae.* Wyniki ukazały morfologiczne i geograficzne spójności niektórych kladów, z licznymi zmianami w makromorfologii kwiatów reprezentowane przez zróżnicowanie indywidualnych cech morfologicznych i typy zapylania. Wnioskowano o monofiletyczności afrykańskich wrzośców. Jednakże z uwagi na ograniczoną rozdzielczość drzewa opartego jedynie na tych sekwencjach, zidentyfikowane klady potraktowano testowo, jako skuteczny przewodnik do dalszego próbkowania powiądzań filogenetycznych z podkreśleniem o konieczności rozszerzenia badań o więcej danych do sformułowania rzetelnych wniosków (Pirie i in. 2011).

Dane te kilka lat później rozszerzono, tłumacząc nieproporcjonalne bogactwo gatunkowe światowej bioróżnorodności w tzw. gorących miejscach (hot-spot) niskim wymieraniem (ewolucyjnym „muzeum") i modelami wysokiej specjacji (hot-bed). Tę hipotezę filogenetyczną oparto na jądrowych i plastydowych sekwencjach DNA 488 gatunków *Erica* – ITS oraz chloroplastowe *trn*T-*trn*L, *trn*L-*trn*F-*ndh*J dla wszystkich prób, oraz inne markery plastydowe intron *trn*L, *atp*I-*atp*H (spacer), intron *trn*K-*mat*K, gen *mat*K, *psb*M-*trn*H spacer, gen *rbc*L, intron *rpl*16 i *trn*L-*rpl*32 (spacer) zostały dodane do analiz jedynie dla wybranych gatunków, bazując na wstępnych badaniach, jako reprezentanci wcześnie dywersyfikujących linii, dla poprawy rozdzielczości węzłów w drzewie plastydowym. Badania te dowiodły endemiczność gatunków CFR, szybkie tempo ewolucji w tym obszarze oraz niższe wskaźniki wymierania. Wskazano, iż różnorodność gatunków wrzośców w południowej Afryce jest wynikiem pojedynczego promieniowania w ciągu ostatnich 15 milionów lat. W zestawieniu z rodowodami przodków na Palearktyce, tempo specjacji przyspieszyło w Afryce i na Madagaskarze, kolejny wybuch specjacji w CFR przekraczał wskaźniki dywersyfikacji netto innych

kladów w tym regionie. Wywnioskowano, iż rodzaj *Erica* ilustruje model formacji roślinnej powstały przez niedawną specjację, co oznacza, że uformowanie współczesnego stanu Cape Floristic Region otworzyło wiele nowych nisz, które sukcesywnie zajmowane były przez dywersyfikację lokalnych gatunków (McGuire i Kron, 2005; Gillespie i Kron, 2010).

W przywołanych przypadkach badania nie uwzględniały szczegółowej, uznanej klasyfikacji rodzaju *Erica*, ani nie zaproponowały nowej jej odsłony. Nie były one przeprowadzone z uwzględnieniem taksonów międzyrodzajowych na podgatunki, sekcje, czy też podgrupy. Były realizowane bez wcześniejszego dogłębnego przeanalizowania, opracowania i poznania genomów, bez sekwencjonowania kompletnych genomów, bez uwzględnienia kompletnych sekwencji DNA, bez odwołania do kompletnych genomów, bazowały jednynie na krótkich fragmentach DNA, co skłania do koncypowania o istnieniu innych miejsc w genomie o znacznie głębszej historii ewolucyjnej rodzaju i dużo bardziej przydatnych do wnioskowania filogenetycznego. Nie opublikowano dotychczas badań dotyczących kompletnych genomów *Erica*, całych genów plastydowych ani analizy regionów NOR w rodzaju *Erica*, ponadto prowadzone do tej pory badania genetyczne nie wyjaśniły w pełni hipotez dotyczących filogenezy wrzośców. Dotychczas badane regiony niekodujące mogą w pełni nie odzwierciedlać rzeczywistej zdolności adaptacyjnej populacji, która jest określana głównie przez geny funkcjonalne (Kardos i in. 2021).

Szczegółowa analiza genów genomów chloroplastowych pozwoli na określenie tempa ewolucji molekularnej poszczególnych genów/rodzin genów z uściśleniem, która rodzina genów charakteryzuje się jej najszybszym tempem, przy okazji możliwa będzie także ich ocena jako markerów filogenetycznych. Szereg szczegółowych badań prowadzi do potwierdzenia wcześniejszych hipotez o przydatności poszczególnych genów jako markerów filogenetycznych, m.in. testując ich informatywność filogenetyczną, tempo mutacji każdego z genów czy poziom wysycenia substytucjami (Soltis i in. 2021; Wang i in. 2022).

Wyjątkowo frapującym zagadnieniem jest określenie stopnia pokrewieństwa między gatunkami z terenu Afryki, a pozostałymi obszarami występowania wrzośców. Filogenetyczne relacje między wybranymi gatunkami *Erica* są analizowane w celu określenia, czy pochodzenie tego rodzaju leży w Europie czy Afryce oraz dla generalnego odzwierciedlenia rozkładu geograficznego historycznej dystrybucji gatunków.

Dotychczasowe badania prowadzone nad rodzajami *Erica* nie uwzględniały tych genów genomu chloroplastowego, które są zaplanowane w niniejszej pracy ani genów rRNA wraz z transkrybowanymi i nietranskrybowanymi regionami międzygenowymi (NOR). Względny udział wszystkich dotychczasowych wyników pozostawia temat pochodzenia rodzaju *Erica* oraz powiązań międzygatunkowych w jego obrębie, wciąż otwartym i jest nadal przedmiotem gorących dyskusji, a wraz z dostępnymi hipotezami filogenetycznymi dla wrzośców stanowi źródło danych do dalszych badań i obszernej konfrontacji naukowej.

Badaniem w niniejszej pracy objęto 2 podrodzaje należące do rodzaju *Erica*, reprezentowane przez 8 gatunków.

- *Syringodea* (7 gatunków, reprezentujących 5 sekcji)
- *Chlamydanthe* (1 gatunek, reprezentujący 1 sekcję)

1.1. Genetyka konserwatorska

Jednym z istotniejszych wyzwań współczesnej nauki jest sprostanie konsekwencjom Szóstego Wielkiego Wymierania Holoceńskiego, do którego głównie przyczyniła się działalność człowieka (Kumaran i Padmalal, 2022; Molnár i in. 2023). Aktualnie z powierzchni Ziemi ustępuje ok. 3 tys. gatunków w ciągu roku. Co godzinę wymierają 3 gatunki, niejednokrotnie nigdy nieopisane, nieodwracalnie zubożając bioróżnorodność biosfery. Wśród roślin istnieje 13% gatunków krytycznie zagrożonych, 19% zagrożonych oraz 38% gatunków wrażliwych. Jedynie 30% gatunków pozostaje w prognozie niskiego ryzyka wyginięcia (Westwood i in. 2021; Cowie i in. 2022).

Rośliny prezentują się pozornie wyjątkową możliwością ochrony. Potencjalnie łatwo rosną na obszarach z odpowiednimi warunkami klimatycznymi; ponadto można przechowywać nasiona większości gatunków przez dziesięciolecia, czasem stulecia, w specjalnie powołanych bankach nasion lub komorach wzrostowych (Ray i Bordolui, 2021; Campbell-Martinez i in. 2022). Kolosalny wzrost industrializacji, wszelakie formy ingerencji człowieka, sprawiły że obszar zajęty przez przyrodę – w tym roślinność - jest znacznie mniejsza niż kiedyś i szybko się kurczy. Antropogeniczne i naturalne zmiany klimatyczne są już postrzegane jako problem globalny, dotychczas podejmowane wysiłki i polityczne larum nie osiągnęły poziomu, na którym trwający wzrost temperatury mógłby zostać ustabilizowany (Ritchie i in. 2021; Molnár i in. 2023).

Ponadto nieprzerwanie odnotowuje się ciągłą, znaczącą utratę różnorodności biologicznej. Przez wiele lat rośliny nie otrzymywały należytej uwagi zarówno medialnej, politycznej jak i protekcyjnej, na jaką niewątpliwie zasługują. W codziennym życiu jesteśmy całkowicie zależni od roślin, są one nie tylko komponentem naszego ekosystemu, ale i elementem naszej żywności, farmaceutyków, materiałów budowlanych, przemysłowych oraz niezliczonych innych produktów codziennego użytku (Ceballos i in. 2015; Panagopoulos, 2022).

Szacuje się od 390 000 do 400 000 opisanych gatunków roślin naczyniowych – widłaki (*Lycopodiophyta*), paprocie (*Polypodiopsida*), nagonasienne (*Gymnospermae*), okrytonasienne (*Angiospermae*) - prawdopodownie około 70 tysięcy gatunków czeka nadal na opisanie, w pewnych przypadkach nawet wciąż na odkrycie. Szacowania, że co najmniej 600 gatunków roślin wymarło w ciągu ostatnich 260 lat są bardzo prawdopodobnie znacząco zaniżone. Wydaje się, że na przestrzeni przyszłych kilkudziesięciu lat uzasadniona jest hipoteza, że 20% ogółu roślin może nieodwracalnie zniknąć, przy czym być może dwukrotnie większa część jest zagrożona utratą egzystencji do końca wieku XXI (Dinerstein i in. 2019; Holz i in. 2022).

Wzrost populacji ludzi, ekspansja i intensyfikacja rolnictwa, działania doprowadzające do trwałych zmian klimatycznych i niepohamowane zapotrzebowanie konsumpcyjne, w tym rosnąca sprzedaż i eksploatacja wielu gatunków roślin znacząco utrudni zachowanie pokaźnej części istniejących formacji roślinnych, w tym lasów tropikalnych, do końca tego wieku (Lovejoy i Hanah, 2019; Raven, 2020).

Szczegółowymi badaniami nad gatunkami zagrożonymi zajmuje się dyscyplina naukowa przyjęta z nomenklatury anglosaskiej jako *conservation genetics*. Termin ten z wolna przenika również do nauki polskiej i jej nomenklatury, przy czym opisywana nim praktyka posiada w Polsce bardzo długie, bo sięgającą okresu międzywojennego, tradycje (Hamm i in. 2013; Willi i in. 2021).

Genetyka konserwatorska pozwala na wykorzystanie pod kątem teoretycznym i praktycznym genetyki w ochronie gatunków jako podmiotów dynamicznie dostosowujących się do zmieniających się warunków środowiska czy działań antropopresyjnych, do radzenia sobie z tymi zmianami tak, aby zminimalizować ryzyko wymarcia gatunku. Poza genetyką klasyczną, populacyjną i molekularną czerpie i łączy osiągnięcia wielu nauk konserwatorskich, molekularnych, ekologii, ewolucjonizmu, bioinformatyki czy statystyki (Avise i in. 1996; Izza Ab Ghani i in. 2022).

Genetyka konserwatorska stanowi teoretyczne i praktyczne wykorzystanie genetyki w ochronie gatunków jako podmiotów dynamicznie dostosowujących się do zmieniających się warunków środowiska, do radzenia sobie z tymi zmianami tak, aby zminimalizować ich ryzyko wymarcia. Obejmuje zarządzanie genetyczne małych populacji w celu maksymalizacji zachowania różnorodności genetycznej i minimalizacji chowu wsobnego, wyjaśnianie taksonomicznych niepewności, wyznaczanie jednostek i praktyk zarządzających ochroną oraz analizy danych molekularnych w dokładnych procedurach badawczych mających na celu zrozumienie istoty biologii gatunków. Głównie genetyka konserwatorska skupia się na procesach zachodzących wewnątrz populacji małych i rozdrobnionych, dotycząc praktycznych metod minimalizujących szkodliwe działanie czynników w ich obrębie. Jej nadrzędnym przesłaniem jest opracowanie i poznanie metod wzmacniających wpływ na los różnorodności genetycznej małych, zagrożonych populacji oraz dużych populacji, w których silnie działa selekcja naturalna. Dlatego genetyka konserwatorska koncentruje się głównie na zachowaniu potencjału ewolucyjnego (Frankham i in 2010; Kershaw i in. 2022).

Dotychczasowe badania, które prowadzono w ramach genetyki konserwatorskiej, opierały się przede wszystkim na analizach sekwencji anonimowych czy niekodujących, natomiast nie skupiano się na analizach sekwencji kodujących, które w główniej mierze odpowiadają za przeżywalność, zdolności adaptacyjne czy prawidłowość i przebieg procesów fizjologicznych (Song i in. 2021; Zhang i in. 2021).

Tego typu sekwencje funkcjonalne zlokalizowane są także w genomach mitochondrialnych i chloroplastowych. Innowacyjnym pomysłem i jednoczesnym przełomem w badaniach z zakresu genetyki konserwatorskiej jest zatem wdrożenie szerokiej analizy genów kodujących funkcjonalne białka z zastosowaniem sekwencjonowania nowej generacji. Pozwoli to między innymi na dogłębne poznanie poziomu i hot spotów zmienności cpDNA (Nakahama, 2021).

1.2. Charakterystyka genomów chloroplastowych roślin

1.2.1. Genom chloroplastowy

Chloroplasty to organelle komórkowe charakterystyczne jedynie dla Królestwa Roślin oraz dla grupy eukariotycznych glonów. Aktualny stan wiedzy sugeruje na ich pochodzenie od sinic, które zostały prawdopodobnie zaatakowane lub pochłonięte przez heterotroficzną, eukariotyczną komórkę gospodarza ok. 1,5 miliarda lat temu. Przypuszcza się, że zarówno pierwotny gospodarz, jak i symbiont przystosowały się do swoistego stanu wewnątrzkomórkowego (endosymbiotycznego) poprzez szereg różnych procesów, takich jak endosymbiotyczny transfer genów czy kurczenie się genomu (Shao i in. 2019; Sato, 2021). Zjawisko to, które obejmuje przeniesienie części genów cyjanobakterii do jądra gospodarza przy jednoczesnej utracie niektórych genów znamy pod terminem pierwotnej endosymbiozy i tworzy ono tzw. pierwotny plastyd. Niektóre gatunki tworzą natomiast tzw. wtórne plastydy w procesie wtórnej endosymbiozy (Chan i in. 2011; Azuma i in. 2023).

Genom chloroplastowy (cpDNA) występuje w chloroplaście w wielu kopiach, od 200 do 300 identycznych cząsteczek DNA. Zbudowany jest na ogół z jednorodnych dwuniciowych kolistych cząsteczek DNA o rozmiarze 110 000 - 200 000 pz, zawierających od 30 do 50 różnych genów RNA i wielu genów kodujących białka. U niektórych gatunków roślin spotyka się jednak także postać liniową i rozgałęzioną genomu chloroplastowego, jak choćby u gatunku koniczyny beczkowatej (*Medicago truncatula*) (Shaver, 2008).

Wśród gatunków zielonych glonów wielkość genomów chloroplastowych cechuje istotna rozbieżność, która nie jest notowana u gatunków roślin lądowych, gdzie oscyluje w granicach 120 000 – 160 000 pz. Wielkość genomów chloroplastowych zielonych glonów waha się bowiem od 37 400 pz w przypadku przedstawiciela rodzaju *Helicosporidium*, do nawet 269 000 pz w rodzaju *Dunaliella* (Smith i in. 2010).

Klasyczny schemat budowy komponentów genomu chloroplastowego to para identycznych powtórzeń odwróconych (Inverted Repeats A and B; IRA, IRB), których długość na ogół mieści się w przedziale od 5 000 do 76 000 pz. Oddzielone są one małym (Small Single Copy Region; SSC) i dużym (Large Single Copy Region; LSC) regionem pojedynczych kopii, którego wielkość na ogół decyduje o rozmiarze całego plastomu. SSC i LSC charakteryzują się obecnością swoistych, unikalnych w budowie sekwencji. Konserwatywne sekwencji regionów IR składają się z 3 podobnych do siebie fragmentów kodujących, m.in. rRNA i tRNA. Od takiego ogólnego wzorca budowy u niektórych gatunków zachodzą odchylenia w postaci skrócenia długości (zmniejszenie obszaru IR charakterystyczne jest dla rodziny sosnowatych (*Pinaceae*) np. u *Pinus thunbergii* – 600 pz, z kolei w rodzajach *Picea* (świerk) i *Nothsuga* (monotypowy rodzaj z Chin) regiony IR zostały zredukowane do ok. 250 pz) bądź całkowitej redukcji powtórzeń odwróconych jak w przypadku euglenin (*Euglena gracilis*) (Shaw, 2007; Chen i in. 2023).

Genów kodowanych przez cpDNA jest około 200, a liczba funkcjonalnych genów i ich wielkość nie koreluje bezpośrednio z rozmiarem genomu (Palmer, 1985; Green, 2011). Geny, które kodują białka można podzielić na dwie zasadnicze grupy: geny, które są zaangażowane w mechanizmy ekspresji i translacji w chloroplastach oraz geny odpowiedzialne za funkcje bioenergetyczne i geny odpowiadające za przebieg procesu fotosyntezy (Rivas i in. 2002). Geny kodowane w cpDNA to 4 geny dla rRNA, 30 genów dla tRNA, geny białek fotosyntezy, geny podjednostki NADH dehydrogenazy, gen podjednostki proteazy (*clp*P), gen podjednostki acetylo-CoA-arboksylazy (*acc*D), geny dla prokariotycznej polimerazy RNA, geny białek aparatu translacyjnego oraz kilka genów o wysokiej zmienności - *ycf*, które stanowią ramki odczytu dla białek o nieznanej dotąd dokładnie funkcji. Dokładniej opisane są geny RNA, ok 6-8 genów rRNA zlokalizowanych jest w IR, natomiast ok. 36-40 genów tRNA rozprzestrzenia się po całym genomie. Interesującym parametrem wśród plastomów jest ich liczba intronów, która waha się od jednego do kilkudziesięciu i nie stanowi predyktora w koncypacjach filogenetycznych, z uwagi na brak regularności liczebnej w obrębie gatunków tego samego taksonu (Zuo i in. 2020; Xiong i in. 2021; Maciszewski i in. 2022).

Zważywszy na wiele cech chloroplastów, w tym specyfikę właściwości polimerazy chloroplastowej DNA, obecność systemów naprawczych, niski stopień rekombinacji pary GC stanowią 30% ogółu par zasad w genomie chloroplastowym, co jest cechą silnie odróżniającą genom chloroplastowy od genomu jądrowego i przekłada się na wolniejsze od niego tempo ewolucji (Han i in. 2019; Dobrogojski i in. 2020). Wśród wielu gatunków, liczne geny cpDNA wciąż wymagają dogłębnej analizy i scharakteryzowania. Dziedziczenie po jednym rodzicu, przeważnie w linii matczynej, i brak rekombinacji stanowi zaletę wykorzystania cpDNA w analizach filogenetycznych (Rochaix, 1999).

Chloroplasty posiadają unikalny system transkrypcyjny złożony z pozostałych komponentów prokariotycznych, tj. prokariotyczna polimeraza RNA (Mohammed i in. 2022). Geny kodowane w cpDNA są niewystarczające, aby regulować skomplikowaną ekspresję genów, stąd wiele komponentów regulatorowych kodowanych w jądrze komórkowym, jak np. w przypadku genów kodujących białko enzymu karboksydysmutazy (ribulose bisphosphate carboxylase-oxygenase; RuBisCo), na które składa się 8 dużych (RuBisCO Large; *rbc*L) i 8 małych (RuBisCo Small; *rbc*S) podjednostek. Niemal wszystkie gatunki roślin mają zlokalizowany gen kodujący duże podjednostki w genomie chloroplastowym, a gen kodujący małe podjednostki znajduje się już w genomie jądrowym. Gen *rbc*L odpowiada nie tylko za przebieg procesu fotosyntezy, ale także za inne procesy metaboliczne, o czym świadczy jego obecność w genomach chloroplastowych gatunków, które nie przeprowadzają fotosyntezy (Wicke i in. 2011).

Kompleksy białkowe odpowiadające za przebieg procesu fotosyntezy to przede wszystkim fotosystemy I i II (PSI i PSII), syntaza ATP i kompleks NADPH (Dobrogojski i in. 2020). Kompleks fotosystemu I składa się z piętnastu podjednostek; z czego sześć jest kodowanych przez genom chloroplastowy : *psa*A, *psa*B, *psa*C, *psa*I, *psa*J i *psa*M, głównie w obrębie LSC. Kompleks fotosystemu II składa się z siedemnastu podjednostek, z czego piętnaście z nich jest kodowanych w obrębie genomu chloroplastowego, w regionie LSC. Obydwa fotosystemy scala kompleks cytochromów b_6/f, w którym sześć na dziewięć podjednostek jest kodowanych w genomie chloroplastowym, w tym *pet*A, *pet*B, *pet*D, *pet*G, *pet*L i *pet*N. Kompleks białkowy syntazy ATP to dziewięć genów, z czego jedynie trzy identyfikuje się w cpDNA – *atp*A, *atp*B i *atp*E.

Kompleks oksydoreduktazy plastochinonowej reprezentowane są w cpDNA u większości gatunków przez jedenaście genów, choć istnieją przypadki zaniku lub występowania niektórych z podjednostek w formie pseudogenów (Wicke i in. 2011; Mohammed i in. 2022). Pomimo występowania uzależnienia od białek pochodzenia jądrowego, chloroplasty wykształciły własny układ odpowiedzialny za ekspresję genów, pochodzący od ancestralnych sinic. Polimeraza RNA odpowiadająca za transkrypcję jest kluczowym elementem tej maszynerii (Palomar i in. 2022).

W chloroplaście można zdetektować jej dwa typy – jeden kodowany w plastydzie (plastid-encoded RNA polymerase; PEP), drugi kodowany w jądrze komórkowym, specyficzny dla roślin okrytozalążkowych (nuclear-encoded RNA polymerase; NEP). Kompleks PEP składa się z 4 podjednostek – α, β, β, β'' kodowanych adekwatnie przez geny - *rpo*A, *rpo*B, *rpo*C1 i *rpo*C2. Zasadniczą różnicą pomiędzy podjednostkami rdzenia PEP jest ich masa cząsteczkowa. Podjednostka α ma masę 38 kDa, β – 120 kDa, β′ – 85 kDa, β″ – 185 kDa. W przypadku większości roślin lądowych gen *rpo*A jest zorganizowany w jeden operon pod kontrolą tego samego promotora. Z kolei geny *rpo*B, *rpo*C1 i *rpo*C2 współtworzą odrębny operon RPOBC. Drugi nadmieniony typ polimerazy – NEP, składa się natomiast tylko z jednej podjednostki. Polimeraza ta ewoluowała poprzez duplikację genu jądrowego kodującego mitochondrialną polimerazę RNA i wykazuje duże podobieństwo do polimeraz RNA typu fagowego, które także zbudowane są z jednej podjednostki. Zarówno PEP, jak i NEP są nieodzowne w procesie transkrypcji białek chloroplastowych. Chociaż NEP i PEP rozpoznają różne promotory, istnieje wiele genów chloroplastowych, które posiadają promotory rozpoznawane zarówno przez jeden jak i drugi typ polimerazy RNA (Tiller i in. 2014; Dobrogojski i in. 2020).

Genomy plastydowe roślin wykazują się wysokim stopniem informatywności i przydatności w badaniach ewolucyjnych (Guo i in. 2020; Liang i in. 2020). Wykazują bowiem zmienność pod względem struktury, zawartości genów, zmienności nukleotydowych regionów kodujących i niekodujących (Mohammed i in. 2022; Palomar i in. 2022).

Prace badawcze skupione na zgłębieniu budowy strukturalnej plastomów rozpoczęły się zdecydowanie wcześniej niż badania tego samego typu odnoszące się do genomów jądrowych. Pierwsza kompletna sekwencja genomu chloroplastowego pojawiła się w roku 1986 i dotyczyła genomu chloroplastowego tytoniu (*Nicotiana tabacum* var. Bright Yellow 4); natomiast podobne dane na temat genomów jądrowych zostały opublikowane około 10 lat później (Shinozaki i in. 1986; Tong i in. 2020).

Było to uwarunkowane między innymi sporymi różnicami w wielkościach genomów oraz kondycją i rozwojem technik sekwencjonowania, co przekłada się także na aktualny stan wiedzy na temat budowy i funkcji plastomów o znaczeniu poznawczym oraz aplikacyjnym (Slatko i in. 2018).

Bazy danych dysponują aktualnie danymi na temat plastomów ponad ośmiu tysiący gatunków roślin z całego świata, z wielu taksonów (Turudić i in. 2022). Zsekwencjonowane genomy chloroplastowe dostępne są dla mniej niż 70 rodzajów z trzech tysięcy gatunków roślin uprawnych i dla około dwóch tysięcy z 350 tysięcy gatunków roślin kwiatowych, co przekłada się na konieczność kontynuowanaia ekspandowania wiedzy odnoszącej się do funkcjonalności i organizacji genomów chloroplastowych, umożliwiając prowadzenie szczegółowych i dogłębnych badań komparatywnych (Daniell i in. 2021). Taki stan rzeczy przyczynił się także do rozwinięcia wielu gałęzi badań biotechnologicznych w oparciu o sekwencje plastomowe, w tym także inżynierii genomowej dotyczącej transferu genów, kontroli ekspresji genów (Liu i Zhang, 2022).

Tworzenie modyfikacji genomów, transformacja konstruktów i ich kontrola jest prostsza w przypadku genomów chloroplastowych, niż w jądrowych (Olejniczak i in. 2016). Znaczny progres w badaniach nad plastomami przyczynił się do rozszerzenia skali wykorzystania dotychczasowych analiz i modyfikacji w obrębie plastomów. Wśród dostępnych potencjalnych przypadków zastosowania inżynierii plastomu najbardziej obiecujące wydaje się produkowanie białek rekombinowanych. Dzięki wysokiej wydajności białek docelowych w plastomach transformowanych i modyfikacjom potranslacyjnym w połączeniu z brakiem wyciszania naukowcom przyświeca wizja zastosowania modyfikacji chloroplastów (transplastomów) do produkcji antygenów, białek terapeutycznych, enzymów czy szczepionek (Singhal i in. 2022).

Modyfikacje plastomów wdrażające zwiększoną odporność na czynniki stresowe są także intrygującą gałęzią eksploracji, przekładającą się w swoich skutkach pozytywnych na zwiększenie plonów, zmniejszenie zakresu stosowanych środków syntetycznych jak np. pestycydy, umożliwienie roślinom życia na terenach, na których dotychczas miałyby problem z przetrwaniem (Singhal i in. 2022; Zaveri i in. 2022).

Dane na temat struktury plastomu mogą być wielce wartościowe w analizach filogenetycznych, stanowiąc wydajne narzędzie molekularne z uwagi na niski współczynnik mutacji oraz łatwość amplifikacji (Mahadani i in. 2022).

Początkowa niewielka liczba zsekwencjonowanych genomów chloroplastowych przekładałą się na niską jakość analiz filogenetycznych. Intensywny wzrost tego typu badań wpłynął na poprawę wiarygodności wyników tego typu projektów. Badania filogenomiczne w oparciu o analizy genomu chloroplastowego są obecnie prężnie rozwijającą się dziedziną nauki, opierającą się o postępujące metody sekwencjonowania następnej generacji (NGS), która początkowo stosowana była jedynie do wysokich jednostek taksonomicznych, by obecnie wspierać dociekania dotyczące pojedynczych taksonów z konkretnej rodziny lub rodzaju. Cechy genomu chloroplastowego przemawiają za zastosowaniem ich pełnych sekwencji jako roślinnych barkodów (Peng i in. 2013, Olejniczak i in. 2016; Li i in. 2020).

1.3. Charakterystyka regionu organizacji jąderka (NOR)

Organizator jąderka (z ang. *nucleolus organizer region*) – NOR - jest to obszar jąderkotwórczy, stanowiący fragment genomu ulokowany w przewężeniu wtórnym niektórych chromosomów. Odkrywcą organizatora jąderka jest Barbara McClintock, która opisała go po raz pierwszy u gatunku *Zea mays* w 1934 roku (McClintock, 1934).

Każda duża jednostka 35S rDNA (45S u zwierząt) zawiera geny rRNA: 18S, 5,8S i 26S, wewnętrzne transkrybowane region (ITS) i intergeniczny odstępnik (IGS). Jednostki 35S są zorganizowane w układach tandemowych na jednym lub kilku loci. 5S rDNA kodujący transkrypt o długości 120 par zasad, uważany był w wielu badaniach tradycyjnie za zajmujący osobne miejsca chromosomalne (typ S) u roślin nasiennych. Fizyczne połączenie genów 5S i 35S przeważa nad organizacją rDNA u glonów i wczesnych roślinach lądowych, takich jak mchy. Nowsze badania te doprowadziły do hipotezy, że „uwolnienie" genu 5S z jednostki 35S mogło wystąpić u wspólnego przodka okrytozalążkowego po oddzieleniu od wcześnie rozchodzących kladów roślin (Wicke i in. 2011).

Połączony układ jednostek 35S-5S (typ L) został znaleziony w kilku gatunkach z rodzaju *Artemisia* (z rodziny *Asteraceae*, uważanej za jedną z najbardziej filogenetycznie wywodzących się grup okrytozalążkowych), najpierw w badaniach cytogenetycznych, następnie potwierdzonych badaniami molekularnymi. Dodatkowe badania wykazały, że aż 25% członków *Asteraceae* mogło mieć nietypowy układ rDNA typu połączonego L, który także został znaleziony w żywej skamieniałości miłorzębu dwuklapowego *Gingko biloba* (Garcia i in. 2009).

Wciąż pozostaje nierozstrzygnięte, czy typ łączny (L) czy osobny (S) stanowi rodowy typ rDNA u gatunków okrytozalążkowych. W większości genomów typu L insercja 5S występuje w IGS w odległości 1 tys zasad poniżej genu 26S, a odpowiedni transkrypt jest kodowany wyłącznie na przeciwnej nici DNA niż 26S rRNA. Pomimo, iż w komórce musi istnieć stechiometryczny stosunek cząsteczek rRNA, istnieją zasadnicze różnice w odniesieniu do regulacji transkrypcji poszczególnych genów.

Duży transkrypt 35S wytwarzany przez polimerazę RNA I (Pol I) jest przetwarzany endonukleolitycznie z wytworzeniem dojrzałych cząsteczek 18S, 5,8S i 26S rRNA. Promotor Pol I jest zlokalizowany w intergenicznym obszarze 26S-18S, wiążąc się z kompleksem czynników transkrypcyjnych. Natomiast transkrypcja genów 5S jest przeprowadzana przez polimerazę RNA III (Pol III), która potrzebuje wewnętrznego promotora w genie, oprócz czynników transkrypcyjnych TFIIIA, TFIIIB i TFIIIC. Trójelementowa konstrukcja wewnętrznego promotora Pol III zawiera element wewnętrzny (tzw. IE), A-box oraz C-box, elementy wysoce konserwatywne zarówno u roślin, jak i zwierząt (Girbig i in. 2022; Jarrous i in. 2022).

Narzędzia epigenetyczne to kolejna odsłona możliwości kontroli ekspresji zaangażowana w regulację obu typów genów rRNA. W wyciszeniu kopii transkrybowanych genów biorą udział złożone mechanizmy epigenetyczne oparte na modyfikacjach chromatyny, w tym metylacji DNA. Nowe badania dowodzą o nietypowym układzie połączonym genów 5S, występującym u co najmniej kilku gatunków roślin, który jest w pełni kompatybilny z ich ekspresją i funkcjonalnością. Badania te wskazują na niezwykłą dynamikę genu 5S przejawiającą się na różnych poziomach, takich jak zmiana pozycji wewnątrz chromosomów, struktura jednostkowa, modyfikacja epigenetyczna i rozważana różnorodność motywów regulacyjnych (Lan i in. 2011, Singh i in. 2021).

Klasyczny NOR zbudowany jest z powtarzających się genów kodujących cząsteczki rRNA o stałych sedymentacji 5,8S, 18S i 28S (na ogół; istnieje wiele odstępstw), przedzielanych intronami (tu: ITS) (Oborská-Oplová i in. 2022). Po zakończeniu podziału komórkowego i odtworzeniu normalnej struktury jądra komórkowego (w interfazie), NOR-y ulegają dekondensacji i tworzy się wokół nich jąderko. Organizatory pochodzące z różnych chromosomów mogą się łączyć, tworząc wspólne jąderko. Pomiędzy genami rRNA 18S i 5,8S znajduje się ITS1, natomiast między 5,8S a 26S znajduje się ITS2. ITS2 jest uznanym markerem analiz filogenetycznych domeny eukariontów, który ma cechy dobrego barkodu (Dutta, 2022; Hilário i in. 2022).

Istnieją pewne przesłanki sugerujące, że sekwencje ITS są zachowywane niezależnie od ewolucji historii życia królestwa roślin. Geny kodujące rybosomalne RNA i elementy dystansowe występują w powtórzeniach tandemowych, w tysiącach kopii długości, oddzielonych od siebie regionami nie transkrybowanego DNA zwanymi międzygenowymi dystansowymi (IGS) lub nietranskrybowanego dystansowego DNA (NTS). Porównanie sekwencji regionu ITS jest szeroko stosowane w taksonomii i filogenetyce, ponieważ jest łatwe do amplifikacji nawet przy niewielkich ilościach DNA (ze względu na dużą liczbę kopii genów rRNA), oraz ma wysoki stopień zmienności nawet w przypadku blisko spokrewnionych gatunków (Hilário i in. 2022). Można to wyjaśnić relatywnie niską presją ewolucyjną działającą na takie sekwencje niefunkcjonalne (Sultanov i Hochwagen, 2022).

Bardzo czułe markery zmian ewolucyjnych, rejestrujące zmiany szybciej niż analizowane często regiony ITS stanowią regiony IGS. Regiony IGS posiadają budowę segmentową. Są zlokalizowane pomiędzy 3’ końcem genu 26S rRNA a końcem 5’ genu 18S rRNA. Zawierają w swojej budowie: 3’-ETS (external spacer), NTS (non-transcribed region) oraz 5’-ETS. W tym mieszczą się także: miejsce inicjacji transkrypcji (TIS) oraz miejsce terminacji transkrypcji (TTS). IGS są inaczej nazywane również przekładkami międzygenowymi. Jądrowe rybosomalnej DNA (rDNA) kodujące podjednostki 5,8S, 18S i 26S rRNA należą do najważniejszych genów utrzymujących porządek w obrębie komórki, odgrywających kluczową rolę w metabolizmie komórkowym, regulacji procesów podstawowych, istotnych w przeżyciu i właściwym funkcjonowaniu komórki. W genomach roślin może znajdować się od kilkuset do kilku tysięcy bardzo jednorodnych kopii każdego z tych genów (Inácio i in. 2014; Matyam i in. 2014). Duża liczba kopii tych genów jest prawdopodobnie ważna celem zapewnienia zwiększonego zapotrzebowania na proces syntezy białek podczas rozwoju roślin, a również z uwagi na inne funkcje jak np. zapewnienie stabilizacji jądra komórkowego (Aguilera i in. 2015; Coutinho i in. 2016).

W niniejszej pracy poddano dogłębnym analizom sekwencje kodujące organellowego DNA oraz sekwencje NOR, które mają swoje zastosowanie w roli markerów genetycznych do badań w obrębie filogenetyki czy taksonomii roślin, a także opracowano kompletne dwa genomy chloroplastowe, pierwsze dla rodzaju *Erica*.

1.4. Rodzina *Ericaceae*

Współczesny system klasyfikacji rodziny *Ericaceae* na 8 podrodzin zawdzięcza swój aktualny stan postępowi badań filogenetycznych, których intensywny rozwój nastąpił w XXI wieku. Wcześniej bowiem wielu reprezentantów rozproszonych było w ujęciach systematycznych innych rodzin. *Ericaceae Juss.* – nazwa systematyczna rodziny roślin - wrzosowatych, należących do gromady roślin okrytozalążkowych (*Angiospermae*) w podkrólestwie roślin naczyniowych (*Tracheophyta*), królestwa roślin (*Plantae*). Zaklasyfikowane są do rzędu wrzosowców (*Ericales*) (Oliver, 1993).

W obrębie tej rodziny wyróżniamy podział na 8 podrodzin, w tym 3 podrodziny monotypowe (tj. z jednym rodzajem). Do podrodzin monotypowych zaliczamy podrodzinę *Enkianthoideae*, którą reprezentuje rodzaj enkiant (*Enkianthus*) dzielący się na 16 gatunków o zasięgu azjatyckim; podrodzinę *Cassiopoideae* z rodzajem kasjopeja (*Cassiope*) i 16 gatunkami o zasięgu himalajskim oraz okołobiegunowym półkuli północnej; a także podrodzinę *Harrimanelloideae* (rodzaj *Harrimanella*, 2 gatunki okołobiegunowe o nieregularnym rozmieszczeniu). Pozostałe podrodziny w obrębie rodziny wrzosowatych to: *Vaccinioidea, Styphelioideae, , Ericoideae, , Arbutoideae oraz Monotropoideae* (Oliver, 1993; Crayn i in. 2020).

Rodzina *Ericaceae* liczy sobie obecnie ponad 4 tysiące gatunków, zaklasyfikowanych w 145 rodzajów szeroko rozpowszechnionych w Himalajach, Nowej Gwinei i Południowej Afryce. W Turcji odnotowano występowanie 12 rodzajów z tej rodziny, z około trzydziestoma gatunkami. Występuje tam między innymi 5 przedstawicieli rodzaju *Erica*: gatunki endemiczne (*Erica bocquetii*) oraz szeroko rozpowszechnione w przybrzeżnych rejonach tureckich (*Erica arborea, Erica manipuliflora*) (Pirie i in. 2011; Köroğlu i in. 2019).

Większość rodzajów należących do rodziny wrzosowatych jest taksonomicznie udokumentowana bardzo dobrze, to jednak wciąż uboga pozostaje wiedza na temat relacji międzygatunkowych, wskazania taksonów monofiletycznych wraz z gatunkami ancestralnymi. Szczególnie wymagającą kwestią pozostaje ta, która bazowałaby na badaniach molekularnych. Niepoznana została wciąż kwestia pochodzenia wielu rodzajów, w tym m.in. rodzaju *Erica*. Podobna sytuacja dotyczy także innych rodzajów. Ponadto w precyzyjnych eksploracjach rodziny *Ericaceae* proponuje się szczegółowe ustalenia względem samego podziału rodziny na rodzaje, który prawdopodobnie wymaga kilku poprawek (Oliver, 1993; Kron i in. 2002).

Poza kwestiami spornymi w temacie samej systematyki, wciąż niezgłębionym i wielce fascynującym zagadnieniem pozostają tematy fizjologiczne oraz temat współegzystowania wrzosowatych z grzybami i tworzenie mikoryzy erikoidalnej (Leopold i in. 2021).

Do rodziny wrzosowatych zaliczamy rośliny w postaci krzewów, krzewinek a także drzew (tu np. siedmiometrową formę drzewiastą *Erica arborea* z rodzaju *Erica*). Zasięg geograficzny *Ericaceae* rozpościera się od Europy po Japonię, alpejskie piętra tropikalne Azji, Afrykę (Prowincja Przylądkowa – Cape Floristic Region), a także Amerykę Południową oraz Północną. Największe rodzaje w obrębie tej rodziny to rodzaj *Rhododendron* (różanecznik) z liczbą gatunków ok. 1000 oraz rodzaj *Erica* (wrzosiec) z liczbą gatunków ok. 865 (Oliver i in. 2003).

Wśród przedstawicieli Polski mamy jedynie kilkanaście rodzajów reprezentujących wrzosowatych w naszej rodzimej florze. Należą do nich rozmaite gatunki, m.in.: bażyna czarna (*Empetrum nigrum L.*), bażyna obupłciowa (*Empetrum hermaphroditum*), borówka bagienna (*Vaccinum uliginosum*), borówka brusznica (*Vaccinium vitis-idaea* L.), borówka czarna (*Vaccinium myrtillus* L.), chamedafne północna (*Chamaedaphne calyculata*), gruszyczka mniejsza (*Pyrola minor L.*), gruszyczka okrągłolistna (*Pyrola rotundifolia* L.), gruszyczka średnia (*Pyrola media*), gruszyczka zielonawa (*Pyrola chlorantha*), gruszycznik jednokwiatowy (*Moneses uniflora*), gruszynka jednostronna (*Orthilia secunda*), korzeniówka pospolita (*Monotropa hypopitys*), mącznica lekarska (*Arctostaphylos uva-ursi*),

modrzewnica pospolita (*Andromeda polifolia*), pomocnik baldaszkowy (*Chimaphila umbellata*), różanecznik żółty (*Rhododendron luteum*), bagno zwyczajne (*Ledum palustre*) oraz wrzosiec bagienny (*Erica tetralix*) (Mirek i in. 2002; Stevens 2008; Kaźmierczakowa i in. 2016).

Przedstawiciele wrzosowatych to często gatunki owocowe, a także liczna grupa roślin ozdobnych. Różaneczniki (*Rhododendron*) czy azalie (*Azalea*) są częstym imponującym udekorowaniem ogrodów i pozostają niezwykle urokliwe w okresie kwitnienia. Bardzo często jednak wrzosowate sprawiają poważne kłopoty w uprawie. Pomijając ich wymogi glebowe – zakwaszanie gleby, w której sadzi się wrzosowate, np. za pomocą kwaśnego torfu wysokiego, których praktyka przynosi rezultaty na pierwszych etapie wzrostu, jednak potem odznaczają się albo słabym kwitnięciem, albo jego absolutnym ustaniem (Kobayashi i in. 2021).

System korzeniowy wrzosowatych jest zbudowany z bardzo licznych, cienkich korzeni, pozbawionych włośników. Włośniki jako cienkościenne, wydłużone uwypuklenia komórek ryzodermy w obszarze strefy włośnikowej, penetrują przestrzenie glebowe, znacznie zwiększając powierzchnię wchłaniania wody z rozpuszczonymi jonami soli mineralnych z roztworu glebowego, u niektórych roślin zwiększają tę powierzchnię nawet kilkunastokrotnie. Brak włośników generuje ograniczenia, z którymi roślina w naturze radzi sobie współżyjąc z grzybami mikoryzowymi (Valgaz i in. 2022).

Symbioza ta według najnowszych doniesień naukowych zawiązała się ponad 140 milionów lat temu i umożliwia przedstawicielom tej rodziny zasiedlanie potencjalnie trudnych siedlisk – kwaśnych i ubogich w związki organiczne. Mikoryza erikoidalna to najmłodszy typ mikoryzy, o znacznie mniejszym zasięgu występowania niż dwa najpowszechniejsze typy – endotroficzna i ektotroficzna (Genre i in. 2020). Jest to typ mikoryzy spotykany tylko u gatunków z rodziny *Ericaceae*. Jednak niektóre źródła podają także występowanie tej mikoryzy w rodzinie *Empetraceae* oraz *Epacridaceae* (z wyjątkiem rodzajów *Arbutus* i *Artcostaphylos*) (Adeoyo, 2021).

Według jednych informacji literaturowych mikoryza erikoidalna jest formą mikoryzy ektotroficznej, według innych doniesień — odrębną. Nazywana jest także ektendomikoryzą ze względu na to, iż strzępki grzyba w tym typie mikoryzy jednocześnie tworzą na zewnątrz opilśń i wnikają do komórek roślinnych. Do najczęściej wymienianych grzybów, które tworzą ten typ mikoryzy, należą workowce z rodzaju *Hymenoscyphus*. Skróty nazwy tej mikoryzy, z którymi można spotkać się w literaturze to ERM lub ER. Mikoryza u wrzosowatych, jak i u innych roślin, które mikoryzują, jest korzystna dla obu stron (Leopold i in. 2021).

Doświadczenia laboratoryjne dowiodły, że rośliny mikoryzowane znacznie szybciej rosną niż te, które rozwijają się w podłożu bez grzybów mikoryzowych. Symbiozie roślin z grzybami towarzyszą często bakterie, pełniąc rolę wspomagającą. Są one nazywane helperami lub Plant Growth Promoting Rhizobacteria (PGPR) (Şirin i in. 2022).

Wspólne występowanie grzybów i bakterii ma swoje znaczenie w procesie odżywiania rośliny oraz w poprawie kompozycji gleby. Grzyby mikoryzowe chronią rośliny przed stresami abiotycznymi (susza, nieodpowiednia kwasowość podłoża, nieodpowiednie światło) i biotycznymi (patogeny glebowe; np. *Phytophthora cinnamomi*). Obecność grzybów zwiększa zaopatrzenie roślin w potas (jego dobór powoduje zwiększenie odporności na stres wywołany brakiem wody, czy suszę), fosfor, azot, mikroelementy. Mikoryza zwiększa wytrzymałość roślin na wahania temperatur, rośliny mikoryzowane są rzadziej uszkadzane przez nicienie, produkują większe ilości cytokinin powstających w wierzchołku wzrostu korzenia, zasolenie, obecność metali ciężkich, zmiany pH podłoża, niekorzystne cechy fizyczne podłoża (np. zbite gleby), a przede wszystkim nawiązując do wspomnianego wyżej braku włośników – w przypadku wrzosowatych nadrzędną rolą jest wspomaganie pobierania wody i soli mineralnych z roztworu glebowego, takie wzbogacenie systemu korzeniowego wrzosowatych łączy się także z ułatwionym ukorzenianiem sadzonek (Lanfranco i Fiorilli, 2022).

Wiedza na temat mikoryzy erikoidalnej została bardzo szybko wykorzystana m.in. w ogrodnictwie. Utrzymanie wrzosowatych w dobrej kondycji i doprowadzenie do zadowalającego kwitnienia stało się przystępne i proste. Grzyby mikoryzowe stanowią

bowiem produkt handlowy w postaci biopreparatów, szczepionek mikoryzowych zwanych inokulum. Mogą one zawierać fragmenty grzybni, zmikoryzowane korzenie roślin lub zarodniki grzybów (z czego zarodniki są najsilniej odporne na niesprzyjające warunki zewnętrzne, jednak są stosowane jedynie w przypadku mikoryzy zewnętrznej ze względu na sposób „podania" i kontakt preparatu z rośliną). Inokulum ma w swoim składzie także nośnik ze stałego podłoża, jak np. ziemia, glina, piasek czy torf. Szczepionki mikoryzowe dla wrzosowatych są dostępne w postaci preparatów typu „żywa grzybnia mikoryzowa", skład różni się w zależności od tego dla jakich konkretnie przedstawicieli zakupujemy szczepionkę. Dla rododendronów, azalii i wrzosów polecane jest stosowanie szczepionki, w składzie której znajdują się grzyby z rodzaju *Oidiodendron* i *Hymenoscyphus* (wyizolowane z dzikich rododendronów, wrzosów i mącznicy lekarskiej), w efekcie uzyskuje się lepszy wzrost i kondycję roślin, obfitszy efekt kwitnienia i podwyższoną tolerancję na stres glebowy (w tym pH) (Jarosz i in. 2021).

W przypadku borówki szczepionka komponowana jest osobno niż dla pozostałych wrzosowatych. Zawiera bowiem grzyby wyizolowane z wielu gatunków borówki *(Vaccinium myrtillus, Vaccinium Vitis-ideae, Vaccinium oxycoccus, Vaccinium uliginosum)*, także z rodzaju *Hymenoscyphus*. Użycie tej szczepionki powoduje bardziej zadowalający wzrost, wyższe plony i lepszą zdrowotność (Zydlik i in. 2019).

Intensywnie prowadzone są badania nad wrzosowatymi mikoryzującymi i ich wymogami środowiskowymi, a także pewnymi wytwarzającymi się zależnościami, identyfikacja grzybów w korzeniach wrzosowatych stała się możliwa szczególnie dzięki rozwojowi nowoczesnych technik molekularnych, w tym sekwencjonowania DNA (Vohník i Réblová, 2023).

Wrzosowate z mikoryzą erikoidalną z powodzeniem kolonizują gleby kwaśne zawierające duże stężenie metali ciężkich. Łączy się to ze zjawiskiem zwiększenia wchłanialności składników mineralnych w przebiegu „infekcji mikoryzowej". W przeprowadzanych badaniach nad *Calluna vulgaris, Vaccinium macrocarpon* i *Rhododendron ponticum* stwierdzono obecność zdecydowanie wyższego stężenia metali ciężkich u osobników z erikoidalną mikoryzą (Bradley i in. 1982).

Wywołało to dyskusję na temat ekologicznego znaczenia unikania metali toksycznych i zmian w dostarczaniu endofitycznych strzępków do mikoryzy, tak aby ułatwić wykluczenie metali z pędów. Jednocześnie taka cecha wytworzona na drodze erikoidalnej mikoryzy sprzyja wrzosowatym zasiedlanie miejsc skażonych wysokim stężeniem metali ciężkich. Mechanizm tej tolerancji nie został wciąż dokładnie poznany. Jedna z możliwości koncentruje się na interakcjach między rodzajem metalu i nierozpuszczalnymi metabolitami grzybowymi. Porównanie szczepów z zanieczyszczonych i czystych gleb oraz ich mobilizacji nierozpuszczalnych, nieorganicznych związków cynku wykazało, że szczepy z zanieczyszczonych gleb rzeczywiście wykazały mniejszą zdolność rozpuszczania związków cynku, przy czym te z czystych gleb charakteryzowały się wyższym potencjałem rozpuszczania. Za rozpuszczanie metali odpowiedzialna jest także indukcja kwasów organicznych (jabłczanu, cytrynianu). Wrzosy mikoryzowe z zanieczyszczonych i czystych gleb mogą więc w różnym stopniu oddziaływać ze związkami metalu. Można spekulować, iż odzwierciedla to konkretne strategie w celu utrzymania homeostazy podstawowych metali w różnych warunkach glebowych. Mechanizmy tolerancji mogą także obejmować wytwarzanie śluzu i zewnątrzkomórkowych substancji (nazywanych pigmentami) przez komórki grzyba, a także usuwanie jonów metalu z cytoplazmy czy wysoką tolerancję wewnątrzkomórkowego stężenia. Niezależnie od mechanizmu, wiele grzybów wykazuje tolerancję stężenia metali ciężkich, które normalnie powodują ciężkie objawy toksyczności u roślin wyższych (Cairney i Meharg, 2003; Casarrubia i in. 2020).

Stwierdzono także bardzo wysokie zdolności regeneracyjne komórek grzyba *Oidiodendron maius* w warunkach wysokiego stężenia cynku, kadmu czy kobaltu. Ponadto podkreśla się także w mechanizmach tejże tolerancji rolę wydzielania i aktywności poligalakturonazy (PG) – enzymu, który hydrolizuje składnik pektyny ściany komórki roślinnej. Prawdopodobnie odpowiedź reakcji PG na obecność metali ciężkich jest wstępnym faktorem adaptacyjnym podczas kolonizacji zanieczyczonych gleb przez *Oidiodendron maius*. Najczęściej identyfikowanym gatunkiem wyizolowanym z korzeni mikoryzujących gatunków z rodziny *Ericaceae* okazuje się *Pezoloma ericae*, należący do gromady workowców (*Ascomycota*) (Stribley i Read, 2006; Baba i in. 2021, Fadaei i in. 2021).

1.5. Charakterystyka rodzaju *Erica*

Rodzaj *Erica* (wrzośce) to największy obok rodzaju *Rhododendron*, rodzaj należący do rodziny *Ericaceae*. W skład rodzaju wrzośców wchodzi około 865 gatunków roślin kwiatowych. Wśród nich możemy wyróżnić krzewinki o wysokości od 20 do 140 cm oraz krzewy. Wyjątek stanowi forma drzewiasta wysokości ok. 7 metrów, wrzosiec drzewiasty *Erica arborea*. Wrzośce to rośliny wiecznie zielone, o drobnych szpilkowatych liściach i bardzo licznych kwiatach o barwach od białych, po różowe, fioletowe i wszelkie odcienie przejściowe. Znane jest działanie prozdrowotne suszonych liści i kwiatów wrzośców, w tym także *Erica arborea*, oraz etnofarmakologiczne stosowanie ziela wrzośców (Nazemiyeh i in. 2008; Amroun i in. 2021).

W medycynie ludowej i naturalnej niektórych regionów (m.in. w Turcji, we Włoszech czy w Hiszpanii) preparaty oparte na składnikach z ziela wrzośców stosowane są powszechnie jako środki moczopędne, ściągające i leczące infekcje dróg moczowych. Kwiaty wrzośca bagiennego (*Erica tetralix*) w medycynie ludowej stosowane były jako środek przeciwzapalny i moczopędny, w mieszankach ziołowych występują pod postacią „mieszanki przeciwnerwowej" jako komponent uspokajający (Castro i Carro, 2022). Wodny ekstrakt z części nadziemnych wrzośca drzewiastego wykazuje potencjał przeciwzapalny, natomiast jego ekstrakt wykazuje zastosowanie w leczeniu zaburzeń lipidowych w niewydolności sercowo-naczyniowej. *Erica arborea* charakteryzuje się także działaniem moczopędnym, przeciwbakteryjnym, przeciwwrzodowym, przeciwcukrzycowym. Napar wrzośca drzewiastego pity po posiłkach wykazuje działanie odchudzające (Suna i in. 2018; Sargin, 2021).

Erica cinerea jest rośliną bardzo bogatą w polifenole, wykazuje działanie przeciwzapalne, antyoksydacyjne / przeciwrodnikowe. Podobnie jak większość przedstawicieli z rodzaju – jest rośliną miododajną, miód z wrzośca popielatego nazywa się miodem portugalskim, wykazuje działania prozdrowotne zbliżone do miodu manuka, w tym przeciwdrobnoustrojowe i przeciwgrzybiczne (zwalcza biofilm *Candida*, wielooporne drożdże) (Fernandes i in. 2021; Feknous i Boumendjel, 2022).

Liście gatunku *Erica multiflora* znajdują z kolei zastosowanie głównie w Hiszpanii jako doskonały okład na rany. Napar w tego gatunku w ziołolecznictwie przyjmuje się jako środek tonizujący napięcia nerwowe, przy leczeniu kaszlu i dolegliwości serca. Z kolei maść sporządzoną z jego ziela stosuje się w kuracji zapalenia stawów i reumatyzmu. Medycyna ludowa uważa tę roślinę za skuteczną w leczeniu hiperlipidemii i pomagającą obniżyć poziom cholesterolu. W średniowieczu uważano, że roślina może rozpuszczać kamienie żółciowe. Gatunki z rodzaju *Erica* zawierają głównie flawonoidy, fenole, antocyjany, glikozydy fenylopropanoidowe, kumaryny i związki triterpenowe. (Anastassakis, 2022; Bekkai i in. 2022). Te aktywne substancje mogą być źródłem skutecznych preparatów farmaceutycznych. Rodzaj *Erica* został zbadany pod kątem ich biologicznych aktywności, takich jak m.in.: przeciwwrzodowe, przeciwbólowe, przeciwbakteryjne, antyseptyczne, przeciwlityczne, cytotoksyczne, etc. Istnieje także kilka raportów fitoterapeutycznych na temat ogromnego potencjału tych roślin jako źródła naturalnych związków przeciwutleniających (Tuzlaci i in. 2001; Tlas i in. 2021; Yaici i in. 2021).

Rodzaj *Erica*, według dotychczasowych danych, prawdopodobnie pochodzi z Europy i migrował na południe przez Afrykę, kolonizując góry Afryki Środkowej, wyspy Oceanu Indyjskiego, góry Smocze w Afryce Południowej i Państwo Przylądkowe na południowym krańcu Afryki Południowej, gdzie wydaje się, że po kolonizacji nastąpiła gwałtowna radiacja adaptacyjna (Murgrabi de Kuppler i in. 2015). Wskaźniki dywersyfikacji, a także obszary przodków i wzorce ewolucji charakteru *Erica* zostały wywnioskowane na podstawie filogenezy jądrowych i plastydowych markerów sekwencji DNA reprezentujących do około 60% znanych gatunków. Specjacja gatunków z rodzaju *Erica* występujących w kladzie Przylądka była szczególnie szybka, co może być związane ze złożonymi wzorami zmienności morfologicznej i regionalnym endemizmem wykazywanym przez gatunki i populacje *Erica* rozmieszczone w górzystym krajobrazie przylądkowym (Pirie i in. 2016). Podczas radiacji kladu *Erica* w południowej Afryce zanotowano liczne zmiany fenotypowe, m.in. morfologia i kolor kwiatów zmieniały się wielokrotnie (Le Maitre, 2017). Przesunięcia między różnymi zespołami zapylania, tj. zestaw cech, które promują określone wektory, jak np. różne gatunki owadów i ptaków, mogły skutkować ograniczeniami przepływu genów między populacjami, które ostatecznie napędzały specjację (Le Maitre i in. 2019). Kolor kwiatów

zmieniał się z różowego plezjomorficznego (przodków), najczęściej występującego w starszych, europejskich liniach (głównie zapylanych przez owady), do białego, zielonego, czerwonego i żółtego, zwłaszcza na obszarze Państwa Przylądkowego, gdzie te kolory są związane z większą różnorodnością zapylaczy, w tym ptaków nektarników (*Nectariniidae*) (Van der Niet i in. 2014; McCarren i in. 2021).

Dotychczas prowadzono badania polegające na opisaniu zmian genotypowych, mogących leżeć u podstaw fenotypowych zmian koloru kwiatów u gatunków z rodzaju *Erica* w regionie Państwa Przylądkowego. Zdentyfikowano i zsekwencjonowano geny szlaku biosyntezy antocyjanów i ich trans-działających czynników transkrypcyjnych u *Erica plukenetii* oraz opracowano test RT-qPCR do określania poziomu ich ekspresji (Le Maitre, 2017). Prowadzono także badania w celu wykazania jakoby zmiana koloru kwiatów była spowodowana: utratą funkcji mutacji enzymu lub wielu enzymów na szlaku, utratą czynników transkrypcyjnych, utratą ekspresji, utratą miejsc wiązania czynników transkrypcyjnych powyżej genu (lub genów) na szlaku poprzez mutacje (Le Maitre i in. 2019). Analiza licznych przykładów niezależnych zmian koloru w różnych gatunkach rodzaju *Erica*, prowadzi do stworzenia modelu łączącego genotyp z fenotypem (antocyjany, kolor kwiatów) i zidentyfikowania ogólnych procesów, które mogły odgrywać rolę w napędzaniu szybkiego różnicowania się rodzaju *Erica* w Cape Floristic Region (Le Maitre, 2017; Le Maitre 2019; Chautá i in. 2022; Erickson i Pessoa, 2022).

Zasięg występowania rodzaju od północnych obszarów Europy, po południowe krańce Afryki, rozciągający się na wschód ku środkowej Turcji i południowo-zachodnią część Półwyspu Arabskiego charakteryzuje się bardzo nierównomierną gęstością taksonów. W Europie, w zależności od źródeł występuje 20 - 21 gatunków. W tropikalnej Afryce wrzośce występują w niewielkich grupach od 2 do 22 gatunków, zazwyczaj związane są z regionami górskimi. Osobne centrum występowania stanowi wyspa Madagaskar, gdzie zanotowano 35 gatunków. CFR - Cape Floristic Region (Państwo Przylądkowe w Afryce Południowej) jest regionem w najbogatszym stopniu skupiającym reprezentantów rodzaju *Erica*, gdzie zliczono ich łącznie około 800 (Pirie i in. 2022). Jednocześnie jest to w ogóle największy rodzaj roślin w obrębie terytorium tego regionu. Wiele gatunków wrzośców bytuje tam na zasadzie koegzystencji, szczególnie gatunki drobnokwiatowe zapylane przez pszczoły miodne z Przylądka Dobrej Nadziei w Afryce.

Gatunki, które wspólnie zakwitają i mają wspólnego zapylacza, mogą być narażone na pewne straty w poziomie stałości kwiatów wraz ze stopniem czystości pyłku, z powodu zakłóceń reprodukcyjnych, czy też z uwagi na odkładanie się heterospecyficznego pyłku, zwłaszcza gdy gatunki te są blisko spokrewnione (Van der Niet i in. 2020).

Wśród wielu południowoafrykańskich przedstawicieli rodzaju prawdopodobnie wielokrotnie dochodziło do ewolucji zapylania przez ptaki, pszczoły miodne, owady o wydłużonym aparacie gębowym czy wiatr. Plan budowy kwiatów wrzośców tworzy idealny system do testowania wpływu zmian zapylaczy na modularność kształtu kwiatów. Temat mechanizmów oraz warunków umożliwiających współistnienie wspólnie kwitnących gatunków *Erica*, które mają tego samego zapylacza oraz badania nad modułowością i ewolucją kształtu kwiatu wrzośców wykorzystujące najnowsze techniki geometrycznej morfometrii 3D, stanowią jedne z, wiodących prymat, motywów podejmowanych aktualnie przez badaczy botaników z Południowej Afryki (Stevens, 2004; Reich, 2020).

Państwo Przylądkowe wyznaczające się wysokim poziomem endemizmu wrzośców, reprezentuje jednocześnie uosobienie bioróżnorodności tego rodzaju. Początki tej bioróżnorodności są od dawna tematem badań. Reprezentowany poziom bogactwa gatunkowego jest normalnie niespotykany poza tropikami, z około 10 000 gatunków roślin, z których ok. 7 000 jest endemicznych w regionie (Glen, 2002). To, które czynniki ukształtowały i napędzały to spektakularną radiację, było od dawna nurtującym topikiem biologii ewolucyjnej. Największym kladem w Cape Floristic Region jest tzw. „Cape Erica" reprezentujące endemiczną radiację. Ten największy klad Państwa Przylądkowego rozpoczął ją już około 10-15 milionów lat temu (Linder, 2003). Pozostałe, nieliczne gatunki wrzośców spotykane są także w niektórych regionach kontynentu Azji, na zachodzie Europy oraz w regionie Śródziemnomorskim (Stevens i in. 2004; Pirie, 2016).

Roślinność regionu CFR objęta jest w przeważającej większości Programem Ochrony Zagrożonych Gatunków w Południowoafrykańskim Narodowym Instytucie Bioróżnorodności (South African National Biodiversity Institute - SANBI). Obejmuje on zintegrowanie działań ochronnych *ex situ* i *in situ*. Strategie ochronne roślin Republiki Południowej Afryki zostały opracowane w odpowiedzi na globalną strategię ochrony

roślin (Hitchcock i in. 2020). Działania ochronne w ramach powyższego programu dotyczą także gatunków wrzośców, które w którymś momencie uznane zostały za wymarłe – tj. *Erica verticillata* (uznano, że wyginęła w roku 1950) czy też *Erica turgida* (wprowadzona na listę „wymarłe na wolności" w roku 1996). Ochrona ex situ w botanice, kolekcje ogrodowe i Milenijny Bank Nasion od tego czasu zezwalają na ochronę *in situ* w krytycznie zagrożonym typie roślinności (Hitchcock i in. 2020). Proces odtwarzania wrzosowisk stawiał wiele wyzwań. Początkowo próby były utrudnione przez ograniczoną dostępną wiedzę na temat odpowiednich siedlisk niszowych. Pionierskie prace przeprowadzone w Rezerwacie Przyrody Rondevlei pozwolił na zidentyfikowanie siedliska o stosownych cechach, które zastosowano później na jedynych naturalnych obszarach pozostających w historii tych gatunków lub w jej pobliżu zakresu jego dystrybucji. Pomyślne odtworzenie tego typu gatunków zależy od jego zdolności do rekrutacji po pożarze, co jest podstawowym procesem ekologicznym w fynbosie (nazwa południowoafrykańskiej formacji roślinnej). Kluczowym wnioskiem w przytoczonych przykładach jest potrzeba uwzględnienia zrównoważonego zarządzania całym ekosystemem podczas odtwarzania gatunków oraz nie ograniczanie go do pojedynczych gatunków. Sukces w odtworzeniu gatunku zależy od zdrowego stanowiska danego typu roślinności, wraz z zapylaczami i inną kluczową fauną oraz pozostałe naturalne procesy ekosystemowe, opiera się nierzadko na wielu dekadach badań, prób i monitorowania pod kątem przeżycia oraz rekrutacji roślin (Hitchcock i in. 2020; Hitchcock i in. 2020; Pirie i in. 2022). W roku 2020 powołano konsorcjum konserwatorskie - The Erica Conservation Consortium – scalające i jednoczące ogrody botaniczne, instytucje ogrodnicze oraz botaniczne, pracujące nad celem koordynacji i nadania priorytetu ochronie *ex situ* wrzosów zagrożonych na Przylądku Afryki Południowej oraz w całym zasięgu geograficznym rodzaju (Nelson i Pirie, 2022).

Jedynym przedstawicielem wrzośców w rodzimej florze Polski jest wrzosiec bagienny (*Erica tetralix*), który objęty jest ścisłą ochroną gatunkową od 1983 roku. W naszym kraju jest gatunkiem rzadkim, możemy spotkać jego stanowiska przeważnie w północno-zachodniej części Polski, szczególnie wzdłuż brzegu Morza Bałtyckiego. Wrzosiec bagienny występuje głównie na terenach wilgotnych wrzosowisk i torfowisk, z uwagi na to, iż rozwija się najszybciej na podłożu torfowym (organicznym). Ponadto można go spotkać w borach bagiennych oraz trzęślicowych. Z racji na sowite zniszczenia,

bądź ewentualne transformacje w obrębie siedlisk *Erica tetralix* wykazuje charakter ustępujący i jest zaklasyfikowany jako gatunek zagrożony. Umieszczony na Czerwonej liście roślin i grzybów Polski (2006) w grupie gatunków narażonych na wyginięcie poza głównym obszarem występowania (kategoria zagrożenia: V). 10 lat później uplasowano go w grupie gatunków wysokiego ryzyka narażonych na wyginięcie z powodu progresywnego spadku populacji, deficytów siedliskowych bądź przesadnej eksploatacji (kategoria zagrożenia: VU) (Mirek i in. 2006; Kaźmierczakowa i in. 2016).



Fotografia 1. *Erica tetralix*, fot. Krzysztof Ziarnek[1]

[1] łąka *Molinia caerulea* (Trzęślica modra), Grzybowo, 3 km od Kołobrzegu, północno-zachodnia Polska; fotografia z dn. 18.07.2014 r., udostępniona na licencji CC BY-SA 4.0 na commons.wikipedia.org

2. Cele pracy

Ze względu na aktualny stan wiedzy dotyczący filogenezy rodzaju *Erica*, który nie wyjaśnia pochodzenia i historii wrzośców oraz brak zsekwencjonowanych genomów chloroplastowych w obrębie tego rodzaju, postawiono następujące cele pracy:

1. poznanie wzorców zmienności genetycznej plastomów w rodzaju *Erica,*

2. poznanie ewolucji jądrowych klastrów rRNA (18S-ITS1-5.8S-ITS2-26S-IGS) w rodzaju *Erica,*

3. rekonstrukcja filogenezy rodzaju *Erica* ze szczególnym uwzględnieniem biogeografii rodzaju w oparciu o kompletne genomy plastydowe, grupy genów chloroplastowych oraz sekwencje jądrowych genów rRNA,

4. opracowanie barkodów DNA dla rodzaju *Erica* opartych o regiony jądrowe i plastydowe,

5. uzyskanie i opracowanie zasobów genomowych zagrożonych i endemicznych gatunków z rodzaju *Erica* mających praktyczne zastosowania w genetyce konserwatorskiej.

Hipotezy badawcze

1. Dla rodzaju *Erica* istnieje grupa barkodów pozwalająca na ich molekularną identyfikację.

2. Europejskie gatunki wrzośców stanowią grupę monofiletyczną, pochodzącą od jednego szeroko rozpowszechnionego przodka o pochodzeniu afrykańskim.

3. Bikontynentalna *Erica arborea* jest gatunkiem ancestralnym dla europejskich i afrykańskich wrzośców.

4. Wysoce wyspecjalizowane endemiczne gatunki wrzośców z Południowej Afryki charakteryzują się większą różnorodnością genetyczną ze względu na izolację, zmiany klimatu, gleb i topografię terenu sprzyjającą specjacji allopatrycznej.

5. Wzorce zróżnicowania sekwencji powtórzonych w obrębie regionu IGS odzwierciedlają procesy ewolucyjne w rodzaju *Erica.*

Realizację wyznaczonych celów partycjonowano na poszczególne zadania badawcze:

1. Sekwencjonowanie genomu chloroplastowego *Erica versicolor* i analiza porównawcza pomiędzy tym cpDNA a innymi dostępnymi w bazie NCBI pochodzącymi z tej samej rodziny – *Ericaceae.*

2. Sekwencjonowanie genomu chloroplastowego *Erica arborea* i analiza porównawcza pomiędzy tym cpDNA a innymi dostępnymi w bazie NCBI pochodzącymi z tej samej rodziny – *Ericaceae*; oraz analiza porównawcza pomiędzy obydwoma pełnymi genomami chloroplastowymi *E. arborea* i *E. versicolor.*

3. Otrzymanie pełnych sekwencji wszystkich genów chloroplastowych przedstawicieli rodzaju *Erica* poprzez zmapowanie bibliotek na otrzymany kompletny genom *Erica versicolor* i ich analiza porównawcza oraz analiza zmienności.

4. Oszacowanie informatywności filogenetycznej oraz analiza filogenomiczna na podstawie genów genomu chloroplastowego w rodzaju *Erica.*

5. Charakterystyka jądrowych regionów NOR rodzaju *Erica.*

6. Analiza filogenomiczna oparta o sekwencje jądrowych regionów NOR w rodzaju *Erica.*

7. Analiza porównawcza regionów odwróconych powtórzeń (IR) kompletnych sekwencji genomów chloroplastowych z rodzaju *Erica* z kompletnymi genomami chloroplastowymi innych przedstawicieli z rodziny *Ericaceae*.

8. Obliczenie i zwizualizowanie wzorców zmienności genomowej wśród genomów chloroplastowych z rodzaju *Erica* mająca przedstawić ich wzajemny stopień podobieństwa sekwencji.

3. Materiał i metody

3.1. Materiał użyty do badań

Materiał wykorzystany w niniejszych badaniach zawierający 8 gatunków z rodzaju *Erica*, reprezentujących następujące 2 podrodzaje: *Syringodea* i *Chlamydanthe*. Wszystkie badane gatunki są zaprezentowane w Tabeli 1. Klasyfikacja rodzaju przytaczana w tej pracy jest zastosowana zgodnie z podziałem zaproponowanym przez Benthama (1839), zmodyfikowanym przez Hansena (1950) i uzupełnionym przez Olivera (2000). Nomenklatura gatunków zaadoptowana z prac Olivera, wieloletniego eksperta taksonomicznego badanego rodzaju (Oliver, 1993; Oliver i in. 2003).

Do badań przeprowadzonych w ramach niniejszej dysertacji wybrano 8 gatunków z rodzaju *Erica*, o różnej charakterystyce i odmiennym rozmieszczeniu geograficznym. Dla dwóch z nich otrzymano i opracowano kompletne genomy chloroplastowe – dla *Erica arborea* oraz *Erica versicolor*.

Erica arborea, gatunek znany często pod popularną nazwą użytkową odmiany ‘Estrella gold’. Zasiedla zaroślowe formacje roślinne specyficzne dla regionu śródziemnomorskiego. Dystrybucyjnie jest bikontynentalnym, najbardziej rozproszonym spośród gatunków poddanych niniejszym badaniom. Zasięg jego występowania obejmuje Europę Południową, cały basen Morza Śródziemnego, Wyspy Kanaryjskie, Maderę; północno-zachodnią Afrykę, a także jej regiony południowe idące pasmem od Sudanu do terytorium Ugandy czy Tanzanii. Drewno wrzośca drzewiastego jest bardzo solidne i mocne. Z uwagi na swoje właściwości na przestrzeni wieków zyskało wiele praktycznych zastosowań. Układ gęstych i delikatnych gałęzi *Erica arborea* miał w tradycji zastosowanie przy wytwarzaniu mioteł. Ponadto drewno wrzośca drzewiastego odnalazło przeznaczenie, m.in. w konstrukcjach palisadowych, konstruowaniu rzeźb, biżuterii, mebli, ornamentów, uchwytów do broni i sztućców. Dzięki swojej żaroodporności i niewpływaniu na aromaty ziół od wieków używane jest do wyrobu bongo, fajek i lufek.

Erica arborea jest rośliną ciepłolubną, niewykazującą wysokiego stopnia mrozoodporności. Charakteryzują go igiełkowate złocistożółte (czasem limonkowe) liście. Kwiaty, w zależności od odmiany, przybierają kolor od najczęściej występującego białego, przez jasny róż aż do odcieni fioletu (Désamoré i in. 2010; Yaici i in. 2021).



Fotografia 2. *Erica arborea,* fot. Javier Martin[2]

[2] Pasmo górskie Sierra Madrona w Hiszpanii; fotografia z dn. 19.03.2009 r., udostępniona na zasadach domeny publicznej na commons.wikipedia.org

Całkowity genom plastydowy otrzymano także dla południowoafrykańskiego gatunku - *Erica versicolor*. Ten wyprostowany, rozłożysty krzew osiąga on nawet 3 metry wysokości, natomiast rozgałęzia się do ok. metra szerokości. Wyróżnia się okazałymi kwiatami rurkowatymi o długości około 3 cm i skrajnym zabarwieniem (stąd też nazwa łacińska), najczęściej czerwono-żółtymi bądź czerwono-zielonymi. Kwiaty lekko zagięte w pokroju rosną pojedynczo lub w grupie po 3 na końcu krótkich, giętkich, czasem prostych, bocznych liściastych gałęzi. Posiada status ochrony najmniejszej troski, ze względu na szerokie rozpowszechnienie występowania. Częstokroć zachodzi problem identyfikacyjny tego gatunku z *Erica discolor* i *Erica unicolor* oraz ich odmian (Booyens, 2012; Hitchcock, 2016).



Fotografia 3. *Erica versicolor, fot.* Stan Shebs[3]

[3] Ogród botaniczny w San Francisco, Kalifornia, Stany Zjednoczone; fotografia z dn. 01.06.2005 r., udostępniona na licencji CC BY-SA 3.0 na commons.wikipedia.org

Dla pozostałych gatunków otrzymano geny chloroplastowe oraz (podobnie jak dla dwóch wspomnianych wyżej gatunków) kompletne sekwencje NOR.

Erica cinerea, zwana wrzoścem dzwonkowatym lub popielatym, występuje w rejonach Europy Zachodniej, najbogatszym dystrybucyjnie miejscem dla niej jest głównie Wielka Brytania i Francja. Wprowadzona na Czerwoną Listę Zagrożonych Gatunków IUCN w kategorii gatunku najmniejszej troski. Jest to wrzosiec rosnący w formie niewysokiego krzewu, osiągający od 20 do 60 cm wysokości. Posiada kwiaty dzwonkowate, najczęściej fioletowo zabarwione, rzadziej białe, nazywane „popiołowymi". Gatunek ten jest bogato wykorzystywany do produkcji miodu wrzosowego, często uprawiany w ogrodach miejskich oraz prywatnych jako roślina ozdobna (Vera i in. 1995; Khela, 2013; Fernandes i in. 2021).



Fotografia 4. *Erica cinerea,* fot. Balles2601[4]

[4] Ourense, Galicja (42° 22′ 31″ N, 7° 36′ 50.7″ W), specjalny obszar ochrony w Hiszpanii; fotografia z dn. 04.08.2015 r., udostępniona na licencji CC BY-SA 4.0 na commons.wikipedia.org

Erica vagans to także gatunek zidentyfikowany jedynie na terenie Europy, gdzie głównie lokuje się na obszarach Irlandii, Francji i Hiszpanii. Nazywany potocznie „wrzoścem rozpierzchłym" – ze względu na łacińską nazwę oraz szerokie rozpowszechnienie okazu; oraz wrzoścem kornwalijskim - z uwagi na krainę geograficzno-historyczną intensywnego zasiedlania, uważany jest za kwiatowy herb (symbol) regionu Kornwalii. Nie przekracza metra wysokości, rozgałęzia się szeroko, na kilkadziesiąt centymetrów. Kwiaty kuliste o barwie wyjątkowo jednolitej - głównie różowej, ewentualnie liliowej, skupione w zjawiskowych gronach, osiągających długość do 15 cm. Prawdopodobnie przybył do Irlandii w czasach polodowcowych i od tego czasu stał się genetycznie odrębny. W dostępnych wynikach badań, prowadzonych głównie na terenie północnej Hiszpanii, gatunek ten charakteryzuje się wysoce skuteczną strategią regeneracji wegetatywnej, co przekłada się m.in. na zdolność odrastania po pożarach (Vera i in. 1995; Durán Gómez i Rodriguez, 2022).



Fotografia 5. *Erica vagans,* fot. Patrice78500[5]

[5] Wyspa Belle-Île-en-Mer, wybrzeże Bretanii, Francja; fotografia z dn. 07.07.2013 r., udostępniona na licencji CC BY-SA 3.0 na commons.wikipedia.org

Erica globiceps, często nazywany wrzoścem kulistym, związany jest jedynie z fynbosem – formacją roślin będącą odpowiednikiem makii, ale oznaczającą zarośla wyłącznie regionu południowej Afryki, gdzie siedliskowo preferuje wyższe zbocza. Jest endemitem południowoafrykańskim o statusie najmniejszej troski, zważywszy na stabilność jego populacji. Gatunek ten rośnie w formie niskich, zwartych krzewów, o wysokości maksymalnie do 30 cm. Charakteryzuje się zredukowaną liczbą pręcików – z 8, u większości kwiatowatych wrzośców, do 4 – wydatnych, brunatnych i wznoszących się ponad korony kwiatów. Kwiaty niewielkie, drobne, rurkowate, o barwie bladoróżowej, przechodzącej do lilowej. Dotychczas zidentyfikowano i opisano trzy jego podgatunki (Oliver, 1993; Brickell, 2003; Raimondo i in. 2009).



Fotografia 6. *Erica globiceps,* fot. David Gwynne-Evans (Cameron McMaster) [6]

[6] Region Overberg, Republika Południowej Afryki na wschód od Kapsztadu; fotografia z dnia 10.08.2008 r., udostępniona na zasadach domeny publicznej na casabio.org

Kolejnym gatunkiem występującym jedynie w tej formacji roślinnej jest *Erica plukenetii,* swoją nazwę zawdzięcza lekarzowi i botanikowi – Leonardowi Plukenetowi, nadwornemu ogrodnikowi Krolowej Anglii Marii II, nadzorcy Ogrodu Hampton Court. Zilustrował on i opisał ponad 2700 rzadkich gatunków roślin egzotycznych, opublikowanych w jego dziele „*Phytographiae*" (wydanym w Londynie w roku 1696). Jako pierwszy nadmienił informacje na temat gatunku *Erica* z Przylądka Południowej Afryki (Gunn i Codd, 1981).

Erica plukenetii występuje na zboczach i równinach górskich wśród skał i na piaszczystych glebach, a także na nizinach w pobliżu morza. Jest stosunkowo zmienny pod względem morfologicznym, w obrębie którego wyróżniono dotychczas 5 podgatunków, osiąga od 900 cm do 2 m wysokości. Tworzy gęste skupiska wiszących kwiatów rurkowatych, o długości od 7mm do 18mm, których kolor zmienia się od białego, różowego, czerwonego, zielonego do żółtego. Kwiaty podgatunków wielobarwnych o okazałych rozmiarach zapylane są przez ptaki, natomiast podgatunki posiadające kwiaty mniejsze o barwie białej zapylane są przez ćmy. Sposób, w jaki zapylacze wiszą na jej kwiatach podczas spijania z nich nektaru, wiąże się z jego potoczną nazwą „mały wieszak". Zważywszy na jego wytrzymałość i ogromną atrakcyjność estetyczną jest doskonałym gatunkiem ogrodowym, obficie przyciągającym ptaki do ogrodu. Jest gatunkiem pospolitym, o szerokim spektrum dystrybucji, występuje na wielu obszarach będących chronionymi rezerwatami przyrody i nie jest uważany za zagrożony (Brown i Duncan, 2006; Van der Niet i in. 2014).



Fotografia 7. *Erica plukenetii,* fot. Joachim Huber[7]

[7] Zachodni Przylądek formacji fynbosu, Afryka Południowa; fotografia z dn. 03.06.2007 r. udostępniona na licencji CC BY-SA 2.0 na commons.wikipedia.org

Z tym samym regionem geograficznym Afryki Południowej ściśle związana jest *Erica imbricata*, rosnąca na suchych obszarach przybrzeżnych oraz na wzgórzach i zboczach gór. Nazwa łacińska pochodzi od słowa „imbricatus" – nakładający się, odnosi się do dominującego kielicha zachodzącego na koronę. Znana jest pod ludową, regionalną nazwą, ker-ker, ze względu na dźwięk wydawany przy lekkim dotknięciu/otarciu tej rośliny. Ten wyprostowany krzew, osiąga do 80 cm wysokości. Posiada kwiaty owalne, głównie białe, występują jednak także w kolorach od różu, przez czerwień aż po brąz, zapylane przez owady. Kielich kwiatów długi tak jak korona, otaczający w taki sposób, że widoczne pozostają przylistki i płatki. Wyróżniające identyfikacyjnie są w tym gatunku ciemnobrązowe, wydłużone i wystające pylniki. Krzew ten często spotykany jest w postaci bardzo bogato rozgałęzionej. Gatunek o statusie „najmniejszej troski" (LC – least concern) wg Czerwonej Listy Roślin Południowej Afryki (Goldblatt i in. 2000; Foden i Potter, 2005; Raimondo i in. 2009).



Fotografia 8. *Erica imbricata,* fot. Andrew Massyn[8]

[8] Góry Stołowe, Kapsztad; fotografia z dn. 07.09.2001 r. udostępniona na licencji CC BY-SA 3.0 na commons.wikipedia.org

Równie imponująco, co *Erica imbricata*, rozgałęzia się gatunek wrzośca wielokwiatowego - *Erica multiflora*, który jest stosunkowo wysoki, osiąga bowiem nawet do 250 cm. Swoją nazwę zawdzięcza dużym kwiatostanom z wieloma kwiatami, zabarwionymi od bieli przez róż do wszystkich odcieni pośrednich prowadzących do fioletu. Z drewna tego wrzośca od wieków wyrabia się narzędzia, meble i fajki, natomiast z korzeni robione są instrumenty muzyczne. Według pradawnych tradycji *Erica multiflora* stosowana była jako roślina rytualna o silnych mocach ochronnych, przynosząca szczęście. To także gatunek wrzośca bikontynentalnego. W Afryce zasiedla głównie jej północne regiony, w Europie rozpowszechniona jest nieco szerzej: większość wysp Balearów, wschodnia część Półwyspu Iberyjskiego, przez basen Morza Śródziemnego, gdzie jest gatunkiem rodzimym zasiedlającym suche i piaszczyste gleby, do granic wschodniej Serbii (Llorens i in. 2003; Sadki i in. 2010).



Fotografia 9. *Erica multiflora,* fot. Thomas Schoch[9]

[9] Majorka, Archipelag Balearów, wyspa hiszpańska na Morzu Śródziemnym; fotografia z dn. 08.10.2007 r. udostępniona na licencji CC BY-SA 3.0 na commons.wikipedia.org

Gatunki z rodziny *Ericaceae*, których kompletne genomy zostały opublikowane i opracowane, a co najważniejsze są dostępne w bazie NCBI, inkorporowano do poszczególnych badań tej pracy jako grupy zewnętrzne oraz sekwencje referencyjne. Użyto ich, ponieważ to dotychczas jedne z niewielu dostępnych sekwencji genomów chloroplastowych gatunków fotosyntetyzujących z rodziny wrzosowatych, które zostały umieszczone w bazie danych i z racji przynależności do jednej rodziny, są to gatunki najbliżej spokrewnione z rodzajem, będącym przedmiotem pracy doktorskiej.

Należą do nich: Chamaedaphne północna (*Chamaedaphne calyculata*; numer akcesyjny GenBank KJ463365.1), żurawina wielkoowocowa (*Vaccinium macrocarpon*; numer akcesyjny GenBank NC_019616), borówka brusznica (*Vaccinium vitis-idaea*; numer akcesyjny GenBank LC521968), borówka bagienna (*Vaccinium uliginosum*; numer akcesyjny GenBank LC521968), borówka morska (*Vaccinium bracteatum*; numer akcesyjny GenBank LC521967), żurawina japońska (*Vaccinium oldhamii*; numer akcesyjny GenBank NO042713), 3 gatunki różaneczników (*Rhododendron pulchrum*; numer akcesyjny GenBank MN182619, *Rhododendron delavayi*; numer akcesyjny GenBank MN711645, *Rhododendron platypodum*; numer akcesyjny GenBank NC053746) drzewo truskawkowe (*Arbutus unedo*; numer akcesyjny GenBank JQ067650.2) oraz gruszyczka okrągłolistna (*Pyrola rotundifolia*; numer akcesyjny GenBank KU833271.1).

Przy czym, do prymarnego otrzymania kompletnego genomu chloroplastowego *Erica versicolor*, referencję stanowił genom *Vaccinium macrocarpon* (numer akcesyjny GenBank NC_019616).

Materiał roślinny został zebrany podczas badań terenowych i ekspedycji naukowych, m.in. związanych z realizacją grantu KBN N303 807840. Materiały zielnikowe prób wykorzystanych w niniejszych badaniach są zdeponowane w kolekcjach herbarium POZ (Zielnik Roślin Naczyniowych Zakładu Taksonomii Roślin, Uniwersytet im. Adama Mickiewicza w Poznaniu). Pochodzenie analizowanych gatunków oraz nazwisko kolektora danej próby są przedstawione w Tabeli 1, natomiast rozmieszczenie geograficzne badanych populacji zawarto na mapie na Rycinie 1.

Tabela 1. Lista gatunków wykorzystanych do analiz z ich przynależnością do podrodzaju i sekcji w obrębie rodzaju *Erica* wraz z miejscem zebrania materiału i kolekcjonerem próby

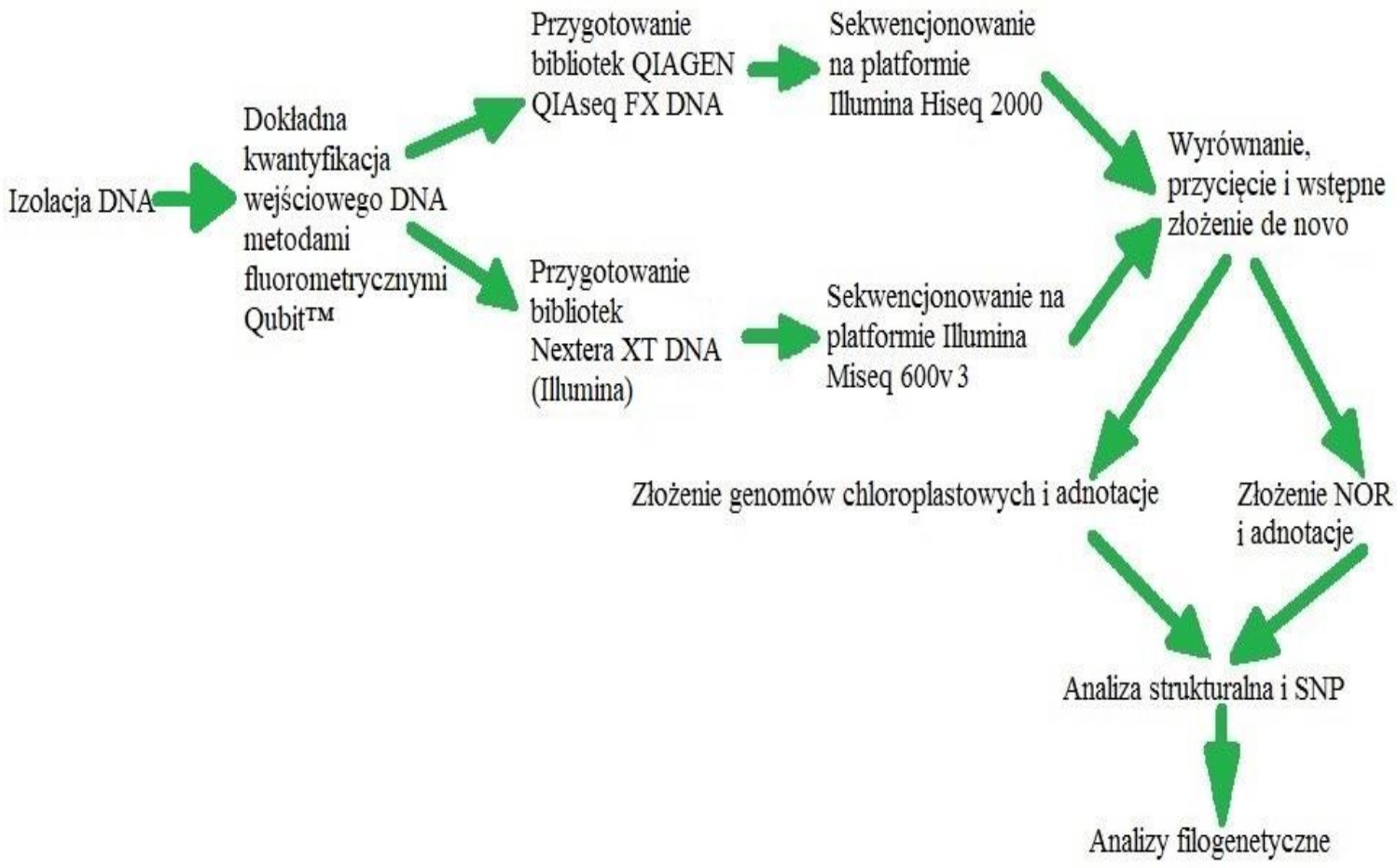
Rodzaj	**Podrodzaj**	**Sekcja**	**Nazwa gatunku**	**Lokalizacja i kolektor**
Erica	*Syringodea*	*Gigandra*	*Erica plukenetii*	RSA, Groot Gabel Kraal[1]
		Evanthe	*Erica versicolor*	Bergfontein[1]
		Gypsocallis	*Erica vagans*	Spain, Navarra[2]
			Erica multiflora	Spain, Velencia[2]
			Erica globiceps	RSA, Groot Hagelkraal[2]
		Arsace	*Erica arborea*	Spain, Cap de Creus[2]
		Brachycallis	*Erica cinerea*	Spain, Navarra[2]
	Chlamydanthe	*Geissostegia*	*Erica imbricata*	RSA, Elandskloof, Wolseley[2]

Wyjaśnienie: kolektorzy prób: [1] Oliver E.G.H., [2]Szkudlarz P.



Rycina 1. Rozmieszczenie geograficzne badanych populacji (kolorowe kropki)

Europa (Hiszpania): kolor różowy – Navarra (*Erica vagans* i *Erica cinerea*), kolor biały – Walencja (*Erica arborea* i *Erica multiflora*). Afryka południowa (RPA): kolor czerwony – Bergfontein (*Erica versicolor*), kolor zielony - Groot Hagelkraal (*Erica globiceps*), kolor żółty - Elandskloof, Wolseley (*Erica imbricata*), kolor niebieski - Groot Gabel Kraal (*Erica plukenetii*)



Rycina 2. Schemat przedstawiający poszczególne etapy prowadzonych prac

3.2.1. Izolacja DNA

Wysokiej jakości DNA jest niezbędne do uzyskania wiarygodnych wyników sekwencjonowania. Najważniejszym warunkiem wstępnym każdego eksperymentu analizy sekwencji DNA jest spójny, wysokiej jakości materiał DNA w każdej próbce eksperymentalnej. W związku z tym procedury pobierania próbek i izolacji DNA są kluczowe dla powodzenia badań. Resztkowe ślady białek, soli lub innych zanieczyszczeń spowodują degradację DNA lub obniżą wydajność aktywności enzymatycznej niezbędnej do optymalnego przygotowania bibliotek.

Liście starto w Mini-Beadbeater-1 tissue disruptor. Izolację materiału DNA przeprowadzono za pomocą sprawdzonego zestawu firmy Zymo Research (isolation kit ZR Plant/Seed DNA MiniPrep™ Kit). Do procesu izolacji użyto materiał wysuszonych liści wrzośców. Odmierzono ok. 100 do max. 150 mg tkanki roślinnej każdego z badanych gatunków.

1. Odważony materiał roślinny przełożono do probówek dedykowanych w zestawie przeznaczonych na etap lizy komórek, wyposażonych w kulki rozdrabniające tkanki roślinne. Dodano 750 µl roztworu do lizy.

2. Zabezpieczone, ściśle zakręcone celem zapobiegnięcia przeciekom, probówki poddano mechanicznej homogenizacji w maksymalnej prędkości przez 10 minut.

3. Następnie odwirowano probówki w mikrowirówce przy obrotach 10 000 rpm przez jedną minutę.

4. Otrzymany w ten sposób supernatant przeniesiono do filtrujących probówek w objętości 400 µl i wirowano przez minutę przy obrotach 7 000 rpm.

5. Do filtratu powstałego po wirowaniu w powyższych probówkach dodano 1 200 µl buforu wiążącego DNA i dokładnie wymieszano.

6. Z tak powstałej mikstury pobrano 800 µl do nowych probówek kolumienkowych i poddano je wirowaniu przez minutę, przy obrotach 10 000 rpm. Powstały przesącz odrzucono. Do tych samych probówek dodano kolejne 800 µl mikstury i ponownie wirowano je przez minutę, przy obrotach 10 000 rpm.

7. Kolumienki przełożono do nowych probówek i dodano do nich 200 µl buforu wstępnie płuczącego DNA. Wirowano je przez minutę, przy obrotach 10 000 rpm. Następnie dodano 500 µl buforu płuczącego DNA i wirowano przy obrotach 10 000 rpm przez minutę.

8. Kolumienki przeniesiono do czystych probówek i dodano 100 µl buforu do elucji centralnie na powierzchnię membrany kolumienki, celem zapewnienia uzyskania maksymalnej wydajności odzysku DNA. Celem elucji DNA wirowano probówki przez 30 sekund przy obrotach 10 000 rpm.

9. Odzyskane DNA przeniesiono do nowych kolumienek w nowych probówkach, które uprzednio przygotowano przez trzyminutowe wirowanie przy obrotach 8 000 rpm. Ostatecznie filtrowanie DNA przeprowadzono przy obrotach 8 000 rpm przez minutę. Tak otrzymany izolat DNA był gotowy do dalszych analiz. Izolaty przechowywano zamrożone w temperaturze -20^{o}C.

3.2.2. Kontrola jakości i ilości wyizolowanego DNA

Otrzymane DNA sprawdzono pod kątem ilościowym oraz jakościowym za pomocą spektofotometru firmy Qubit ™ zgodnie z zaleceniami proporcji producenta i jedynie z użyciem dedykowanego sprzętu. Barwnik Quant-iT™ wymieszano wraz buforem Quant-iT™ do otrzymania mieszaniny reakcyjnej Quant-iT™ Working Solution. Objętość przeliczono na konkretną liczbę prób, tak aby na 200µl mieszaniny reakcyjnej 1µl stanowił reagent, a resztę bufor. Mieszaninę rozdysponowano na 2 próbki ze standardami pochodzącymi z zestawu i służącymi do skalibrowania pomiarów spektofotometru oraz próbek z izolatami. Kompletna objętość pojedynczej próbki wynosiła 200 µl. Do próbek kontroli DNA dodano 4 µl izolatów i 196 µl mieszaniny reakcyjnej. Do próbek standardów po 10 µl każdego z nich oraz 190 µl mieszaniny reakcyjnej. Tak otrzymane probówki worteksowano przez kilka sekund, następnie próby inkubowano przez 2 minuty w temperaturze pokojowej, po czym przystąpiono do pomiarów. Dla każdego z izolatów przeprowadzono 3 niezależne pomiary, dla większej wiarygodności wyników. Wyniki ostateczne stanowiły średnią wartość wyciągniętą z pomiarów kontrolnych. Preparaty uznano za dobre pod względem jakościowym, gdy stosunek absorbancji wyizolowanego DNA wyniósł od 1,0 do 2,0. Dla przykładu stężenie DNA dla *Erica versicolor* wyniosło 13,4 ng/µl.

Jakość izolatów DNA została także dodatkowo oceniona wizualnie za pomocą elektroforezy przeprowadzonej w 1% żelu agarozowym.

3.2.3. Sekwencjonowanie chloroplastowego DNA z zastosowaniem metody nowej generacji NGS

Sekwencjonowanie nowej generacji (NGS) jest siłą napędową wielu nowych i ekscytujących zastosowań, w wielu typach badań, tj. badania medyczne nad komórkami macierzystymi czy nowotworowymi, metagenomika, genetyka populacyjna i konserwatorska. Podczas gdy technologia NGS stale się rozwija, przygotowanie bibliotek pozostaje jednym z największych „wąskich gardeł" w przepływie pracy NGS i obejmuje kilka czasochłonnych kroków, które mogą spowodować znaczną utratę próbki i odpowiadają za potencjalne ryzyko wprowadzenia błędów.

Technologia QIAseq FX firmy QIAGEN została zastosowana do otrzymania genomu *Erica versicolor*, na platformie HiSeq 2000. Wielokrotne próby podejść do otrzymania tego genomu, nielinearność w pokryciu danych, sprawiły, że należało zastosować inny zestaw, głęboką genomikę w postaci potężnego, wysokoprzepustowego systemu sekwencjonowania. Technologia QIAGEN wykorzystuje usprawnioną enzymatyczną fragmentację DNA, zoptymalizowany protokół, który nie wymaga czyszczenia próbki pomiędzy fragmentacją a podwiązaniem adaptera, oszczędzając tym samym czas i minimalizując możliwość wystąpienia błędów. Zoptymalizowany skład enzymów i buforów zapewniają wysoką wydajność sekwencjonowania bibliotek. Usprawniony protokół konstrukcji bibliotek umożliwia również ułatwioną automatyzację przygotowywania bibliotek na różnie obsługiwane platformy.

Zestaw bibliotek QIAGEN QIAseq FX DNA zapewnia szybką, całkowicie enzymatyczną procedurę. Fragmentacja DNA do bibliotek NGS bez etapu oczyszczania, aż do momentu, gdy zostają wprowadzone adaptery ligowane do próbki DNA. W przypadku, gdy próbki składają się z dłuższych fragmentów DNA, takich jak DNA genomowe lub amplikony z dalekiego zasięgu PCR, są najpierw enzymatycznie cięte na mniejsze fragmenty. Mediana wielkości fragmentu zależy od aplikacji i długości odczytu sekwencji i można ją dostosować za pomocą zmiany warunków reakcji fragmentacji DNA QIAseq FX. Fragmentowane DNA jest bezpośrednio naprawiane na końcu, „A"

jest dodawane do 3'końca podczas reakcji FX, dzięki czemu fragmenty DNA są gotowe do związania adaptera.

Następnym krokiem adaptery specyficzne dla platformy są ligowane do obu końców fragmentów DNA. Adaptery te zawierają sekwencje niezbędne dla wiązania bibliotek z tzw. podwójnym kodem kreskowym do komórki przepływowej w celu sekwencjonowania, umożliwiając amplifikację PCR fragmentów ligowanych z adapterem i wiążących standardowe startery do sekwencjonowania Illumina. W celu zapewnienia maksymalnej wydajności z ograniczonych ilości materiału wyjściowego, można przeprowadzić wysoce dokładny etap amplifikacji przy użyciu odczynników zawartych w zestawie bibliotecznym DNA QIAseq FX. Zastrzeżona technologia HiFi PCR Master Mix może w równomierny sposób amplifikować regiony DNA charakteryzujące się zmienną zawartością GC, minimalizując błąd sekwencjonowania wywołany przez PCR.

Adaptery w formacie płytkowym z podwójnym barkodem identyfikacyjnym wchodzą w skład zestawu bibliotecznego QIAseq FX DNA. Każda studzienka zawiera adapter jednorazowego użytku składający się z unikalnej kombinacji dwóch ośmionukleotydowych barkodów identyfikacyjnych. Łącząc jeden kod kreskowy D5 i jeden kod kreskowy D7 w każdy gotowy do użycia adapter, zestawy QIAseq FX obsługują pulę do 24 lub do 96 studzienek łączących. Po skonstruowaniu bibliotek dokonuje się oczyszczenia reakcji i usuwa adaptery-dimery dzięki zastosowaniu kulek Agencourt® AMPure® XP (Beckman Coulter), co umożliwia łatwą automatyzację na różnych platformach o wysokiej przepustowości.

Budowanie bibliotek genomowych i procedura sekwencjonowania, pozostałych siedmiu gatunków, zostały przeprowadzone w oparciu o doświadczenia z wcześniej opracowanymi metodami, stosowanymi podczas otrzymania genomu chloroplastowego *Chamaedaphne calyculata* z użyciem platformy MiSeq 600v3 oraz zestawu Nextera XT DNA Library Preparation Kit (Illumina, San Diego, CA, USA). Nextera XT DNA Library Preparation wykorzystuje zmodyfikowany transpozom do znakowania genomowego DNA, w procesie fragmentowania DNA, a następnie znakuje DNA sekwencjami adaptorowymi w jednym kroku. Limitowany cykl PCR używa adapterów do amplifikacji wstawianego DNA. Etap PCR dodaje również indeksowane sekwencje

adaptera na obu końcach DNA, co pozwala na podwójne indeksowane sekwencjonowanie połączonych bibliotek na platformach sekwencjonowania Illumina. Technologia Nextera daje możliwość na symultaniczny przebieg procesu fragmentacji i przyłączenia adaptorowych sekwencji w reakcji enzymatycznej, w jednym kroku. Przygotowanie bibliotek przebiega z zastosowaniem ultra niskiej puli materiału wyjściowego – zgodnie z zaleceniami obsługuje wkład 1 ng DNA. Nextera XT przewiduje innowacyjną normalizację próbek opartą na kulkach, która eliminuje potrzebę kwantyfikacji biblioteki przed łączeniem i sekwencjonowaniem. Biblioteki przygotowane za pomocą zestawów Nextera XT są kompatybilne ze wszystkimi sekwenatorami Illumina. Zestawy do indeksowania Nextera XT, typu A, B, C i D suponują aż do 384 unikalnych podwójnych indeksów, co zapewnia klarowne przyporządkowanie odczytów i efektywne zastosowanie kuwety przepływowej. Te niepowtarzalne podwójne kody indeksowe wykorzystują kody 10 pz. Ta zmiana w kodach indeksowych par zasad potrzebuje akomodacji konfiguracji w przebiegu sekwencjonowania (Aird i in. 2011; Saunders i in. 2012).

Biblioteki genomowe do sekwencjonowania HiSeq zostały otrzymane za pomocą zastosowania zestawu QIAGEN QIAseq FX DNA (Rycina 2). Do tego celu wykorzystano 1 ng DNA zgodnie z procedurami przedstawionymi w protokole. Liczba i poprawność bibliotek genomowych została zweryfikowana z użyciem starterów, których sekwencje podano w przewodniku Sequencing Library qPCR Quantification Guide (Illumina).

3.2.4. Procedura i etapy przygotowania bibliotek *Erica versicolor*

Etap fragmentacji i naprawy końców DNA

Przed przystąpieniem do tego etapu konieczna była dokładna znajomość wyjściowej wielkości DNA każdej z badanych prób. Skorzystano z tabel załączonych do protokołu, aby określić czas wymagany do fragmentacji wprowadzonego DNA do pożądanego rozmiaru (tu 450 pz). Jeśli wielkość DNA była mniejsza niż 10 ng, dodawano FX Enhancer i przeprowadzano reakcję w połowie czasu reakcji podanego dla 10 ng wejściowego DNA. Odczynniki odtajano na lodzie. Po rozmrożeniu odczynników szybko mieszano bufory, wirowano w celu uniknięcia miejscowych stężeń. Wszystkie odczynniki z tego samego powodu krótko odwirowywano przed użyciem.

1. Przygotowano mieszaninę reakcyjną FX w płytce na lodzie zgodnie z tabelą dla > 10 ng wprowadzanego DNA – 5 µl buforu FX + 2,5 µl enhancera + 2,5 µl wody wolnej od nukleaz + oczyszczone DNA. Totalna objętość bez mieszaniny enzymatycznej FX to 40 µl. Komponenty wymieszano, delikatnie pipetując bez wirowania. Przy braku konieczności zastosowania enhancera, odpowiednią mu objętość dorównano wodą wolną od nukleaz.

2. Dodano 10 µl mieszaniny enzymatycznej FX do każdej sporządzonej próby i dokładnie wymieszano, pipetując w górę i w dół po 6–8 razy. Bardzo ważne było utrzymywanie reakcji na lodzie przez cały czas jej przebiegu.

3. Krótko odwirowano płytkę i natychmiast przeniesiono ją do wstępnie schłodzonego do temperatury 4 °C termocyklera.

4. Po minucie w 4°C mieszaninę reakcyjną inkubowano we wcześniej przygotowanym programie, w następujących warunkach: 5 minut w temperaturze 32°C, następnie 30 minut w temperaturze 65°C i powrót ponownie do temperatury 4°C. Po schłodzeniu w 4 °C próbki przeniesiono z powrotem na lód.

Etap ligacji adapterów

Przed przystąpieniem do tej części doprowadzono, potrzebne do kolejnego etapu, kulki Agencourt® AMPure® XP do temperatury pokojowej. Zworteksowano i zwirowano rozmrożoną płytkę adaptera.

1. Zdjęto pokrywkę ochronnej powłoki adaptera, przebijając uszczelki foliowe dla każdej studzienki, którą przeznaczono do użycia. Kolejno przeniesiono 5 µl z każdej jednej studzienki adaptera DNA do każdych 50 µl próbki z poprzedniego etapu.

2. Przygotowano mieszaninę ligacji (na próbkę DNA) w osobnej płytce PCR na lodzie, dobrze wymieszano ją przez wielokrotne pipetowanie. Skład mieszaniny na jedną próbkę stanowiło 20 µl buforu ligującego, 10 µl ligazy DNA i 15 µl wody wolnej od nukleaz.

3. Dodano 45 µl tak przygotowanej mieszaniny ligacyjnej do każdej próbki, do całkowitej objętości 100 µl, i dobrze wymieszano przez wielokrotne pipetowanie. Po wymieszaniu inkubowano reakcję ligacji w 20 °C przez 15 min.

Etap oczyszczania DNA po ligacji adapterów

1. Po etapie inkubacji natychmiastowo przystąpiono do oczyszczenia adaptera ligacji za pomocą kulek 0.8x (80 µl) Agencourt® AMPure® XP. Dodano po 80 µl zawieszonych kulek Agencourt® AMPure® XP do każdej ligowanej próbki i dokładnie wymieszano przez pipetowanie. Mieszaninę inkubowano przez 5 minut w temperaturze pokojowej, po czym przeniesiono płytkę na czas 2 minut na statyw magnetyczny DynaMag, a następnie ostrożnie odrzucono supernatant na tyle ostrożnie, by nie naruszyć kulek z fragmentami DNA.

2. Kulki przemyto dodając 200 µl sporządzonego wcześniej 80% etanolu. Ponownie wstawiono płytkę z kulkami na statywie magnetycznym i odrzucono supernatant. Proces przemywania etanolem powtórzono jeszcze raz. Usunięto możliwie jak najwięcej nadmiaru etanolu.

3. Inkubowano kulki w płytce na statywie magnetycznym, aż do momentu wyschnięcia kulek, przez ok. 10 minut, uważając, by kulki nie były nadmiernie przesuszone, co mogłoby skutkować obniżeniem odzysku DNA.

4. Po zdjęciu płytki ze statywu magnetycznego przeprowadzono elucję poprzez zawieszenie kulek w 52,5 µl roztworu 10 mM Tris-HCl o pH 8,0. Po umieszczeniu płytki ponownie na statywie magnetycznym DynaMag i krótkiej inkubacji w

temperaturze pokojowej, ostrożnie przeniesiono 50 µl supernatantu na nową, czystą płytkę, uważając by nie naruszyć kulek.

5. Wykonano drugie oczyszczenie przy użyciu 1x (50 µl) kulek Agencourt® AMPure® XP powtarzając kroki dla etapów wiązania DNA i przemywania. Eluowano DNA, dodając 26 µl buforu 10mM Tris-HCl o pH 8,0. Po wstawieniu płytki na statyw magnetyczny, uważnie zebrano 23,5 µl oczyszczonego DNA.

   Przeniesiono próbki do probówek DNA LoBind do amplifikacji biblioteki.

Etap amplifikacji bibliotek DNA

1. Przygotowano mieszaninę reakcyjną na lodzie w składzie: 23,5 µl bibliotek genomowych, 25 µl HiFi PCR MasterMix, 2x oraz 1,5 µl starteru Primer Mix. Wymieszano składniki w 96-studzienkowej płytce PCR, którą przeniesiono następnie do termocyklera.

2. Reakcję PCR przeprowadzono w następujących warunkach termicznych: wstępna denaturacja przez 2 minuty w 98ºC, denaturacja przez 20 sekund w 98ºC, hybrydyzacja odcinków starterowych przez 30 sekund w 60ºC, elongacja 30 sekund w 72 ºC i końcowa elongacja przez 1 minutę w 72ºC. 3 ostatnie etapy zostały powtórzone 12 razy.

Etap oczyszczania bibliotek po procesie amplifikacji

1. Po zakończeniu reakcji PCR dodano 50 µl zawieszonych kulek Agencourt® AMPure® XP do każdego dołka, zawartość wymieszano przez wielokrotne pipetowanie.

2. Mieszaninę inkubowano przez 5 minut w temperaturze pokojowej. Umieszczono płytkę z kulkami na statywie magnetycznym DynaMag, po czym ostrożnie odrzucono supernatant, nie naruszając kulek z fragmentami DNA.

3. Nie zdejmując płytki ze statywu, przemyto kulki, dodając 200 µl świeżo przygotowanego, 80% etanolu, po czym ponownie odrzucono supernatant. Płukanie etanolem powtórzono jeszcze raz, starając usunąć jak najwięcej nadmiaru etanolu. Inkubowano płytkę w temperaturze pokojowej do wyschnięcia, ale nie przesuszenia kulek, następnie płytka została zdjęta ze statywu magnetycznego.

4. Eluowano kulki przez ponowne zawieszenie w 25 µl roztworu 10 mM Tris-HCl o pH 8,0. Płytkę znowu wstawiono do statywu magnetyczego. Ostrożnie, uważając na nienaruszenie kulek, przeniesiono 23 µl supernatantu do czystych probówek.

3.2.5. Procedura i etapy przygotowania bibliotek pozostałych gatunków z rodzaju *Erica*

Biblioteki genomowe do sekwencjonowania MiSeq zostały otrzymane za pomocą zastosowania zestawu Nextera XT DNA Library Preparation.

Etap fragmentacji i znakowania DNA

Przed przystąpieniem do tego etapu wymagane było ustalenie wyjściowej wielkości DNA dla każdej z badanych prób. Określono ilościowo wejściowy DNA i dokonano oceny jakościowej przed rozpoczęciem przygotowania biblioteki. Celem uzyskania najlepszych wyników, użyto zaproponowane w tabelach dołączonych do protokołu, dane wejściowe. Niższe ilości wejściowe skutkowałyby niską wydajnością i zwiększoną liczbą duplikatów. Określenie ilościowe wejściowego DNA przeprowadzono zgodnie z zalecieniami w protokole. Do dokładnej kwantyfikacji wejściowego DNA użyto metod fluorometrycznych Qubit™.

Etap ten wykorzystuje transpozom Nextera do przeprowadzenia znakowania genomowego DNA, stanowiącym proces polegający na jednoczesnej fragmentacji i znakowaniu puli DNA z sekwencjami adapterów.

1. Do każdego dołka 96-dołkowej płytki PCR dodano kolejno: 10 µl buforu TD (Tagment DNA Buffer) oraz 1 ng DNA (5 µl). Składniki wymieszano za pomocą pipety. Następnie dodano do każdego dołka po 5 µl mieszaniny ATM (Amplicon Tagment Mix). Zawartość przepitowano minimum dziesięciokrotnie, po czym zamknięto płytkę.

2. Wirowano następnie przy 280 × g przez 1 minutę, w temperaturze 20^{o}C.

3. Płytkę umieszczono w zaprogramowanym termocyklerze i uruchomiono program TAG (opcja podgrzewania pokrywy na 100 ^{o}C, objętość reakcji 50 µl, 55 ^{o}C przez 5 minut i przejście do 10 ^{o}C). W momencie, gdy program osiągnął 10^{o}C, natychmiast przystąpiono do kolejnych kroków, ponieważ transpozom był wciąż aktywny.

4. Dodano po 5 µl buforu NT (Neutralize Tagment Buffer) do każdego dołka płytki. Zawartość przepitowano minumum dziesięciokrotnie, po czym zamknięto szczelnie płytkę.

5. Wirowano następnie przy 280 × g przez 1 minutę, w temperaturze 20^{o}C, po czym inkubowano płytkę w temperaturze pokojowej przez 5 minut.

Etap amplifikacji bibliotek DNA

Na tym etapie przeprowadza się amplifikację znakowanego DNA przy zastosowaniu programu PCR o limitowanym cyklu. Etap PCR dodaje indeksowane adaptery i sekwencje wymagane do generowania klastrów sekwencjonowania.

1. W pierwszej kolejnojności przygotowano materiały eksploatacyjne zgodnie z protokołem: mieszaninę NPM (Nextera PCR Master Mix) rozmrażaną na lodzie w przedziale temperatur od -25 ºC do - 25 ºC oraz adaptery indeksowe rozmrażane w temperaturze pokojowej. Płytkę zwirowano krótko przed użyciem.

2. Dodano po 5 µl adaptera indeksu na każdą próbkę. Następnie do każdego dołka dodano 15 µl mieszaniny NPM. Zawartość przepitowano minumum dziesięciokrotnie, po czym zamknięto szczelnie płytkę.

3. Wirowano następnie przy 280 × g przez 1 minutę, w temperaturze 20ºC, po czym umieszczono płytkę zaprogramowanym termocyklerze i uruchomiono program NXT PCR: wstępne nagrzanie pokrywy na 100 ºC, objętość reakcji 50 µl. Następnie ustawienie 72 ºC przez 3 minuty i 96 ºC przez 30 sekund, po czym 12 cykli w warunkach: denaturacja przez 10 sekund w 95 ºC, hybrydyzacja odcinków starterowych przez 30 sekund w 55 ºC i elongacja przez 30 sekund w 72 ºC. Końcowa elongacja przez 5 minut w temperaturze 72 ºC i przejście do 10 ºC.

Etap oczyszczania bibliotek po procesie amplifikacji

Przed przystąpieniem do tego etapu zestaw kulek oczyszczających AMPure XP został odpowiednio zworteksowany i wymieszany, by upewnić się że kulki są ułożone regularnie i równomiernie. Zostały one także wyjęte z lodówki na 30 minut przed rozpoczęciem etapu oczyszania, celem osiągnięcia przez nie temperatury pokojowej. Bufor RSB rozmrożono i pozostawiono także do osiągnięcia temperatury pokojowej. Przygotowano 80% roztwór EtOH z etanolu absolutnego.

1. Płytkę zwirowano przy 280 × g przez 1 minutę, w temperaturze 20^{o}C, aby zebrać zawartość na dnie studzienek.

2. Przeniesiono po 50 µl supernatantu z każdej pojedynczej studzienki w płytce PCR do analogicznych studzienek w nowo przygotowanej płyce. Następnie do każdej studzienki z supernatantem dodano po 30 µl zawiesiny kulek magnetycznych AMPure XP i uszczelniono płytkę.

3. Płytkę wytrząsano 1800 obrotów na minutę, przez czas 2 minut, po czym inkubowano ją w temperaturze pokojowej przez 5 minut.

4. Płykę ustawiono na statywie magnetycznym i odczekano kilka minut aż płyn będzie klarowny. Natępnie nie naruszając kulek magnetycznych usunięto i odrzucono cały zgromadzony supernatant.

5. Do płytki umieszczonej wciąż w statywie magnetycznym dodano 200 µl świeżego 80% etanolu, nie mieszając prób. Inkubowano prze 30 sekund, po czym usunięto

i odrzucono cały zgromadzony supernatant. Czynność płukania etanolem powtórzono dwukrotnie, starając się usunąć jak najwięcej etanolu.

6. Pipetą 20 µl usunięto i odrzucono pozostałą zawartość etanolu, po czym płytkę w statywie magnetycznym suszono na powietrzu, w temperaturze pokojowej przez czas 15 minut.

7. Po upływie 15 minut płytkę zdjęto ze statywu magnetycznego. Do każdego dołka z kulkami dodano po 52,5 µl buforu RSB (Resuspension Buffer) i szczelnie zamknięto płytkę, którą następnie wytrząsano 1800 obrotów na minutę, przez czas 2 minut.

8. Po wytrząsaniu płytkę inkubowano w temperaturze pokojowej przez czas 2 minut, po czym przeniesiono ją na statyw magnetyczny i odczekano kilka minut aż płyn będzie klarowny.

9. Do nowej 96-dołkowej płytki PCR przeniesiono odpowiednio po 50 µl supernatantu.

Jakość wszystkich otrzymanych bibliotek genomowych oceniono za pomocą urządzenia do elektroforezy kapilarnej QIAxcel (Qiagen). Elektroforeza została zrobiona z użyciem zestawu QIAxcel High Resolition Kit markerem wyrównania 15–1000 pz. (Qiagen) i markerem wielkości 25–1000 pz DNA (Qiagen). Jako program do elektroforezy zastosowano ustawienia standardowe OL500. Zwalidowane biblioteki zostały połączone zgodnie z zaleceniami protokołu.

Biblioteki genomowe siedmiu gatunków zsekwencjonowano za pomocą platformy Miseq 600v3, który wspiera otrzymanie odczytów 2 × 300 pz przy zestawie 600 cykli, zwiększając liczbę odczytów do 25 milionów. Natomiast biblioteki genomowe *Erica versicolor* zsekwencjoowano za pomocą platformy HiSeq 2000, wpierając otrzymanie odczytów 2 x 150 pz, dając ponad 300 milionów odczytów w cyklu. Powstałe odczyty zostały wstępnie zmontowane przy użyciu Velvet de novo assembler (Zerbino i in. 2008) w usłudze Illumina BaseCloud.

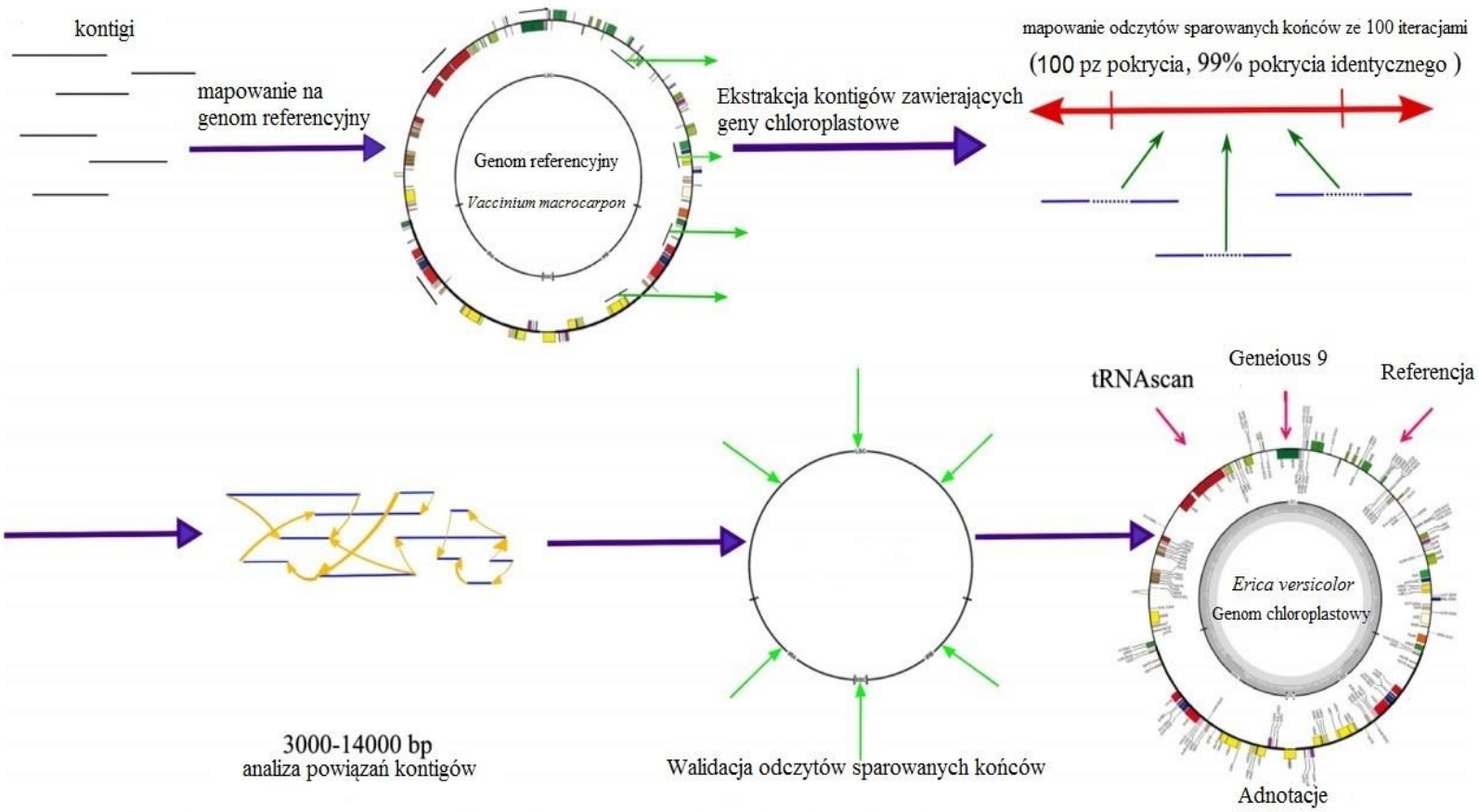
Otrzymane kontigi zostały zmapowane na genomy z rodziny *Ericaceae* z zastosowaniem ustawień średniej wrażliwości dostępnych w programie Geneious 9 (Drummond i in. 2011). Najdłuższe zmapowane kontigi o długości 6-12 kb zostały wykorzystane do rekonstrukcji kompletnego genomu chloroplastowego *Erica versicolor* (Rycina 3), który z kolei stał się referencją dla genomu *Erica arborea*.

Wszystkie odczyty końcowe zostały zmapowane w kolejności za pomocą zmodyfikowanego algorytmu średniej czułości (z minimalnym nakładaniem się 100 pz i 99% nakładania identyczności) ze 100–1000 iteracji. Uzyskane duże kontigi były montowane de novo przy użyciu domyślnych ustawień oprogramowania Geneious 9 (Drummond i in. 2011). Złożony jeden genom plastydowy nieposiadający żadnych luk i z nakładającymi się dużymi kontigami wahał się od 705 pz do 3670 pz. Schemat poglądowy dla konstrukcji genomu plastydowego *Erica versicolor* jest przedstawiony na Rycinie nr 4, a dla genomu plastydowego *Erica arborea* na Rycinie nr 5.

Adnotacja genomu została wykonana przy użyciu programów Geneious 9 (Drummond i in. 2011), CpGavas (Liu i in. 2012) oraz aplikacji DOGMA (Wyman i in. 2004). Złożone i opisane genomy plastydowe *Erica versicolor* i *Erica arborea* zostały zdeponowane w bazie GenBank, gdzie nadano im nr akcesyjne, kolejno: MW282955 oraz MW338736. Mapa okrągłego genomu plastomu dla obu gatunków została sporządzona w programie Organellar Genome DRAW (Greiner i in. 2019).

Wyrównanie genomu przeprowadzono w programie Mauve 2.3.1 (Darling i in. 2010).

Wyrównane genomy zastosowano do obliczenia odległości p w programie MEGA11 (Tamura i in. 2021) w celu oceny użyteczności pełnego sekwencjonowania genomu plastydowego dla filogetyki i identyfikacji gatunków. Luki zostały wykluczone z analizy.



Rycina 3. Schemat składania genomu *Erica versicolor* in silico i jego adnotacja

3.2.6. Analiza bioinformatyczna sekwencji genomowych otrzymanych metodą NGS

Informatywność filogenetyczną (PI) genów oszacowano za pomocą serwera PhyDesign, a profile informatywności filogenetycznej wykreślono w odniesieniu do nieskalibrowanego drzewa (Lopez-Giraldez i Townsend, 2011). PATHd8 użyto do drzewa o ultrametryzacji stosowanego do nałożenia historycznych zmian w współczynnikach substytucji, które uzyskano za pomocą ML (Britton i in. 2007). Korzeń drzewa ustalono na czas ewolucyjny 1, a wskazówki na czas 0 dla uzyskania względnego wieku kladów. Obliczenia PI dla zestawów danych nukleotydowych oszacowano również przy użyciu programu HyPhy, wykorzystującego empiryczne częstotliwości podstawowe i odwracalny w czasie model substytucji (Pond i in. 2005).

Analizę filogenetyczną wykonano implementując dwie metody: wnioskowanie bayesowskie (Bayesian Inference; BI) oraz metodę maksymalnego prawdopodobieństwa (Maximum Likelihood; ML). Program MrBayes v. 3.2.2 został wykorzystany do analizy filogenetycznej wnioskowania bayesowskiego z wcześniejszymi ustawieniami według danych wyjściowych z programu jModelTest 2.1.7 (Hulsenbeck i in. 2001; Posada 2008; Darriba i in. 2012). Program ten wskazuje najbardziej optymalny model substytucji nukleotydowej dla konkretnego rekordu genów jako kryterium BIC (Bayesian Information Criterion). Oszacowania wnioskowania bayesowskiego dokonano przez uruchomienie sześciu przyrostowo ogrzewanych łańcuchów (algorytm MCMC) dla 1 500 000 pokoleń, próbkując jeden na każde 100 pokoleń losowych drzew. Wykonano test, aby określić liczbę pokoleń, które należy wyeliminować z obliczeń drzewa konsensusu. W pierwotnym badaniu pierwsze 30 000 pokoleń zostało odrzuconych jako „wypalone”. Drzewo konsensusu bayesowskiego zostało zbudowane z wykorzystaniem pozostałych pokoleń. Program FigTree v1.4.2 zastosowano do graficznej korekty wygenerowanych drzew filogenetycznych (Rambaut, 2010).

Do przeprowadzania analizy ML oraz MP (w przypadku analizy odwróconych powtórzeń) zastosowano oprogramowanie MEGA11 (Tamura i in. 2021). Do uzyskania drzewa ML ukazującego zmienności między badanymi gatunkami zastosowano zaawansowane oprogramowanie RAxML (Stamatakis, 2014). Jako pierwszy krok dokonano wyboru najbardziej odpowiedniego modelu substytucji nukleotydowej i zastosowano go jako podstawę do analiz ML (Christelova i in. 2011; Felsenstein, 1985). Otrzymany kompletny genom chloroplastowy *Erica versicolor* posłużył jako referencja do otrzymania genów chloroplastowych pozostałych sześciu (poza *Erica arborea*, dla którego otrzymano także cały genom chloroplastowy) przedstawicieli z rodzaju *Erica*, dla których przeprowadzono szereg analiz filogenetycznych. Proces mapowania odczytów bibliotek genomowych tych gatunków na kompletne geny z genomu chloroplastowego *Erica versicolor* z zastosowaniem ustawień średniej wrażliwości przeprowadzono w programie Geneious 9 (Drummond i in. 2011). W analizie zmienności sekwencji między genami przeprowadzono test, celem wybrania najodpowiedniejszego modelu, służącego do kalkulowania dystansu genetycznego polimorfizmu nukleotydów. Test ten przeprowadzono z zastosowaniem programu MEGA11, przy użyciu narzędzia „model selection", wybierając wybór drzewa typu domyślnego, za pośrednictwem metody statystycznej ML (Maximum Likelihood, największego podobieństwa) z uwzględnieniem pominięcia indeli. Model wytypowano zgodnie z założeniem, że model substytucji najtrafniej opisywany jest przez te modele, które osiągnęły najniższą wartość BIC (Bayesian Information Criterion; bayesowskie kryterium informacyjne Schwartza). Wnioskowanie z drzewa przeprowadzono za pomocą funkcji Close Neighbor Interchange (CNI; w metodzie maksymalnej parsymonii w analizach IR) w metodzie MP, algorytm na poziomie wyszukiwania 3 i liczbie drzew początkowych równy 10. Filogeneza była badana zarówno dla maksymalnego prawdopodobieństwa (ML) i maksymalnej parsymonii przy użyciu metody bootstrap z liczbą bootstrap replikacji na poziomie 2000 (Tamura i in. 2021). W przypadku analizy genów cpDNA dla reprezentantów z rodzaju *Erica* zaimplementowano metodę wygenerowania sieci filogenetycznej przeprowadzoną w programie SplitsTree 4.19.0 (Huson i Bryant, 2006). Metoda ta pomaga wyodrębnić sygnały filogenetyczne, które mogły zostać pominięte przez metody oparte na drzewach filogenetycznych. Nawet przy zastosowaniu zbiorów danych genomowych z dziesiątkami tysięcy znaków informacyjnych niektóre klady mogą mieć niskie bądź

niejednoznaczne wartości wsparcie, pozostawiając pewne relacje filogenetyczne niepewnymi. Sieć konsensusu dostarcza głównie informacji o wartościach wsparcia, które korespondują z podziałem. Nie niosą informacji o rzeczywistej rozbieżności pomiędzy taksonami zdefiniowanymi przez dany podział (Xu i in. 2019). Program TASSEL 5 posłużył do wygenerowania wartości zróżnicowania nukleotydowego dla poszczególnych zestawów danych genomowych oraz połączonych sekwencji cpDNA w rodzinie *Ericaceae* i w samym rodzaju *Erica* (Bradbury i in. 2007).

Krótkie powtórzenia tandemowe zostały określone za pomocą programu MIcroSAtellite (MISA) (Beier i in. 2017) z minimalnym progiem 7 nukleotydów dla powtórzeń mononukleotydowych, 4 nukleotydy dla powtórzeń dinukleotydowych oraz z minimalnym progiem o wartości 3 dla powtórzeń tri-, tetra-, penta- i heksanukleotydowych.

Program REPuter (Kurtz i in. 2001) został wykorzystany do analizy powtórzeń oligonukleotydowych w przód (F - forward), w tył (R - reverse), komplementarnych (C - complement) i palindromowych (P - palindromic) o minimalnej wielkości powtórzeń 30 pz i podobieństwie 90%.

Ponadto powtórzenia tandemowe zostały ocenione przez program Tandem Repeats Finder wersja 4.09 (Benson, 1999) przy użyciu parametrów wyrównania: dopasowanie - 2, niedopasowanie - 7, indele - 7, z minimalnym wynikiem wyrównania do zraportowania powtórzeń – 50 i maksymalną wielkością okresu 500. Prawdopodobieństwo dopasowania pM oraz prawdopodobieństwo idel pI są domyślne i nie można ich modyfikować w tej wersji programu (pM= 0,80, pI = 0,10). Program ten został stworzony do lokalizowania i wyświetlania powtórzeń tandemowych w sekwencjach DNA bez potrzeby określania wzoru, rozmiaru ani żadnego innego parametru. Otrzymane dane wyjściowe zawierają dwa dokumenty: plik tabeli powtórzeń ze szczegółowymi informacjami o każdym powtórzeniu, takimi jak jego lokalizacja, rozmiar, liczba powtórzeń, zawartość nukleotydów oraz plik wyrównania.

Celem wizualizacji porównawczej otrzymanych powtórzeń tandemowych w obrębie zsekwencjonowanych genomów chloroplastowych posłużono się narzędziem wizualizacyjnym Circos, które ułatwia identyfikację i analizę podobieństw oraz różnic w zestawieniu komperatywnym plastomów (Krzywinski, 2009). Skutecznie obrazuje zmienności strukturalne genomu, relacje pozycyjne między poszczególnymi komponentami. Liczba powtórzeń sekwencji tandemowych i ich wysokie zróżnicowanie są skorelowane z licznymi zdarzeniami rekombinacyjnymi. Oprogramowanie to opiera się na okrągłym układzie ideogramów, co upraszcza nakreślenie relacji między parami genów za pomocą wstążkowych linii, określających pozycję, rozmiar oraz orientację przestrzenną powiązanych ze sobą ingredientów genomowych. Otrzymane obrazy stworzono z danych wejściowych wygenerowanych za pomocą zautomatyzowanych narzędzy, dzięki czemu platforma ta doskonale sprawdza się w sprawnych wdrażaniu, analizie i raporcie danych. Etap ten był szczególnie istotny w kwestii poszukiwania wzorców, odpowiedniego kontekstu interpretacji i weryfikacji stawianych hipotez, bowiem możliwości wizualizacyjne stanowią swoistą furtkę pomiędzy algorytmami a naukowymi umiejętnościami poznawczymi ujawniając subtelne struktury danych. Dane biologiczne dysponują zazwyczaj wysoką rozdzielczością – zmiany na poziomie podstawowej pary danych mogą mieć znaczenie, odległości między zmianami mogą być większego rzędu, odległe regiony łączą się przez relacje przylegania, fuzje czy przegrupowania genów. W tradycyjnych przeglądarkach genomowych – powszechnych narzędziach do przedstawienia cech genomów, wyodrębnienia genów, polimorfizmów, etc. - stosowany jest głównie tak zwany układ liniowy, który między innymi powyższe kwestie uwzględnia w znacząco utrudnionym stopniu. Tonują to, przycinają i porządkują segmentowo właśnie wizualizacje genomu, mogąc prowadzić korektę dowolnie do analizowanego segmentu czy części. Circos służy zobrazowaniu rearanżacji genomu, gdzie relacja między dwoma pozycjami genomu reprezentuje fuzję strukturalną. (Krzywinski, 2009; Nusrat i in. 2019).

Na przestrzeni ostatniej dekady opracowano wiele metod automatyzacji analizy danych genomowych. Niemniej jednak wizualna eksploracja zmian w genomach plastydowych, nowotworowych, epigenomach i transkryptomach w wielowymiarowych zestawach danych oraz relacji między tymi zmianami stanowi szczególne wyzwanie. Powszechnym sposobem wizualizacji danych genomowych jest pokazanie zmian

związanych z ich loci genomowymi. Podejście to dobrze nadaje się do udzielania odpowiedzi na pytania dotyczące topografii genomowej zmian lub do badania konkretnych loci genomu – jak tu sekwencji wielopowtórzeniowych. Rozróżniamy dwa główne podejścia do wizualizacji, które wykorzystują współrzędne genomowe: przeglądarki genomu i wykresy kołowe. Circos, który z powodzeniem był stosowany w wielu badaniach i publikacjach naukowych na całym świecie, jak i w publikacjach o charakterze ogólnym i opisowym. Jako skuteczne narzędzie wizualizacyjne wykazał impakt w generalnym postrzeganiu zmian genomowych i ich tendencji. Dzięki jego zastosowaniu można zaprezentować graficznie zarówno różnice pomiędzy minimum dwoma analizowanymi genomami, jak i zmiany zachodzące w czasie w konkretnym analizowanym przypadku. Najbardziej sztandarowym przykładem użycia Circos jest wykreowanie danych obrazujących dywergencję pomiędzy zdrowymi genomami a genomami nowotworowymi oraz genami przed i po leczeniu dla bardziej przemawiającej do wyobraźni wykładni prezentowanych zmian (Jhuraney i in. 2015; Chauss i in, 2022).

Możliwości generowania danych sekwencjonowania w ostatnich latach stale rośnie, przeprowadzenie dokładnej analizy jest potencjalnie etapem ograniczającym szybkość w badaniach genomicznych. Wizualizacja danych genomowych jest niezbędna do prawidłowej interpretacji oraz formułowania przydatnych hipotez. Stanowi także cenną pomoc w zakomunikowaniu odkryć naukowych różnego kalibra. Narzędzia wizualne wypełniają lukę pomiędzy podejściem czysto algorytmicznym a umiejętnościami kognitywnymi badaczy.

Narzędzia służące do wizualizacji danych genomowych, ułatwiają wykonywanie zadań analitycznych, umożliwiają efektywną eksplorację, interpretację oraz manipulowanie danymi, w niektórych przypadkach także wykonywanie obliczeń. Dostępne metody graficzne zostały zaprojektowane do analizy sekwencji złożonych de novo, dopasowań odczytu, przeglądania genomu oraz genomiki porównawczej. Różne typy danych genomowych mogą być kodowane jako zestawy cech. Wizualizacja genomowa składa się z jednego lub wielu układów współrzędnych o określonej budowie, podziałach, zestawach (w przypadku wielu osi) współrzędnych sekwencji. Zestawy funkcji są kodowane jako ścieżki i umieszczane w układadzie współrzędnych. Wiele ścieżek można wyrównać przez układanie w grupki lub nakładanie na siebie.

Wizualizacja może składać się z jednego lub wielu widoków, z których każdy zawiera zestaw wyrównanych ścieżek. Wiele widoków może pokazywać dane w jednej lub wielu skalach czy też ogniskach. Oś sekwencji może być wyświetlana w układzie liniowym, kołowym, przestrzennym, przy różnych ustawieniach dwóch osi. Można wizualizować ją w całości, niektóre jej części lub wyodrębnić całą sekwencję. Osobne segmenty sekwencji mogą być wizualizowane jako całość (ciągłe) lub być oddzielone. Poszczególne cechy w wizualizacji genomu mogą być kodowane poprzez kolory, wielkości bloków czy też zajmowane pozycje. Graficznego zestawienia dwóch sekwencji otrzymanych genomów chloroplastowych z rodzaju *Erica*, dokonano przy zastosowaniu prawdopodobnie najstarszego sposobu porównania dwóch sekwencji poprzez wizualizację danych na wykresach punktowych. Dot Plot jest metodą stworzoną przez Gibbs i McIntyre w roku 1970, koncepcja tej metody analitycznej zakłada, iż osie wykresu punktowego odpowiadają dwóm porównywanym sekwencjom - tu kompletnym genomom chloroplastowym, a punkty na planszy uskazują identyczność sekwencji (Thaenkham i in. 2022).

Kolejność i orientację sekwencji w obu genomach można odczytać bezpośrednio: ukośne linie wskazują odpowiednie segmenty sekwencji, a poziome przesunięcia podkreślają zmiany w kolejności. Wykresy punktowe umożliwiają charakteryzację złożonych lub powtarzających się sekwencji, wizualne wykrywanie motywów DNA i identyfikację modułowych podobieństw między sekwencjami. Wykresy dot plot wykonano przy użyciu programu EMBOSS Dot Matcher z rozmiarem okna 20 i wynikiem progowym 39, wykresy zweryfikowano w programie Geneious 9 (Rice i Longden, 2000; Drummond i in. 2011).

Narzędzie IR Scope wykorzystano do zwizualizowania genów na granicach miejsc połączeń genomu chloroplastów, co daje najbardziej obiecujące wyniki dla plików organizmów blisko spokrewnionych (Amiryousefi i in. 2018). Kompletna macierz danych zawierała 12 akcesji, w tym 11 gatunków z rodziny *Ericaceae* i sekwencję gatunku *Camellia azalea* (kamelia azalia z rodziny kameliowatych/herbatowatych *Theaceae*, rzędu *Ericales* - wrzosowców) jako grupę zewnętrzną.

Sekwencje wykorzystanych gatunków pobrano z repozytorium GenBank. Aby ułatwić dopasowanie sekwencji, wszystkie IR zostały podzielone. Początkowo zostały one wyrównane za pomocą programu Clustal W z ustawieniami domyślnymi, następnie wyregulowane indywidualnie, manualnie, w programie BioEdit Sequence Alignment Editor oraz w MUSCLE v5, nie wszystkie regiony były bowiem tożsame dla analizowanych gatunków (Thompson i in. 1994; Hall, 1999; Edgar, 2004).

Narzędziem porównawczym wykorzystanym do niniejszych badań w przypadku zestawienia dwóch kompletnych genomów chloroplastowych *Erica arborea* i *Erica versicolor* oraz w porównaniu regionów odwróconych IR sporządzonego zestawu danych z rodziny *Ericaceae* był program mVISTA. W przypadku porównania regionów IR, ponobnie jak przy konstruowaniu drzewa ML i MP, macierz danych zawierała 14 akcesji, w tym 13 gatunków z rodziny *Ericaceae* i sekwencję gatunku *Actinidia arguta* jako grupę zewnętrzną. Sekwencje wykorzystanych gatunków pobrano z repozytorium GenBank. Serwer mVISTA pozwala na porównanie sekwencji DNA różnych gatunkówi stanowi podstawową metodę identyfikacji elementów funkcjonalnych w genomach. Umożliwia wizualizację globalnych dopasowań sekwencji DNA o dowolnej wielkości, zapewniając identyfikację podobieństw, konfigurowanie wizualizacji wyrównań na różnych poziomach rozdzielczości wraz z powiązanymi adnotacjami funkcjonalnymi. Zapewnia porównanie i dopasowanie od dwóch do stu sekwencji jednocześnie (Mayor i in. 2000; Frazer i in. 2004).

Dopasowania sekwencji odwróconych powtórzeń, przegrupowań i inwersji w ich obszarze na przykładzie wybranych gatunków z rodziny *Ericaceae* oraz gatunku *Camellia azalea* wykonano w programie Mauve 2.3.1, który wygenerował bloki dopasowania oraz rearanżacji w obrębie regionów odwróconych powtórzeń (Darling i in. 2010).

3.2.7. Otrzymanie i analiza sekwencji jądrowych NOR

Budowanie regionów NOR zaczęto od mapowania uzyskanych bibliotek genomowych na sekwencje ITS otrzymanych za pomocą sekwencjonowania metodą Sangera (sekwencje *Erica plukenetii* dostępne w bazie danych NCBI). Zawierały w swoim składzie ITS1, 5.8S rRNA, ITS2 (internal transcribed spacer 1, partial sequence, 5,8S ribosomal RNA gene, complete sequence and internal transcribed spacer 2, partial sequence) – o łącznej długości 638 par zasad linearnego DNA. Wielokrotne iteracje rozbudowywały sekwencje aż do zamknięcia pętli i otrzymania kompletnej sekwencji regionów NOR dla badanych osobników. Zmontowane sekwencje zostały wyregulowywane i wyrównywane manualnie w programach BioEdit 7 (Hall, 1999) oraz MUSCLE v5 (Edgar, 2004).

Kompletne sekwencje IGS wyekstrahowano z jednostek rDNA i dopasowano przy użyciu programu MUSCLE (Multiple Sequence Compare by Log-Expectation). W programie EMBOSS Dot Matcher dla otrzymanych sekwencji IGS wykonano wykresy dot plot odstępów międzygenowych, gdzie rozmiar ramki zawierającej blok między powtórzeniami odpowiadał długości między powtórzeniami, otrzymane wykresy zweryfikowano w programie Geneious 9 (Rice i Longden, 2000; Drummond i in. 2011).

Aby znaleźć, określić i przeanalizować powtarzalne motywy w regionach IGS, wykorzystano pakiet MEME Suite 5.0.1. (Bailey i in. 2009).

Do ustalenia sekwencji motywów użyto narzędzia MEME, weryfikacja motywów została przeprowadzona za pomocą MAST. Motywy zostały ręcznie dostosowane za pomocą edytora BioEdit Sequence Alignment Editor (Hall, 1999; Edgar, 2004; Bailey i in. 2009). Zakres minimalnej i maksymalnej szerokości pojedynczego motywu iteracyjnego dla gatunku *Erica* ustalono na 100-150 pz oraz 20-200 pz w przypadku *Chamaedaphne calyculata*. Na podstawie wartości E-value poszczególnych wygenerowanych motywów wybrano najbardziej odpowiednią długość motywów.

Przypisanie przypuszczalnych miejsc TIS (miejsca inicjacji transkrypcji) i TTS (miejsca terminacji transkrypcji) przeprowadzono po kompleksowej analizie porównującej sekwencje z danymi literaturowymi w odniesieniu do różnych gatunków w królestwie roślin.

Informatywność filogenetyczną (PI) regionów w obrębie rDNA oszacowano, podobnie jak w przypadku genów cpDNA, za pomocą serwera PhyDesign, a profile informatywności filogenetycznej wykreślono w odniesieniu do nieskalibrowanego drzewa (Lopez-Giraldez i Townsend, 2011). PATHd8 użyto do drzewa o ultrametryzacji stosowanego do nałożenia historycznych zmian w współczynnikach substytucji, które uzyskano za pomocą ML o ultrametryzowanym obliczeniu dla pełnej jednostki rDNA (Britton i in. 2007). Korzeń drzewa ustalono na czas ewolucyjny 1, a wskazówki na czas 0 dla uzyskania względnego wieku kladów.

Obliczenia PI dla zestawów tych danych nukleotydowych oszacowano również przy użyciu programu HyPhy, wykorzystującego empiryczne częstotliwości podstawowe i odwracalny w czasie model substytucji (Pond i in. 2005). Wygenerowano dwa typy profili: profil ogólny sieciowy oraz drugi, powstały w przeliczeniu na miejsca filogenetycznej informatywności dla wybranych regionów w obrębie rDNA. Wyniki informatywności filogenetycznej zostały rozszerzone o analizę kladogramów obliczoną osobno dla regionu 3'ETS + NTS oraz dla kompletnej sekwencji ITS (ITS1 + ITS2) z zastosowaniem metody maksymalnego podobieństwa (ML) w programie MEGA11, za pomocą którego przeprowadzono także szczegółową analizę zmienności genetycznej w obrębie poszczególnych regionów NOR oraz detekcję mutacji typu SNP (Tamura i in. 2021).

4. Wyniki

4.1. Wyniki sekwencjonowania bibliotek i organizacji otrzymanych genomów chloroplastowych

Biblioteki genomowe siedmiu gatunków zsekwencjonowanych za pomocą platformy Miseq 600v3, osiągnęły bardzo rozbieżne wielkości. Wielkość zsekwencjonowanej biblioteki dla gatunku *Erica multiflora* wyniosła wartość najwyższą 713 808 000 pz, z łączną liczbą 2 379 360 odczytów. Natomiast dla gatunku *Erica imbricata* wielkość zsekwencjonowanej biblioteki wyniosła zaledwie 1 484 600 pz, z łączną liczbą 4 948 odczytów.

Z kolei biblioteka genomowa *Erica versicolor* zsekwencjonowana za pomocą platformy HiSeq 2000, osiągnęła wielkość 472 134 800 pz, z łączną liczbą 3 147 565 odczytów.

Zestawienie wielkości wszystkich zsekwencjonowanych bibliotek dla ośmiu badanych reprezentantów z rodzaju *Erica* wraz z łączną liczbą odczytów zaprezentowane zostało w Tabeli 2.

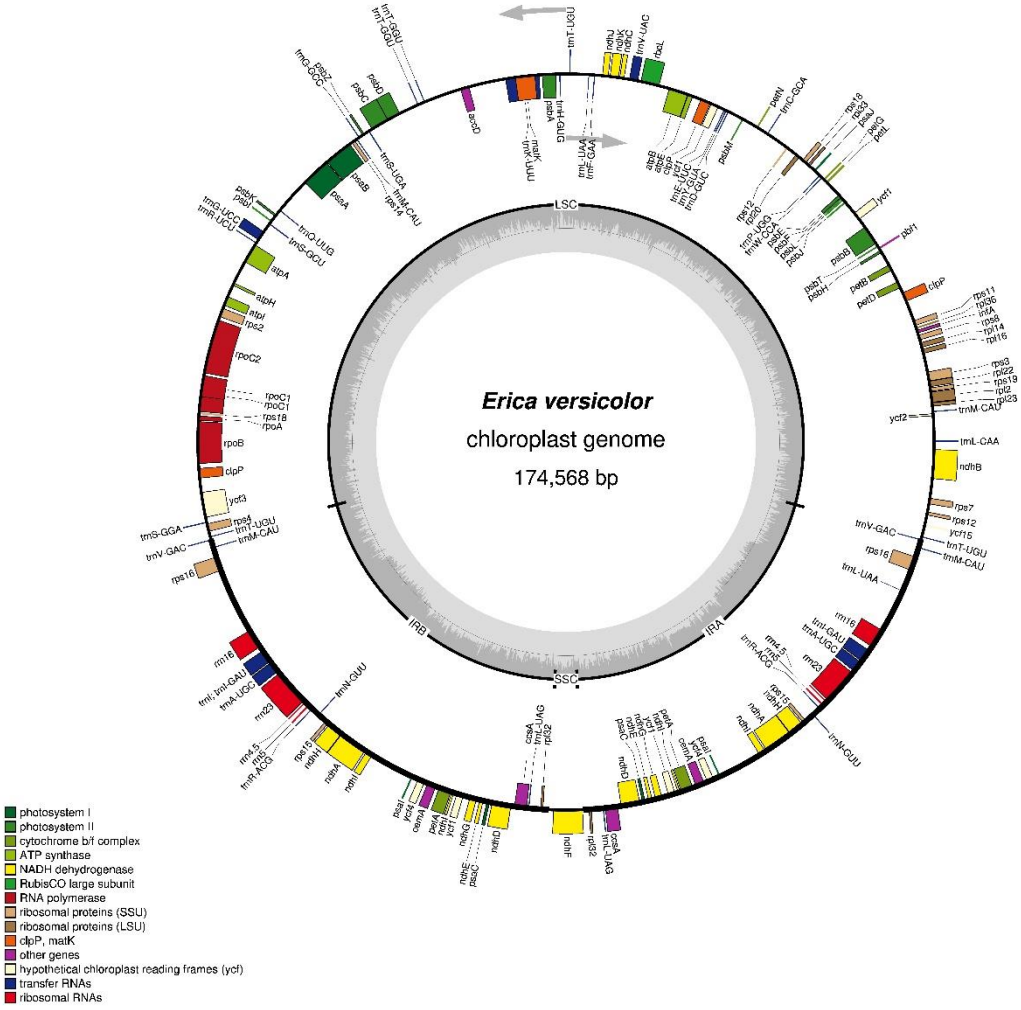
Tabela 2. Długość i liczba odczytów dla poszczególnych osobników, uzyskanych z sekwencjonowania NGS (całkowity otrzymany zasób genomowy)

Nazwa gatunku	**Liczba odczytów**	**Wielkość zsekwencjonowanej biblioteki [pz]**
Erica arborea	1 853 478	556 043 600
Erica cinerea	188 106	56 432 000
Erica globiceps	1 306 262	391 878 800
Erica imbricata	4 948	1 484 600
Erica multiflora	2 379 360	713 808 000
Erica plukenetii	281 505	84 451 500
Erica vagans	413 016	123 905 000
Erica versicolor	3 147 565	472 134 800

4.1.1. Genom chloroplastowy *Erica versicolor* – wyniki sekwencjonowania i organizacja

Kompletny genom plastydowy *Erica versicolor* ma długość 174 587 par zasad (Rycina 4), włączając region LSC o długości 101 464 unikalnych par zasad, oddzielony od regionu SSC długości 2 571 pz przez dwa regiony odwróconych powtórzeń o długości 35 276 pz każdy. Region IR występuje pomiędzy *trn*T-UGU a *rpl*32. Wszystkie regiony kodujące, wliczając w to introny (95 234 pz), stanowią 54,54% całego cpDNA, natomiast rejony międzygenowe (79 353 pz) stanowią pozostałe 45,46%. Plastom koduje 112 genów, wykluczając drugi region powtórzeniowy, korespondujących do 80 genów kodujących białka, 4 rybosomalne RNA (rRNA) i 28 unikalnych genów transportujących RNA (tRNA) (Tabela 3). W rejonach IR znajduje się 30 zduplikowanych genów, wśród których m.in. występuje 6 genów z grupy podjednostki dehydrogenazy NADH, 4 geny z grupy rRNA, 8 genów z grupy tRNA oraz pojedyncze z innych grup. Potencjalnymi genami kodującymi białka są matK w LSC i *ycf*15 w IR.

Geny zidentyfikowane w genomie plastydowym są przedstawione w Tabeli 4. Sekwencję genomu chloroplastowego *Erica versicolor* zdeponowano w bazie danych GenBank, otrzymując dla niej numer akcesyjny MW282955.



Rycina 4. Mapa genomu chloroplastowego *Erica versicolor*. Geny umieszczone po wewnętrznej stronie okręgu są transkrybowane w kierunku przeciwnym wobec ruchu wskazówek zegara, natomiast kierunek transkrybowania genów ulokowanych po zewnętrznej stronie okręgu zachodzi zgodnie z ruchem wskazówek zegara. Strzałki wskazują kierunek transkrypcji. Wewnętrzny okrąg wskazuje duży region pojedynczych kopii (LSC), regiony odwróconych powtórzeń (IRA, IRB) i mały region pojedynczych kopii (SSC). Zwizualizowano na nim zawartość G/C (odcień ciemnoszary) i A/T (odcień jasnoszary). Geny są oznaczone kolorami na podstawie ich funkcji, przynależności do odpowiednich rodzin przytoczonych w legendzie poniżej genomu.

Tabela 3. Dystrybucja genów w regionie plastydowym u *Erica versicolor*

	LSC	IR	SSC	Razem
ORF	62 (włączając *mat*K)	17	1	80
tRNA	20 (+2*)	8	0	28
rRNA	0	4	0	4
Razem	82	29	1	112

*występujące zarówno w LSC, jak i w IR

Tabela 4. Geny ustalone w genomie chloroplastowym *Erica versicolor* ([a]geny zlokalizowane w regionach pojedynczych kopii. [b]pseudogeny, [c]geny zlokalizowane w regionie odwróconych powtórzeń,[d]2 kopie, 1 funkcjonalna i 2 pseudogen)

Nazwy genów	**Rodzaj genu**
*acc*D[a]	Karboksylaza acetylo-CoA
*atp*A[a], *atp*B[a], *atp*E[a], *atp*F [a], *atp*H[a], *atp*I[a]	Podjednostki syntazy ATP
*pet*A[c], *pet*B[a], *pet*D[a], *pet*G[a], *pet*L[a], *pet*N[a]	Podjednostki kompleksu b/f cytochromu
*ccs*A[c]	Białko cytochromu c
*cem*A[c]	Białko jądrowej membrany
*mat*K[a]	Maturaza K
*rrn*4.5[c], *rrn*5 [c], *rrn*16 [c], *rrn*23 [c]	Rybosomowe RNA

Tabela 4. Ciąg dalszy.

Nazwy genów	Rodzaj genu
*ndh*A[c], *ndh*B[a], *ndh*C[a], *ndh*D[c], *ndh*E[c], *ndh*F[c], *ndh*G[c], *ndh*H[c], *ndh*I[c], *ndh*J[a], *ndh*K[a]	Podjednostki dehydrogenazy NADH
*psa*A[a], *psa*B[a], *psa*C[c], *psa*I[c], *psa*J[a]	Podjednostki fotosystemu I
*psb*A[a], *psb*B[a], *psb*C[a], *psb*D[a], *psb*E[a], *psb*F[a], *psb*H[a], *psb*I[a], *psb*J[a], *psb*K[a], *psb*L[a], *psb*M[a], *psb*T[a], *psb*Z[a]	Podjednostki fotosystemu II
*rpo*A[a,d], *rpo*B[a], *rpo*C1[a], *rpo*C2[a]	Podjednostki polimerazy RNA
*rbc*L[a]	Duża podjednostka Rubisco
*rps*2[a], *rps*3[a], *rps*4[a], *rps*7[a], *rps*8[a], *rps*11[a], *rps*12[a], *rps*14[a], *rps*15[c], *rps*16[c], *rps*18[a], *rps*19[a]	Białka małej podjednostki rybosomu
*rpl*2[a], *rpl*14 [a], *rpl*16 [a], *rpl*20[a], *rpl*22[a], *rpl*23[a], *rpl*32[c], *rpl*33[a], *rpl*36[a]	Białka dużej podjednostki rybosomu
*ycf*1[a,c,d], *ycf*2[a], *ycf*3[a], *ycf*4[c], *ycf*15[c]	Hipotetyczne geny

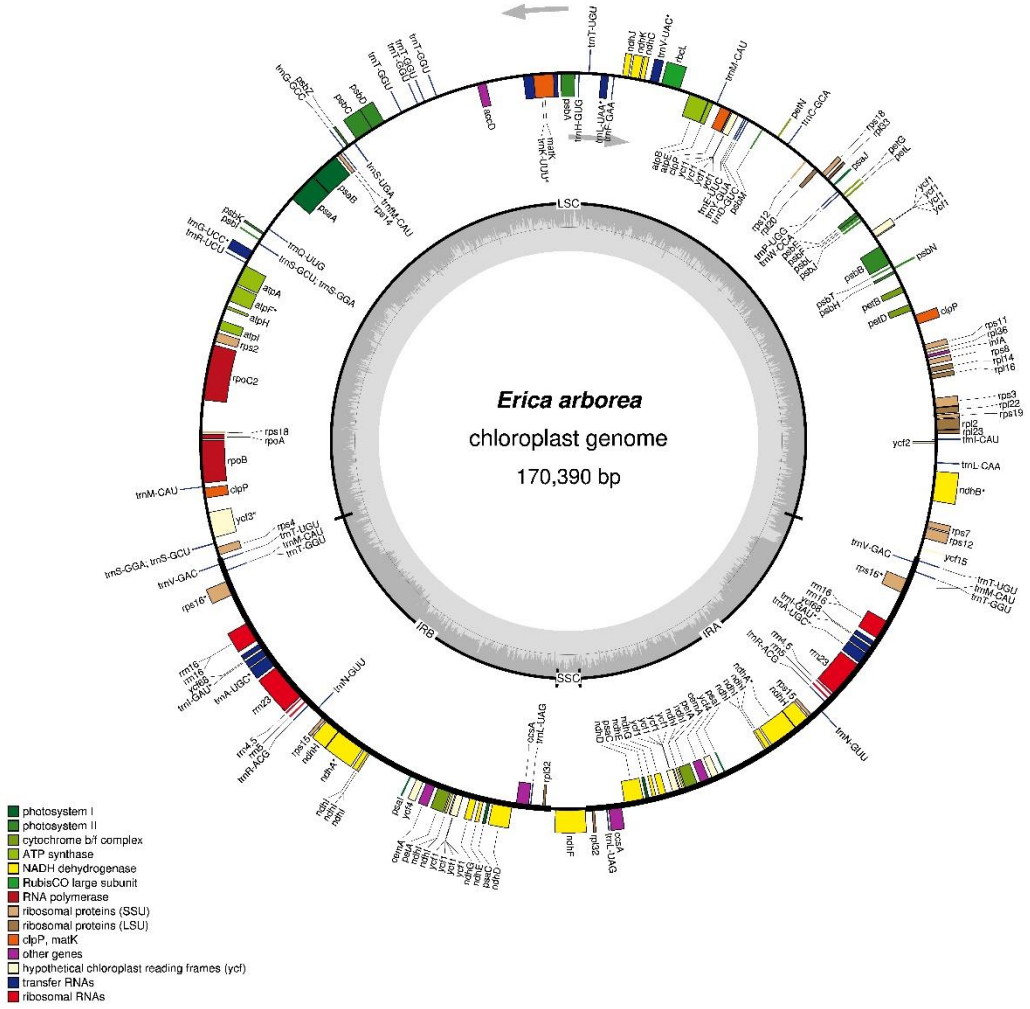
Tabela 4. Ciąg dalszy.

Nazwy genów	Rodzaj genu
*trn*A-UGC [c], *trn*C-GCA [a], *trn*D-GUC [a], *trn*E-UUC [a], *trn*F-GAA [a], *trn*G-GCC [a], *trn*G-UCC [a], *trn*H-GUG [a], *trn*I-GAU [c], *trn*K-UUU [a], *trn*L-CAA [a], *trn*L-UAA [a], *trn*L-UAG [c], *trn*M-CAU [a, c], *trn*N-GUU [c], *trn*P-UGG [a], *trn*Q-UUG [a], *trn*R-ACG [c], *trn*R-UCU [a], *trn*S-GGA [a], *trn*S-GCU [a], *trn*S-UGA [a], *trn*T-GGU [a], *trn*T-UGU [a, c], *trn*V-GAC [c], *trn*V-UAC [a], *trn*W-CCA [a], *trn*Y-GUA [a]	Transportujące RNA
*clp*P [a, d], *inf*A [a]	Pozostałe geny

4.1.2. Genom chloroplastowy *Erica arborea* – wyniki sekwencjonowania i organizacja

Kompletny genom plastydowy *Erica arborea* ma długość 170 390 pz (Rycina 5), włączając region LSC o długości 97 240 unikalnych par zasad, oddzielony od regionu SSC długosc 2 570 pz przez dwa regiony odwróconych powtórzeń o długości 35 290 pz każdy. Region IR występuje pomiędzy *trn*T-UGU a *rpl*32. Wszystkie regiony kodujące, wliczając w to introny (96 063 pz), stanowią 56,38% całego cpDNA, natomiast rejony międzygenowe (74 327 pz) stanowią pozostałe 43,62%. Plastom koduje 115 genów, wykluczając drugi region powtórzeniowy, korespondujących do 82 genów kodujących białka, 4 rybosomalne RNA (rRNA) i 29 unikalnych genów transportujących RNA (tRNA) (Tabela 5). W rejonach IR znajduje się zduplikowanych 32 genów, wśród których m.in. występuje 6 genów z grupy podjednostki dehydrogenazy NADH, 4 geny z grupy rRNA, 9 genów z grupy tRNA oraz pojedyncze z innych grup. Potencjalnymi genami kodującymi białka są matK i LSC i ycf68 w IR.

Geny zdetektowane w genomie plastydowym są przedstawione w Tabeli 6. Sekwencję genomu chloroplastowego *Erica arborea* zdeponowano w bazie danych GenBank, otrzymując dla niej numer akcesyjny MW338736.



Rycina 5. Mapa genomu chloroplastowego *Erica arborea*. Geny umieszczone po wewnętrznej stronie okręgu są transkrybowane w kierunku przeciwnym wobec ruchu wskazówek zegara, natomiast kierunek transkrybowania genów ulokowanych po zewnętrznej stronie okręgu zachodzi zgodnie z ruchem wskazówek zegara. Strzałki wskazują kierunek transkrypcji. Wewnętrzny okrąg wskazuje duży region pojedynczych kopii (LSC), regiony odwróconych powtórzeń (IRA, IRB) i mały region pojedynczych kopii (SSC).. Zwizualizowano na nim zawartość G/C (odcień ciemnoszary) i A/T (odcień jasnoszary). Geny są oznaczone kolorami na podstawie ich funkcji, przynależności do odpowiednich rodzin przytoczonych w legendzie poniżej genomu.

Tabela 5. Dystrybucja genów w regionie plastydowym u *Erica arborea*

	LSC	IR	SSC	Razem
ORF	63 (włączając *mat*K)	18	1	82
tRNA	21 (+2*)	8	0	29
rRNA	0	4	0	4
Razem	84	30	1	115

*występujące zarówno w LSC, jak i w IR

Tabela 6. Geny ustalone w genomie chloroplastowym *Erica arborea* ([a]geny zlokalizowane w regionach pojedynczych kopii. [b]pseudogeny, [c]geny zlokalizowane w regionie odwróconych powtórzeń, [d]2 kopie, 1 funkcjonalna i 2 pseudogen)

Nazwy genów	**Rodzaj genu**
*acc*D[a]	Karboksylaza acetylo-CoA
*atp*A[a], *atp*B[a], *atp*E[a], *atp*F [a], *atp*H[a], *atp*I[a]	Podjednostki syntazy ATP
*pet*A[c], *pet*B[a], *pet*D[a], *pet*G[a], *pet*L[a], *pet*N[a]	Podjednostki kompleksu b/f cytochromu
*ccs*A[c]	Białko cytochromu c
*cem*A[c]	Białko jądrowej membrany
*mat*K[a]	Maturaza K
*ndh*A[c], *ndh*B[a], *ndh*C[a], *ndh*D[c], *ndh*E[c], *ndh*F[c], *ndh*G[c], *ndh*H[c], *ndh*I[c], *ndh*J[a], *ndh*K[a]	Podjednostki dehydrogenazy NADH
*psa*A[a], *psa*B[a], *psa*C[c], *psa*I[c], *psa*J[a]	Podjednostki fotosystemu I

Tabela 6. Ciąg dalszy.

Nazwy genów	Rodzaj genu
*psb*A[a], *psb*B[a], *psb*C[a], *psb*D[a], *psb*E[a], *psb*F[a], *psb*H[a], *psb*I[a], *psb*J[a], *psb*K[a], *psb*L[a], *psb*M[a], *psb*N[a], *psb*T[a], *psb*Z[a]	Podjednostki fotosystemu II
*rrn*4.5[c], *rrn*5[c], *rrn*16[c], *rrn*23[c]	Rybosomowe RNA
*rpo*A[a, d], *rpo*B[a], *rpo*C1[a], *rpo*C2[a]	Podjednostki polimerazy RNA
*rbc*L[a]	Duża podjednostka Rubisco
*rps*2[a], *rps*3[a], *rps*4[a], *rps*7[a], *rps*8[a], *rps*11[a], *rps*12[a], *rps*14[a], *rps*15[c], *rps*16[c], *rps*18[a], *rps*19[a]	Białka małej podjednostki rybosomu
*rpl*2[a], *rpl*14[a], *rpl*16[a], *rpl*20[a], *rpl*22[a], *rpl*23[a], *rpl*32[c], *rpl*33[a], *rpl*36[a]	Białka dużej podjednostki rybosomu
*ycf*1[a,c, d], *ycf*2[a], *ycf*3[a], *ycf*4[c], *ycf*15[a], *ycf*68[c]	Hipotetyczne geny

Tabela 6. Ciąg dalszy.

Nazwy genów	Rodzaj genu
*trn*A-UGC [c], *trn*C-GCA [a], *trn*D-GUC [a], *trn*E-UUC [a], *trn*F-GAA [a], *trn*G-GCC [a], *trn*G-UCC [a], *trn*H-GUG [a], *trn*I-CAU [a], *trn*I-GAU [c], *trn*K-UUU [a], *trn*L-CAA [a], *trn*L-UAA [a], *trn*L-UAG [c], *trn*M-CAU [a, c], *trn*N-GUU [c], *trn*P-UGG [a], *trn*Q-UUG [a], *trn*R-ACG [c], *trn*R-UCU [a], *trn*S-GGA [a], *trn*S-GCU [a], *trn*S-UGA [a], *trn*T-GGU [a], *trn*T-UGU [a, c], *trn*V-GAC [c], *trn*V-UAC [a], *trn*W-CCA [a], *trn*Y-GUA [a]	Transportujące RNA
*clp*P[a,d], *inf*A[a]	Pozostałe geny

4.2. Charakterystyka otrzymanych genomów

4.2.1. Analiza sekwencji powtórzeniowych

Analizę sekwencji powtórzeniowych w otrzymanych dwóch kompletnych plastomach z rodzaju *Erica* przeprowadzono w trzech etapach. Pierwszym etapem było określenie liczby i rozmieszczenia prostych powtórzeń sekwencji (SSR), w drugiej kolejności zlokalizowano powtórzenia oligonukleotydowe, trzecim krokiem było natomiast wyznaczenie liczby i lokalizacji powtórzeń tandemowych (TD).

Detekcja za pomocą programu MISA ujawniła 874 proste powtórzenia sekwencji (SSR) w genomie chloroplastowym *Erica versicolor*, które obejmowały następującą liczbę w typach powtórzeń: 475 powtórzeń mononukleotydowych, 66 dinukleotydowych, 319 trinukleotydowych, 13 tetranukleotydowych i 1 powtórzenie pentanukleotydowe (Rycina 7). Ten rozkład SSR w całym genomie chloroplastowym wykazał obecność 72% SSR w regionach niekodujących, 26,2 % w regionach CDS oraz 1,8 % w genach tRNA i rRNA (16 SSR).

W genomie chloroplastowym *Erica arborea* wykryto 814 prostych powtórzeń sekwencji (SSR), które obejmowały następującą liczbę w typach powtórzeń: 439 powtórzeń mononukleotydowych, 56 dunukleotydowych, 307 trinukleotydowych, 11 tetranukleotydowych i 1 powtórzenie pentanukleotydowe (Rycina 7). Taki rozkład SSR w całym genomie chloroplastowym wykazał obecność 70,4% SSR w regionach niekodujących, 28,1 % w regionach CDS oraz 1,5 % w genach tRNA i rRNA (13 SSR).

Porównania powtórzeń SSR w analizowanych genomach wykazało, że najbardziej rozpowszechnione były powtórzenia mononukleotydowe (z dużą liczbą motywów A/T), następnie trinukleotydowe (z dużą liczbą motywów AAG/CTT), a następnie powtórzenia dinukleotydowe (z dużą liczbą motywów AT/TA), pentanukleotydy i heksanukleotydy znaleziono w niewielkich liczbach bez wyraźnie wyodrębniających się powtarzalnych motywów nukleotydowych (tabele Suplementów 7-10).

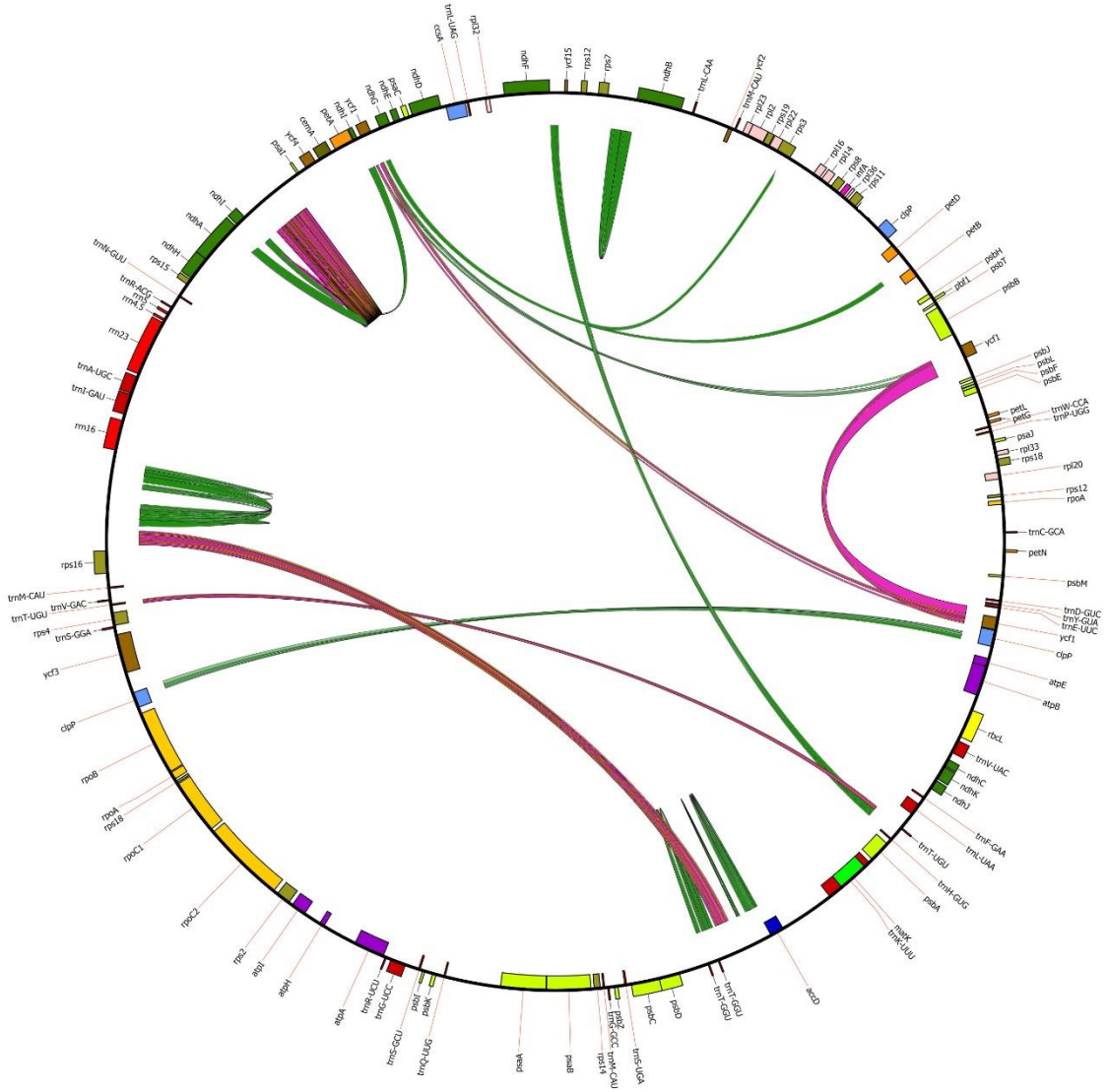
Wykryte SSR w genomach chloroplastowych *Erica arborea* i *Erica versicolor* reprezentują wysoki udział nukleotydów A (27%, 27,4 %) oraz nukleotydów T (24,7%, 24,5%). Większość SSR wykryta została w regionie LSC, a następnie w regionie IR, z wyłączeniem regionu SSC.

Analiza programem REPuter po zignorowaniu powtórzeń i zbędnych powtórzeń tRNA ujawniła po 50 powtórzeń oligonukleotydowych (powyżej 30 pz), z których u *Erica arborea* 24 były do przodu (F), a 26 powtórzeń było odwrotnych (R). Nie wykazano występowania powtórzeń palindromowych. Natomiast u *Erica versicolor* 25 powtórzeń było do przodu (F), a 25 powtórzeń odwrotnych (R), również nie wykryto powtórzeń typu palindromowego. Większość powtórzeń oligonukleotydowych obecna była w regionie LSC, przy czym największa liczba tych powtórzeń występowała w regionach niekodujących, podczas gdy tylko kilka powtórzeń oligonukleotydowych znaleziono w genach kodujących białka, głównie w *ycf*1 i *ndh*I. Długość ujawnionych powtórzeń oligonukleotydowych wahała się od 187 pz do 704 pz u *Erica arborea* oraz odpowiednio od 193 pz do 706 pz u *Erica versicolor*. Największe powtórzenia znaleziono w regionach niekodujących *pet*B i *pet*D, *rps*7 i *ndh*B oraz *ycf*15 i *ndh*F. Dystrybucję powtórzeń oligonukleotydów w dwóch gatunkach z rodzaju *Erica* zestawiono według długości powtórzeń i frekwencji występowania w tabelach Suplementu 11 i 12.

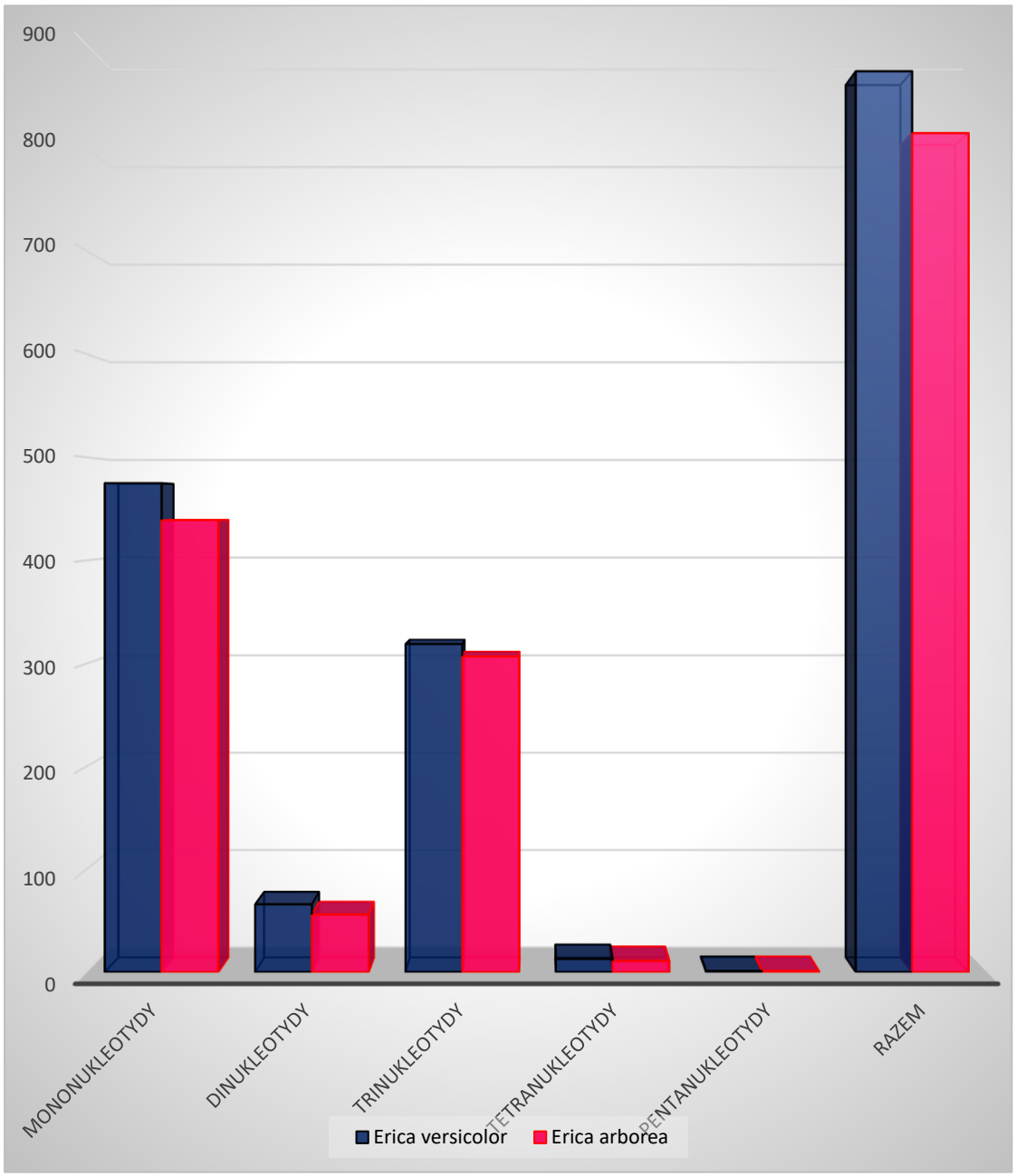
Za pomocą programu TRF (Tandem Repeats Finder) 4.09 przeprowadzono zaawansowane wyszukiwania każdej z sekwencji plastomów. Zidentyfikowano 92 powtórzenia tandemowe w plastydowym genomie *Erica versicolor*, rozmieszczone praktycznie całkowicie w regionach niekodujących oraz jedynie jednostkowo w regionach CDS. Wielkość okresu powtórzeń tandemowych stanowiła rozpiętość od 2 pz do 330 pz. Spośród 92 powtórzeń zanotowano jedynie 8 o małym rozmiarze wzorca powtórzenia tandemowego (od 2 do 6 pz), większość ma duże rozmiary wzorca powtórzenia tandemowego – wielkości bardzo różnorodne, nie tworzące charakterystycznych motywów, o wartościach rzędu 24, 35, 63, 99, 131 pz. Stwierdzono, iż genom chloroplastowy *Erica versicolor* posiada głównie powtórzenia tandemowe z liczbą kopii w granicach od 2 do 6, choć zanotowano występowanie wyjątkowych powtórzeń powielonych kilkunastokrotnie, a nawet kilkudziesięciokrotnie.

W genomie plastydowym *Erica arborea* zidentyfikowanych zostało 99 powtórzeń tandemowych, również rozlokowanych w głównej mierze w regionach niekodujących plastydu oraz jedynie w pojedynczych przypadkach w regionach CDS. Wielkość okresu powtórzeń tandemowych w tym gatunku plasuje się w granicach od 3 pz do 330 pz, natomiast powtórzenia tandemowe występują z liczbą kopii od 2 do 63. Spośród 99 powtórzeń zanotowano jedynie 12 o małym rozmiarze wzorca powtórzenia tandemowego (od 2 do 6 pz), większość ma duże rozmiary wzorca powtórzenia tandemowego – wielkości bardzo różnorodne, podobie jak u *Erica versicolor* nie wyodrębniają charakterystycznych układów motywów (21, 33, 39, 63, 84 99, 164 pz). Odsetek dopasowań między sąsiednimi sekwencjami wahał się od 79 do 100, podczas gdy odsetek indeksów wahał się od 0 do 20. Sugeruje to możliwe mutacje w ewolucji gatunków z rodzaju *Erica*.

Zestawienie sekwencji powtórzeń tandemowych w genomach chloroplastowych *Erica arborea* i *Erica versicolor* przedstawiono w tabelach Suplementu 13 oraz 14, zaś graficznie zwizualizowano je za pomocą programu Circos (Rycina 6).



Rycina 6. Wykres Circos oparty na całkowitym dopasowaniu genomów chloroplastowych między dwoma gatunkami rodzaju *Erica* przedstawionymi w pracy. Mapa połączeń dotyczy wizualizacji zdetektowanych powtórzeń tandemowych w genomach chloroplastowych, których wartość dopasowania wyniosła minimum 90%.



Rycina 7. Zestawienie liczbowe sekwencji powtórzeniowych mikrosatelitarnych w genomach chloroplastowych rodzaju *Erica*

4.2.2. Analiza regionów odwróconego powtórzenia (IR)

Badania najnowszej filogenetyki plastydowej prowadzone na przestrzeni ostatnich lat wyjaśniające filogenezę roślin okrytozalążkowych sugerują, że region odwróconych powtórzeń (IR) może stanowić cenne źródło danych rozwiązujących problemy wnioskowania filogenetycznego w niektórych, głównie problematycznych, węzłach, pomagając np. z sukcesem rozdzielić poszczególne klady. IR są w analizach szczególnie przydatne do szacowania głębszych poziomów filogenetycznych, co ma kluczowe znaczenie dla pełnego zrozumienia rozwoju, morfologii i historii ekologicznej roślin. Stanowi zatem obiecujący marker służący do zobrazowania wysokiej rozdzielczości głębokich rozbieżności genetycznych, mogący pomóc rozwiązać pozostające wciąż nierozwikłane obszary problemowe w filogenezie konkretnych grup organizmów.

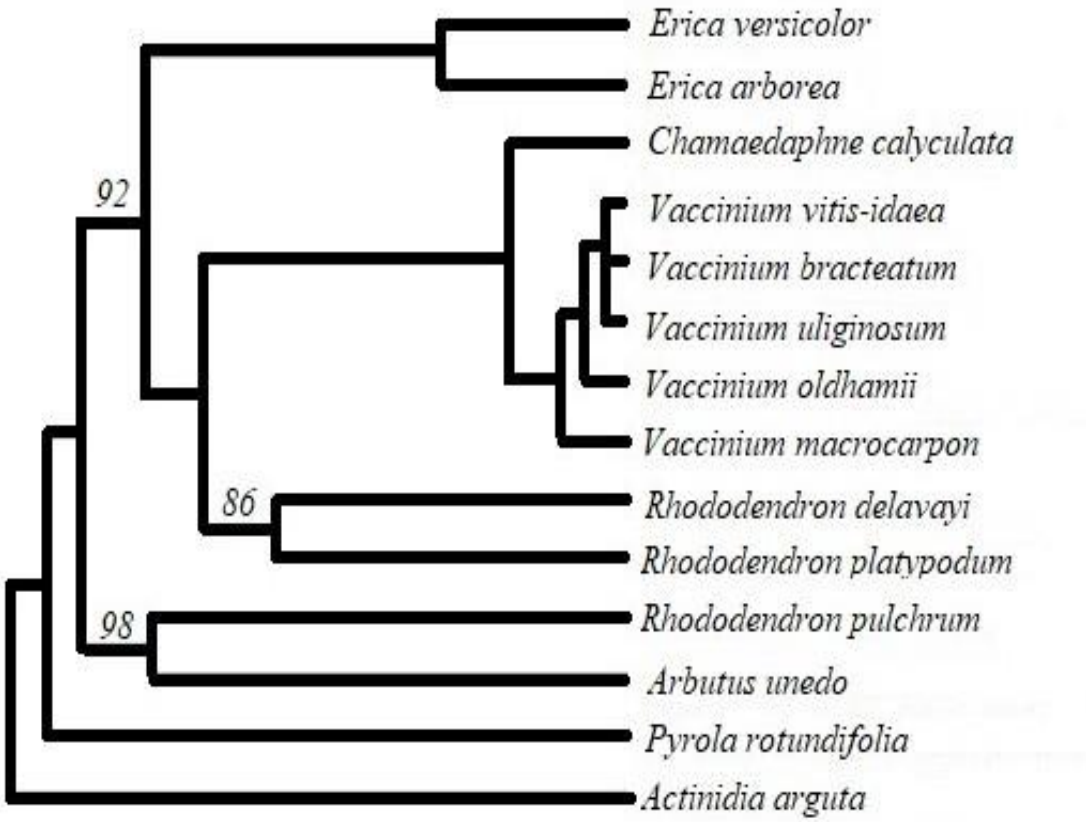
Wspólną cechą charakterystyczną niemal wszystkich genomów plastydowych roślin lądowych jest podział genomu plastydu na dużą i małą pojedynczą kopię przez region IR. Znacząca większość roślin okrytozalążkowych posiada regiony IR o zbliżonej długości, na zmiany w ich rozmiarach wpływa rozszerzanie i kurczenie się końców IR, proces pseudogenizacji poniektórych genów, np. genów *ycf* 2, *ndh* B czy *rpl* 23, czy też sporadyczna utrata intronów w wielu taksonach, np. gen *ndh* B i introny *rpl* 2 (Asaf i in. 2020; Wang i in. 2021; Wu i in. 2021).

Region odwróconych powtórzeń (IR) charakteryzuje się wysokim stopniem konserwatywności w strukturze oraz stosunkowo niskim tempem ewolucji w zestawieniu z regionami pojedycznych kopii, co wykazuje pozytywny impakt na ocenę filogenetyczną, szczególnie wyższych jednostek taksonomicznych, zmniejszając wpływ wysycenia mutacjami. Pomimo niskiego tempa ewolucji w IR badania dowodzą, że analizy maksymalnego podobieństwa (ML) i maksymalnej parsymonii (MP) regionu IR wykazały wysoką zgodność z analizami filogenomicznymi plastydów. Jedynym uchybieniem były głębsze węzły drzew maksymalnej parsymonii (MP) wykazujące błąd wynikający z długości gałęzi.

Ogólnie niższe wsparcie obserwowane w drzewach IR dla gałęzi, które były trudne do rozwiązania w innego typu analizach, prawdopodobnie odzwierciedlają stosunkowo niski współczynnik ewolucji w obrębie IR, skutkujący mniejszą liczbą znaków obsługujących poszczególne gałęzie. Niemniej jednak ten niski wskaźnik okazał się korzystny w rekonstrukcji pokrewieństwa w obrębie niektórych grup (Asaf i in. 2020; Wang i in. 2021).

Zsekwencjonowane całkowite regiony odwróconego powtórzenia (IR) plastydów z rodzaju *Erica* dla dwóch gatunków, łącznie z istniejącymi danymi IR pozwoliły na zestawienie dużego, bogatego w znaki zestawu danych. Kompletna macierz danych zawierała 14 akcesji, w tym 13 gatunków z rodziny *Ericaceae* i sekwencję gatunku *Actinidia arguta* (aktinidia ostrolistna/ małe kiwi z rodziny *Actinidiaceae*, rzędu *Ericales* - wrzosowców) jako grupę zewnętrzną. Połączenie tych sekwencji może złagodzić problemy z niską gęstością taksonów obserwowaną w zestawie danych genomów plastydowych, zapewniając wyższą rozdzielczość danych oraz wsparcie dla relacji filogenetycznych na głębszym poziomie, ze szczególnym uwzględnieniem relacji w obrębie rodziny *Ericaceae*.

Relacje w drzewie MP były w dużej mierze zgodne z tymi w drzewie ML, nie wykazano kluczowych różnic topologicznych (Rycina 8).

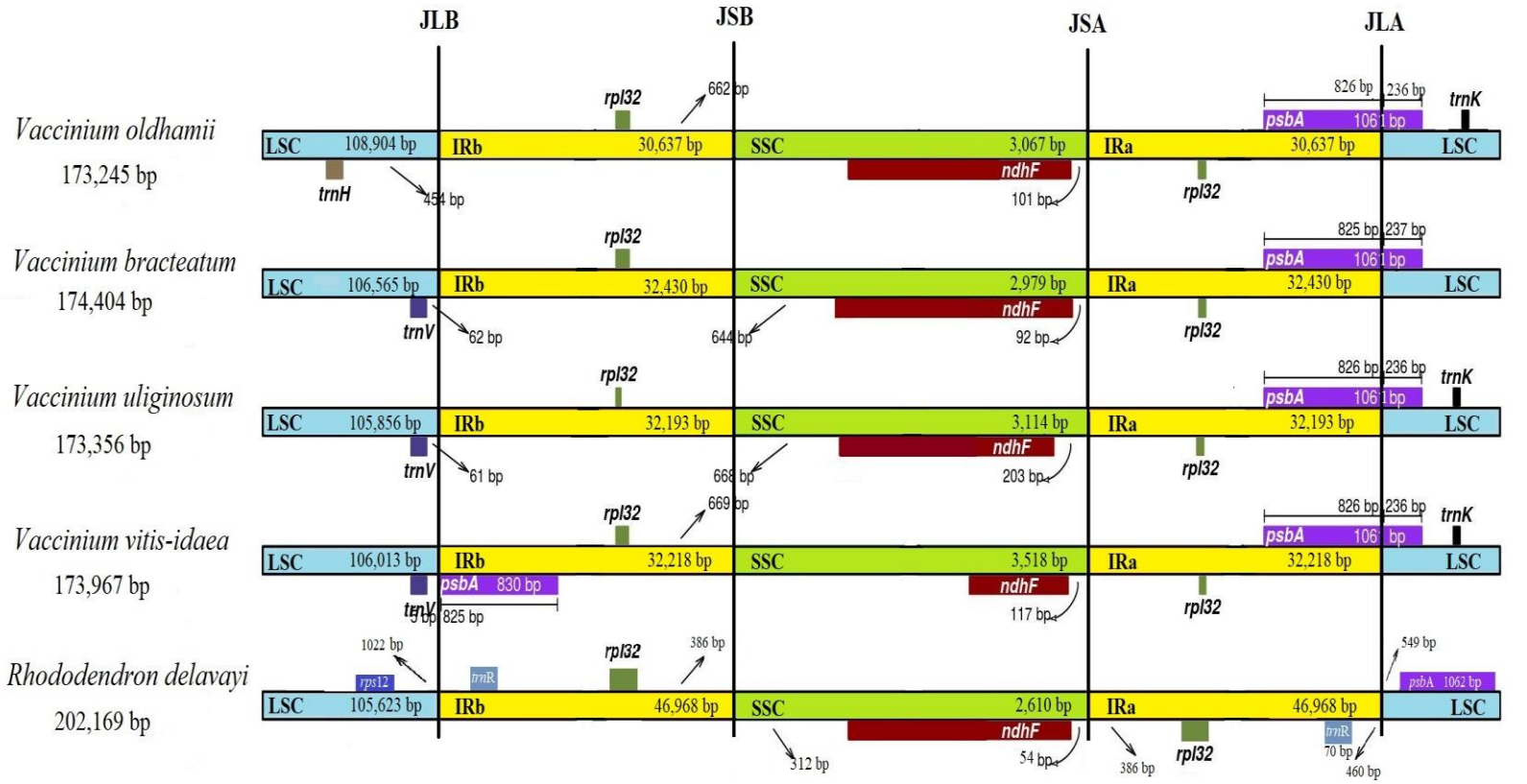


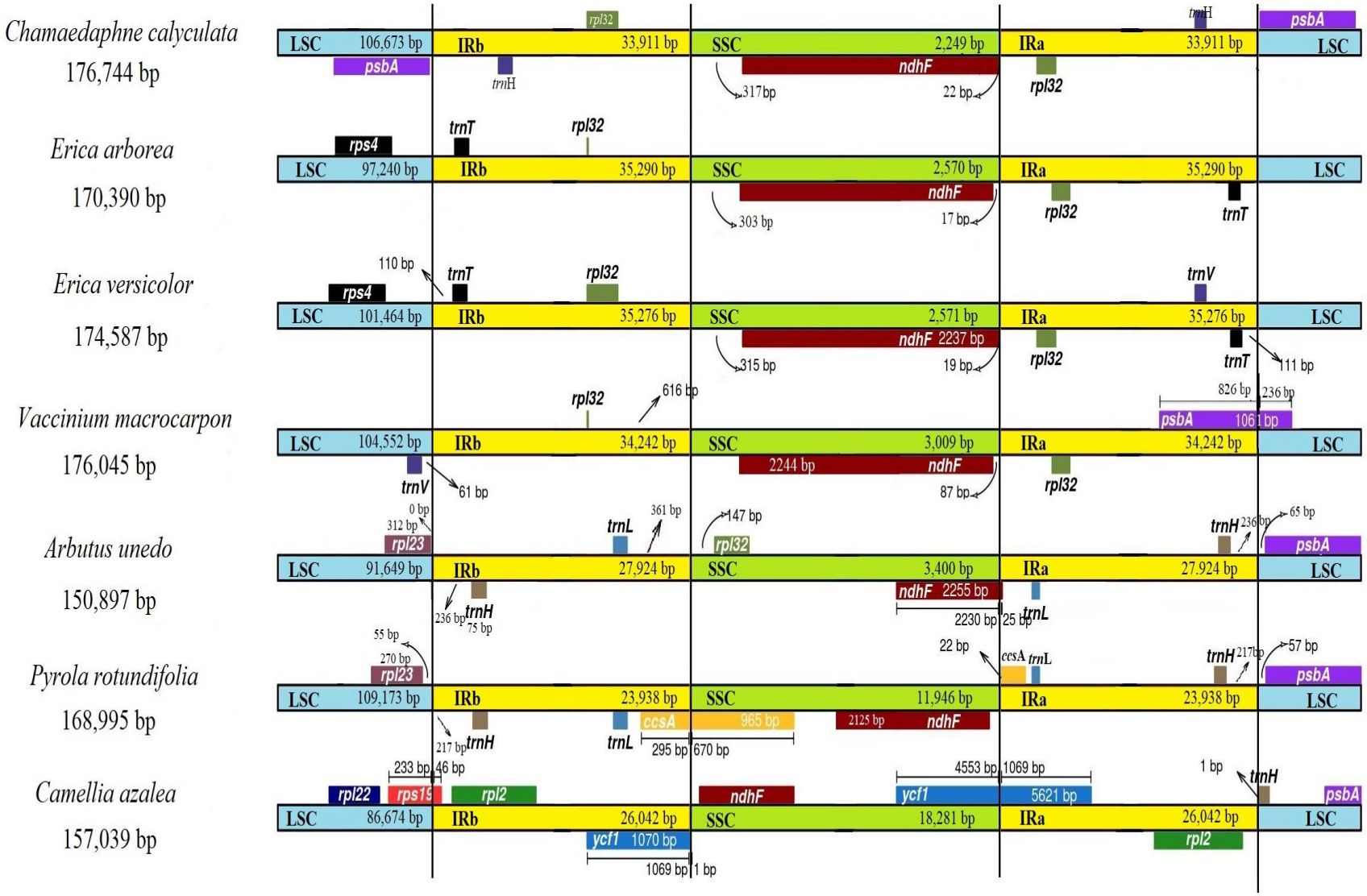
Rycina 8. Drzewo zbudowane na regionach IR gatunków z rodziny *Ericaceae*. Grupę zewnętrzną stanowił gatunek *Actinidia arguta*. Wartości bootstrap poniżej 50% zostały ukryte.

Narzędzie IR Scope służące do wizualizacji genów na granicach miejsc połączeń genomu chloroplastów, daje najbardziej obiecujące wyniki dla plików organizmów blisko spokrewnionych. Pozwala na zaprezentowanie architektury genetycznej maksymalnie dziesięciu genomów chloroplastowych w pobliżu miejsc łączących region odwróconych powtórzeń (IR) z krótkim oraz długim regionem pojedynczej kopii (SSC i LSC), miejsc tych jest cztery (JS – junction site). Każde z miejsc JS między konkretnymi regionami w genomie ma własną nazwę i symbol: JLB to miejsce między IRb a LSC, JSB oznacza miejsce łączące IRb z SSC, JSA to połączenie regionu SSC z IRa, natomiast JLA odpowiada miejscu łączącemu region IRa z LSC.

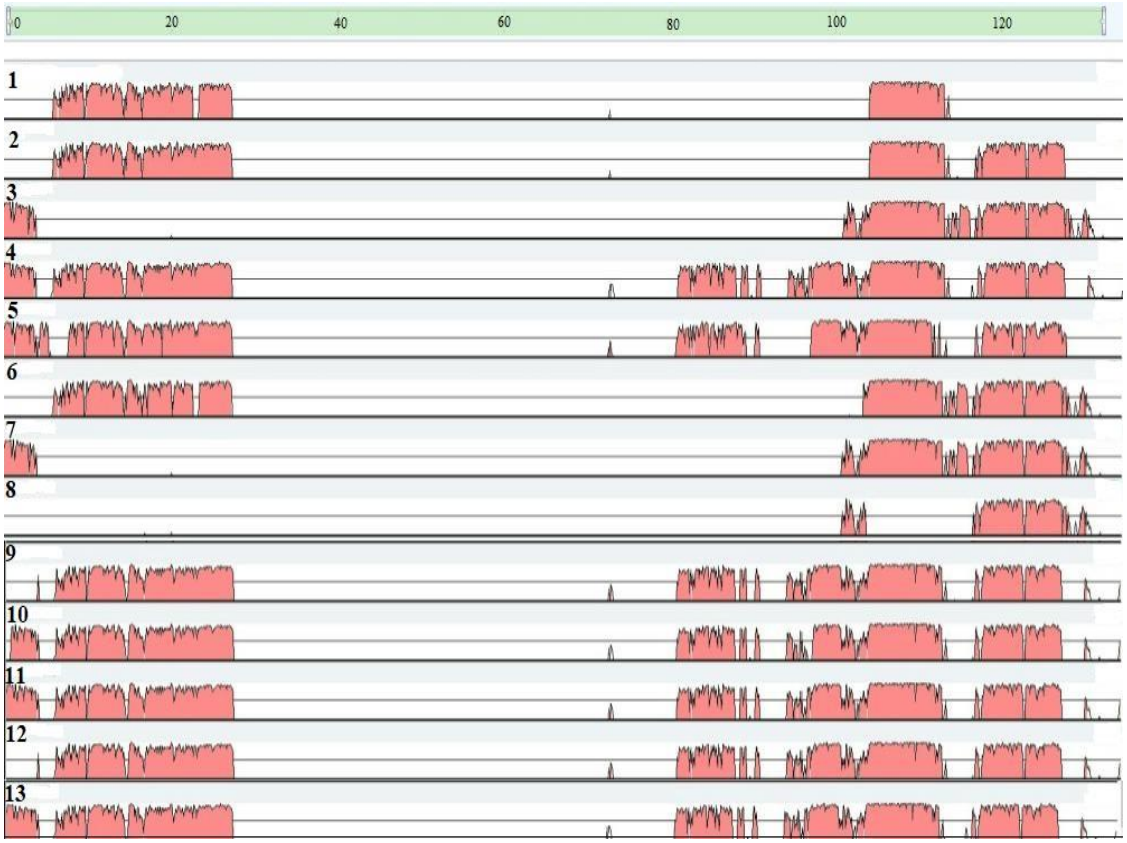
Organizacja genomu w tych wzdłuż tych miejsc ma doniosłe znaczenie może bowiem wskazywać np. dryf ewolucyjny w analizowanych liniach. Obszary te są korygowane na drodze ewolucji, a ich rozrastanie się i kurczenie (obszaru granicznego IR i SSC) wpływa na zmienność wielkości genomów plastydowych (Wang i in. 2021).

Wykres wygenerowany w narzędziu IR Scope przeskalowany jest do realistycznego rozmiaru par zasad nukleotydów w obszarach połączenia, jego interpretacja zdaje się być pomocną w formułowaniu celnych hipotez dotyczących homologii między różnymi liniami, w analizie strukturalnej i porównawczej genomów. Obrazuje rozdzielone wyraźne cztery sekcje: LSC, IRb, SSC i IRa z nadanymi konkretnymi kolorami (Rycina 9).

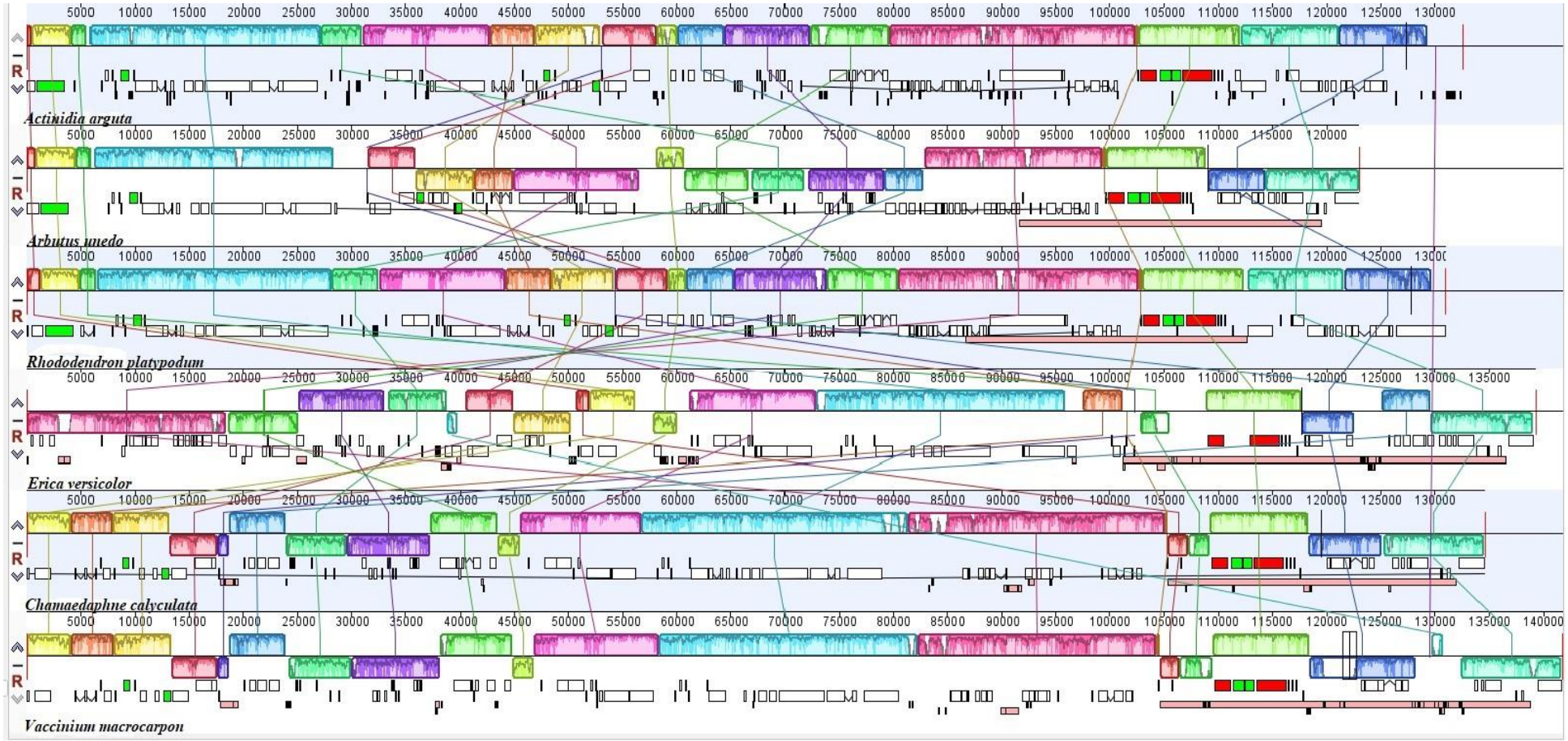

JLB
JSB
JSA
JLA
Vaccinium oldhamii
173,245 bp
LSC 108,904 bp
IRb 30,637 bp
SSC 3,067 bp
IRa 30,637 bp
LSC
rpl32
662 bp
826 bp
236 bp
psbA 1061 bp
trnK
trnH
454 bp
ndhF
101 bp
rpl32
Vaccinium bracteatum
174,404 bp
LSC 106,565 bp
IRb 32,430 bp
SSC 2,979 bp
IRa 32,430 bp
LSC
rpl32
825 bp
237 bp
psbA 1061 bp
trnV
62 bp
644 bp
ndhF
92 bp
rpl32
Vaccinium uliginosum
173,356 bp
LSC 105,856 bp
IRb 32,193 bp
SSC 3,114 bp
IRa 32,193 bp
LSC
rpl32
826 bp
236 bp
psbA 1061 bp
trnK
trnV
61 bp
668 bp
ndhF
203 bp
rpl32
Vaccinium vitis-idaea
173,967 bp
LSC 106,013 bp
IRb 32,218 bp
SSC 3,518 bp
IRa 32,218 bp
LSC
rpl32
669 bp
826 bp
236 bp
psbA 1061 bp
trnK
psbA 830 bp
825 bp
ndhF
117 bp
rpl32
Rhododendron delavayi
202,169 bp
LSC 105,623 bp
IRb 46,968 bp
SSC 2,610 bp
IRa 46,968 bp
LSC
1022 bp
rps12
trnR
rpl32
386 bp
549 bp
psbA 1062 bp
ndhF
312 bp
54 bp
386 bp
rpl32
trnR
70 bp
460 bp


Chamaedaphne calyculata
176,744 bp
Erica arborea
170,390 bp
Erica versicolor
174,587 bp
Vaccinium macrocarpon
176,045 bp
Arbutus unedo
150,897 bp
Pyrola rotundifolia
168,995 bp
Camellia azalea
157,039 bp
LSC
IRb
SSC
IRa
rpl32
psbA
trnH
ndhF
trnT
rps4
trnV
rpl23
trnL
ccsA
rpl22
rps19
rpl2
ycf1

Rycina 9. Odległość między sąsiednimi genami a połączeniami dużej pojedynczej kopii (LSC), małej pojedynczej kopii (SSC) i dwóch odwróconych powtórzeń (IR) wśród genomów cp z dwunastu gatunków z rodziny *Ericaceae*. Każdy gatunek i odpowiadająca mu długość sekwencji genomu chloroplastów są przedstawione po lewej stronie każdej ścieżki. Geny transkrybowane na nici dodatnie i ujemne są przedstawione powyżej i poniżej odpowiadających im ścieżek odpowiednio w kierunkach od prawej do lewej i od lewej do prawej. Strzałki pokazują odległość pz początkowej lub końcowej współrzędnej danego genu od odpowiedniego miejsca połączenia. W przypadku genów rozciągających się z jednego regionu do drugiego, pasek T na górze lub poniżej pokazuje zasięg ich części wraz z odpowiadającymi im wartościami. Geny w pobliżu skrzyżowań to realistycznie wyskalowane projekcje odległości pz dla każdego miejsca.



Rycina 10. Porównanie sekwencji regionów odwróconych IR trzynastu gatunków z rodziny *Ericaceae*. Graficzna informacja o podobieństwie oparta na mVISTA przedstawia identyczność sekwencji IR z *Actinidia arguta* jako regionem odniesienia. Dla wykresów zastosowano punkt odcięcia 50% identyczności. Na każdym wykresie oś w skali Y reprezentuje identyczność procentową (50 do 100%). Przerywane przestrzenie wskazują bardzo rozbieżne regiony w porównaniu z regionem odniesienia. Cyfry odpowiadają kolejno następującym gatunkom: 1 – *Erica versicolor*, 2 – *Erica arborea*, 3 – *Arbutus unedo*, 4 – *Chamaedaphne calyculata*, 5 – *Pyrola rotundifolia*, 6 – *Rhododendron delavai*, 7 – *Rhododendron platypodum*, 8 – *Rhododendron pulchrum*, 9 – *Vaccinium bracteatum*, 10 – *Vaccinium macrocarpon*, 11 – *Vaccinium oldhamii*, 12 – *Vaccinium uliginosum*, 13 – *Vaccinium vitis-idaea*.



Rycina 11. Syntenie i rearanżacje wykryte w kompletnych regionach odwróconych powtórzeń genomów chloroplastowych sześciu gatunków: przedstawicieli z rodziny *Ericaceae* oraz *Actinidia arguta*. Jednokolorowe pola połączone liniami wskazują lokalne bloki współliniowe i reprezentują regiony synteniczne. Profile podobieństwa tożsamości sekwencji są reprezentowane przez histogramy w każdym bloku, a sekwencje transkrybowane w odwrotnych kierunkach są oznaczone kolorowymi blokami powyżej i poniżej linii środkowej odwróconych powtórzeń badanych gatunków.

Porównano szczegółowo strukturę graniczną 12 genomów chloroplastowych z rodziny *Ericaceae* (Rycina 9). Regiony IR *Vaccinium bracteatum* i *Vaccinium vitis-idaea* zawierają w obszarze IRA gen *rpl*32 i na granicy IRA/LSC zawierają część genu *psb*A. Wyróżniające w analizie porównawczej jest, iż *Vaccinium vitis-idaea* zawiera dwie kopie genu *psb*A: Jeden na granicy IRA/LSC, a drugi w regionie IRB. Dwie kopie tego genu posiada także *Chamaedaphne calyculata*, z tą różnicą, że u tego gatunku jest on zlokalizowany w regionie LSC po obu stronach granicznych. W rodzaju *Vaccinium* zlokalizowano gen *ndh*F w regionie SSC, od 92 do 203 pb od granicy z IRA. Gen *rpl*32 przebywający w IRB, 616-669 bp dalej od granicy SSC/IRA. U *Vaccinium uliginosum* i *Vaccinium vitis-idaea* gen *trn*V-UAC o wielkości 38 pz zlokalizowany był w region LSC, podczas gdy *Vaccinium macrocarpon* posiada wariant 661 pz, a *Vaccinium bracteatum* nie posiada w ogóle tego genu.

Genomy z rodzaju *Erica* wykazują podobieństwo do współanalizowanych gatunków pod względem występowania genu *rpl*32 w regionach odwróconych powtórzeń. Kwestią indywidualną jest występowanie w LSC na granicu JBL genu *rps*4, zarówno u *Erica arborea*, jak i *Erica versicolor* oraz genu *trn*T w regionach IR. Ogólna struktura dla tych dwóch gatunków jest zbliżona. Różnicą są niuanse np. w odległości od granic czy długości genów. W przypadku chociażby regionu SSC odległość genu *ndh*F od granic wynosiła odpowiednio do granicy JSB 303 pz dla *Erica arborea* i 315 pz dla *Erica versicolor* oraz do granicy JSA 17 pz dla *Erica arborea* i 19 pz dla *Erica versicolor*.

Gatunek wybrany jako swego rodzaju referencja dla porównania tych regionów, czyli *Camellia azalea* wykazał największe aberracje i odstępstwa w budowie oraz składzie genowym struktury granicznej. Część genu *rps*19 położona jest w regionie LSC, część przechodzi na region IRb. Regiony IR tego gatunku posiadają geny *rpl*2 oraz *ycf*1, przy czym w przypadku regionu IRa część genu *ycf*1 położona jest w regionie SSC. Obecność genu *trn*H w regionie LSC na granicy z IRa (JLA) ukazuje częściowe podobieństwo do aranżacji genów w tej strukturze dla gatunków *Arbutus unedo* oraz *Pyrola rotundifolia*, które ów gen mają umiejscowiony w regionach IR granicznie do LSC, w odległości odpowiednio 217 pz (*Pyrola rotundifolia*) i 236 pz (*Arbutus unedo*)

od granicy z LSC. Jednakże zlokalizowany jest on w regionie LSC, bliżej granicy z IRb. Gen *trn*H w strukturze granicznej wystąpił także u *Vaccinium oldhamii,* 454 pz od JLB.

Ponadto w regionach IR tych dwóch gatunków występuje gen *trn*L, ze strony granicy z SSC. Gen *rpl*32 nie występuje u nich, jak u wszystkich analizowanych w obszarze IR. U *Arbutus unedo* gen ten został transportowany w obszar SSC (odległość 147 pz od JSB), natomiast u *Pyrola rotundifolia* w obrębie struktury granicznej nie występuje wcale. Posiada on natomiast jako jedyny z dwunastu, gen *ccs*A ulokowany w IRb i przechodzący w region SSC, oraz w regionie IRa w odległości 22 pz od SSC na granicy JSA. Unikalną cechą tych gatunków jest także występowanie genu *rpl*23 w regionie LSC na granicy z IRb, u *Arbutus unedo* dosłownie na styku (zerowa odległość od granicy), u *Pyrola rotundifolia* odległość 55 pz od JLB. Jedynym wspólnym i stałym komponenetem struktury granicznej dla wszystkich gatunków poddanych tej analizie jest gen *ndh*F zlokalizowany w regionie SSC, który zmienia jedynie swoją wielkość oraz odległość od granic z IR. Pozostałe ulegają rozmaitym rearanżacjom i translokacjom. Struktura *Rhododendron delavai* jest najbliższa ogólnej aranżacji w rodzaju *Vaccinium* . Pozostałe struktury zachowują specyfikę rodzajową i wykazują charakterystyczną architektonikę tych obszarów.

Analiza wyrównania Mauve 2.3.1. została przeprowadzona na wybranych reprezentantach: *Erica versicolor, Chamaedaphne calyculata, Vaccinium macrocarpon, Rhododendron platypodum, Arbutus unedo* oraz – będąca referencją przy sporządzaniu drzewa ML – *Actinidia arguta*. Wykazano wyraźne przegrupowania międzygatunowe z nieznacznym stopniem podobieństwa w poszczególnych regionach. Największy model podobieństwa zaobserwowano pomiędzy gatunkami *Vaccinium macrocarpon* a *Chamaedaphne calyculata*. Odkryto, że kolejność genów plastomów *Chamaedaphne* jest współliniowa z *Vaccinium,* ale nie jest współliniowa z kolejnością *Rhododendron*. Kolejność genów plastomów z pięciu rodzajów z rodziny wrzosowatych jest współliniowa, jednak rzędy można wyjaśnić przez więcej niż dziesięć inwersji. Widoczne rearanżacje i rekombinacje w obrębie poddanych wyrównaniu sekwencji stanowią o wysokiej odrębności rodzaju *Erica* na tle innych analizowanych gatunków, gdzie zanotowano wystąpienie unikalnych segmentów specyficznych dla tego rodzaju (Rycina 11).

Procesy ewolucyjne widoczne są tu w postaci rekombinacji segmentów: część jest wprowadzana poprzez transfer poziomy, niektóre fragmenty są usuwane poprzez delecje. Sama kolejność poszczególnych segmentów jest tasowana pomiędzy gatunkami. Globalna struktura przegrupowań wskazuje, iż największe podobieństwo bloków syntenicznych występuje w finalnych sekwencjach regionów odwróconych powtórzeń, co koreluje z wynikiem analizy przeprowadzonej za pomocą programu mVISTA.

Analiza w programie mVISTA została przeprowadzona na większej grupie osobników: *Erica versicolor, Erica arborea, Arbutus unedo, Chamaedaphne calyculata, Pyrola rotundifolia, Rhododendron delavai, Rhododendron platypodum, Rhododendron pulchrum, Vaccinium bracteatum, Vaccinium macrocarpon, Vaccinium oldhamii, Vaccinium uliginosum* oraz *Vaccinium vitis-idaea*.

Ukazała ona ogólną identyczność sekwencji i rozbieżne regiony w sekwencjach IR wymienionych gatunków. Stosunkowo niski stopień zachowania syntenii i kolejności genów wskazuje na ewolucyjną dynamiczność na poziomie plastomu (Rycina 10). Poziom podobieństwa analizowanych sekwencji jest stosunkowo niski, nie wykazuje nigdzie pokrycia wyższego niż 70%. Najwyższy stopień odrębności ujawnia się u gatunku *Rhododendron pulchrum*.

Największa dywergencja obserwowana jest w pobliżu pozycji ok. 90 000 pz, w okolicy sekwencji obejmujących regiony IR oraz SSC, co wiąże się z faktem, iż region IR-SSC uznany jest nierzadko za region hipermutacyjny i porównywane pomiędzy gatunkami granice IRb-SSC wykazują wielokrotnie większe różnice niż region IRa-LSC.

4.2.3. Analiza porównawcza genomów *Erica*

Celem zbadania zróżnicowania pomiędzy genomami chloroplastowymi *Erica versicolor* i *Erica arborea*, ich sekwencje zostały porównane za pomocą programu Mauve Genome 2.3.1., Geneious Alignment v 9 oraz mVISTA (Ryciny 12-14). Subsydialnie przeprowadzono wizualizację metodą graficzną dot plot (Rycina 16). Powstałe niewyrównane regiony są często albo powtarzalnymi, albo regionami specyficznymi dla linii, uzyskanymi poprzez transfer poziomy lub w inny sposób. Pomocą w klasyfikacji niewyrównanych regionów obok szczegółowej analizy sekwencji powtórzeniowych jest zestawienie polimorfizmów w sekwencjach kodujących i niekodujących. Wykresy tożsamości oparte na mVISTA oraz Mauve ujawniły rozbieżności sekwencji między genomami. Rzeczywiście, różnicowanie nukleotydów dotyczy głównie regionów międzygenowych / niekodujących, gdzie między innymi genomy wykazały duże obszary delecji (ponad 500 pz; poniżej 50% identyczności).

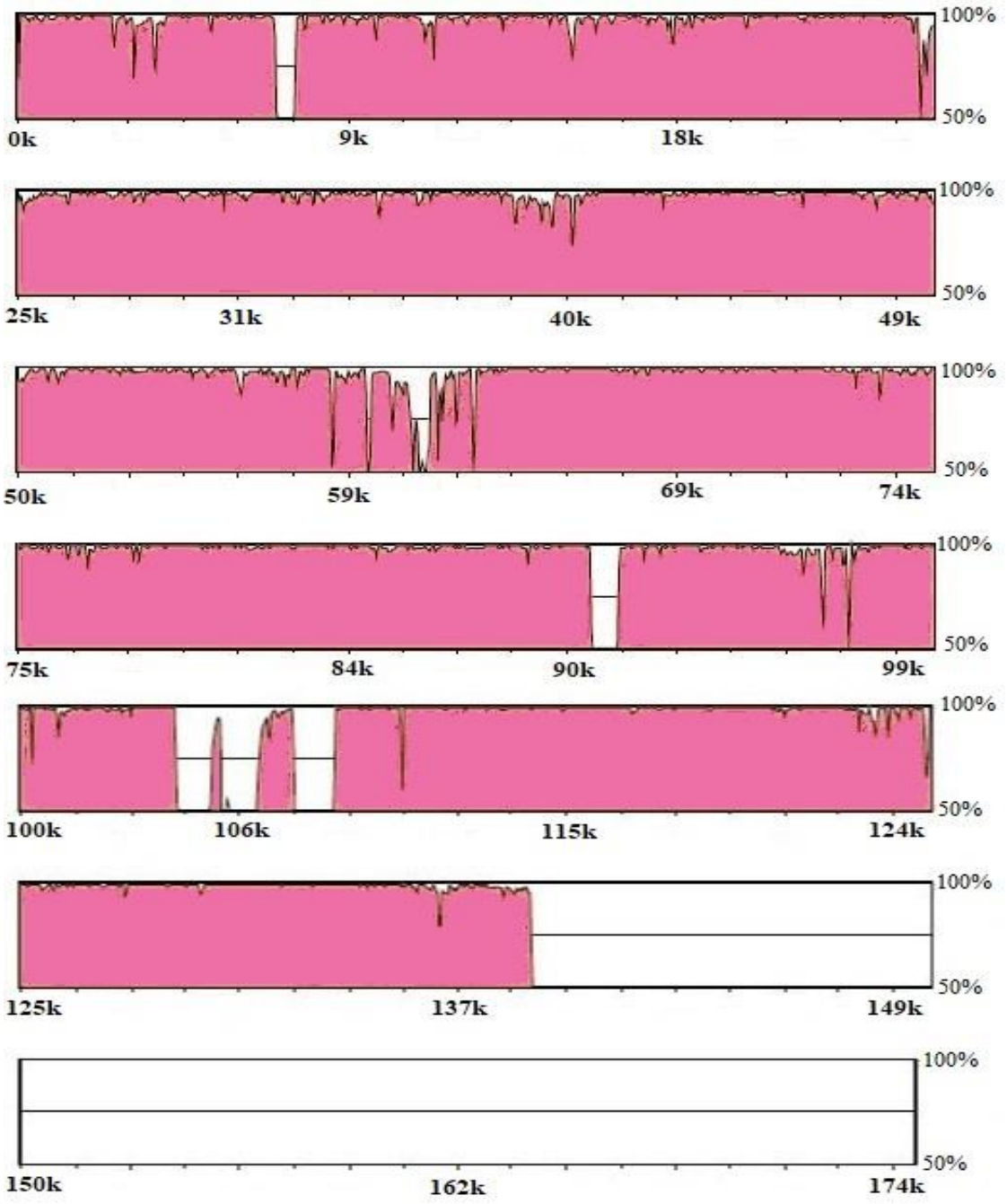
W zestawionych kompletnych genomach chloroplastowych z rodzaju *Erica* zidentyfikowano 9 160 miejsc zmiennych (5,25%), w tym miejsc informatywnych 9 157 (5,23%) w całym genomie chloroplastowym. Wśród tych regionów IR wykazał różnorodność nukleotydów na poziomie 0,11962, a SSC wykazał najniższą rozbieżność (0,00209) (tabela Suplementu 16).

Genom chloroplastowy *Erica versicolor* jest dłuższy od *Erica arborea* o 4 197 pz; rejon LSC jest dłuższy o 4 224 pz, rejony SSC są niemal identyczne w długości (odpowiednio 2 571 pz i 2 570 pz), natomiast rejony odwróconych powtórzeń są dłuższe u *Erica arborea* o 14 pz każdy (Tabela 11).

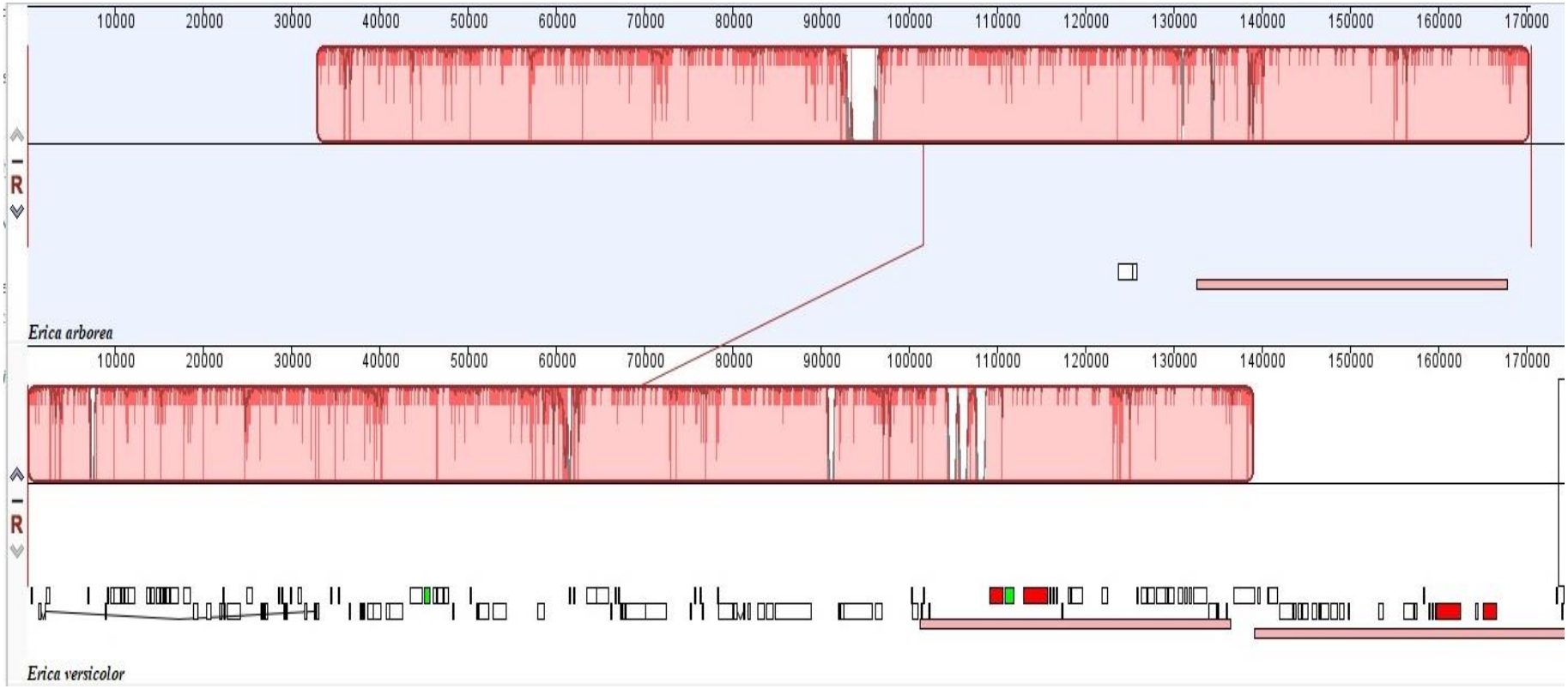
Porównując genomy, zgodnie z zestawieniem wykonanym za pomocą Geneious Alignment, zidentyfikowano 131,517 miejsc identycznych pomiędzy sekwencjami analizowanych genomów o poziomie identyczności pairwise identity 91,6%.

Frekwencja nukleotydów po wyrównaniu genomów rozkłada się następująco: A (32,4%), T (32%), C (17,8%) i G (17,8%). Natomiast średnia zawartość GC 98,552 pz (35,6%).

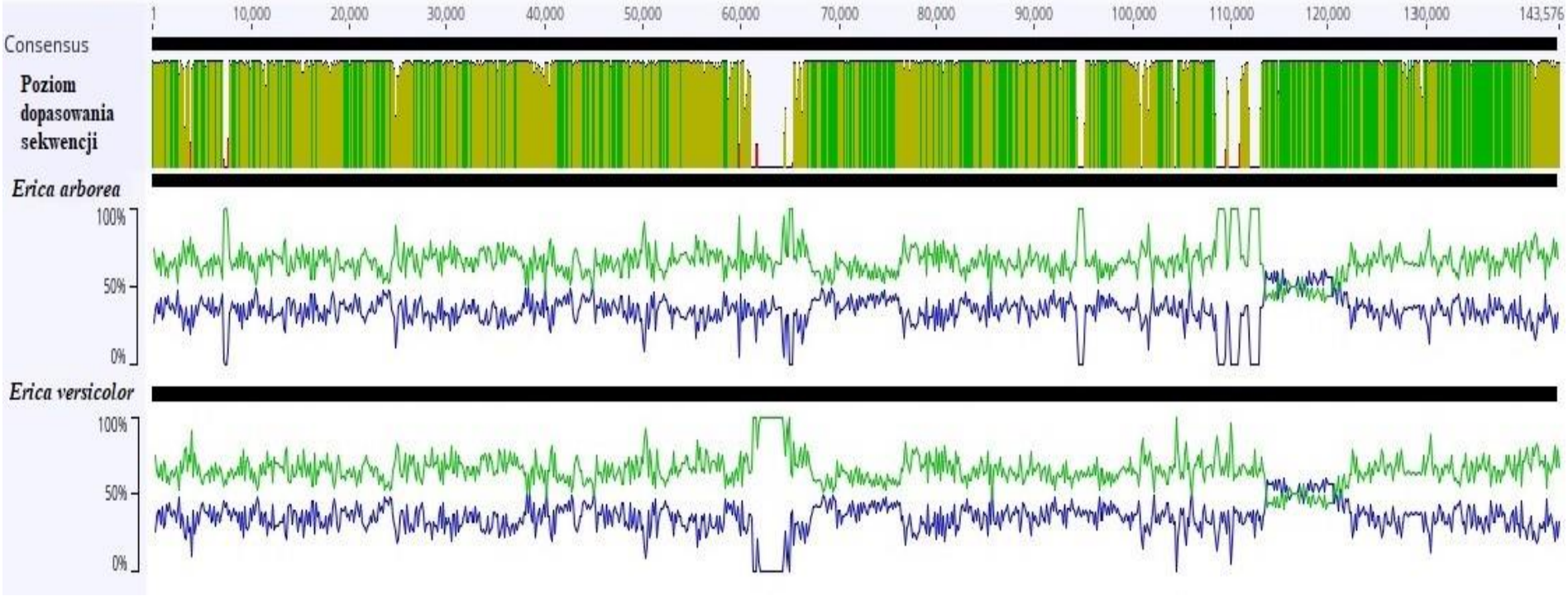
Po zestawieniu porównawczym kompletnych genomów *Erica* względem siebie, wyłączając drugi region odwróconego powtórzenia IRB, łącznie zidentyfikowano 5 570 miejsc polimorficznych. W regionach kodujących zidentyfikowano 454 miejsc polimorficznych, w tym 49 insercji/delecji, 233 mutacje synonimiczne i 172 mutacje niesynonimiczne. Stosunek mutacji synonimicznych do niesynonimicznych (S/N) wynosi 1,35. W regionach niekodujących zostało zidentyfikowanych 5116 miejsc polimorficznych, w tym 3556 insercji/delecji oraz 960 mutacji typu SNP. Średnia odległość genetyczna dla całych, porównywanych genomów wynosi p=0,01385 (model trójparametryczny Tamury).



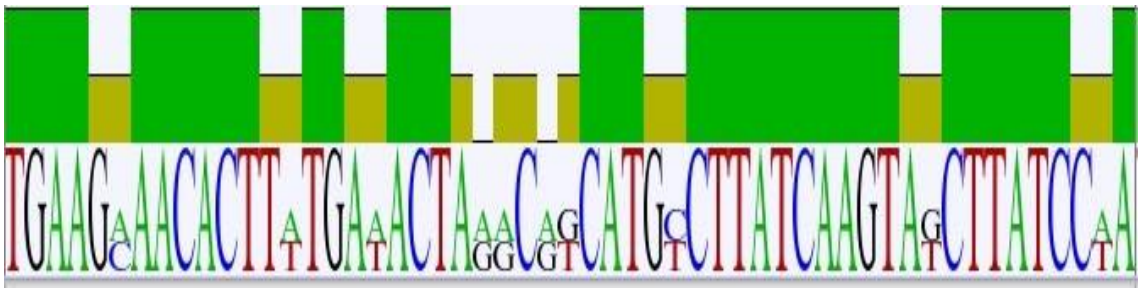
Rycina 12. Porównanie sekwencji genomów chloroplastowych z rodzaju *Erica*. Graficzna informacja o podobieństwie oparta na mVISTA przedstawia poziom identyczności sekwencji plastomu Erica versicolor z Erica arborea jako odniesieniem. Dla wykresów zastosowano punkt odcięcia 70% identyczności.Oś Y reprezentuje procent identyczności w zakresie 50%-100% Przerywane przestrzenie wskazują rozbieżne regiony w zestawieniu z regionem odniesienia.



Rycina 13. Porównanie genomu chloroplastowego *Erica arborea* z genomem chloroplastowym *Erica versicolor*. Dane uzyskane w programie Mauve 2.3.1. Jednokolorowe pola połączone liniami wskazują lokalne bloki współliniowe i reprezentują regiony synteniczne. Widoczny profil podobieństwa tożsamości sekwencji reprezentowany jest przez histogramy wewnątrz bloków.



Rycina 14. Porównanie sekwencji genomów chloroplastowych *Erica arborea* i *Erica versicolor*. Kolor intensywnie zielony oznacza 100% poziom identyczności, kolor zielonkawo-brunatny to podobieństwo niższe niż 100%, ale powyżej 30%, czerwony przypisany jest dla poziomu podobieństwa poniżej 30%. Histogramy poniżej schematu dopasowania sekwencji to zawartość procentowa AT (kolor zielony) oraz GC (kolor niebieski) w poszczególnych genomach.



Rycina 15. Przykładowe logo sekwencji konsensusowej; wykonane dla fragmentu genu *rpl*22 (11 243-11 295 pz). Kolor intensywnie zielony odpowiada dopasowaniu sekwencji genu między genomami *Erica arborea* a *Erica versicolor* na poziomie 100%. Kolor zielonkawo-brunatny to miejsca wystąpienia mutacji.

Wśród regionów kodujących odległość genetyczna wykazuje rozpiętość od p = 0,000 (identyczność 100%) do p = 0,21795 dla genu *rpl*22 (Rycina 15).

Nie wykryto zmienności genów białkowych *psb*I, *psb*J, *psb*L, *psb*M, *psb*T, *psa*C, *psa*I, *psa*J. Sekwencje te odpowiadają za kodowanie podjednostek fotosystemu I oraz II, pełniona przez nie funkcja sprawia, że reprezentują charakter silnie konserwatywny. Pozostałe geny należące do rodzin *psa* i *psb* wykazują także niski stopień polimorfizmu (Tabela 7). Ponadto stuprocentowe podobieństwo sekwencji pomiędzy *Erica arborea* a *Erica vesricolor* wykazały także geny z rodziny białek rybosomalnych (RP): *rps*7, *rps*15, *rps*16, *rps*18 oraz *rpl*20. Są one niezbędne w syntezie białek, odgrywając kluczową rolę w różnorodnych procesach rozwojowych, jak biogeneza rybosomów, synteza białek, apoptoza, wzrost, czy chociażby reakcja na stres. Plastomy roślinne najczęściej kodują 12 białek dla małych podjednostek rybosomów (geny *rps*) i dziewięć dużych białek podjednostek rybosomów (geny *rpl*). (Saha i in. 2017).

Poza tym identycznymi genami były także geny kodujące składniki kompleksu tylakoidów NDH, działającym na zasadzie zasilania elektronów w celu dostosowania poziomu redoks cyklicznych fotosyntetycznych transporterów elektronów, kodujące podjednostki dehydrogenazy NADH: *ndh*E, *ndh*G, *ndh*I, *ndh*J oraz geny kodujące

podjednostki kompleksu cytochromów b_6f pełniące główną rolę w fotosyntetycznym łańcuchu transportu elektronów: *pet*A, *pet*G oraz *pet*N.

Największą odległość genetyczną zidentyfikowano dla genu *rpl*22 (p = 0,21795), należącym do wspomnianej w powyższym akapicie rodzinie genów kodujących duże białka podjednostek rybosomów, o długości 459 pz, w którym wykryto 21 SNP, w tym 3 mutacje niesynonimiczne. W sekwencjach regionów kodujących znane białka, drugim najbardziej odmiennym jest *rpo*C2 (p = 0,14040) o długości 4169 pz u *Erica arborea* i 4164 pz u *Erica versicolor*. Zostało w nim zidentyfikowanych 43 SNP, w tym 15 mutacji niesynonimicznych. Ponadto dużą różnicę zaobserwowano w genie z tej samej rodziny genów kodujących podjednostki plastydowej polimerazy RNA (PEP), w genie *rpo*C1 (p = 0,13339) o długości sekwencji 2060 pz u *Erica arborea* i 2058 pz u *Erica versicolor*. Zidentyfikowano w nim 11 SNP, w tym 2 mutacje niesynonimiczne. Inne geny z tej rodziny (*rpo*A i *rpo*B) wydają się być mniej polimorficzne, aczkolwiek liczba SNP w *rpo*B wyniosła 12 przy 8 mutacjach niesynonimicznych (p = 0,05905). Geny *rpo* są często propowanymi sekwencjami do analiz filogenetycznych i stosowane są do tworzenia wiarygodnych barkodów dla wielu plastomów roślinnych (Zhang i in. 2020). Genami, w których zanotowano także dużą różnicę były m.in. geny kodujące podjednostki dehydrogenazy NADH: *ndh*A (p = 0,13089) i *ndh*B (p = 0,10137); sekwencje prawdopodobnie kodujące białka z rodziny *ycf* : *ycf*1 (p = 0,05787) i *ycf*2 (p = 0,02130), które są rejonami również sugerowanymi w badaniach filogenetycznych roślin; oraz gen *rpl*23 (p = 0,08865).

Spośród regionów międzygenowych o długości przekraczającej 100 pz najbardziej podobne o stuprocentowym stopniu identyczności (p = 0, 0000) były rejony: *psb*H-*psb*N (101 pz w obu genomach), *pet*L-*pet*G (171 pz w obu genomach), *pet*G-*trn*W (121 pz w obu genomach), *trn*W-*trn*P (164 pz w obu genomach), *ndh*K-*ndh*J (103 pz w obu genomach), *psb*A-*trn*H (212 pz w obu genomach), *ndh*G-*ndh*E (225 pz w obu genomach), *ndh*E-*psa*C (246 pz w obu genomach), *psa*C-*ndh*D (164 pz w obu genomach) oraz rejon *ndh*D-*ccs*A (240 pz w obu genomach). Do rejonów najbardziej odrębnych wśród regionów międzygenowych należą odpowiednio: *clp*P-*ycf*3 (p = 0,69801), *ndh*I-*psa*I (p = 0,45007), *rps*16-*rrn*16 (p = 0,43884) oraz *rpl*32-*ndh*F (p = 0,27802). Dwa rejony najbardziej polimorficzne wśród sekwencji niekodujących są jednocześnie rejonami

najbardziej odrębnymi w całym genomie. Rejon *clp*P-*ycf*3 ma długość 1046 pz u obu genomów, wykryto w nim 144 idele oraz 34 SNP. Rejon *ndh*I-*psa*I osiągnął długosc 3329 pz u *Erica arborea* i 3354 pz u *Erica versicolor*, zidentyfikowano w nim 50 SNP i 77 ideli. Sekwencje niekodujące w plastomach *Erica* wykazują kilka okazałych obszarów indeli, które powodują przesunięcia pomiędzy genomami, zawierają ponad trzykrotnie więcej mutacji SNP niż sekwencje kodujące i wykazują dużo wyższy poziom polimorfizmu niż sekwencje kodujące. Długość sekwencji kodujących w przeważającej większości są konserwatywne pomiędzy genomami *Erica arborea* a *Erica versicolor*. Zmiany długości odnotowano w 7 sekwencjach CDS (*ndh*A, *ndh*F, *rpo*C1, *rpo*C2, *clp*P, *ycf*1,*rps*11), które wahają się od 2 pz (*rpo*C1) do 83 pz (*ycf*1), wartości pośrednie różnicy długości sekwencji 5 pz, 6 pz i 7 pz należą odpowiednio do *ndh*A, *rpo*C2, *ndh*F, *rps*11 i *clp*P. Najwyższą wartość zróżnicowania nukleotydowego w regionach kodujących osiągnął gen *rpo*C1 (π = 0,3502). Natomiast najwyższą wartość zróżnicowania nukleotydowego w regionach niekodujących wykazał region *rps*16-*rrn*16 (π = 0,6918). Średnia zróżnicowania nukleotydowego dla regionów kodujących wyniosła π = 0,015, z kolei dla regionów niekodujących π = 0,067.

Średnia wartość dystansu genetycznego dla grupy białek aparatu fotosyntetycznego wynosi p = 0,00845 (s = 0,32941). Wartości wahają się tu od p = 0,0000 (*psb*I, *psb*J, *psb*L, *psb*M, *psb*T, *psa*C, *psa*I, *psa*J, *ndh*E, *ndh*G, *ndh*I, *ndh*J, *pet*A, *pet*G oraz *pet*N) do p = 0,13089 (*ndh*A). Najwyższą średnią wartość p = 0,03142 (s = 0,84858) wykazują geny związane z ekspresją genów. Najbardziej podobnymi rejonami są tu *rps*7, *rps*15, *rps*16, *rps*18 oraz *rpl*20 (p = 0,000), natomiast najbardziej odmiennym *rpl*22 (p = 0,21795). Średnia wartość odległości genetycznej dla genów metabolizmu fotosyntetycznego wynosi p = 0,00398 (s = 0,02792) i jest najniższa spośród wszystkich grup genów kodujących białka. Wartości mieszczą się w przedziale od p = 0,00131 (*atp*A) do p = 0,00819 (*atp*H).

Klasyfikacja genów została dokonana wg Changa i in. (2006). Szczegółowe informacje dotyczące liczby i rodzajów mutacji punktowych w genach białkowych zostały zawarte w Tabeli 7, natomiast dane dla regionów niekodujących zestawiono w Tabeli 8.

Tabela 7. Zestawienie polimorfizmów w sekwencjach kodujących pomiędzy badanymi genomami *Erica arborea* a *Erica versicolor*. Długość sekwencji w genomie (pz), liczba mutacji oraz insercji w genach i rejonach kodujących (CDS) u *Erica arborea* względem *Erica versicolor*. SNP – liczba substytucji, InDel – liczba insercji/delecji w CDS, π – wartości zróżnicowania nukleotydowego, p – odległość genetyczna, NS – mutacje niesynonimiczne, kategorie genu wg Changa i in. (2006): GE – geny ekspresji genów, GA – geny aparatu fotosyntetycznego, GM – geny metabolizmu fotosyntetycznego, R – różne. Kolejność genów zgodna z wynikiem porównania otrzymanego za pomocą Geneious Alignment v 9. Wykluczono z tabeli geny identyczne pomiędzy genomami oraz drugie powtórzenie odwrócone IRB.

Nazwa genu	**Kat. genu**	***Erica arborea* [pz]**	***Erica versicolor* [pz]**	**Liczba SNP**	**Liczba NS**	**Liczba InDel**	**π**	**p**
*rps*12	**GE**	785	785	6	6	0	0,0089	0,00898
*ndh*B	**GA**	2217	2217	8	7	1	0,0112	0,10137
*ycf*2	**R**	192	192	4	1	0	0,0208	0,02130
*rpl*23	**GE**	300	300	2	0	1	0,0833	0,08865
*rpl*2	**GE**	804	804	3	2	0	0,0037	0,00375
*rps*19	**GE**	303	303	1	0	0	0,0066	0,00664
*rpl*22	**GE**	459	459	21	3	1	0,1873	0,21795
*rps*3	**GE**	714	714	8	5	0	0,0112	0,01134
*rpl*16	**GE**	360	360	1	0	0	0,0027	0,00279
*rpl*14	**GE**	384	384	3	1	0	0,0078	0,00786
*rps*8	**GE**	399	399	6	2	0	0,0150	0,01520
*inf*A	**GE**	246	246	13	8	0	0,0569	0,05965
*rpl*36	**GE**	123	123	1	0	0	0,0081	0,00825
*rps*11	**GE**	423	417	5	1	6	0,0283	0,08746
*pet*D	**GA**	525	525	4	3	0	0,0076	0,00766
*pet*B	**GA**	489	489	1	1	0	0,0020	0,00205

Tabela 7. Ciąg dalszy.

Nazwa genu	Kat. genu	*Erica arborea* [pz]	*Erica versicolor* [pz]	Liczba SNP	Liczba NS	Liczba InDel	π	p
*psb*H	GA	222	222	2	2	0	0,0090	0,00909
*psb*N	GA	132	132	1	1	0	0,0075	0,00769
*psb*B	GA	1527	1527	4	4	0	0,0026	0,00262
*ycf*1	R	541	624	40	13	0	0,0587	0,05787
*psb*F	GA	120	120	2	2	0	0,0166	0,01681
*psb*E	GA	282	282	1	0	0	0,0035	0,00357
*pet*L	GA	132	132	1	1	0	0,0075	0,00769
*rpl*33	GE	213	213	1	0	0	0,0046	0,00471
*rps*18	GE	348	348	3	0	0	0,0086	0,00866
*clp*P	R	817	810	47	5	13	0,0941	0,06143
*atp*E	GM	402	402	3	1	0	0,00746	0,00751
*atp*B	GM	1497	1497	3	1	0	0,0020	0,00201
*rbc*L	GM	1428	1428	4	4	0	0,0028	0,00281
*ndh*C	GA	363	363	1	1	0	0,0027	0,00276
*ndh*K	GA	678	678	1	1	0	0,0014	0,00148
*psb*A	GA	1062	1062	5	5	0	0,0047	0,00473
*mat*K	R	1521	1521	15	8	0	0,0098	0,00993
*acc*D	R	687	687	3	2	0	0,0058	0,00585
*psb*D	GA	1062	1062	4	4	0	0,0037	0,00371
*psb*C	GA	1404	1404	1	1	0	0,0007	0,00071
*psb*Z	GA	189	189	1	0	0	0,0052	0,00531

Tabela 7. Ciąg dalszy.

Nazwa genu	Kat. genu	*Erica arborea* [pz]	*Erica versicolor* [pz]	Liczba SNP	Liczba NS	Liczba InDel	π	p
*rps*14	GE	303	303	1	0	0	0,0033	0,00332
*psa*B	GA	2205	2205	9	8	0	0,0040	0,00409
*psa*A	GA	2253	2253	5	5	0	0,0022	0,00222
*psb*K	GA	186	186	2	1	0	0,0107	0,01082
*atp*A	GM	1527	1527	2	2	0	0,0013	0,00131
*atp*F	GM	1271	1271	6	2	0	0,0047	0,00474
*atp*H	GM	246	246	2	2	0	0,0081	0,00819
*atp*I	GM	744	744	1	1	0	0,0013	0,00135
*rps*2	GE	708	708	3	2	0	0,0042	0,00426
*rpo*C2	GE	4169	4164	43	15	5	0,0115	0,14040
*rpo*C1	GE	2060	2058	11	2	2	0,3502	0,13339
*rpo*A	GE	375	375	2	0	8	0,008	0,00804
*rpo*B	GE	3210	3210	12	8	0	0,0059	0,05905
*ycf*3	GA	507	507	2	1	0	0,0019	0,00198
*rps*4	GE	627	627	2	1	0	0,0031	0,00320
*ndh*F	GA	2232	2238	44	18	6	0,0232	0,02096
*ycf*15	R	138	138	3	0	0	0,0217	0,02211
*ndh*H	GA	1182	1182	1	1	0	0,0016	0,00170
*ndh*A	GA	2197	2192	10	4	6	0,0054	0,13089
*ycf*4	GA	555	555	1	1	0	0,0018	0,00108
*cem*A	R	690	690	3	1	0	0,0043	0,00436
*ndh*D	GA	1467	1467	2	0	0	0,0013	0,00136

Tabela 7. Ciąg dalszy.

Nazwa genu	**Kat. genu**	***Erica arborea* [pz]**	***Erica versicolor* [pz]**	**Liczba SNP**	**Liczba NS**	**Liczba InDel**	**π**	**p**
*ccs*A	**R**	963	963	4	1	0	0,0041	0,00416
*rpl*32	**GE**	216	216	4	0	0	0,0185	0,01872

Tabela 8. Zestawienie polimorfizmów w sekwencjach niekodujących pomiędzy genomem chloroplastowym *Erica arborea* a genomem chloroplastowym *Erica versicolor*. Długość sekwencji w genomie [pz], liczba mutacji oraz insercji w sekwencjach międzygenowych u *Erica arborea* względem *Erica versicolor*. SNP – liczba substytucji, InDel – liczba insercji/delecji, p – odległość genetyczna, π – wartości zróżnicowania nukleotydowego. Wykluczono z tabeli sekwencje identyczne pomiędzy genomami oraz drugie powtórzenie odwrócone IRB.

Region w genomie	***Erica arborea* [pz]**	***Erica versicolor* [pz]**	**Liczba SNP**	**Liczba InDel**	**π**	**Odległość genetyczna (p)**
*rps*7-*ndh*B	1494	1 463	55	226	0,1	0,06147
*ndh*B-*trn*L	580	580	4	0	0,0069	0,00693
*trn*L-*ycf*2	1339	1862	9	531	0,3197	0,04982
*ycf*2-*trn*I	101	101	1	0	0,0099	0,00996
*trn*I-*rpl*23	353	350	6	1	0,0310	0,01743
*rps*19-*rpl*22	50	50	1	0	0,02	0,02024
*rps*3-*rpl*16	1409	1412	10	7	0,0162	0,00716
*rpl*14- *rps*8	194	195	4	1	0,0410	0,03711
*rps*8-*inf*A	131	139	10	12	0,1888	0,09276
*rpl*36-*rps*11	124	124	3	0	0,0242	0,02545
*rps*11-*clp*P	1540	1540	21	1	0,0162	0,01248

Tabela 8. Ciąg dalszy.

Region w genomie	*Erica arborea* [pz]	*Erica versicolor* [pz]	Liczba SNP	Liczba InDel	π	Odległość genetyczna (p)
*clp*P- *pet*D	389	388	5	1	0,0128	0,01039
*pet*D- *pet*B	931	937	3	6	0,0096	0,00323
*pet*B-*psb*H	1056	1054	6	2	0,0075	0,00572
*psb*T-*psb*B	183	183	2	0	0,0109	0,01107
*psb*B-*ycf*1	637	695	28	93	0,1676	0,05172
*ycf*1-*psb*J	985	992	12	6	0,0191	0,01230
*psb*E-*pet*L	1239	1232	15	7	0,0185	0,01311
*trn*P-*psa*J	411	409	8	2	0,0243	0,01984
*rpl*33-*rps*18	173	165	2	8	0,0578	0,01229
*rps*18-*rpl*20	294	295	5	5	0,0336	0,01743
*rpl*20-*trn*C	2604	2623	26	20	0,0156	0,00773
*trn*C-*pet*N	817	821	13	6	0,0243	0,01986
*pet*N-*psb*M	1146	1147	11	8	0,0173	0,00970
*psb*M-*trn*D	1137	1137	9	0	0,0079	0,00796
trnE-*ycf*1	460	457	8	15	0,0602	0,03405
*clp*P-*atp*E	600	624	9	38	0,0683	0,01531
*atp*B-*rbc*L	780	774	10	8	0,0230	0,01308
*trn*V-*ndh*C	380	382	3	2	0,0130	0,00794
*ndh*C-*ndh*K	52	59	1	8	0,1355	0,02032
*ndh*J-*trn*F	536	536	7	0	0,0130	0,01318
*trn*F-*trn*L	363	371	6	9	0,0377	0,01674
*trn*L-*trn*T	840	838	15	8	0,0272	0,01821
*trn*T-*trn*H	665	665	11	0	0,0165	0,01673
*trn*H-*psb*A	260	264	4	5	0,0339	0,01167
*mat*K-*acc*D	3521	3544	62	45	0,0298	0,01760
*acc*D-*trn*T	3061	2341	117	609	0,3918	0,09383
*trn*T- *psb*D	4234	2863	40	775	0,5179	0,18238
*trn*S-*psb*Z	349	349	4	0	0,0114	0,01160

Tabela 8. Ciąg dalszy.

Region w genomie	***Erica arborea* [pz]**	***Erica versicolor* [pz]**	**Liczba SNP**	**Liczba InDel**	**π**	**Odległość genetyczna (p)**
*psb*Z-*trn*G	256	255	3	1	0,0156	0,01187
*trn*G-*trnf*M	145	144	1	1	0,0206	0,01412
*rps*14-*psa*B	155	155	1	0	0,0064	0,00648
*psa*A-*trn*Q	2762	2783	36	23	0,0222	0,01354
*trn*Q- *psb*K	380	380	3	0	0,0078	0,00798
*psb*K-*psb*I	394	395	4	1	0,0126	0,01022
*psb*I-*trn*S	139	140	6	1	0,05	0,04455
*trn*S-*trn*G	664	676	12	13	0,0355	0,01832
*trn*G-*trn*R	237	241	4	4	0,0332	0,01710
*trn*R-*atp*A	116	115	6	1	0,0689	0,06492
*atp*A-*atp*F	104	104	2	0	0,0192	0,01963
*atp*F-*atp*H	377	377	6	2	0,0901	0,09669
*atp*H-*atp*I	894	894	3	0	0,0033	0,00337
*atp*I-*rps*2	210	210	1	0	0,0047	0,00478
*rps*2-*rpo*C2	236	236	2	0	0,0084	0,00852
*rpo*C2-*rpo*C1	182	174	3	8	0,0604	0,01746
*rps*18- *rpo*A	103	110	1	7	0,0636	0,00321
*rpo*A-*rpo*B	66	61	1	6	0,0901	0,01658
*rpo*B-*trn*M	202	202	6	0	0,0297	0,03050

Tabela 8. Ciąg dalszy.

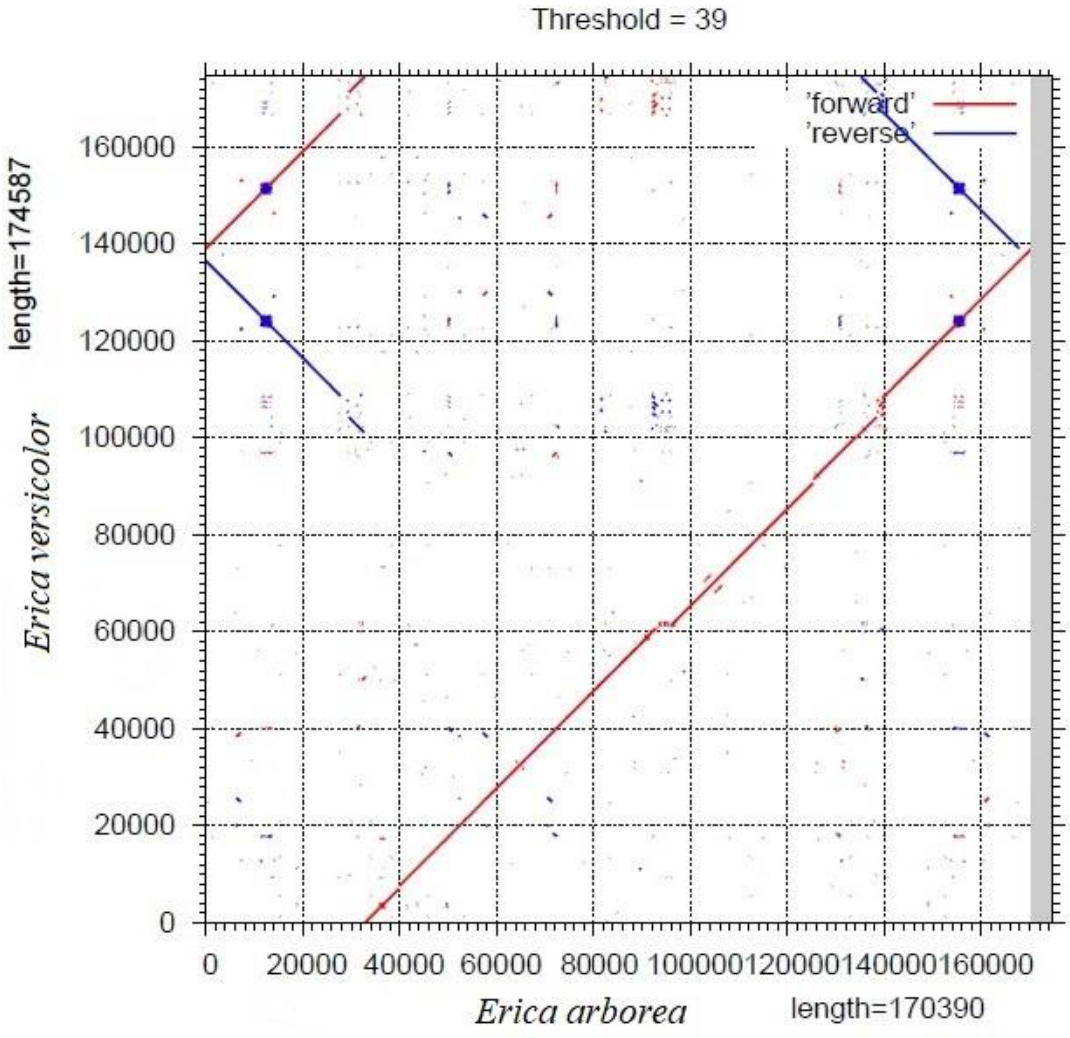
Region w genomie	***Erica arborea* [pz]**	***Erica versicolor* [pz]**	**Liczba SNP**	**Liczba InDel**	**π**	**Odległość genetyczna (p)**
*trn*M-*clp*P	63	63	3	0	0,0317	0,03296
*clp*P-*ycf*3	1046	1046	34	144	0,5717	0,69801
*ycf*3-*trn*S	334	335	3	1	0,0059	0,00300
*trn*S-*rps*4	282	237	3	48	0,2447	0,98486
*rps*4-*trn*T	356	356	9	13	0,0794	0,03255
*trn*T-*trn*V	162	170	7	8	0,0823	0,03801
*trn*V- *trn*T	537	537	2	0	0,0037	0,00373
*trn*T- *rps*16	686	687	8	1	0,0160	0,01472
*rps*16-*rrn*16	2309	5009	110	700	0,6918	0,43884
*trn*A-*rrn*23	152	152	1	0	0,0065	0,00660
*trn*R-*trn*N	607	606	4	1	0,0082	0,00663
*trn*N-*rps*15	659	659	3	0	0,0045	0,00457
*ndh*I-*psa*I	3329	3354	50	77	0,0376	0,45007
*psa*I-*ycf*4	427	428	4	1	0,0011	0,00942
*ycf*4-*cem*A	221	221	1	0	0,0045	0,00454
*cem*A-*pet*A	233	238	1	5	0,0252	0,00430
*pet*A-*ndh*G	1472	1473	8	1	0,0061	0,00546
*trn*L-*rpl*32	841	841	5	0	0,0059	0,00597
*rpl*32-*ndh*F	741	752	20	23	0,0567	0,27802

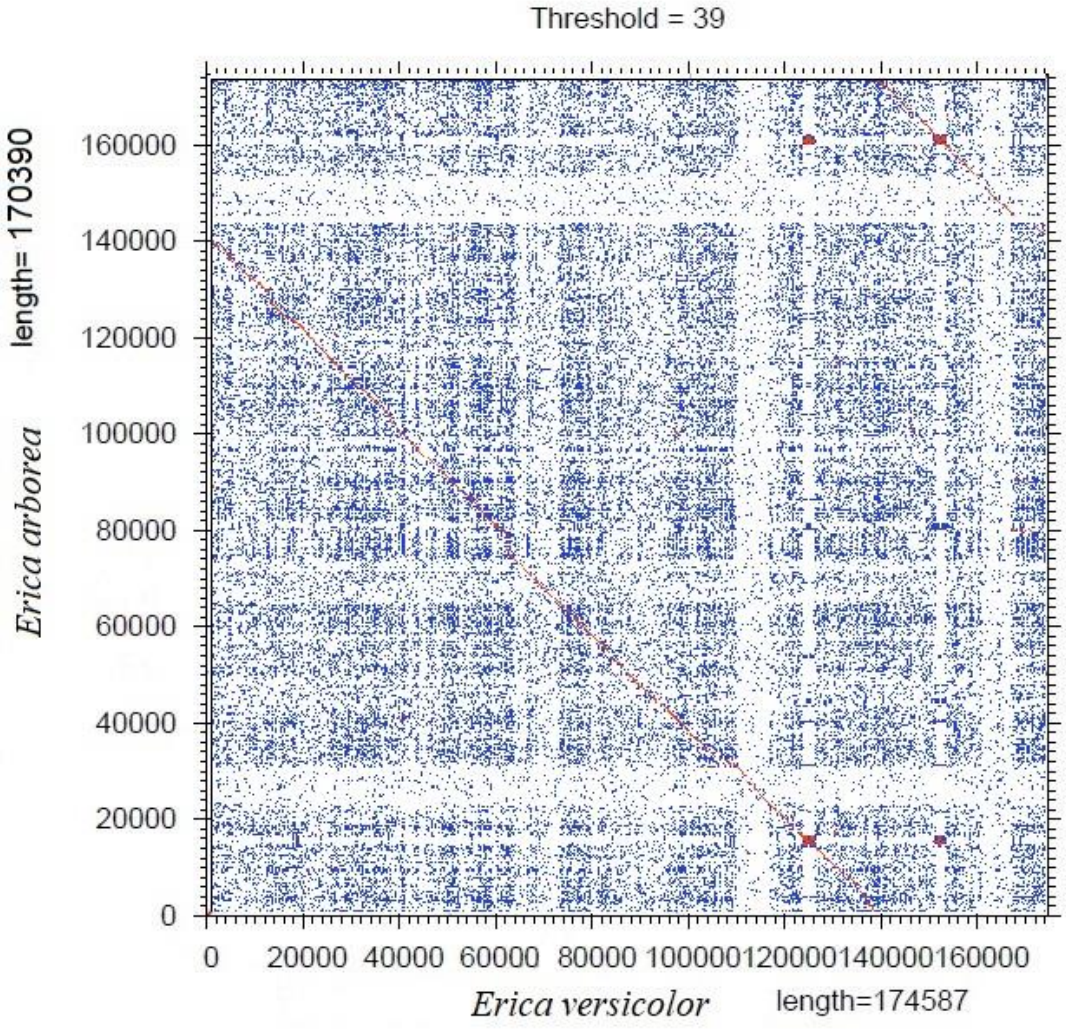
Sekwencje tRNA w badanych genomach chloroplastowych z rodzaju *Erica* w większości wykazują silny stopień konserwatywności. 20 genów nie zawiera żadnych polimorfizmów, z kolei w 8 sekwencjach zostały zidentyfikowane polimorfizmy wpływające na dojrzałą cząsteczkę tRNA (tabela Suplementu 15).

Najbardziej odrębną sekwencją (p = 0,08051) odpowiadającą dojrzałej cząsteczce tRNA okazał się gen *trnL-UAA*, w którym odnotowano 5 mutacji SNP. Długość tego rejonu w obu genomach wyniosła odpowiedno 75 pz. Ponadto dużą zmienną wykazały także *trnT-UGU (p* =0,05769) zawierający 4 SNP o długości sekwencji 73 pz oraz *trnT-GGU* (p = 0,02009) zawierający 4 SNP i 2 indele. Długość tego rejonu wyniosła 73 pz u *Erica arborea* i 71 pz u *Erica vesricolor*. Sekwencje pozostałych genów tRNA, które wykazały zmienność posiadały od 1 do 2 SNP. Pozostałe sekwencje tRNA: *trnL-CAA, trnW-CCA, trnP-UGG, trnC-GCA, trnD-GUC, trnY-GUA, trnE-UUC, trnF-GAA, trnH-GUG, trnK-UUU, trnS-UGA, trnG-GCC, trnQ-UUG, trnS-GCU, trnS-GGA, trnR-UCU, trnV-GAC, trnI-GAU, trnN-GUU, trnL-UAG* były identyczne.

Geny rRNA podobnie jak tRNA wykazały silny stopień konserwatywności sekwencji. Rejony *rrn*5 oraz *rrn*4.5 nie wykazały żadnych polimorfizmów w obrębie badanych genomów i posiadają stuprocentową identyczność sekwencji. W sekwencji genu *rrn*16 (p = 0,03223) zanotowano 3 SNP i jeden idel, długość tej sekwencji wyniosła odpowiednio 1493 pz u *Erica arborea* oraz 1492 pz u *Erica versicolor*. W sekwencji *rrn*23 (p = 0,00179) odnotowano 5 SNP, jego długość w obu analizowanych genomach wyniosła 2803 pz.

Szczegółowe informacje dotyczące liczby i rodzajów mutacji punktowych w genach tRNA oraz rRNA zostały zawarte w tabeli Suplementu 15.


Threshold = 39
'forward'
'reverse'
length=174587
Erica versicolor
160000
140000
120000
100000
80000
60000
40000
20000
0
0
20000
40000
60000
80000
100000
120000
140000
160000
Erica arborea
length=170390



Rycina 16. Graficzne zestawienie metodą dot plot dwóch kompletnych sekwencji plastomów *Erica arborea* i *Erica versicolor* celem identyfikacji regionów bliskiego podobieństwa między nimi.

Macierz podobieństwa przedstawia wizualizację podobieństwa pomiędzy dwiema sekwencjami nukleotydowymi kompletnych genomów chloroplastowych *Erica arborea* i *Erica versicolor* (Rycina 16). W jednym z grafów oś X obszaru prostokątnego reprezentuje *Erica arborea*, w drugim układ jest zamienny. Analizując te ryciny należy mieć na względzie fakt, iż przy długich sekwencjach wydajność tej metody wizualizacji jest znacząco niższa, niż w przypadku sekwencji krótkich, dodatkowo sporym ograniczeniem jest swoisty szum, tj. gromadzenie licznych nieistotnych dopasowań poza przekątną. Ponadto metoda ta ma głównie na celu unaocznić potencjalne zestawenia pomiędzy analizowanymi sekwencjami, zidentyfikować regiony o wysokim stopniu podobieństwa, odnaleźć transpozycje w obrębie sekwencji, lokalizację podobnych genów oraz ewentualnego niesekwencyjnego alignmentu. Kropki wykreślane na poszczególnych rezyduach i rzędach wykazują podobieństwa pomiędzy bazami. Główna czerwona przekątna ukazuje układ sekwencji identycznych. Minimalne przesunięcia ramy obrazują zmiany w obszarze sekwencji, zachodzące w wyniku insercji, delecji bądź mutacji. Niebieska linia symbolizuje powtórzenia odwrócone. Wysoki stopień podobieństwa sekwencji obrazuje długa ukośna linia przebiegająca przez macierz, natomiast występowanie przesunięć i mutacji w obrębie analizowanych sekwencji unaocznia się przez przeskoki i przesunięcia tej linii, co pokrywa się z wynikami otrzymanymi z użyciem innych narzędzi (Ryciny 12-14).

4.2.4. Porównanie genomów chloroplastowych z rodzaju *Erica* względem plastomów fotosyntetyzujących przedstawicieli rodziny *Ericaceae*

Ogólny poziom konserwatywności kompletnych sekwencji analizowanych genomów chloroplastowych *Erica arborea* i *Erica versicolor* wygenerowany w programie MEGA11 plasował się na poziomie 63%, co dało 132 158 pz o wysokim pokryciu. Poddając podobnemu zestawieniu genomy z rodzaju *Erica* wraz z zestawem kompletnych genomów fotosyntetyzujących z rodziny *Ericaceae* poziom konserwatywności osiągnął poziom 34% z 143 277 miejscami konserwatywnymi (Tabela 9). Obliczono odległość genetyczną (p) pomiędzy badanymi taksonami w programie MEGA11 z użyciem parametrów modelu o najniższej wartości BIC (Tamura i in. 2021).

W celu wizualizacji zmienności pomiędzy fotosyntetyzującymi przedstawicielami rodziny *Ericaceae* skonstruowano także konsensusowe drzewo filogenetyczne metodą bayesowską z podaniem wartości otrzymanych w analizie ML (Rycina 17).

Dokładny rozkład odległości między parami, odległości genetycznej (p) między taksonami; w plastomach fotosyntetyzujących rodziny *Ericaceae* przedstawiono w Tabeli 10 , gdzie rozrzut tych wartości mieści się w zakresie zmienności od p = 0.0087 dla pary gatunków z rodzaju *Vaccinium*: między *Vaccinium bracteatum* a *Vaccinium oldhamii* (najniższy poziom zmienności) do p = 0,8838 dla pary gatunków z rodzaju *Rhododendron*: między *Rhododendron platypodum* a *Rhododendron pulchrum* (najwyższy poziom zmienności). Oszacowana odległość genetyczna między samymi przedstawicielami z rodzaju *Erica: Erica versicolor* i *Erica arborea* wyniósł 0.0139. Wartości zróżnicowania nukleotydowego (π i θ) oraz test TajmaD dla połączonych sekwencji cpDNA wygenerowano w programie TASSEL 5 (Bradbury i in. 2007). Wartości π wyniosły odpowiednio 0,3733 dla zestawu danych z rodzaju *Erica* oraz 0,29514 dla zestawu danych plastomów fotosyntetyzujących z rodziny *Ericaceae*.

Tabela 9. Charakterystyka wyrównanych sekwencji genomowych w odniesieniu do obu kompletnych sekwencji z rodzaju *Erica* oraz wszystkich kompletnych sekwencji plastomów fotosyntetyzujących przedstawicieli rodziny *Ericaceae*

	Erica versicolor* x *Erica arborea	**Plastomy fotosyntetyzujących *Ericaceae***
Długość wyrównania [pz]	210 979	421 599
Długość sekwencji [pz]	170 390 – 174 587	136 249 – 202 169
Pokrycie [100%]	133 997	20 276
Procent konserwatywny CpG	34 363	581
Znaki zmienności (V)	9 160 (5,23 %)	113 351 (27 %)
pi (π) diversity na pz	0,3733	0,29514
theta (θ) diversity na pz	0,37312	0,31577
Tajma D test	0,03221	-0,30247
Znaki informatywności parsymonii (Pi)	9 157 (5,25 %)	52 932 (12 %)
Miejsca pojedyncze (S)	-	54 441 (13 %)
Odległość między parami	0.0139	0.0087 – 0,8838
Miejsca konserwatywne (C)	132 158 (63 %)	143 277 (34 %)

Tabela 9. Ciąg dalszy.

	Erica versicolor* x *Erica arborea	**Plastomy fotosyntetyzujących *Ericaceae***
Miejsca zdegenerowane 0-krotnie	137 175	262 235
Miejsca zdegenerowane 2-krotnie	46 994	72 668
Miejsca zdegenerowane 4-krotnie	24 466	36 802

Tabela 10. Odległość genetyczna (p) między badanymi gatunkami w plastomach fotosyntetyzujących *Ericaceae*

	Erica versicolor	*Erica arborea*	*Chamaedaphne calyculata*	*Vaccinium vitis-idaea*	*Vaccinium bracteatum*
Erica versicolor					
Erica arborea	0,0139				
Chamaedaphne calyculata	0,2602	0,2918			
Vaccinium vitis-idaea	0,2529	0,2880	0,0401		
Vaccinium bracteatum	0,2532	0,2881	0,0413	0,0107	
Vaccinium uliginosum	0,2526	0,2875	0,0401	0,0088	0,0113
Vaccinium oldhamii	0,2566	0,2968	0,0404	0,0095	0,0087
Vaccinium macrocarpon	0,2615	0,2986	0,0398	0,0146	0,0163
Rhododendron delavayi	0,1874	0,2098	0,2515	0,2562	0,2561
Arbutus unedo	0,1881	0,1766	0,2815	0,2678	0,2684
Pyrola rotundifolia	0,4456	0,5207	0,2920	0,3113	0,3120
Rhododendron pulchrum	0,2085	0,2063	0,3587	0,3544	0,3539
Rhododendron platypodum	0,4171	0,5520	0,3417	0,3620	0,3625

Tabela 10. Ciąg dalszy.

	Vaccinium uliginosum	***Vaccinium oldhamii***	***Vaccinium macrocarpon***	***Rhododendron delavayi***	***Arbutus unedo***
Erica versicolor					
Erica arborea					
Chamaedaphne calyculata					
Vaccinium vitis-idaea					
Vaccinium bracteatum					
Vaccinium uliginosum					
Vaccinium oldhamii	0,0093				
Vaccinium macrocarpon	0,0151	0,0157			
Rhododendron delavayi	0,2560	0,2582	0,2701		
Arbutus unedo	0,2680	0,2647	0,2540	0,1949	
Pyrola rotundifolia	0,3112	0,3005	0,3021	0,4443	0,2610
Rhododendron pulchrum	0,3543	0,3661	0,3652	0,2305	0,1155
Rhododendron platypodum	0,3618	0,3481	0,3666	0,3006	0,4066

Tabela 10. Ciąg dalszy.

	Pyrola rotundifoliaa	*Rhododendron pulchrum*
Erica versicolor		
Erica arborea		
Chamaedaphne calyculata		
Vaccinium vitis-idaea		
Vaccinium bracteatum		
Vaccinium uliginosum		
Vaccinium oldhamii		
Vaccinium macrocarpon		
Rhododendron delavayi		
Arbutus unedo		
Pyrola rotundifolia		
Rhododendron pulchrum	0,5967	
Rhododendron	0,3809	0,8838

Dokonano zestawienia dotyczącego szczegółów budowy genomów chloroplastowych fotosyntetyzujących przedstawicieli z rodziny *Ericaceae* wraz z reprezentantami z rodzaju *Erica*.

Porównano wielkości genomów, długości regionów LSC, SSC oraz IR, łączną liczbę genów, liczbę genów kodujących białka, liczbę genów kodujących tRNA i rRNA oraz procentową zawartość GC dla 13 wybranych reprezentantów z rodziny *Ericaceae*. Wielkości genomów chloroplastowych wahają się od 136 249 pz dla *Rhododendron pulchrum* do 201 047 dla *Rhododendron platypodum*. Największy region LSC należy do *Rhododendron delavayi* (160 234 pz), najkrótszy natomiast region LSC posiada *Arbutus unedo* (93 609 pz). Najdłuższy region SSC posiada *Pyrola rotundifolia* (11 945 pz), najkrótszy *Chamaedaphne calyculata* (2 249 pz). Wszystkie gatunki zawierają w budowie regiony odwróconych powtórzeń poza *Rhododendron pulchrum*. Poziom GC w badanych genomach chloroplastowych rozkłada się w podobnych wartościach od 35,7% u *Pyrola rotundifolia* do 37,3% u *Arbutus unedo*. Łączna liczba genów u badanych osobników waha się od 73 (*Rhododendron pulchrum*) do 147 (*Vaccinium macrocarpon*).

Dane z analizowanych plastomów zaprezentowano w Tabeli 11.

Tabela 11. Porównanie szczegółów budowy genomów chloroplastowych *Erica arborea*, *Erica vesricolor* oraz plastomów fotosyntetyzujących przedstawicieli rodziny *Ericaceae*

Nazwa gatunku	***Erica arborea***	***Erica versicolor***	***Chamaedaphne calyculata***	***Arbutus unedo***	***Pyrola rotundifolia***
Nr akcesyjny GenBank	MW338736	MW282955	KJ463365	JQ067650	KU833271
Wielkość genomu [pz]	170 390	174 587	176 744	150 897	168 995
Długość regionu LSC [pz]	97 240	101 464	106 673	93 609	109 174
Długość regionu SSC [pz]	2 570	2 581	2 249	3 400	11 945
Długość regionu IR [pz]	35 290	35 276	33 911	26 944	23 937
Łączna liczba genów	115	112	113	114	101

Tabela 11. Ciąg dalszy.

Nazwa gatunku	***Erica arborea***	***Erica versicolor***	***Chamaedaphne calyculata***	***Arbutus unedo***	***Pyrola rotundifolia***
Liczba genów kodujących białka	82	80	79	75	67
Liczba genów kodujących rRNA	8	8	8	8	8
Liczba genów kodujących tRNA	29	28	30	30	30
Zawartość GC [%]	36,1	36,1	36,3	37,3	35,7
Nazwa gatunku	***Vaccinium vitis-idaea***	***Vaccinium uliginosum***	***Vaccinium macrocarpon***	***Vaccinium oldhamii***	***Vaccinium bracteatum***
Nr akcesyjny GenBank	LC521969	LC521968	NC019616	NC042713	LC521967
Wielkość genomu [pz]	173 967	173 356	176 045	173 245	174 404
Długość regionu LSC [pz]	106 013	105 856	104 552	108 904	106 565

Tabela 11. Ciąg dalszy.

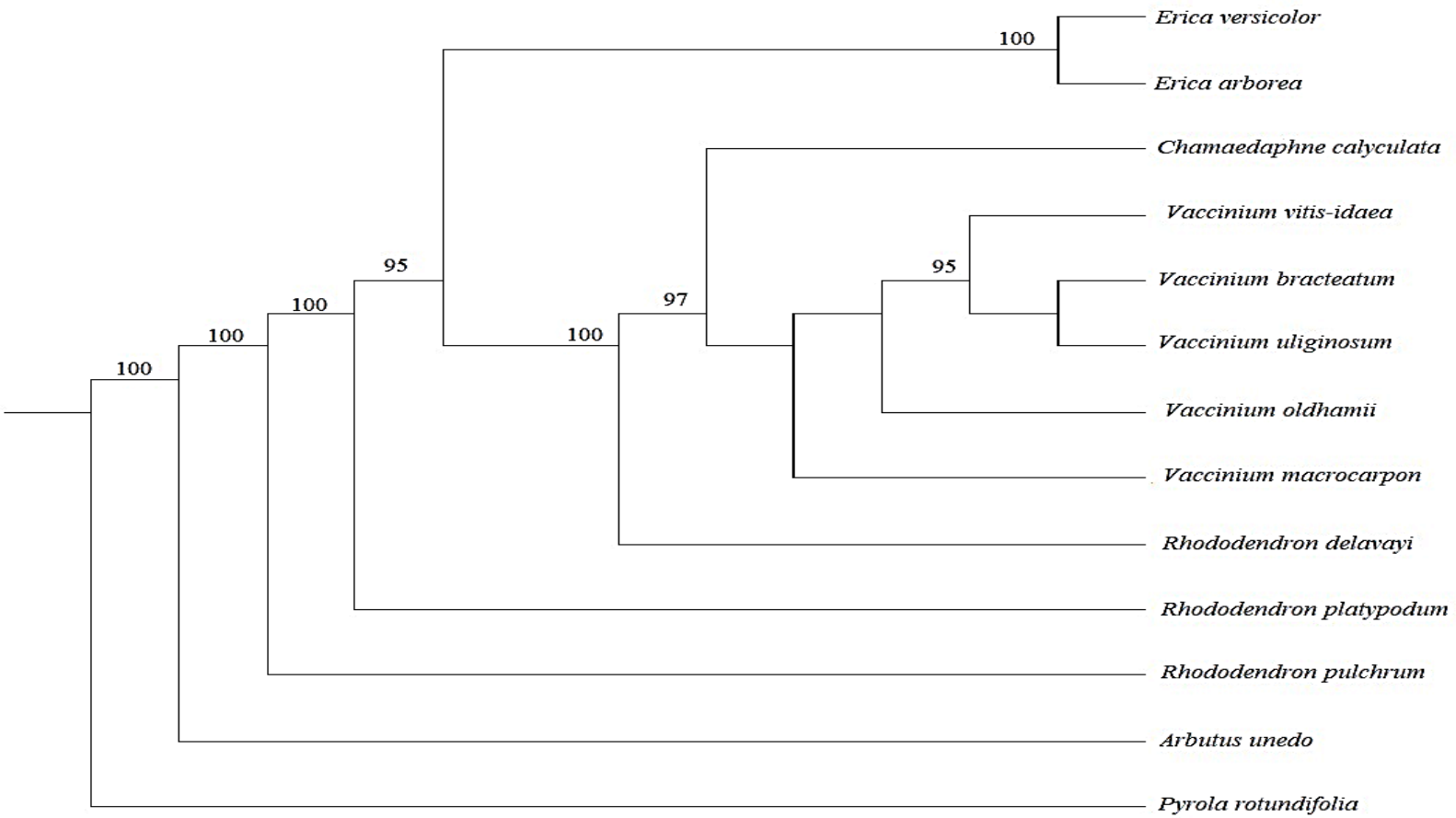
Nazwa gatunku	***Vaccinium vitis-idaea***	***Vaccinium uliginosum***	***Vaccinium macrocarpon***	***Vaccinium oldhamii***	***Vaccinium bracteatum***
Długość regionu SSC [pz]	3 518	3 114	3 009	3 067	2 979
Długość regionu IR [pz]	32 218	32 193	34 242	30 637	32 430
Łączna liczba genów	125	125	147	130	117
Liczba genów kodujących białka	79	79	75	85	79
Liczba genów kodujących rRNA	8	8	8	8	8
Liczba genów kodujących tRNA	38	38	36	37	30
Zawartość GC [%]	36,7	36,8	37,1	37,2	36,8

Tabela 11. Ciąg dalszy.

Nazwa gatunku	***Rhododendron pulchrum***	***Rhododendron delavayi***	***Rhododendron platypodum***
Nr akcesyjny GenBank	MN182619	MN711645	NC053746
Wielkość genomu [pz]	136 249	193 798	201 047
Długość regionu LSC [pz]	-	160 234	109 134
Długość regionu SSC [pz]	-	2 576	2 613
Długość regionu IR [pz]	-	15 494	44 650
Łączna liczba genów	73	123	143

Tabela 11. Ciąg dalszy.

Nazwa gatunku	***Rhododendron pulchrum***	***Rhododendron delavayi***	***Rhododendron platypodum***
Liczba genów kodujących białka	42	80	93
Liczba genów kodujących rRNA	2	8	8
Liczba genów kodujących tRNA	29	35	42
Zawartość GC [%]	35,98	35,99	35,9



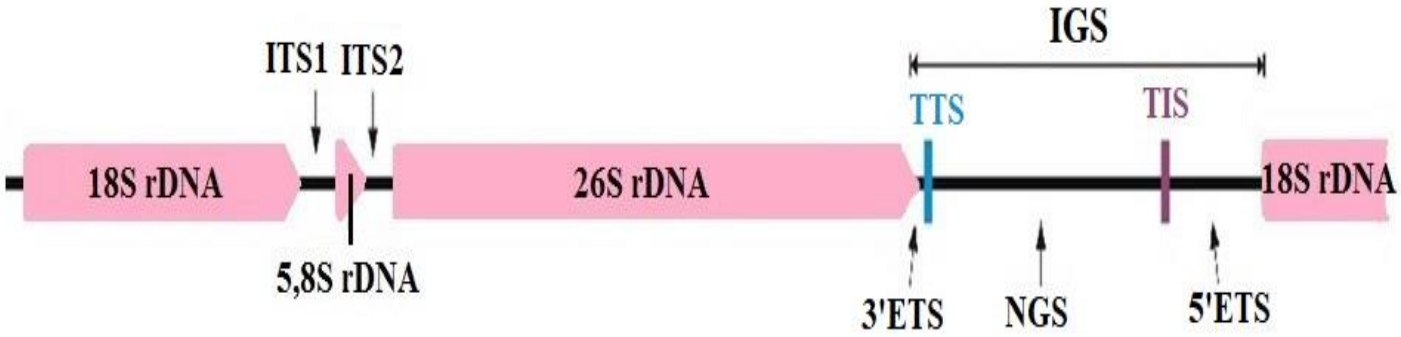
Rycina 17. Konsensusowe drzewo filogenetyczne przedstawicieli z rodziny *Ericeae* skontruowane analizą bayesowską dla kompletnych genomów chloroplastowych. Wartości bootstrap poniżej 50% są ukryte. Liczby nad gałęziami reprezentują wsparcie bootstrapowe uzyskane przez maksymalne prawdopodobieństwo (wartość ML)

4.3. Otrzymanie sekwencji NOR

Otrzymany klaster genów rRNA u analizowanych przedstawicieli z rodzaju *Erica* zawiera 7 komponentów:

- 5’ETS –
- -18S rRNA egzon –
- -ITS1 –
- -5.8S rRNA egzon –
- -ITS2 –
- -26S rRNA oraz
- IGS (Rycina 18).

Wielkość otrzymanych klasterów genów rDNA dla ośmiu reprezentantów z rodzaju *Erica*, których sekwencje zostały zdeponowane w bazie danych GenBank, przedstawiono w Tabeli 12.



Rycina 18. Schemat organizacji strukturalnej otrzymanego klastera genów rDNA rodzaju *Erica*

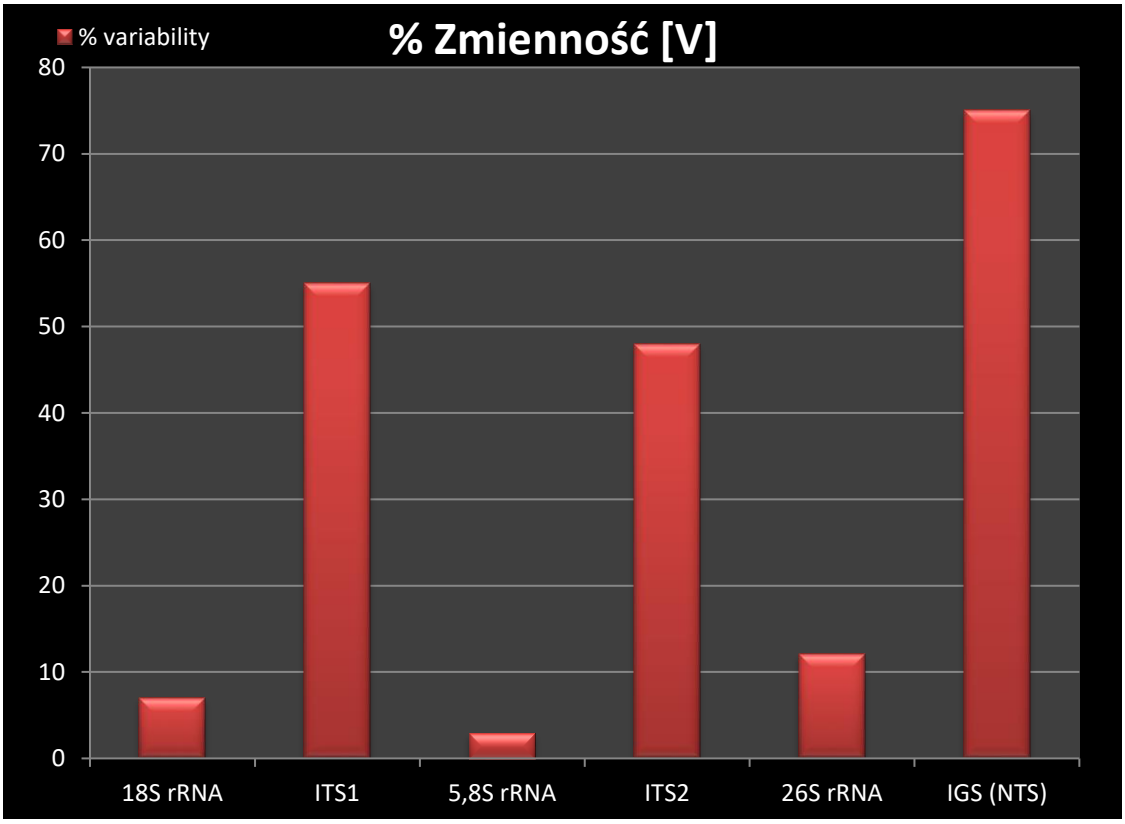
Tabela 12. Całkowita długość otrzymanych sekwencji klasterów genów rDNA wraz z nadanymi numerami akcesyjnymi bazy danych GenBank

Nazwa gatunku	**Długość sekwencji [pz]**	**Numer akcesyjny GenBank**
Erica arborea	14 456	MW132695
Erica cinerea	11 435	MW132696
Erica globiceps	11 694	MW132698
Erica imbricata	11 602	MW132702
Erica multiflora	9 580	MW132699
Erica plukenetii	11 289	MW132701
Erica vagans	8 144	MW132700
Erica versicolor	9 437	MW132697

4.3.1. Zmienność rDNA w rodzaju *Erica*

Analiza zmienności ogólnej sekwencji modelu NOR rodzaju *Erica* wykazała najniższy stopień zmienności dla całego rodzaju (najmniej miejsc zmiennych; variable sites) w genach rRNA: 5,8S rRNA, następnie w 18S rRNA i nieco wyższy, ale wciąż niski w 26S r RNA (zmienność/variability [V] na poziomie ok. 12%).

Dużo wyższa zmienność genetyczna zgromadziła się w obrębie regionów międzygenowych; w internal transcribed spacer (ITS) – więcej mutacji typu SNP zanotowano w ITS1 niż w ITS2. Najwięcej zmian zaobserwowano w obrębie intergenic spacer (IGS/NTS), gdzie poziom zmienności przekraczał 70% (Rycina 19).



Rycina 19. Procentowy rozkład zmienności w obrębie poszczególnych fragmentów funkcjonalnych regionów NOR

Analiza sekwencji obejmująca osiem gatunków z rodzaju *Erica* ujawniła niski poziom zmienności genetycznej w regionach genów rDNA (Tabela 13). Tożsamość parami w analizowanym zestawie sekwencji wyniosła 91,7% w ITS1 i 90,9% w ITS2, przy najniższym poziomie 31,2% w 3'ETS + NTS. Udział znaków zmiennych (V) i informatywności parsymonii (Pi) w ITS1 wyniósł odpowiednio tylko 21% i 8,6%. W obrębie ITS2 znaleziono 51 znaków zmiennych (V = 21%), w tym tylko 16 miejsc informatywności parsymonii (Pi = 6%). Wskazuje to, że region ITS2 ma wyższy poziom parsymonii niż region 5'ETS.

Najbardziej konserwatywnym regionem okazały się geny rDNA. W porównaniu do obu regionów ITS, IGS charakteryzował się znacznie wyższym poziomem zmienności genetycznej przy V = 58%, Pi = 25% i 31% miejsc pojedynczych /singletonowych (S). Jednak zróżnicowanie w ramach IGS nie było jednorodne, o czym świadczy porównanie 5'ETS z 3'ETS + NTS. Podczas gdy region 5'ETS był bardziej konserwatywny, na podobnym poziomie, co w przypadku ITS1. Analiza 3'ETS wraz z NTS była najbardziej zmiennym fragmentem całego rDNA. Dla pary 3'ETS + NTS tożsamość parowa wyniosła 31,2%, podczas gdy miejsca parsymonii (Pi) stanowiły 18,5% z 48,9% zmiennych znaków (V).

Tabela 13. Charakterystyka wybranych elementów funkcjonalnych i domen w ramach rDNA w rodzaju *Erica*. Jeśli w gatunku wykryto więcej niż jeden przypuszczalny TIS, rozważano opcję z krótszym 5'ETS.

Region	**Długość wyrównania [pz]**	**Długość sekwencji [pz]**	**Tożsamość par [%]**	**Znaki zmienności (*V*)**	**Znaki informatywności parsymonii (*Pi*)**	**Miejsca pojedyncze**
rDNA	5323	5320-5323	99,6%	57 (1%)	21 (0,4%)	36 (0,67%)
18S rRNA	1809	1809	99,8%	11 (0,6%)	6 (0,33%)	5 (0,27%)
5,8S rRNA	167	167	99,5%	1 (0,6%)	1 (0,6%)	-
26S rRNA	3350	3347-3350	99,4%	45 (13%)	14 (0,41%)	31 (9%)
ITS1	254	247-250	91,7%	53 (21%)	22 (8,6%)	31 (12%)

Tabela 13. Ciąg dalszy.

Region	Długość wyrównania [pz]	Długość sekwencji [pz]	Tożsamość par [%]	Znaki zmienności (*V*)	Znaki informatywności parsymonii (*Pi*)	Miejsca pojedyncze
ITS2	247	245-246	90,9%	51 (21%)	16 (6%)	35 (14%)
IGS	9417	2890-8620	44,1%	5472 (58%)	2349 (25%)	2976 (31%)
3'ETS+NTS	7262	1092-5629	31,2%	3553 (48,9%)	1349 (18,5%)	2075 (28,5%)
5'ETS	6238	41-6023	36%	1559 (25%)	272 (4,4%)	1165 (18,6%)

4.3.2. Struktura jednostek rDNA i porównania międzygatunkowe

U badanych przedstawicieli rodzaju *Erica* jednostki powtórzeniowe nrDNA konsekwentnie reprezentowały ten sam schemat (Rycina 18). Pojedynczy klaster zawiera geny kodujące rDNA: 18S, 5.8S i 26S oraz regiony niekodujące, takie jak: IGS (Intergenic Spacer Region; międygenowy region dystansowy), ITS 1 oraz ITS 2 (Internal Transcribed Spacer 1 i 2; wewnętrzny transkrybowany region dystansowy). Przekładki międzygenowe, którymi są regiony IGS składają się z następujących elementów funkcjonalnych: 3'ETS (3'External Transcribed Spacer; zewnętrzny transkrybowany odstępnik), NTS (Non-Transcribed Spacer; nietranskrybowany odstępnik) i 5'ETS (5'External Transcribed Spacer). 3'ETS jest dość krótką sekwencją, która znajduje się między 3'-końcem genu 26 S rDNA a TTS - miejscem zakończenia transkrypcji (transcription termination site).

Dokładną lokalizację zasięgu motywów pirymidyny swoistych dla TTS pokazano w Tabeli 14. W regionach IGS w rodzaju *Erica* znaleziono jeden rzekomy TTS. 5'ETS jest sekwencją nukleotydową położoną na drugim końcu regionu IGS, znajduje się między 5'-końcem 18S rDNA a TIS - miejscem inicjacji transkrypcji (transcription initiation site). Liczba rzekomych TIS w badanym rodzaju *Erica* osiągnęła aż liczbę sześciu u *Erica cinerea* (Tabela 15). Nietranskrybowany region (NTS) zlokalizowany między 3'ETS a 5'ETS wykazuje wysoki poziom zmienności długości sekwencji, w rodzaju *Erica* długości NTS osiągają od 236 pz (*Erica cinerea*) do 5610 pz (*Erica arborea*).

Tabela 14. Zestawienie przypuszczalnego TTS u przedstawicieli rodzaju *Erica*. Pozycja pierwszego nukleotydu TTS dla przypuszczalnej sekwencji TTS w odniesieniu do końca 3’ 26S rDNA

Gatunek	Sekwencja TTS	Pozycja TTS
Erica arborea	TCCCTCCCCCTTT	21
Erica cinerea	TCCCTCCCCCTTC	21
Erica globiceps	TCCCTCCCCCTTC	21
Erica imbricata	TCCCTCCCCCTTC	21
Erica multiflora	TCCCTCCCCCTTC	21
Erica plukenetii	TCCCTCCCCCTTC	21
Erica vagans	TCCCTCCCCCTTT	21
Erica versicolor	TCCCTCCCCCTTC	21

Tabela 15. Zestawienie przypuszczalnego TIS u przedstawicieli rodzaju *Erica*. Pozycja pierwszego nukleotydu TIS dla przypuszczalnej sekwencji TIS w odniesieniu do końca 5’ 18S rDNA

Gatunek	Sekwencja TIS	Pozycja TIS
Erica arborea	TCTATAGGGGGG	5631
Erica cinerea	TCTTTAGGGGGA	257
	TCATTGGGGGTG	3683
	TCCTTCGGGGGA	4167
	TCCTTCGGGGGA	4442
	TCATTGGGGGTG	4681
	TCTTTGGGGGTC	5577
Erica globiceps	TCTATAGGGGGG	3681
Erica imbricata	TCTATAGGGGGG	1088
Erica multiflora	TCTATAGGGGGG	2232
Erica plukenetii	ACTATAGGGGGG	2503
Erica vagans	TCTATAAGGAGG	301
	TCAATGGGGGGC	2848
Erica versicolor	ACTATAGGGGGG	2670

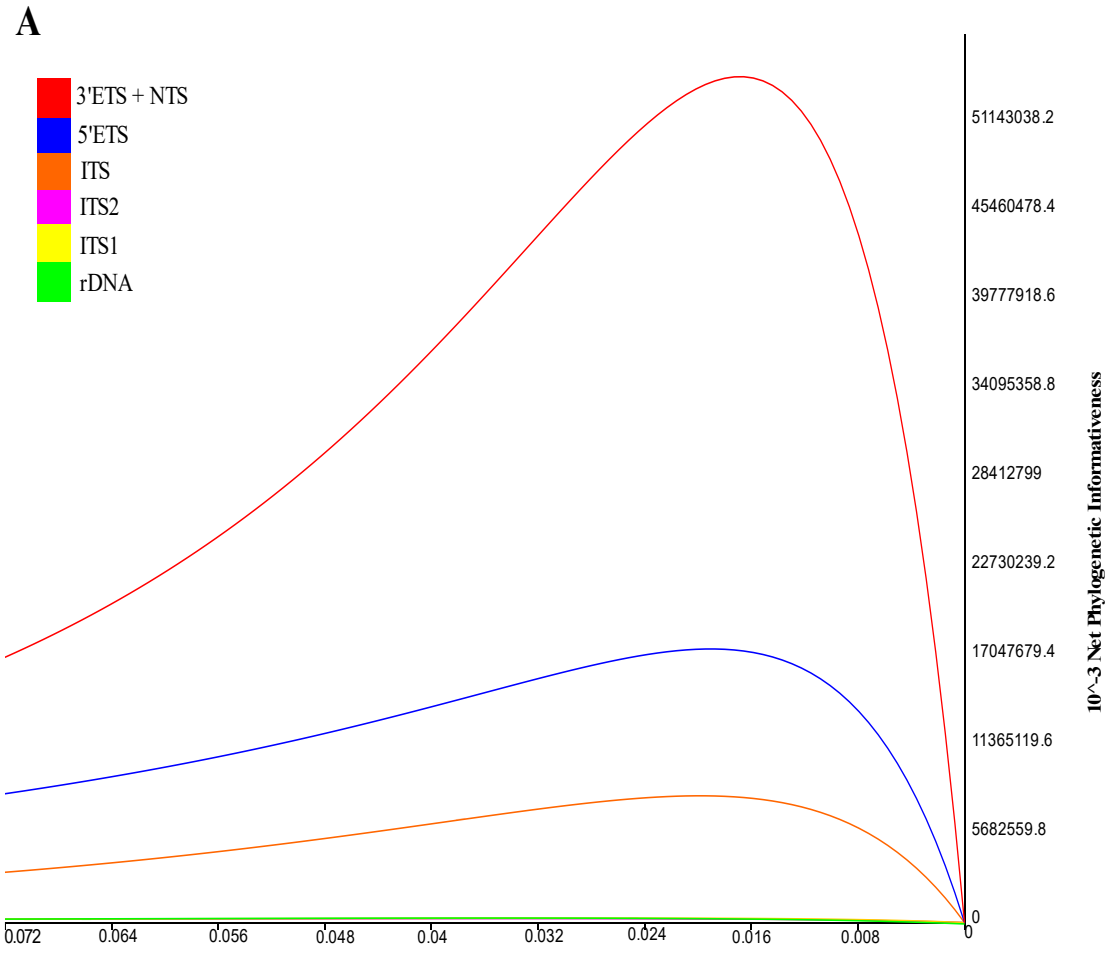
4.4. Analiza filogenetyczna

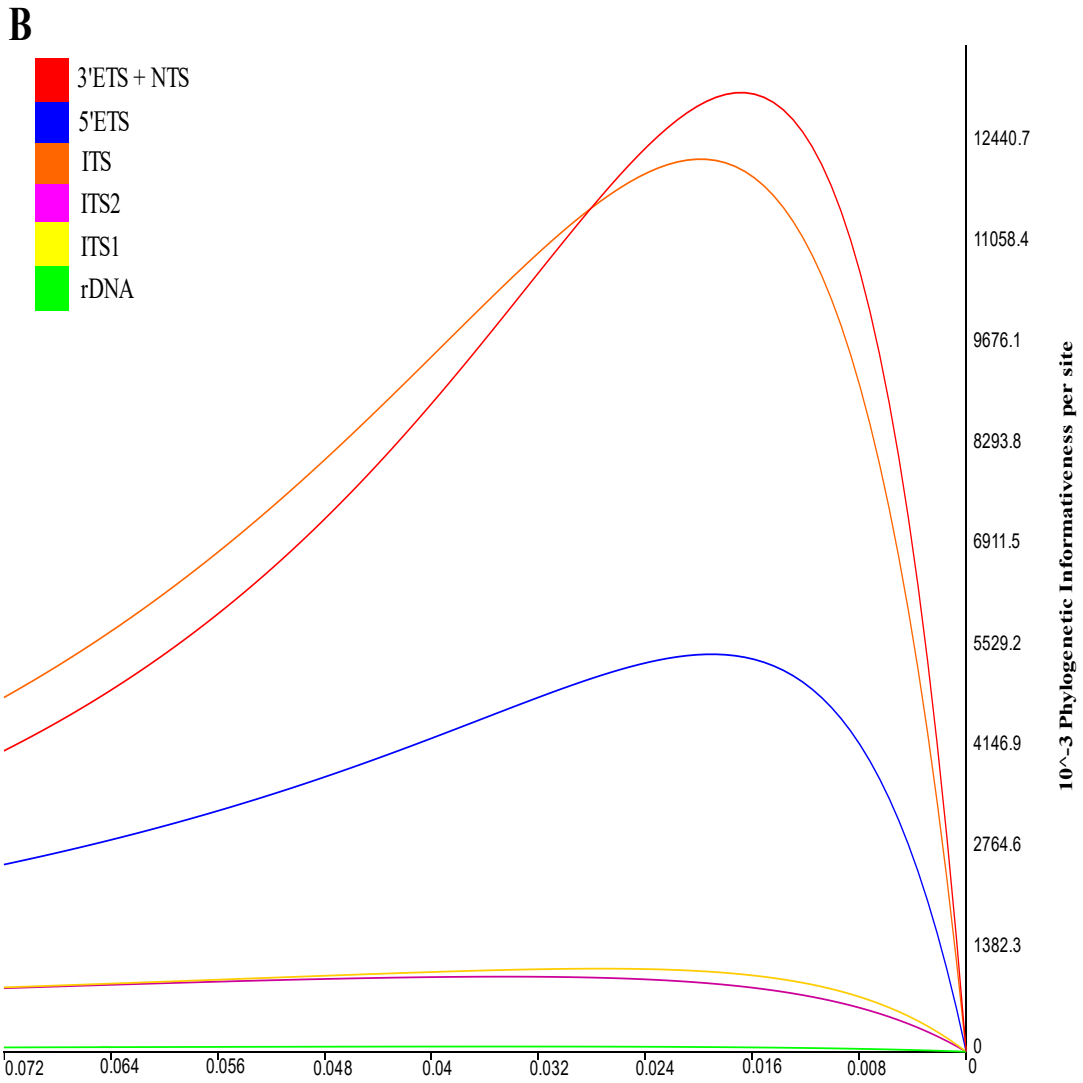
4.4.1. Filogenetyczna informatywność

Badano składowe funkcjonalne jednostki powtarzalne rDNA w celu opisania profilu informatywności filogenetycznej (PI). Zestaw danych z rodzaju *Erica* został zbadany osobno: dla genów rDNA, ITS1, ITS2, kompletnego regionu ITS z 5,8 S rDNA, 3'ETS, w tym NTS oraz dla 5'ETS.

Profil badanej informatywności filogenetycznej sieciowej dla wybranych domen w obrębie rDNA, obejmujący kompleksową informatywność regionów nukleotydowych, wykazał bardzo niski sygnał filogenetyczny podtrzymywany przez geny rDNA. Niewystarczająco wysoki stopień PI ogólnego ujawniono również w regionie ITS1. Piki tego profilu PI dla rDNA i ITS1 nakładały się na siebie. Najwyższy zakres PI netto uzyskano dla regionu 3'ETS z dużą różnicą w stosunku do poziomu następnego regionu w wartości sieciowej PI (Rycina 20 A).

Rozdzielczość profilu PI w przeliczeniu na miejsce potwierdza najniższy poziom informatywności dla genów rDNA i najwyższy odpowiednio dla ITS oraz 3'ETS + NTS (Rycina 20 B). W związku z tym, jeśli chodzi o wynik informatywności filogenetycznej w przeliczeniu na miejsce i ogólnej, najbardziej obiecującym markerem filogenetycznym dla rodzaju *Erica* okazał się region IGS obejmujący 3'ETS i region NTS.

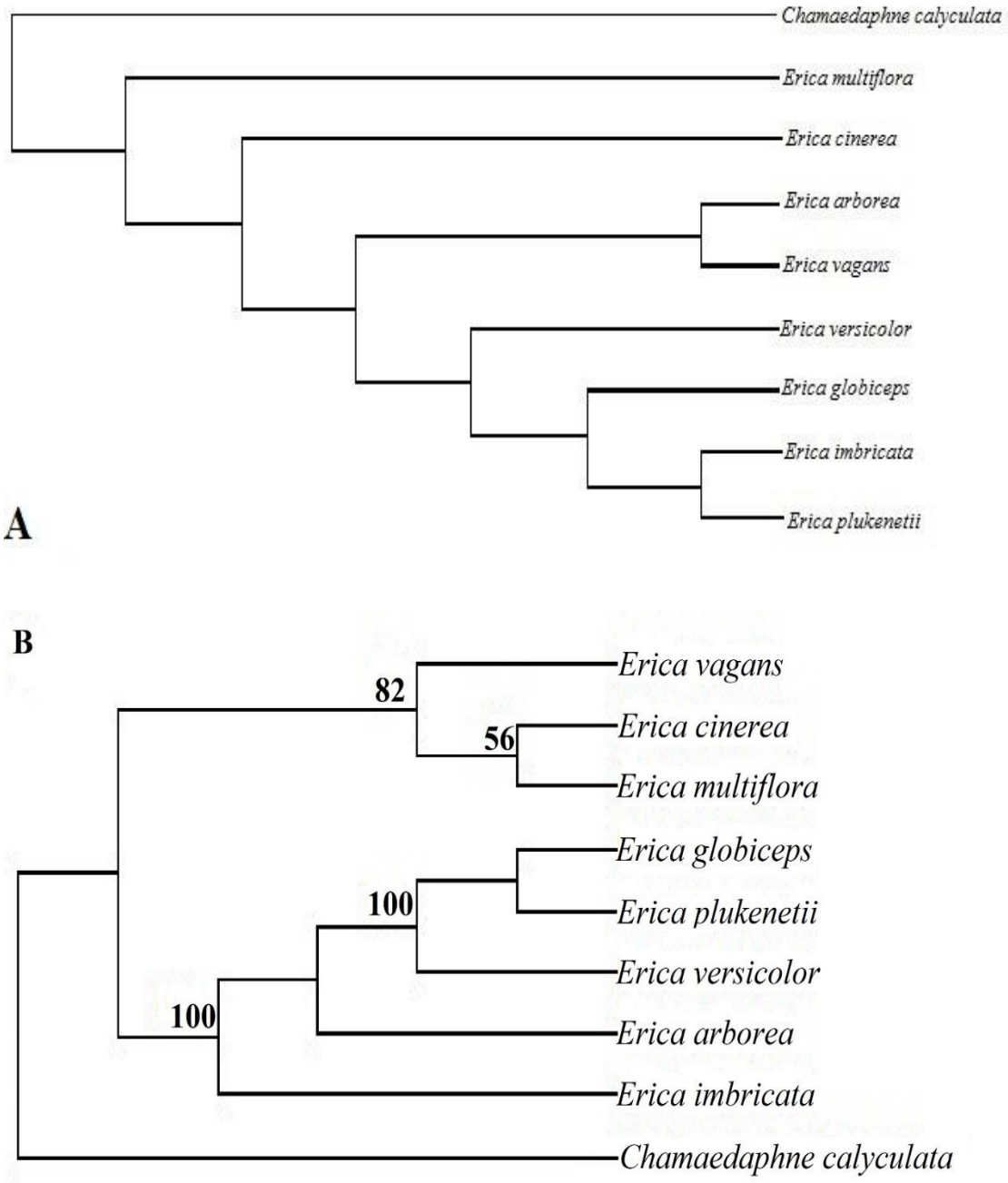

A
3'ETS + NTS
5'ETS
ITS
ITS2
ITS1
rDNA
51143038.2
45460478.4
39777918.6
34095358.8
28412799
22730239.2
17047679.4
11365119.6
5682559.8
0
10^-3 Net Phylogenetic Informativeness
0.072
0.064
0.056
0.048
0.04
0.032
0.024
0.016
0.008
0



Rycina 20. Informatywność filogenetyczna. Profil pierwszy (A) jest ogólny sieciowy, drugi (B) powstał w przeliczeniu na miejsca filogenetycznej informatywności dla wybranych regionów w obrębie rDNA. Oś czasu została skalibrowana do drzewka ML o ultrametryzowanym obliczeniu dla pełnej jednostki rDNA. Czas ewolucji wynoszący 1,0 odpowiada korzeniu drzewa, a wartość 0 jego wierzchołkom

Otrzymane rezultaty zostały potwierdzone przez analizę kladogramów obliczoną osobno dla regionu 3'ETS + NTS i pełnego ITS (ITS1 + ITS2) pochodnej metodą maksymalnego podobieństwa (ML) (odpowiednio: Rycina 21B i 21A).

Drzewo filogenetyczne zbudowane w regionie 3'ETS + NTS jest utworzone z silnie wspieranych kladów z wartościami ładowania początkowego sięgającymi 100%. Z drugiej strony drzewo filogenetyczne zbudowane z sekwencji ITS składa się z niskiego poziomu obsługiwanych kladów ze wszystkimi wartościami bootstrapu poniżej 60%. Ponadto skonfrontowane drzewa filogenetyczne wskazują na znaczną rozbieżność w odniesieniu do lokalizacji przedstawicieli rodzaju. Na przykład rekonstrukcja filogenetyczna pochodząca z ITS wykazuje ścisły związek między *Erica imbricata* i *Erica plukenetii*, podczas gdy rozdzielczość kladogramu skonstruowanego w regionie 3'ETS + NTS umieszcza te gatunki w osobnych liniach o wysokiej wiarygodności.



Rycina 21. Zestawienie dla regionów ITS i 3'ETS + NTS. Kladogramy konsensusowe otrzymane z analizy ML całego ITS (A) i regionu 3'ETS + NTS (B). Wartości bootstrap poniżej 55% nie zostały ujęte

4.4.2. Wzorce zróżnicowania sekwencji powtórzonych w obrębie regionu IGS – zdetektowane sekwencje motywów

W celu szczegółowego badania wariancji IGS, analizę sekwencji IGS przeprowadzono za pomocą metody dot plot, które są zaprezentowane osobno dla każdego badanego gatunku na rycinach Suplementu . Badanie to wykazało większą konserwatywność sekwencji zlokalizowanych początkowo i na końcu struktury IGS. Tyle, że analiza dot plot pokazuje obecność wielu motywów powtarzających się w badanych gatunkach, z bardzo wyraźnym wizualnym odzwierciedleniem odpowiednio liczby motywów u osobnika. Rozpoznanie liczby i długości powtórzonych sekwencji w *Chamaedaphne calyculata* pozwoliło na wyznaczenie dziewięciu wariantów motywów o długości od 20 do 108 pz (Tabela 16, Suplement 4) i miało charakter porównawczy. Zestawienie otrzymanych motywów u innego gatunku z rodziny *Ericaceae* odzwierciedliło niepowtarzalność aranżacji sekwencji powtórzonych w rodzaju *Erica*.

W rodzaju *Erica* wykryto czternaście różnych motywów o długości od 100 do 141 pz (Tabela 16). Podobnie jak u *Chamaedaphne calyculata*, u przedstawicieli *Erica* zauważono polimorfizm pojedynczego nukleotydu (SNP) pomiędzy kopiami poszczególnych motywów w obrębie gatunku. Liczba powtórzeń motywów wynosiła od 2 do 7 na gatunek w przypadku *Chamaedaphne calyculata*. Natomiast w rodzaju *Erica* wykryto bardzo zmienną liczbę kopii między powtórzeniami. Niektóre gatunki nie mają pewnych motywów, jedynie ER10 występuje u każdego gatunku *Erica,* zawsze w jednym powtórzeniu. Najwięcej motywów wykryto u *Erica arborea*, gdzie zanotowano również największą liczbę powtórzeń na gatunek; w przypadku ER5 – aż 26 powtórzeń, co istotnie koreluje z długością całej sekwencji IGS u tego gatunku. Najmniej motywów odkryto u *Erica vagans*, gdzie również ujawniono najmniejszą liczbę powtórzeń poszczególnych motywów oraz brak czterech motywów (ER2, ER3, ER9 i ER13), co także odpowiada długości całej sekwencji IGS *Erica vagans*.

Porównanie regionów IGS wskazuje na niezmiernie wysoki poziom różnorodności sekwencji w regionie wewnętrznych powtórzeń (IR – inter-repeats) i wysoce zróżnicowany rozkład wartości Pi w poszczególnych motywach w osobnych gatunkach z rodzaju *Erica* (Tabela 17). Schemat powtórzeń u gatunków *Erica* nie jest bardzo prosty ze względu na dużą liczbę wykrytych motywów (14) i ich liczne powtórzenia (do 26). Najmniej zróżnicowany moduł IR zidentyfikowano w *Erica cinerea*, gdzie wykryto brak czterech motywów, pozostałe cztery zanotowano jedynie pojedynczo, a wartość zróżnicowania nukleotydów (Pi) dwóch innych motywów była równa 0, ale wszystkie predyktory nie oznaczają, że schemat *Erica cinerea* można scharakteryzować jako prosty. Ponieważ w tym gatunku odkryto nawet 17 kopii jednego typu powtórzeń, każdy o długości 114 pz, nazwanego ER3. Skromny schemat ustalono w regionie IGS *Erica versicolor*, gdzie siedem motywów nukleotydowych wystąpiło tylko w jednej kopii, 3 motywy w dwóch powtórzeniach, a tylko jeden motyw nukleotydowy powtórzony jest dziesięciokrotnie – ER6 (100 pz). W *Chamaedaphne calyculata* wykazano występowanie dziewięciu różnych powtórzeń w zakresie od 20 do 108 pz. Wspólny wzór regionu organizacji wewnętrznych powtórzeń w każdym gatunku *Erica* obejmuje wąską część położoną w pobliżu genu 26S rRNA, gdzie wykryto motyw ER10 oraz część zlokalizowaną w pobliżu genu 18S rRNA, obejmującą konfigurację 3 motywów odpowiednio: ER8, ER7 i ER14. U wszystkich przedstawicieli rodzaju *Erica* domniemany TIS znajdował się w sekwencji ER9. Zestawienie wykrytych motywów w obrębie IGS u przedstawicieli rodzaju *Erica* wraz z liczbą ich powtórzeń przedstawione zostało w tabelo Suplementu 3, natomiast zestawienie wykrytych motywów w *Chamaedaphne calyculata* wraz z ich wartością Pi w tabeli Suplementu 4.

Ogólną aranżację wizualną rozkładu motywów u reprezentantów z rodzaju *Erica* oraz u gatunku *Chamaedaphne calyculata* przedstawiono odpowiednio na rycinach Suplementu 5 oraz 6.

Tabela 16. Zestawienie regionów wewnętrznych powtórzeń (motywów) określonych dla gatunków poddanych analizom

	Motyw	Wartość E-value	Liczba powtórzeń w przeliczeniu na gatunek	Kompletna liczba powtórzeń	Szerokość motywu [pz]
Chamaedaphne calyculata	**Motif 1**	2.5e-096	7	7	92
	Motif 2	6.7e-064	4	4	69
	Motif 3	2.7e-064	7	7	77
	Motif 4	1.0e-048	5	5	55
	Motif 5	4.9e-029	2	2	39
	Motif 6	1.6e-017	4	4	108
	Motif 7	2.5e-012	5	5	55
	Motif 8	3.2e-011	6	6	20
	Motif 9	2.4e-003	2	2	28
Erica	**ER1**	4.3e-1513	2-8	31	130
	ER2	4.6e-1321	0-12	35	141
	ER3	5.2e-955	0-17	27	114
	ER4	1.7e-944	0-7	24	141
	ER5	3.1e-847	0-26	35	100
	ER6	6.6e-629	0-15	46	100
	ER7	1.5e-431	1-2	10	141
	ER8	2.2e-515	2-4	23	140

Tabela 16. Ciąg dalszy.

	Motyw	Wartość E-value	Liczba powtórzeń w przeliczeniu na gatunek	Kompletna liczba powtórzeń	Szerokość motywu [pz]
Erica	**ER9**	3.5e-342	0-4	19	137
	ER10	9.3e-141	1	8	100
	ER11	7.9e-060	0-3	11	100
	ER12	2.0e-041	0-1	6	100
	ER13	1.2e-033	0-3	7	100
	ER14	2.7e-002	0-2	11	100

Tabela 17. Wykres wykrytych motywów w rodzaju *Erica* wraz z ich wartością Pi

Motyw	Nazwa gatunku i wartość zróżnicowania nukleotydowego (Pi) || '*' - jeden motyw, '-' – brak motywu								Kompletna wartość Pi danego motywu
	E.arborea	***E. cinerea***	***E.imbricata***	***E.multiflora***	***E.plukenetii***	***E.vagans***	***E.versicolor***	***E.globiceps***	
ER1	0,09487	0	0,03049	0,16462	0,15385	0,37179	0,40769	0,26154	0,15795
ER2	0,36099	-	0,15130	-	0,42514	-	0,26950	0,31746	0,31544
ER3	-	0,03805	*	0,08647	-	-	*	-	0,27123
ER4	0,33097	0	0,04458	0,33546	0,59574	0,42790	*	-	0,27212
ER5	0,22972	-	*	-	0,31333	0,17333	*	*	0,29466
ER6	0,61389	-	-	*	0,35489	*	0,22911	0,19429	0,40780
ER7	*	*	0	*	*	0,60993	*	*	0,15603
ER8	0,65000	0,66429	0,52381	0,63810	0,64286	0,67143	0,64048	0,63333	0,47230
ER9	0,45499	-	0,59124	0,54258	0,54258	-	0,53041	0,35036	0,46707

Tabela 17. Ciąg dalszy.

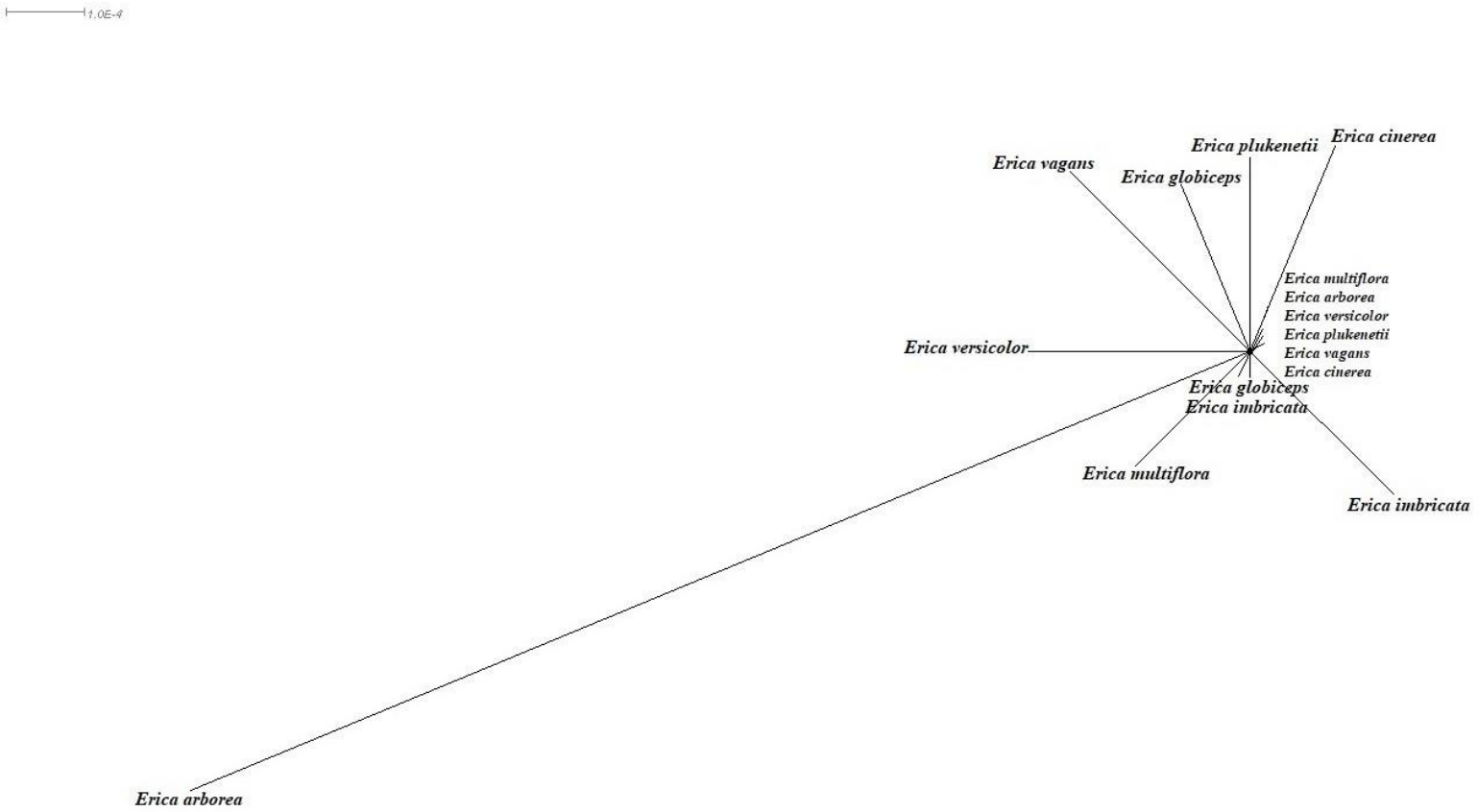
Motyw	Nazwa gatunku i wartość zróżnicowania nukleotydowego (Pi) || '*' - jeden motyw, '-' – brak motywu								Kompletna wartość Pi danego motywu
	E.arborea	***E. cinerea***	***E.imbricata***	***E.multiflora***	***E.plukenetii***	***E.vagans***	***E.versicolor***	***E.globiceps***	
ER10	*	*	*	*	*	*	*	*	0,27714
ER11	0,52862	*	-	*	0,68000	*	*	0,80000	0,54215
ER12	*	*	*	-	-	*	*	*	0,43467
ER13	-	0,34000	0,31000	-	-	-	-	0,47000	0,48190
ER14	-	0,71000	*	0,52000	*	*	0,65000	0	0,49927

4.5. Filogenetyka rodzaju *Erica*

Dostępność w bazach danych kompletnych sekwencji genomów różnych gatunków roślin pozwala na porównywanie i analizowanie ich toku ewolucji poprzez korelację genów w odniesieniu do ich względnych pozycji w genomach oraz zrozumienie ich relacji filogenetycznych.

Otrzymanie całego genomu cpDNA dla dwóch gatunków z rodzju *Erica* oraz obfite zasoby genomiczne pozostałych reprezentantów, umożliwiły w pierwszym kroku analizę całego genomu, a następnie przeprowadzenie badań porównawczych wybranych rodzin/grup genów chloroplastowych. Informacja o sekwencjach tych genów chloroplastowych widnieje w Suplemencie 1.

Wszystkie zidentyfikowane sekwencje genów z poszczególnych grup/rodzin zostały poddane przyrównaniu wielu sekwencji (MSA – multiple sequence alignment), który został wykonany w celu przeprowadzenia dalszych analiz oraz skonstruowania drzewa filogenetycznego przy użyciu parametrów domyślnych. W pierwszym kroku zestaw danych oparty na całkowitym zasobie sekwencji genów cpDNA badanych przedstawicieli z rodzaju *Erica*, posłużył do zrekonstruowania sieci filogenetycznej w programie SplitsTree 4.19.0 w formie sieci konsensusu, tworząc swego rodzaju zestaw ufności zaprezentowanych dalej drzew filogenetycznych (Rycina 22). Drzewa filogenetyczne zostały skonstruowane oddzielnie dla poszczególnych rodzin/grup genów przy użyciu ustawień domyślnych i 2000 replik typu bootstrap, celem zapewnienia wysokiego zakresu ufności i dokładności danych. Grupę zewnętrzną stanowił gatunek *Vaccinium macrocarpon* oraz *Chamaedaphne calyculata* w zależności od dostępności pełnych danych referencyjnych dla konkretnej rodziny/grupy genów w repozytorium GenBank. Topologia prezentowanych dalej drzew filogenetycznych jest całkowicie zgodna z wynikami wnioskowania bayesowskiego. Drzewa ML uzyskane z analiz połączonych sekwencji nukleotydowych poszczególnych genów i grup genów ukazały podobną topologię do drzew bayesowskich. Dla wszystkich badanych genów i grup genów przedstawiono skompilowane wyniki analiz filogenetycznych.

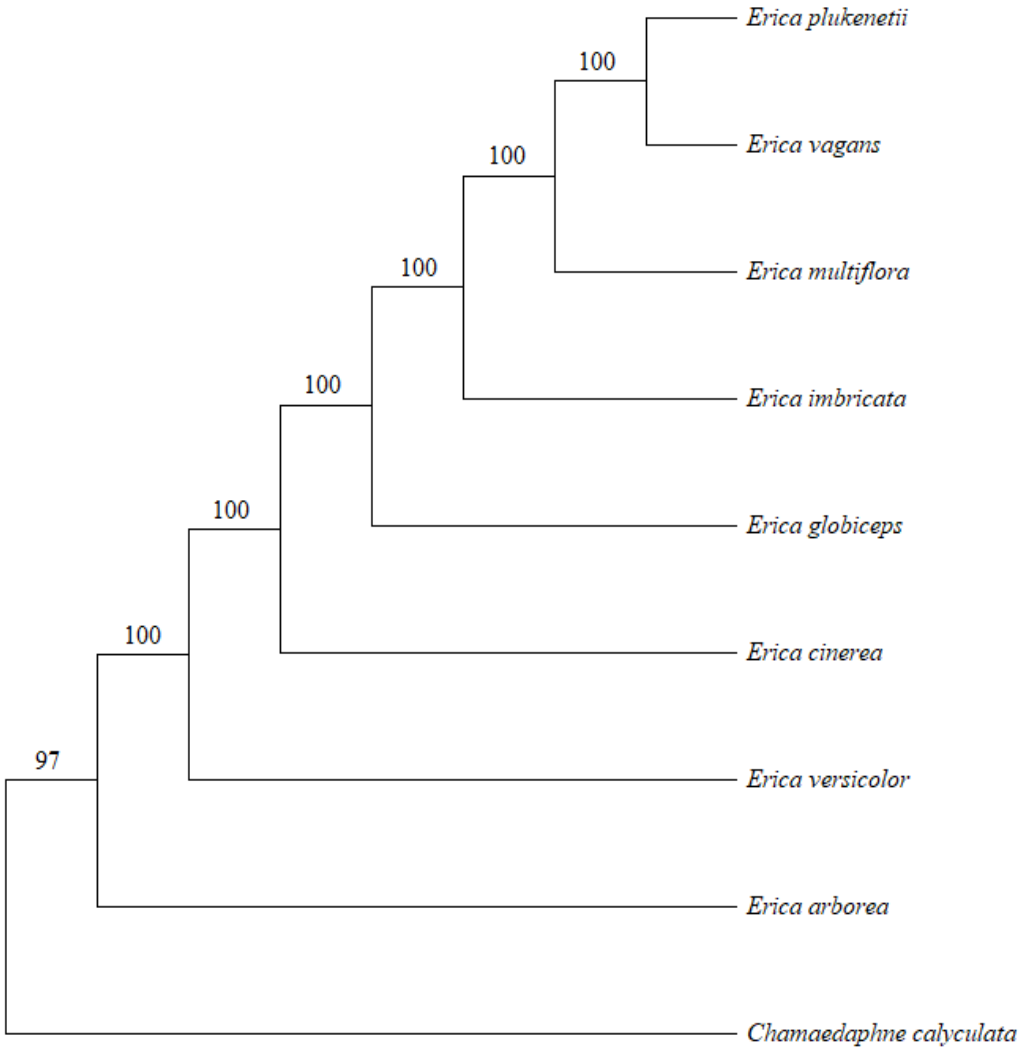


Rycina 22. Konsensusowa sieć filogenetyczna wygenerowana w Splits Tree, zrekonstruowana przy użyciu inkluzywnego zestawu danych sekwencji kompletnych genów cpDNA przedstawicieli z rodzaju *Erica*

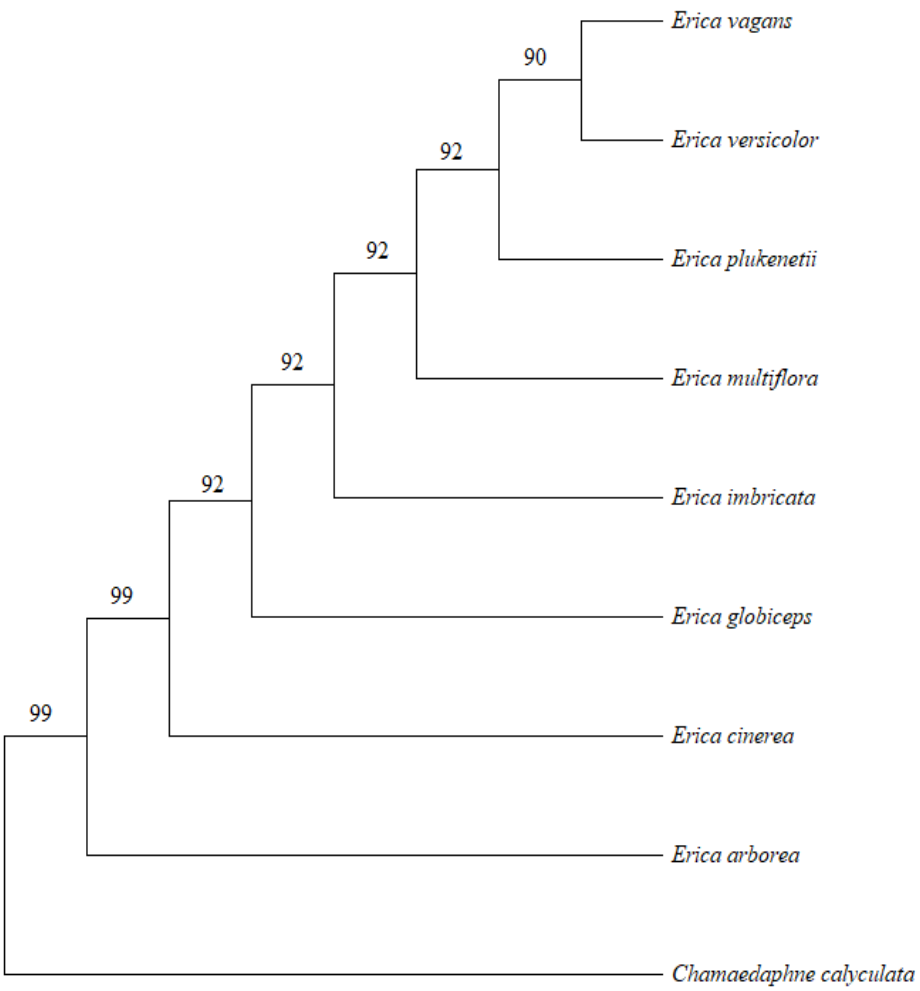
Wygenerowanie konsensusowej sieci filogenetycznej obejmowało wszystkie geny z genomów chloroplastowych badanych ośmiu reprezentantów z rodzaju *Erica*. Podczas gdy analiza drzew ML każdego genu zapawnia statystyczne wsparcie dla sprzecznych filogenez, a oczywistym jest że analizy ML są obarczone błędem systematycznym, posłużyło to do głębszej analizy.

Brak rozszczepionych sieci stanowi o niewystępowaniu wyraźnej i jawnej relacji ewolucyjnej, a brak równoległych krawędzi reprezentuje brak wyraźnych podziałów. Po skonstruowaniu sieci filogenetycznej określono, że nie różni się ona znacząco od wygenerowanych drzew, co sugerować może przeprowadzenie analizy filogenetycznej opartej na drzewie filogenetycznym skontruowanym analizą bayesowską.

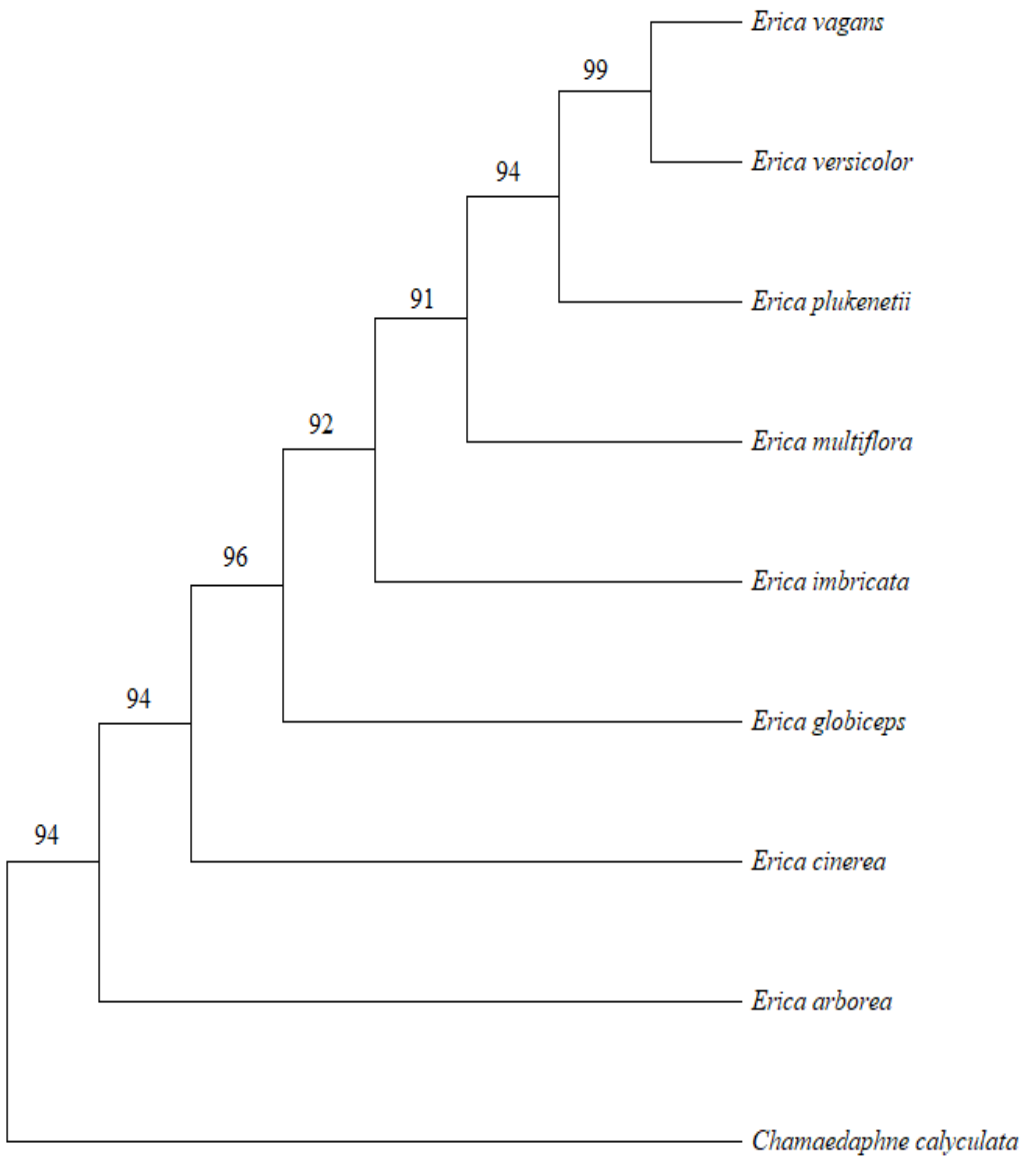
Wyniki zostały potwierdzone przez analizę dendrogramów obliczoną osobno dla każdej wymienionej grupy/rodziny genów i osobno dla poszczególnych genów *acc*D, *cem*A, *ccs*A, *mat*K oraz *rbc*L, metodą maksymalnego podobieństwa (ML) (Ryciny 23-38).



Rycina 23. Konsensusowe drzewo filogenetyczne przedstawicieli z rodzaju *Erica* skontruowane analizą bayesowską dla kompletnych genów *rbc*L. Wartości bootstrap poniżej 50% są ukryte. Liczby nad gałęziami reprezentują wsparcie bootstrapowe uzyskane przez maksymalne prawdopodobieństwo (wartość ML)



Rycina 24. Konsensusowe drzewo filogenetyczne przedstawicieli z rodzaju *Erica* skontruowane analizą bayesowską dla kompletnych genów *mat*K. Wartości bootstrap poniżej 50% są ukryte. Liczby nad gałęziami reprezentują wsparcie bootstrapowe uzyskane przez maksymalne prawdopodobieństwo (wartość ML)



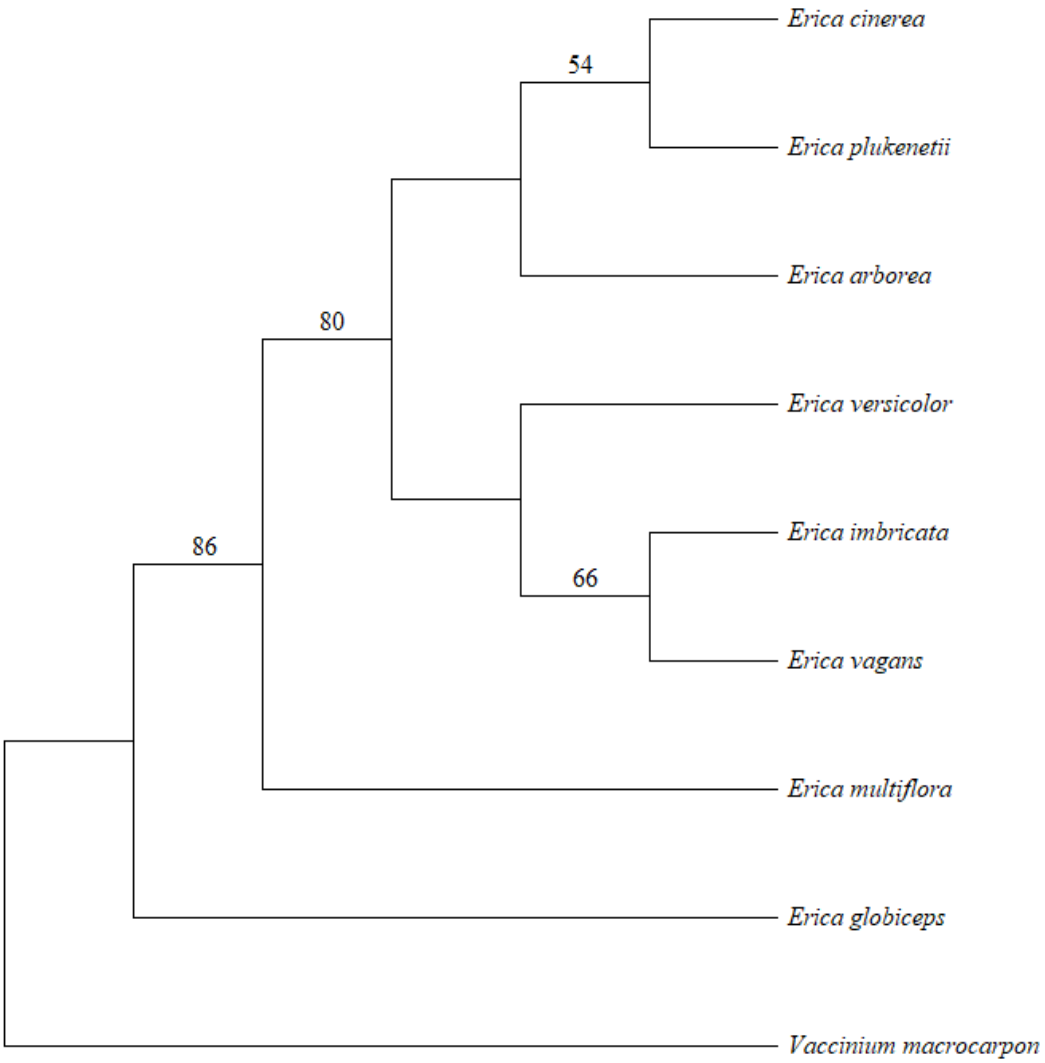
Rycina 25. Konsensusowe drzewo filogenetyczne przedstawicieli z rodzaju *Erica* skontruowane analizą bayesowską dla kompletnych genów *ccs*A. Wartości bootstrap poniżej 50% są ukryte. Liczby nad gałęziami reprezentują wsparcie bootstrapowe uzyskane przez maksymalne prawdopodobieństwo (wartość ML)



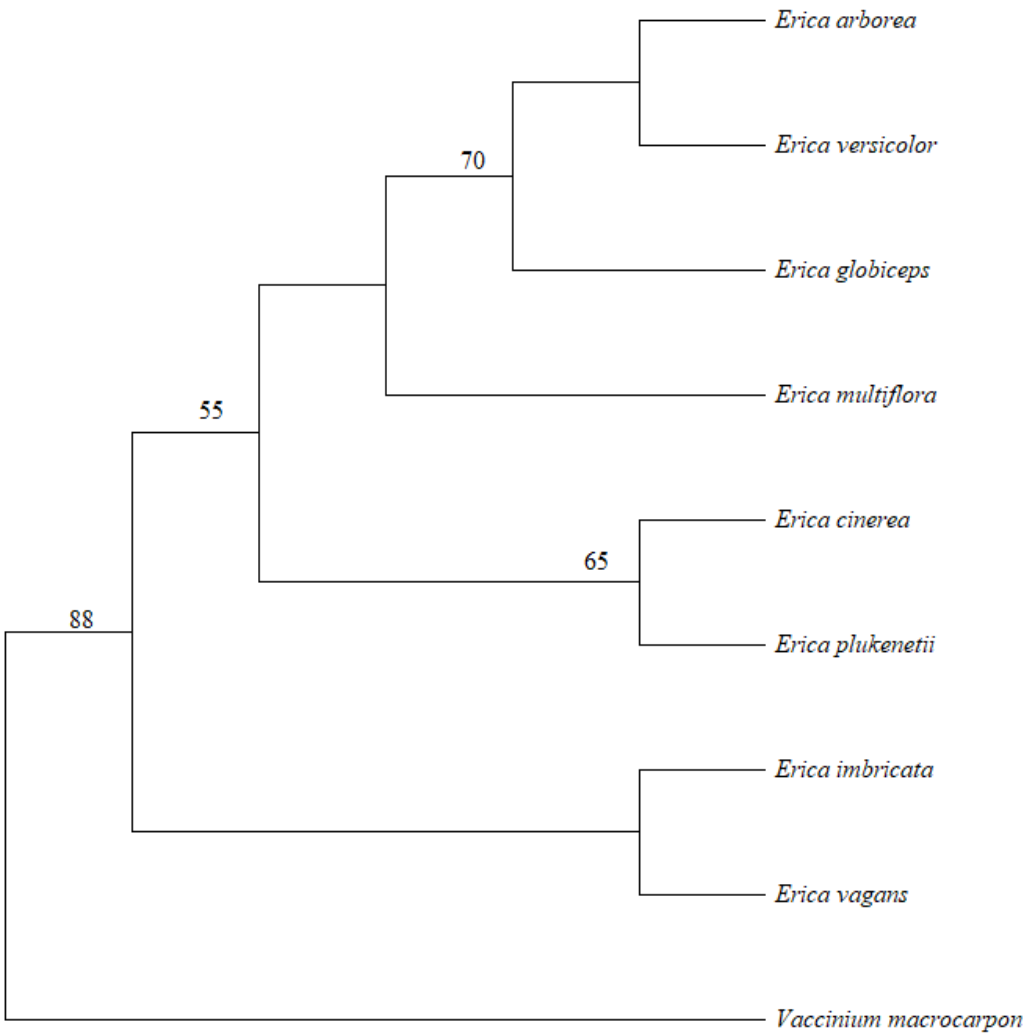
Rycina 26. Konsensusowe drzewo filogenetyczne przedstawicieli z rodzaju *Erica* skontruowane analizą bayesowską dla kompletnych genów *cem*A. Wartości bootstrap poniżej 50% są ukryte. Liczby nad gałęziami reprezentują wsparcie bootstrapowe uzyskane przez maksymalne prawdopodobieństwo (wartość ML)



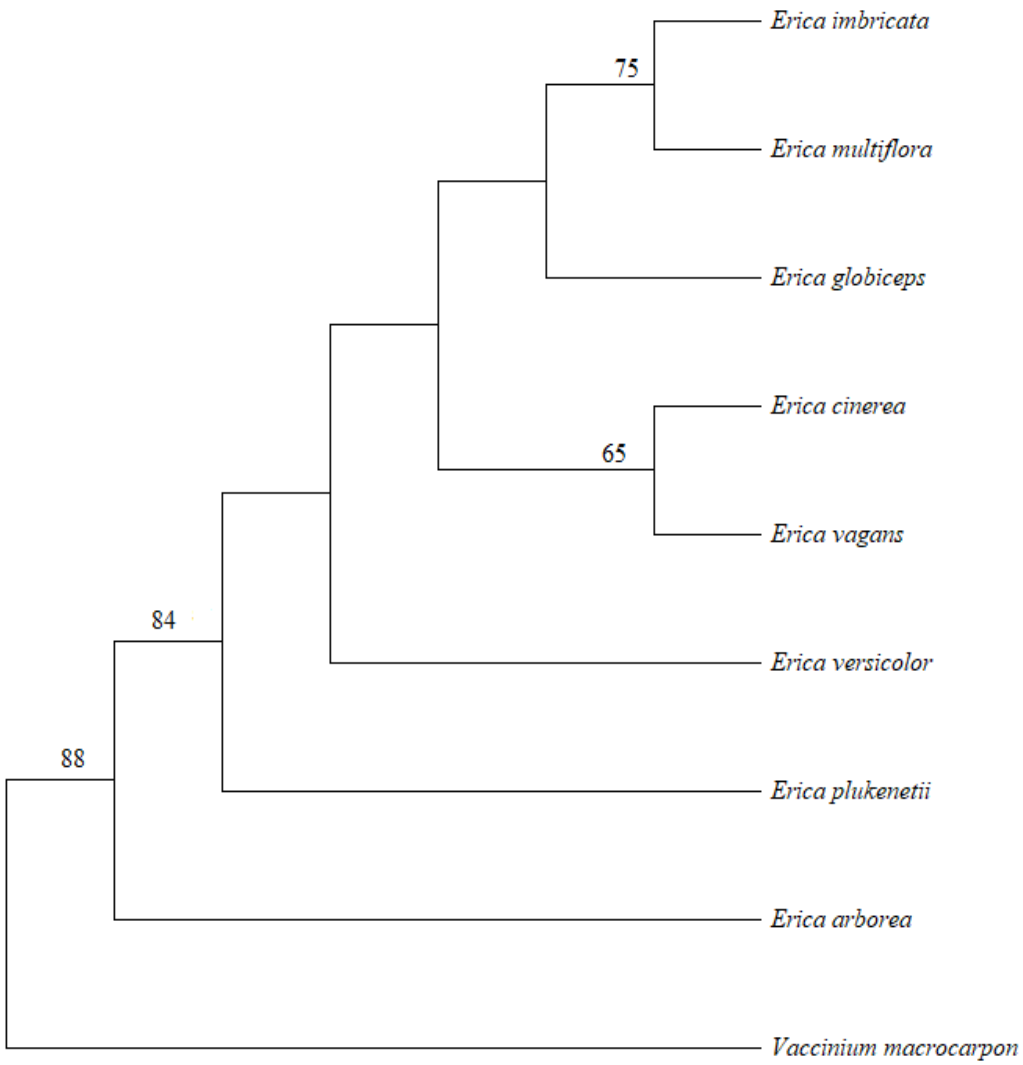
Rycina 27. Konsensusowe drzewo filogenetyczne przedstawicieli z rodzaju *Erica* skontruowane analizą bayesowską dla kompletnych genów *acc*D. Wartości bootstrap poniżej 50% są ukryte. Liczby nad gałęziami reprezentują wsparcie bootstrapowe uzyskane przez maksymalne prawdopodobieństwo (wartość ML)



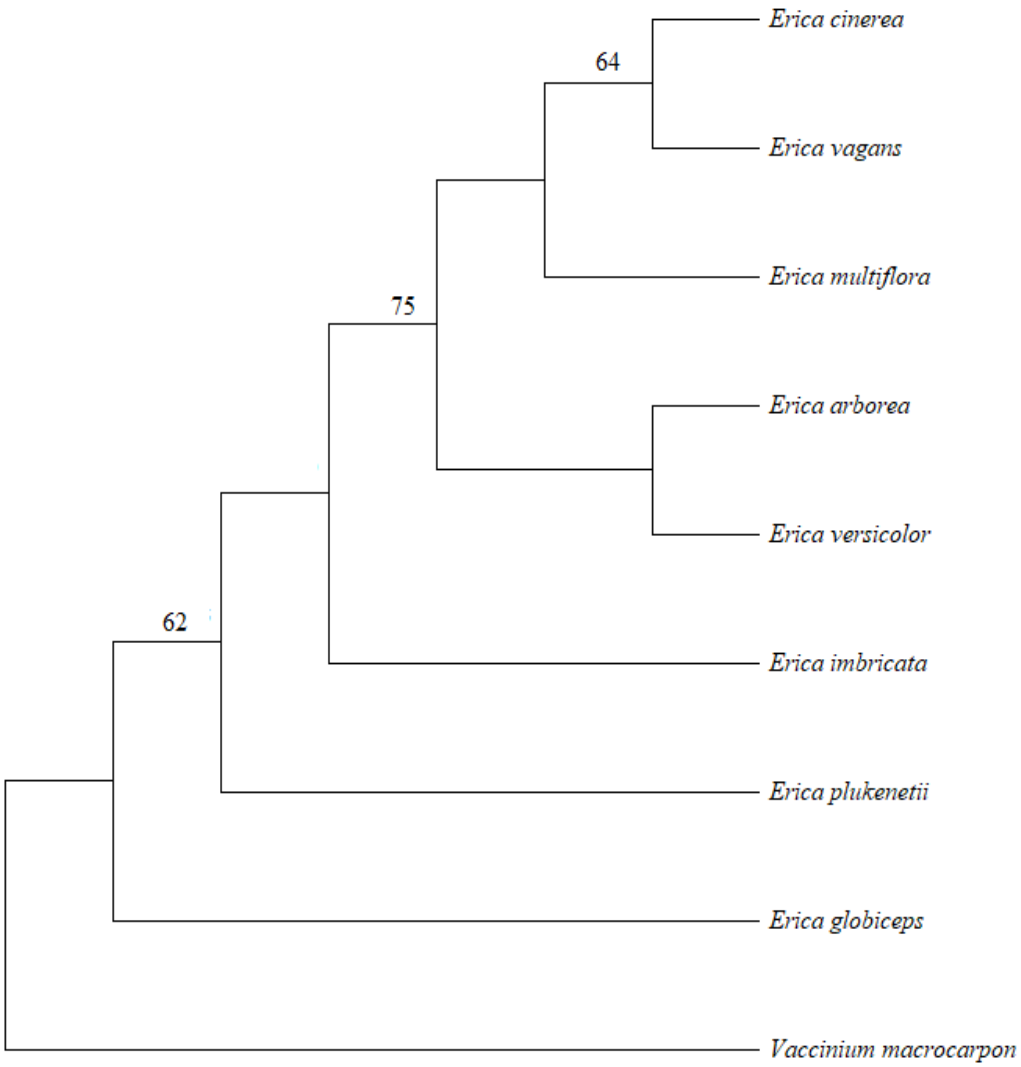
Rycina 28. Konsensusowe drzewo filogenetyczne przedstawicieli z rodzaju *Erica* skontruowane analizą bayesowską dla kompletnych genów *atp*. Wartości bootstrap poniżej 50% są ukryte. Liczby nad gałęziami reprezentują wsparcie bootstrapowe uzyskane przez maksymalne prawdopodobieństwo (wartość ML)



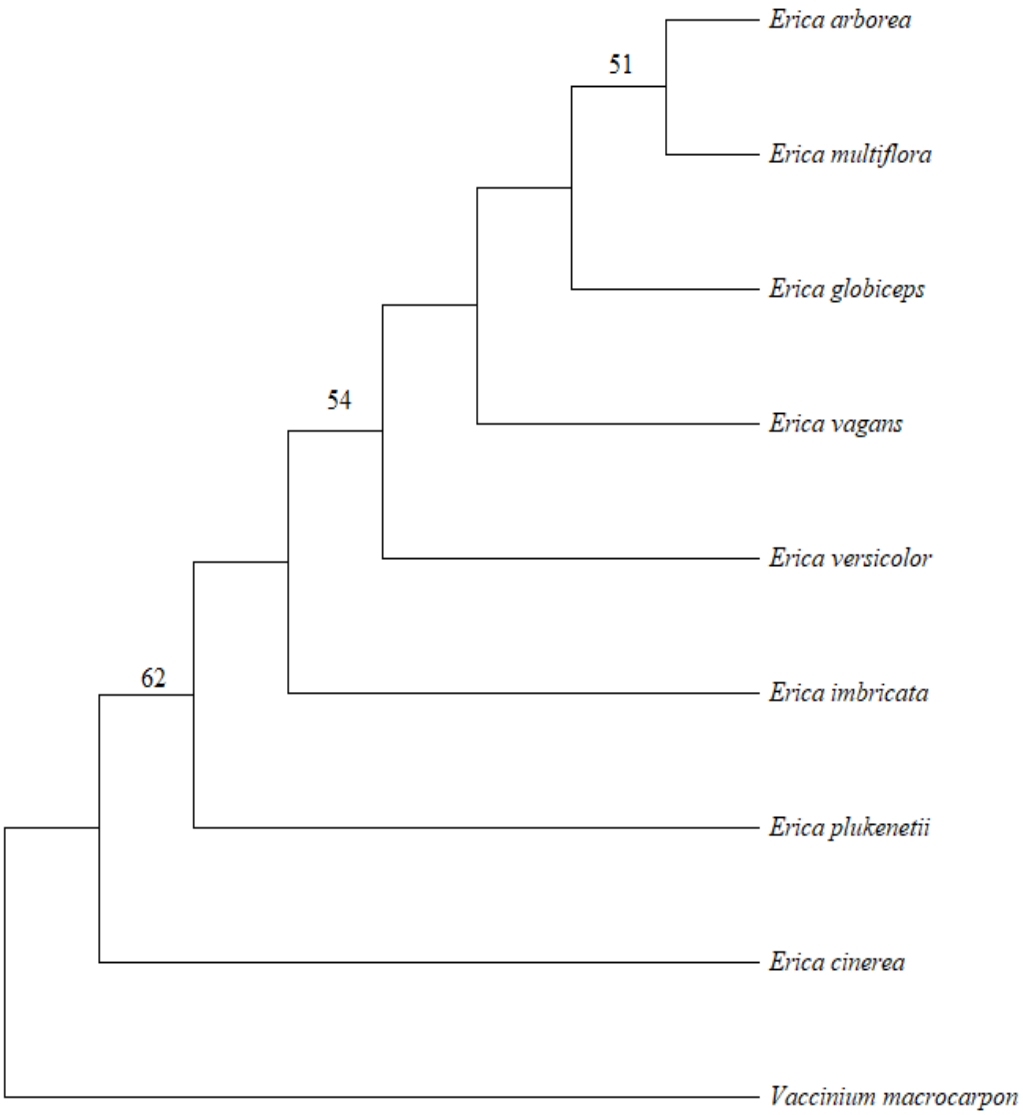
Rycina 29. Konsensusowe drzewo filogenetyczne przedstawicieli z rodzaju *Erica* skontruowane analizą bayesowską dla kompletnych genów *ndh*. Wartości bootstrap poniżej 50% są ukryte. Liczby nad gałęziami reprezentują wsparcie bootstrapowe uzyskane przez maksymalne prawdopodobieństwo (wartość ML)



Rycina 30. Konsensusowe drzewo filogenetyczne przedstawicieli z rodzaju *Erica* skontruowane analizą bayesowską dla kompletnych genów *pet*. Wartości bootstrap poniżej 50% są ukryte. Liczby nad gałęziami reprezentują wsparcie bootstrapowe uzyskane przez maksymalne prawdopodobieństwo (wartość ML)



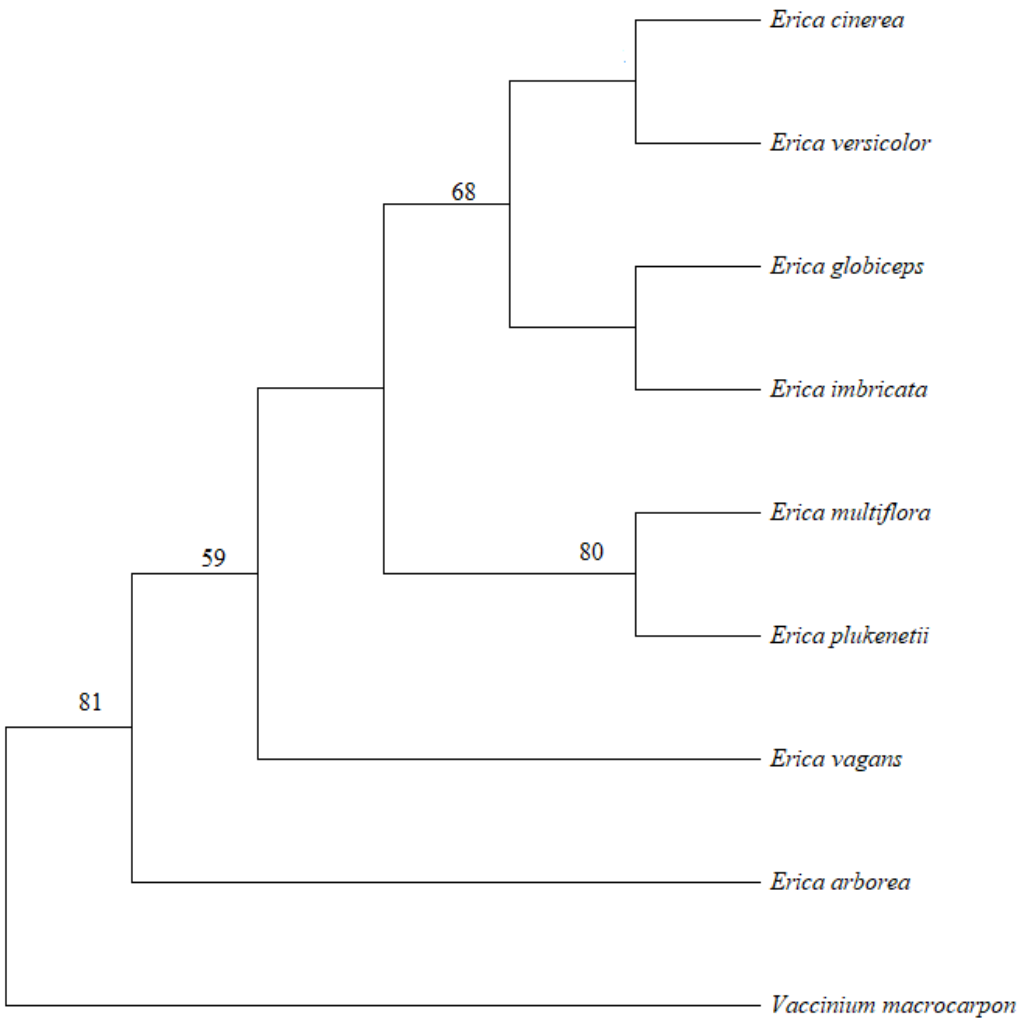
Rycina 31. Konsensusowe drzewo filogenetyczne przedstawicieli z rodzaju *Erica* skontruowane analizą bayesowską dla kompletnych genów *rpo*. Wartości bootstrap poniżej 50% są ukryte. Liczby nad gałęziami reprezentują wsparcie bootstrapowe uzyskane przez maksymalne prawdopodobieństwo (wartość ML)



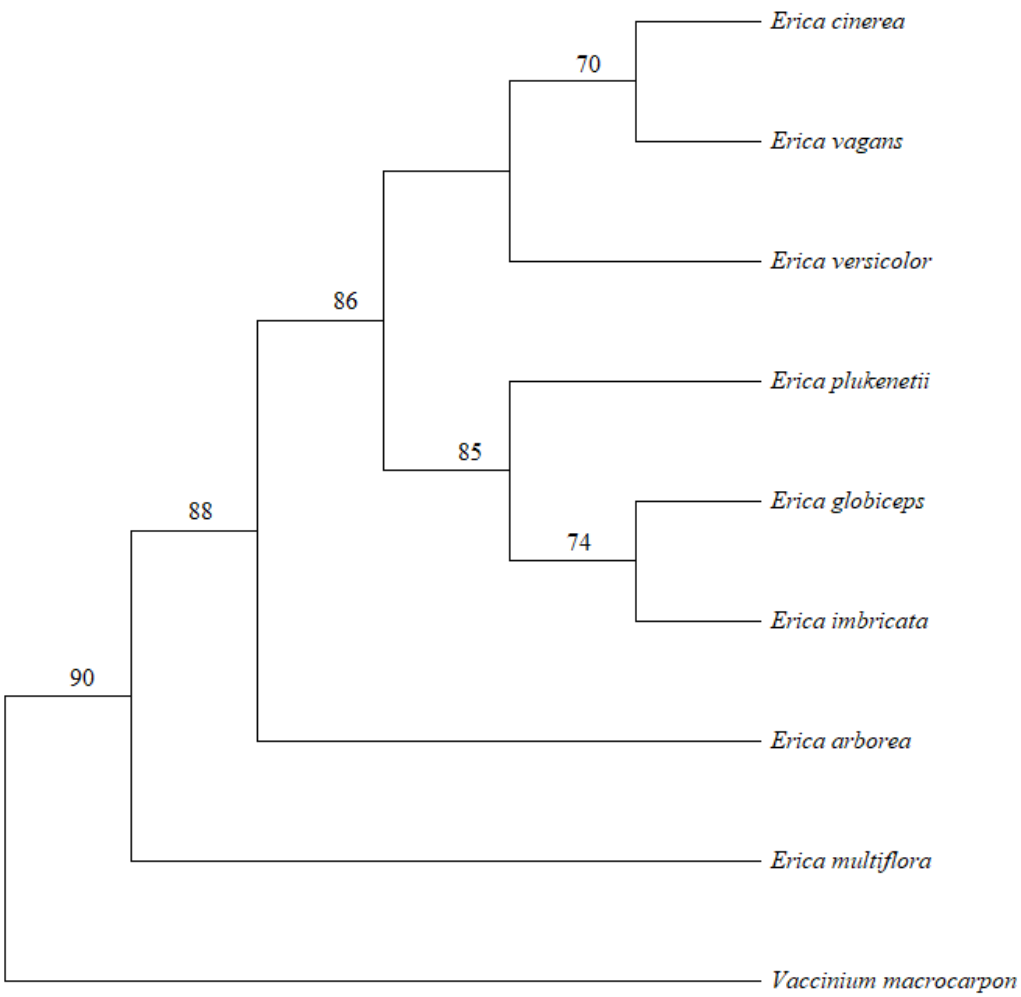
Rycina 32. Konsensusowe drzewo filogenetyczne przedstawicieli z rodzaju *Erica* skontruowane analizą bayesowską dla kompletnych genów *r*RNA. Wartości bootstrap poniżej 50% są ukryte. Liczby nad gałęziami reprezentują wsparcie bootstrapowe uzyskane przez maksymalne prawdopodobieństwo (wartość ML)



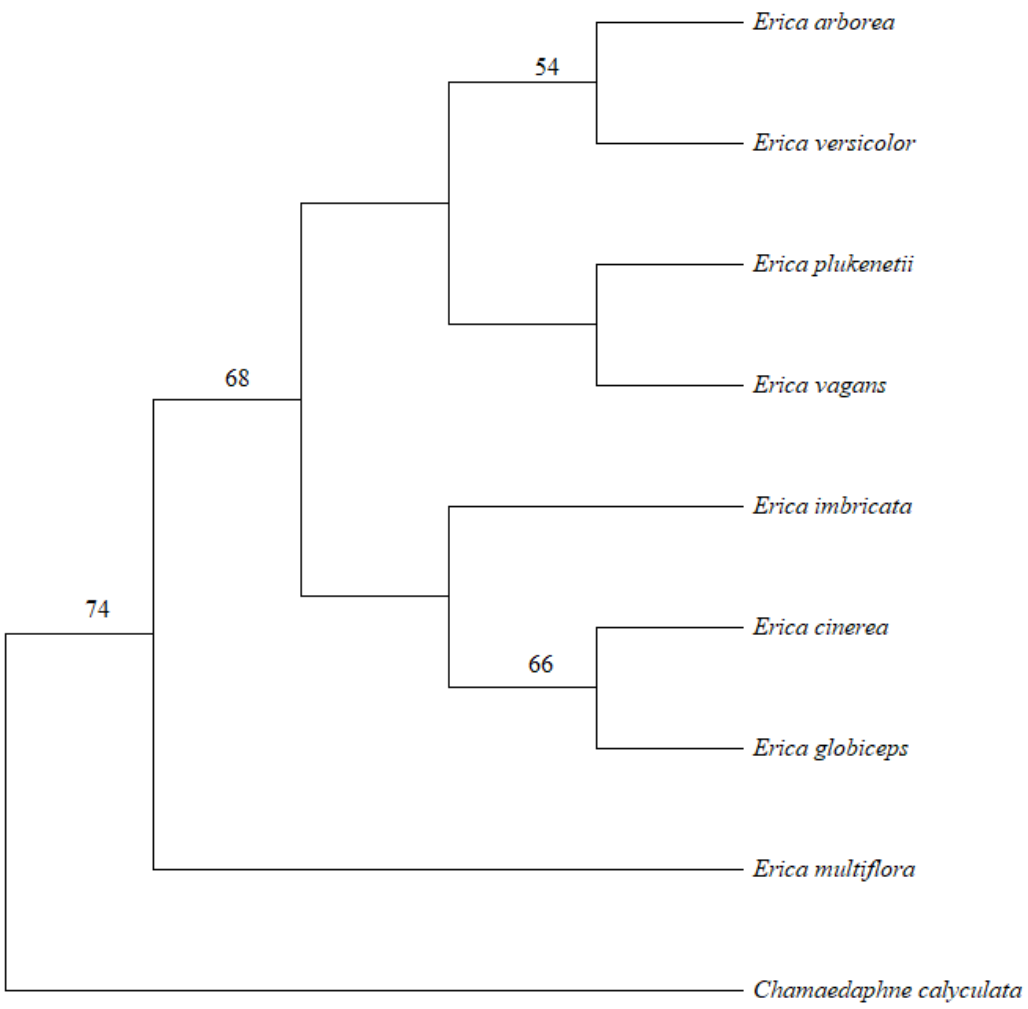
Rycina 33. Konsensusowe drzewo filogenetyczne przedstawicieli z rodzaju *Erica* skontruowane analizą bayesowską dla kompletnych genów *ycf*. Wartości bootstrap poniżej 50% są ukryte. Liczby nad gałęziami reprezentują wsparcie bootstrapowe uzyskane przez maksymalne prawdopodobieństwo (wartość ML)



Rycina 34. Konsensusowe drzewo filogenetyczne przedstawicieli z rodzaju *Erica* skontruowane analizą bayesowską dla kompletnych genów *psa*. Wartości bootstrap poniżej 50% są ukryte. Liczby nad gałęziami reprezentują wsparcie bootstrapowe uzyskane przez maksymalne prawdopodobieństwo (wartość ML)



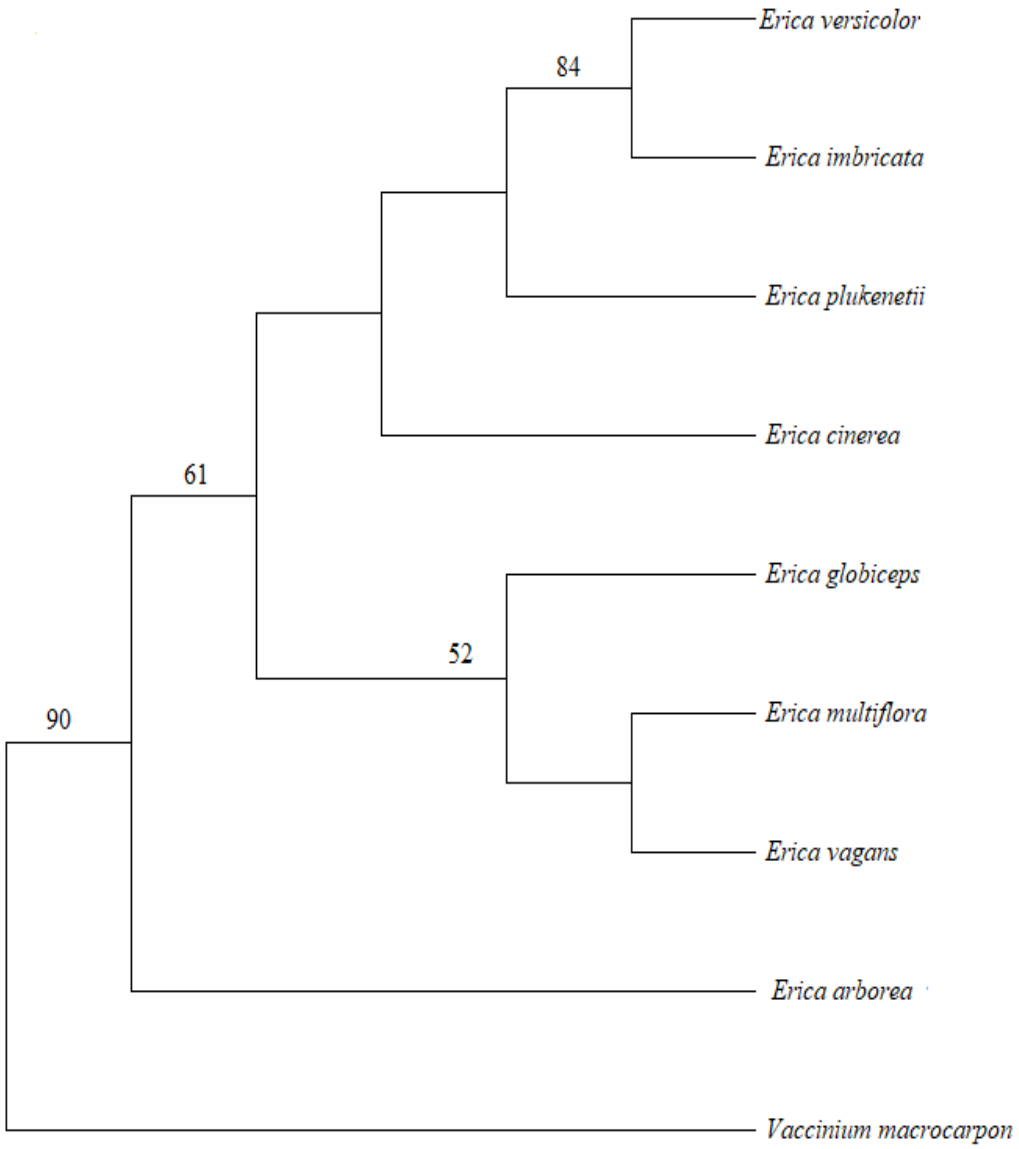
Rycina 35. Konsensusowe drzewo filogenetyczne przedstawicieli z rodzaju *Erica* skontruowane analizą bayesowską dla kompletnych genów *rpl*. Wartości bootstrap poniżej 50% są ukryte. Liczby nad gałęziami reprezentują wsparcie bootstrapowe uzyskane przez maksymalne prawdopodobieństwo (wartość ML)



Rycina 36. Konsensusowe drzewo filogenetyczne przedstawicieli z rodzaju *Erica* skontruowane analizą bayesowską dla kompletnych genów *rps*. Wartości bootstrap poniżej 50% są ukryte. Liczby nad gałęziami reprezentują wsparcie bootstrapowe uzyskane przez maksymalne prawdopodobieństwo (wartość ML)



Rycina 37. Konsensusowe drzewo filogenetyczne przedstawicieli z rodzaju *Erica* skontruowane analizą bayesowską dla kompletnych genów *psb*. Wartości bootstrap poniżej 50% są ukryte. Liczby nad gałęziami reprezentują wsparcie bootstrapowe uzyskane przez maksymalne prawdopodobieństwo (wartość ML)



Rycina 38. Konsensusowe drzewo filogenetyczne przedstawicieli z rodzaju *Erica* skontruowane analizą bayesowską dla kompletnych genów *trn*. Wartości bootstrap poniżej 50% są ukryte. Liczby nad gałęziami reprezentują wsparcie bootstrapowe uzyskane przez maksymalne prawdopodobieństwo (wartość ML)

Drzewo filogenetyczne zbudowane dla genów *rbc*L, *mat*K i *ccs*A jest utworzone z silnie wspieranych kladów z wartościami ładowania początkowego sięgającymi 100% (Ryciny 23, 24 i 25).

W odróżnieniu od drzewa filogenetycznego zbudowanego z sekwencji ITS (Rycina 21A), pomimo występowania niskiego poziomu obsługiwanych kladów w poniektórych przypadkach, nigdzie nie wystąpiło takie drzewo filogenetyczne, by wszystkie jego wartości bootstrap sięgały poniżej 60%.

Istotnym jest, iż skonfrontowane drzewa filogenetyczne wskazują na zauważalną rozbieżność w odniesieniu do lokalizacji przedstawicieli rodzaju. Na przykład rekonstrukcja filogenetyczna pochodząca z genu *acc*D wykazuje ścisły związek między *Erica globiceps* i *Erica vagans*, podczas gdy rozdzielczość kladogramu skonstruowanego w obrębie genów *cem*A lokuje te gatunki w osobnych liniach o znacznej wiarygodności (Ryciny 26 i 27).

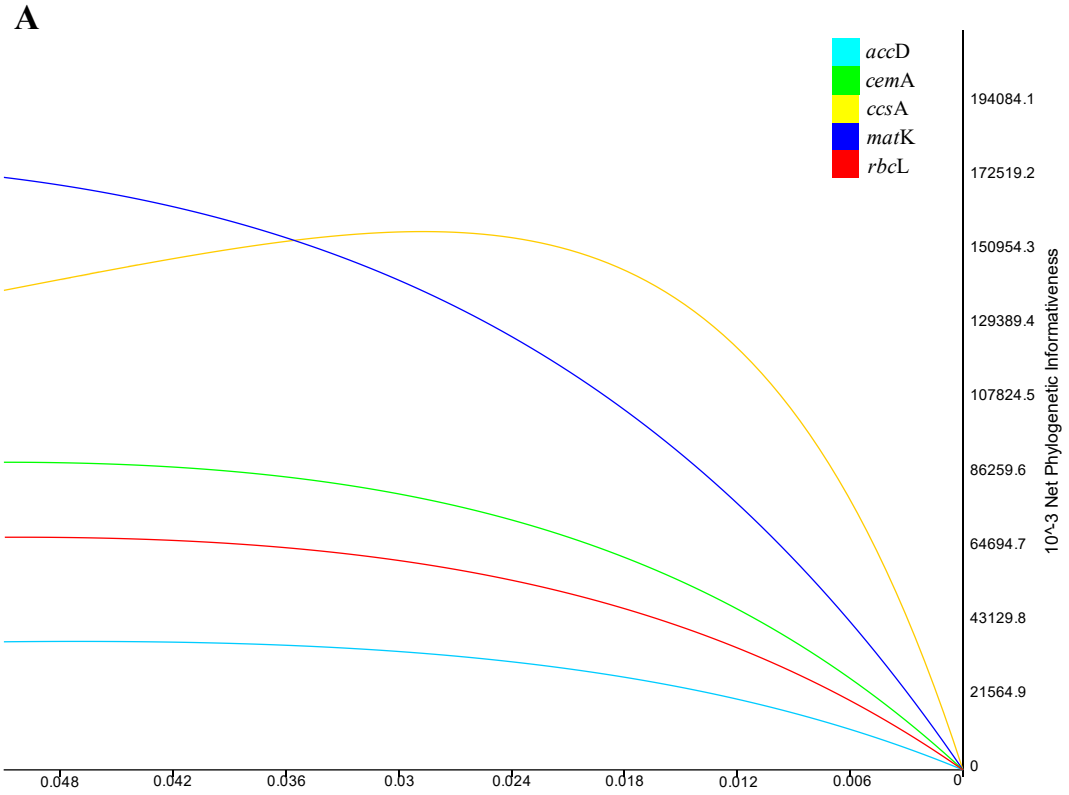
Drzewa filogenetyczne skonstruowane na pojedycznych genach, ze znaczącym wsparciem dla kładów, wykazały jednoznacznie, iż gatunkiem ancestralnym może być *Erica arborea*. Analizy przeprowadzone na poszczególnych grupach genów chloroplastowych potwierdziły tę hipotezę w przypadkach analiz opartych na genach: *pet* (Rycina 30), *psa* (Rycina 34), *psb* (Rycina 37) oraz *trn* (Rycina 38), statystycznie jest to najczęściej wyłaniany gatunek, będący w najwcześniejszej linii ewolucyjnej.

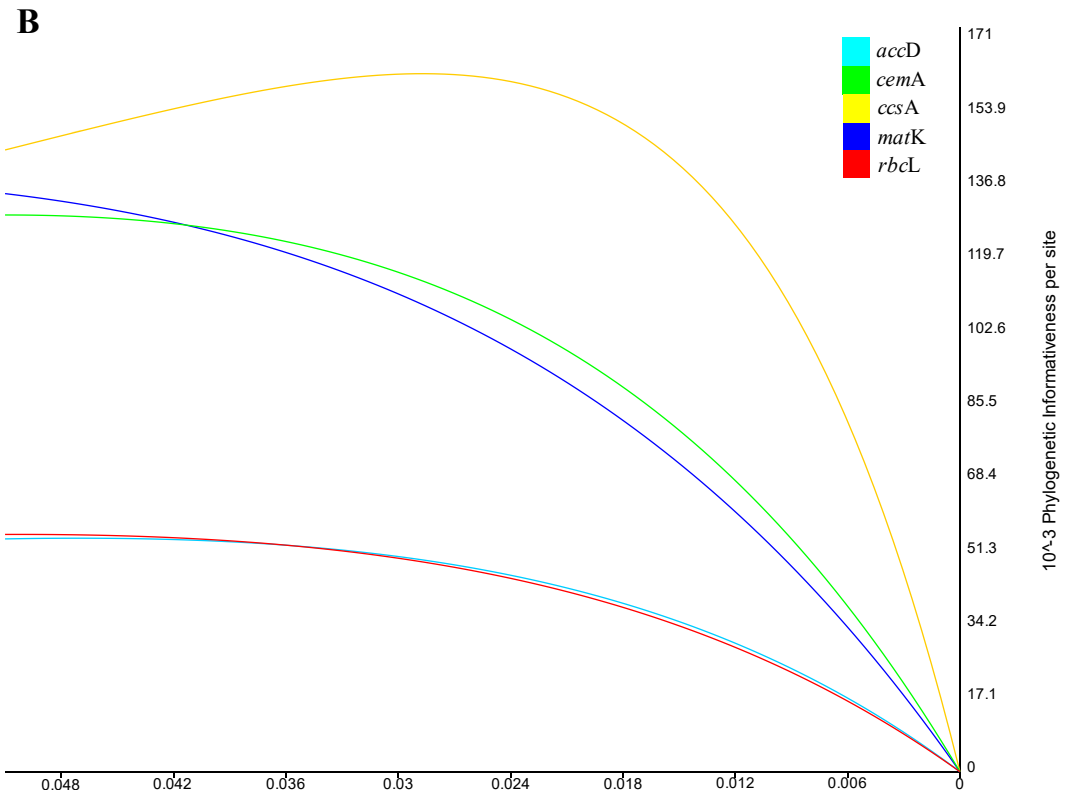
4.5.1. Informatywność filogenetyczna poszczególnych grup genów chloroplastowych w rodzinie *Ericaceae*

Badano kompletne sekwencje kodujące genomów chloroplastowych rodzaju *Erica* dla gatunków *Erica arborea, Erica cinerea, Erica globiceps, Erica imbricata, Erica multiflora, Erica plukenetii, Erica vagans* i *Erica versicolor*, w celu opisania profilu informatywności filogenetycznej (PI).

Zestaw danych z rodzaju *Erica* został zbadany osobno: dla genów *acc*D, *cem*A, *ccs*A, *mat*K i *rbc*L, dla genów kodujących rybosomalne RNA, genów kodujących podjednostki polimerazy RNA, genów kodujących podjednostki syntazy ATP, genów kodujących podjednostki kompleksu b/f cytochromu, genów kodujących podjednostki dehydrogenazy NADH, genów kodujących białka dużej podjednostki rybosomu, genów kodujących białka małej podjednostki rybosomu, genów kodujących podjednostki fotosystemu I i II, genów kodujących transportujące RNA oraz dla hipotetycznych genów (Ryciny 39-50).

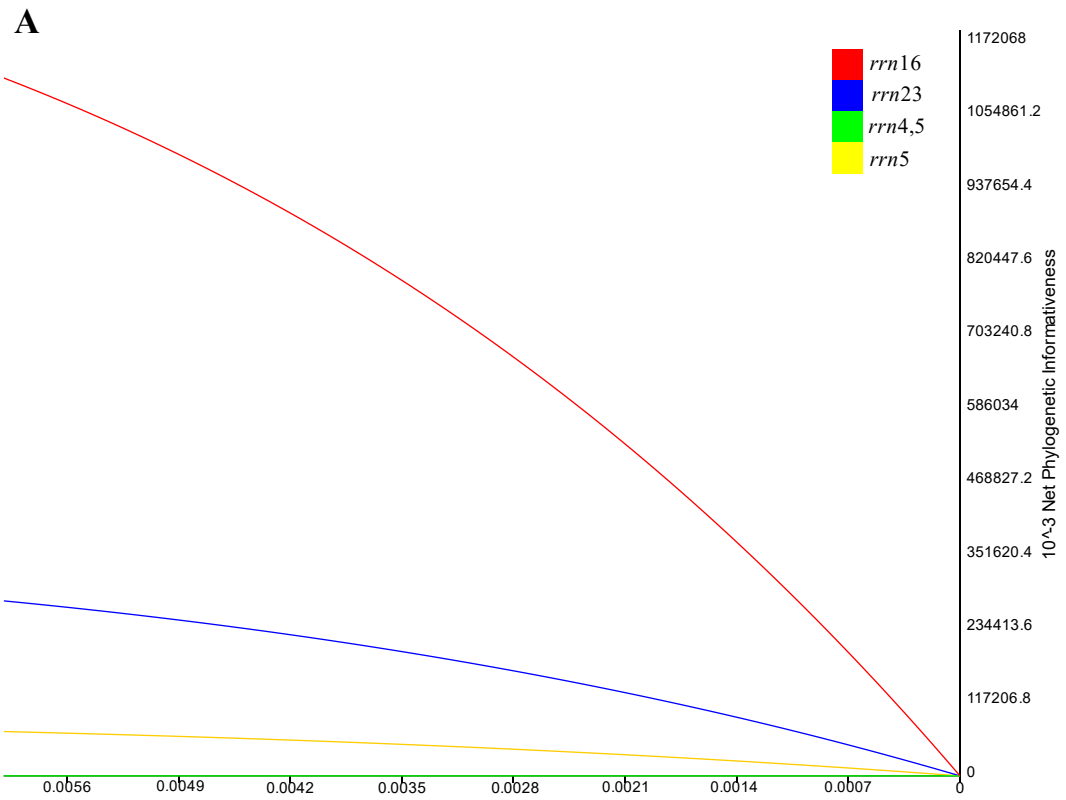
Grupa genów kodujących karboksylazę acetylo-CoA, białko jądrowej membrany, Białko cytochromu c, maturazę K i dużą podjednostkę Rubisco

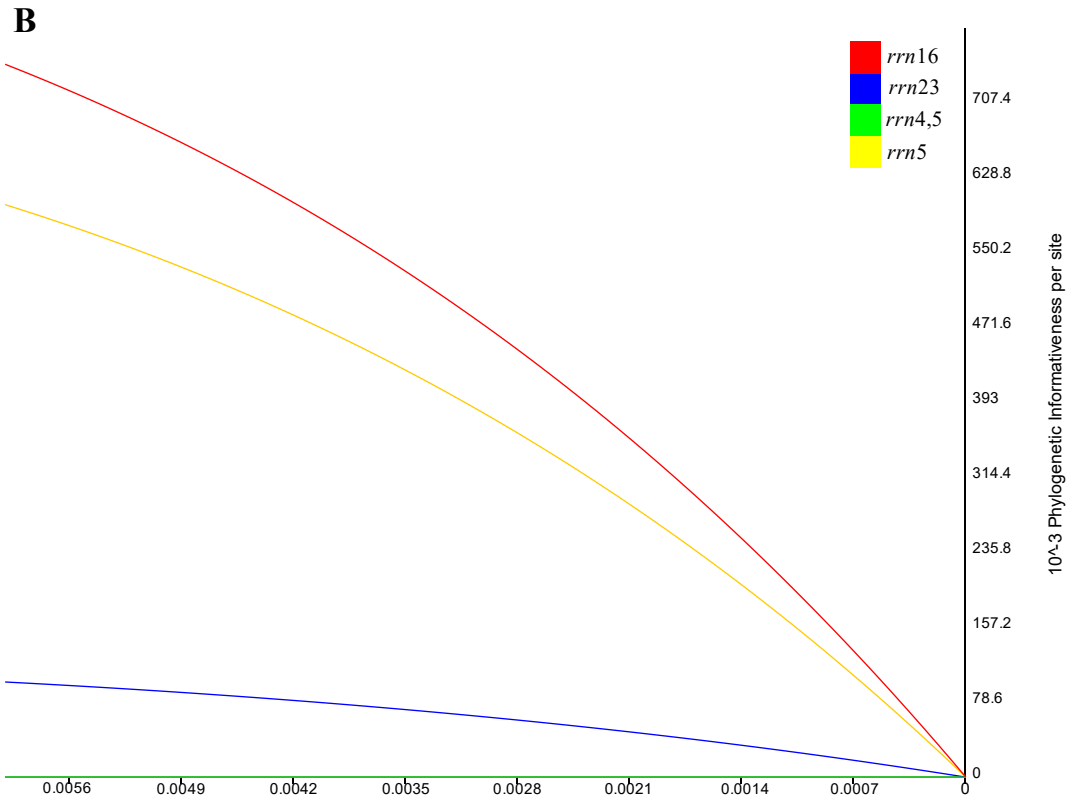




Rycina 39. Informatywność filogenetyczna. Profil pierwszy (A) ogólny sieciowy, drugi (B) powstał w przeliczeniu na miejsca filogenetycznej informatywności dla wybranych regionów w obrębie genów *acc*D, *cem*A, *ccs*A, *mat*K, *rbc*L. Oś czasu została skalibrowana do drzewka ML o ultrametryzowanym obliczeniu wyprowadzony ze zbioru danych wszystkich badanych genów wykorzystanych do kalibracji osi czasu. Czas ewolucji wynoszący 1,0 odpowiada korzeniu drzewa, a wartość 0 jego wierzchołkom

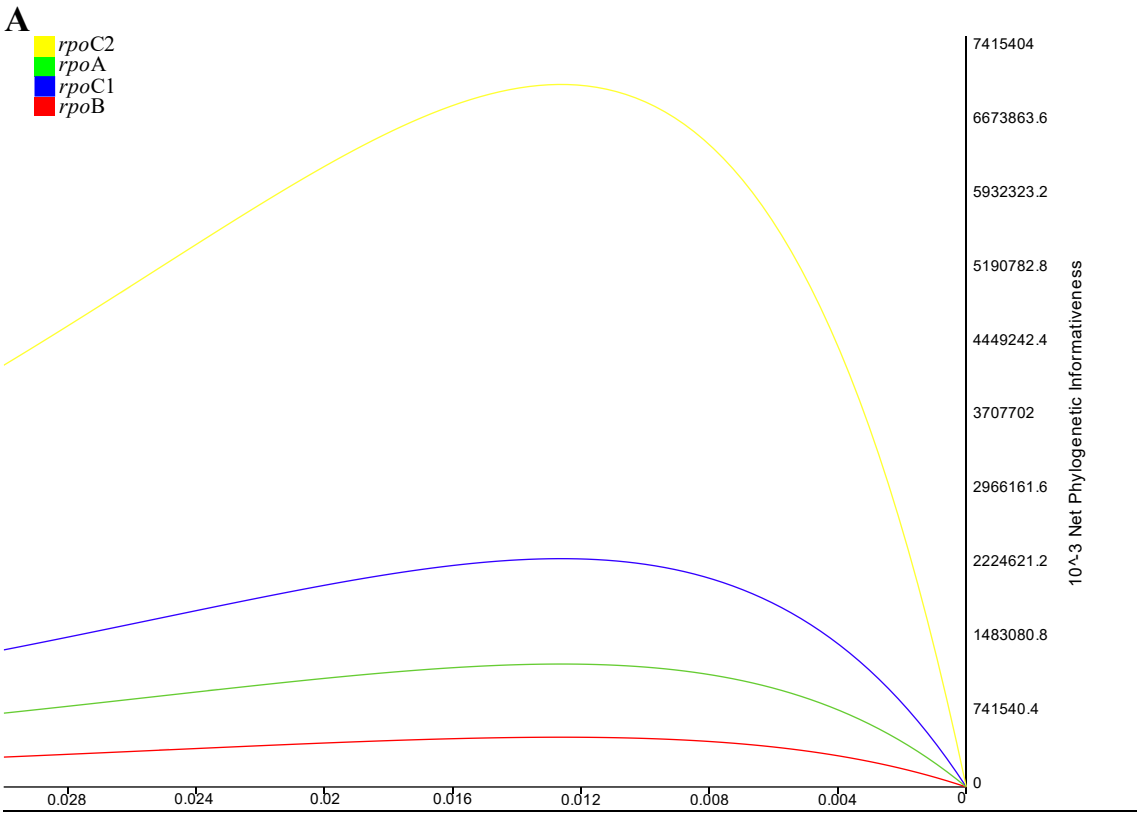
Geny kodujące rybosomowe RNA

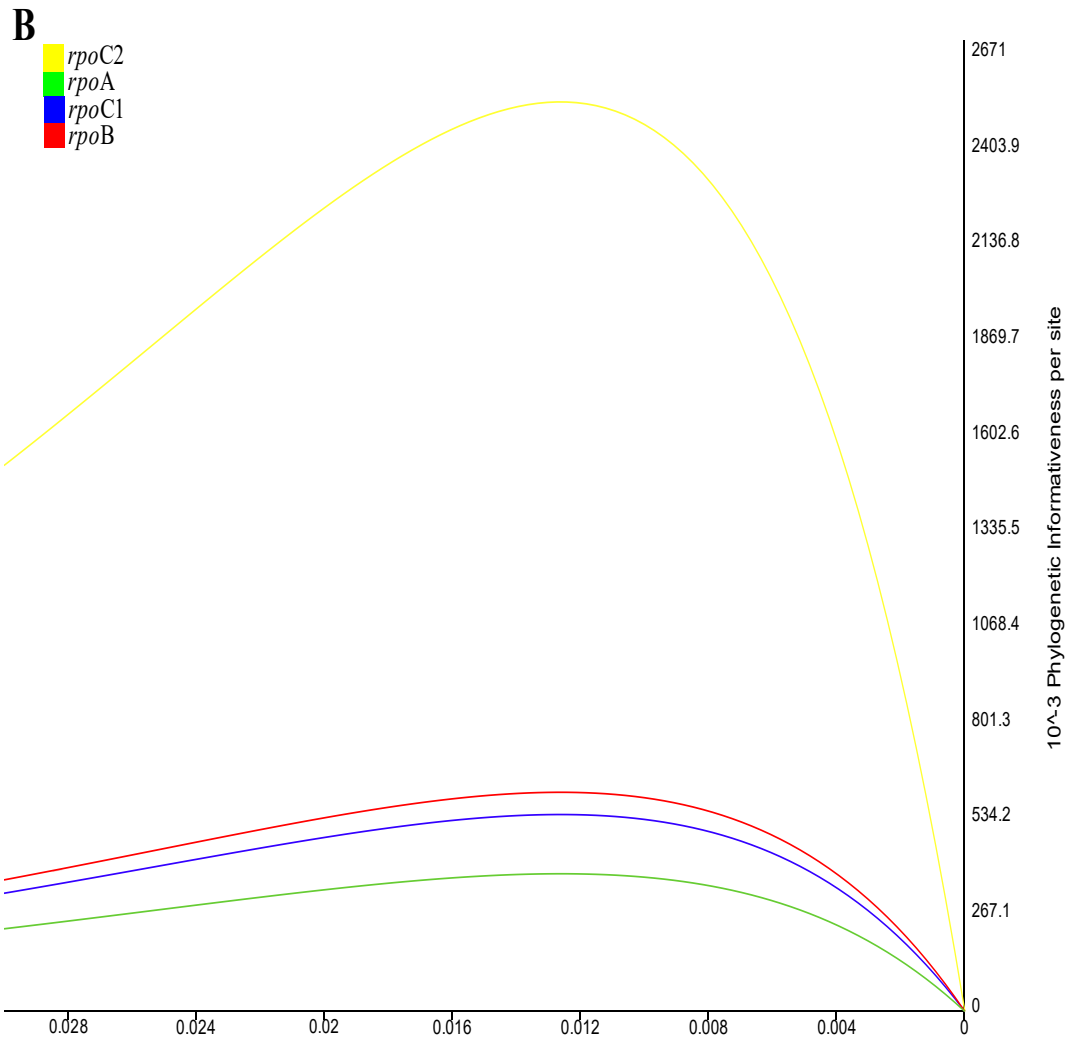




Rycina 40. Informatywność filogenetyczna. Profil pierwszy (A) ogólny sieciowy, drugi (B) powstał w przeliczeniu na miejsca filogenetycznej informatywności dla wybranych regionów w obrębie genów *rrn*. Oś czasu została skalibrowana do drzewka ML o ultrametryzowanym obliczeniu wyprowadzony ze zbioru danych wszystkich badanych genów wykorzystanych do kalibracji osi czasu. Czas ewolucji wynoszący 1,0 odpowiada korzeniu drzewa, a wartość 0 jego wierzchołkom

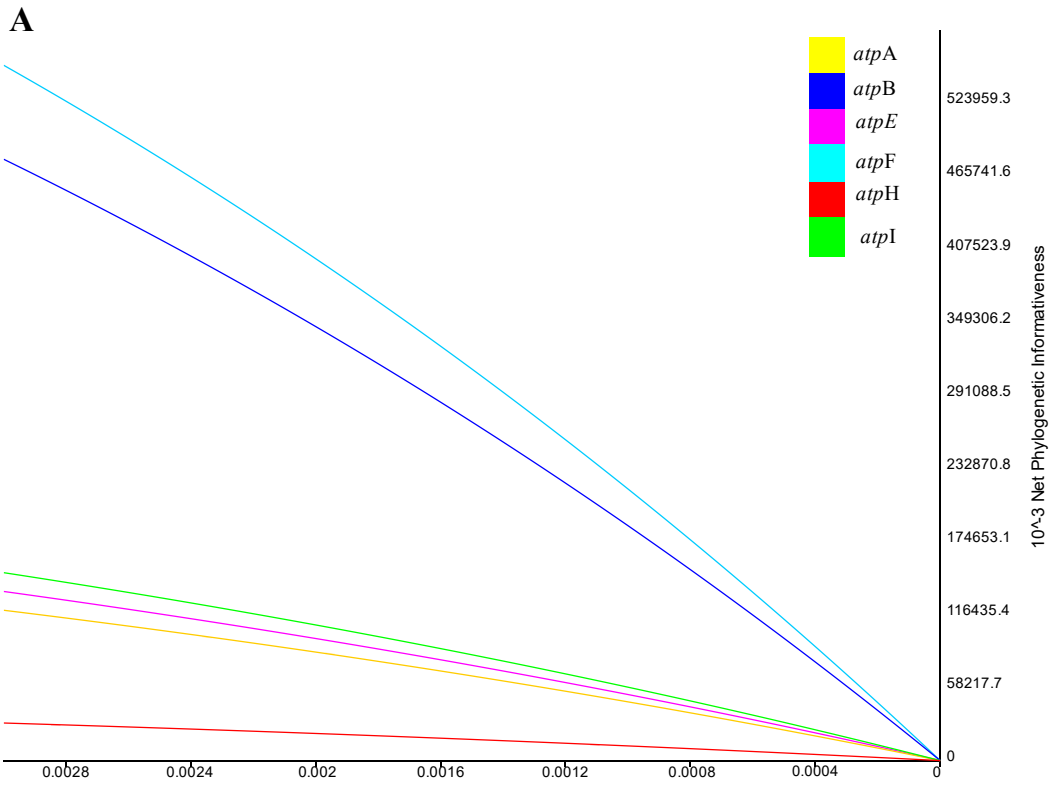
Geny kodujące podjednostki polimerazy RNA

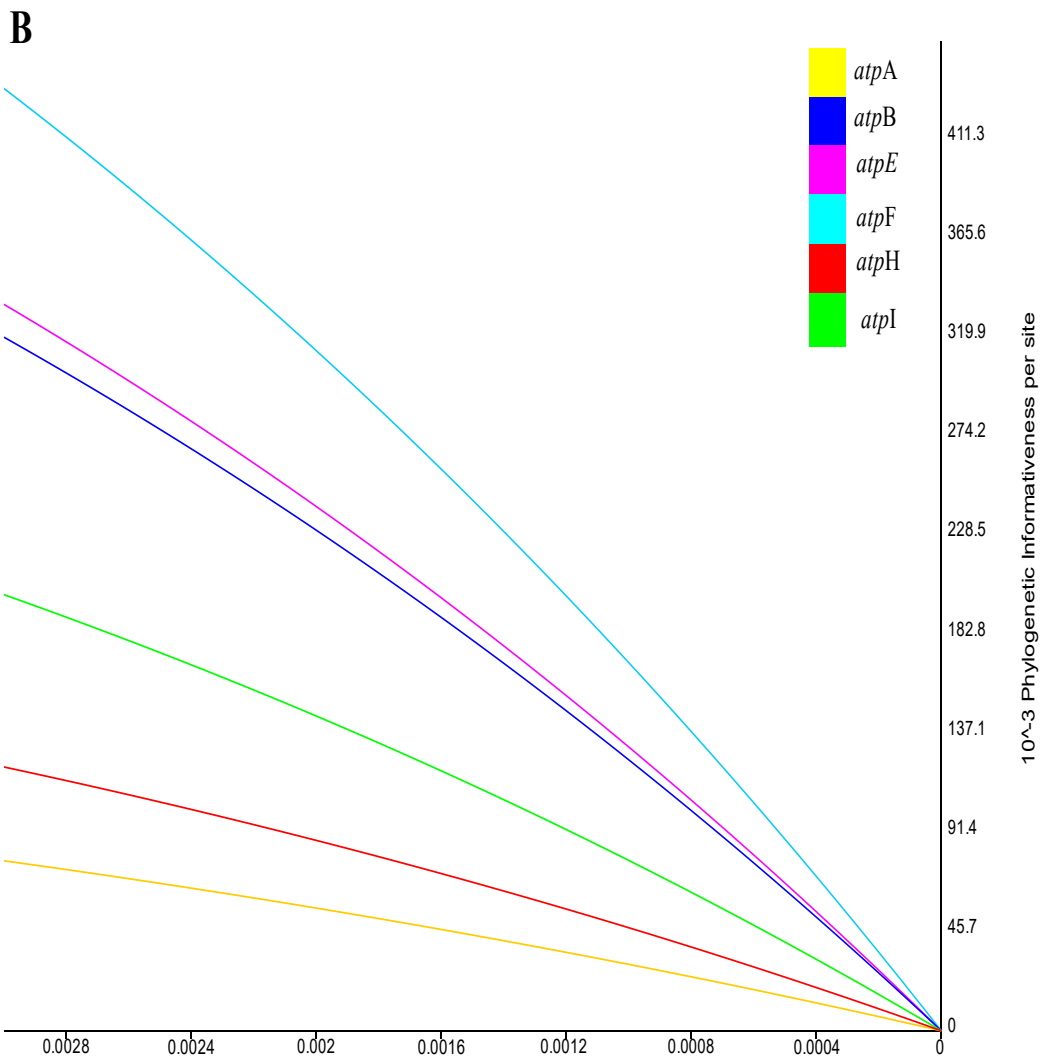




Rycina 41. Informatywność filogenetyczna. Profil pierwszy (A) ogólny sieciowy, drugi (B) powstał w przeliczeniu na miejsca filogenetycznej informatywności dla wybranych regionów w obrębie genów *rpo*. Oś czasu została skalibrowana do drzewka ML o ultrametryzowanym obliczeniu wyprowadzony ze zbioru danych wszystkich badanych genów *rpo* wykorzystanych do kalibracji osi czasu. Czas ewolucji wynoszący 1,0 odpowiada korzeniu drzewa, a wartość 0 jego wierzchołkom

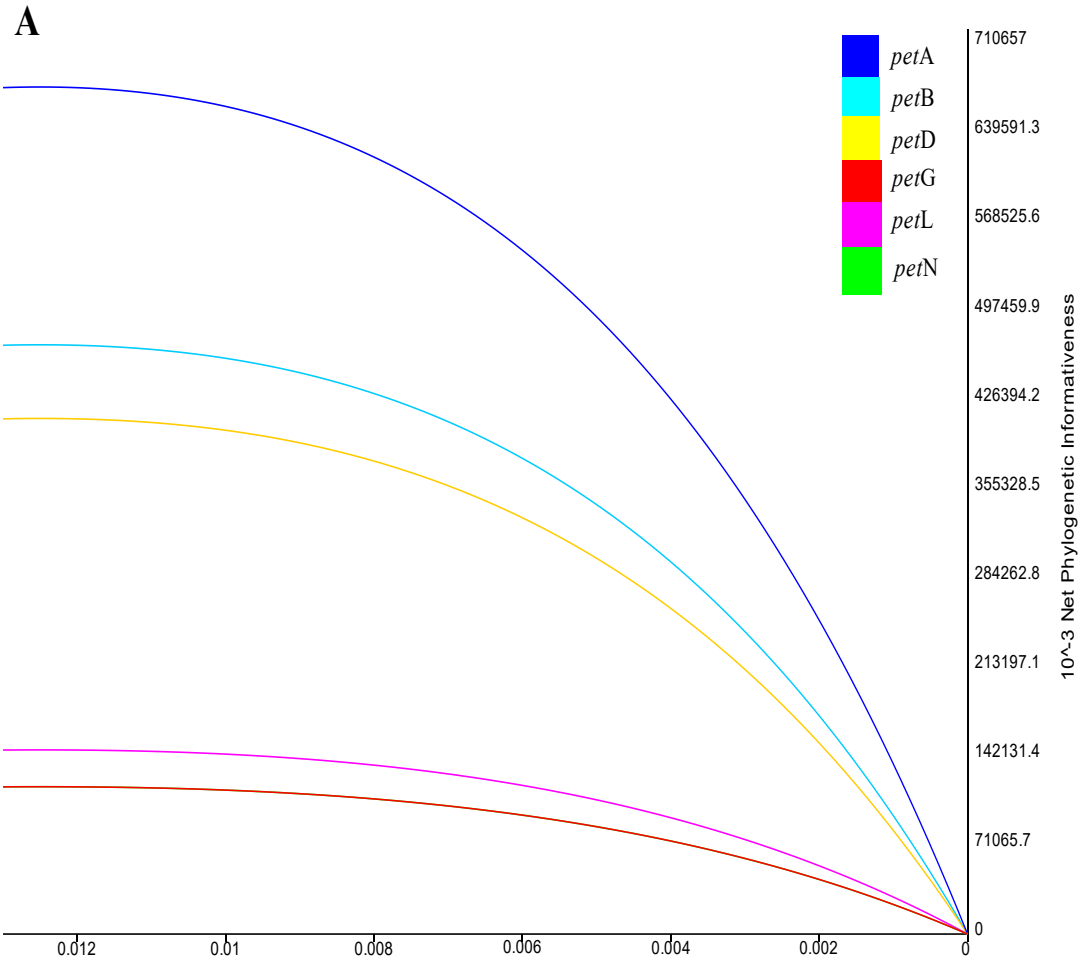
Geny kodujące podjednostki syntazy ATP

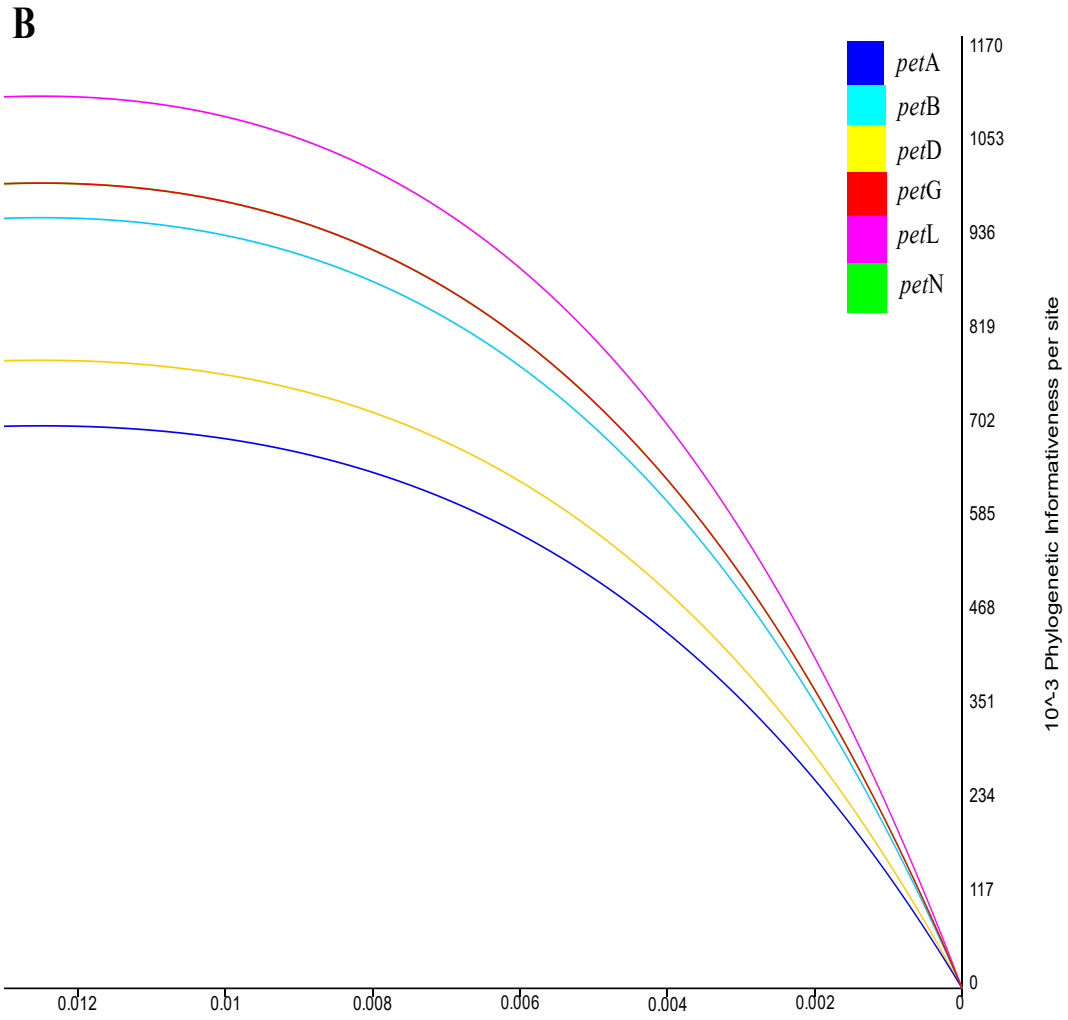




Rycina 42. Informatywność filogenetyczna. Profil pierwszy (A) ogólny sieciowy, drugi (B) powstał w przeliczeniu na miejsca filogenetycznej informatywności dla wybranych regionów w obrębie genów *atp*. Oś czasu została skalibrowana do drzewka ML o ultrametryzowanym obliczeniu wyprowadzony ze zbioru danych wszystkich badanych genów *atp* wykorzystanych do kalibracji osi czasu. Czas ewolucji wynoszący 1,0 odpowiada korzeniu drzewa, a wartość 0 jego wierzchołkom

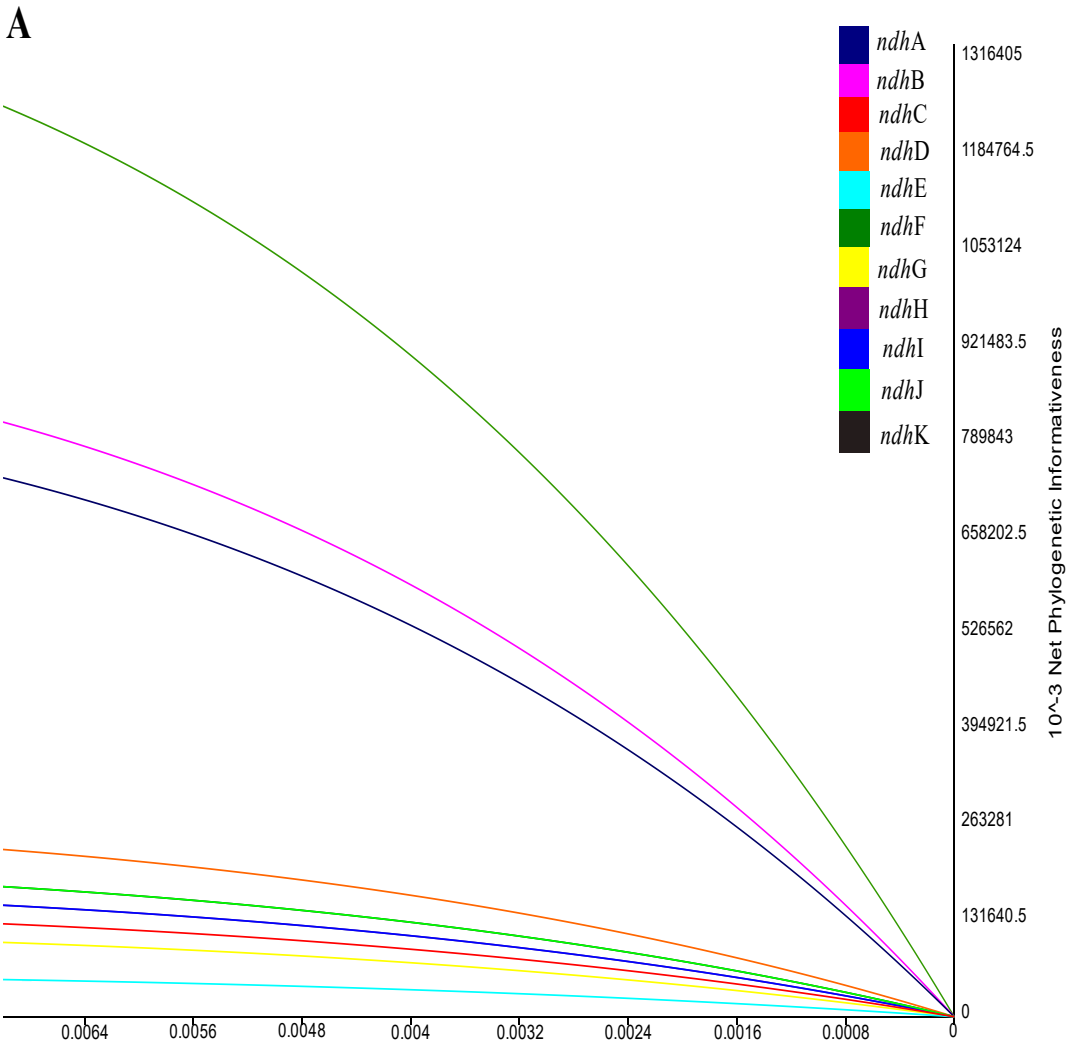
Geny kodujące podjednostki kompleksu b/f cytochromu

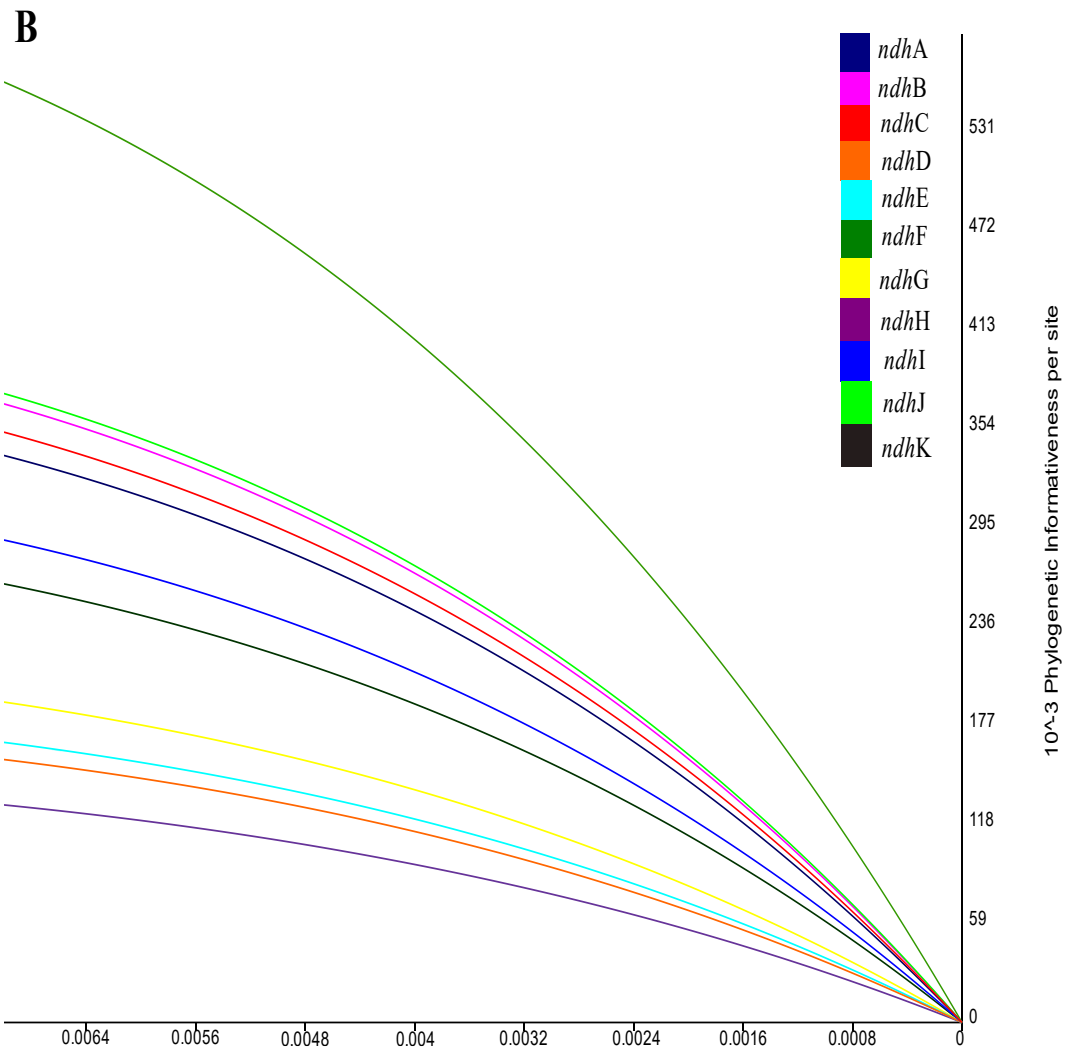




Rycina 43. Informatywność filogenetyczna. Profil pierwszy (A) ogólny sieciowy, drugi (B) powstał w przeliczeniu na miejsca filogenetycznej informatywności dla wybranych regionów w obrębie genów *pet*. Oś czasu została skalibrowana do drzewka ML o ultrametryzowanym obliczeniu wyprowadzony ze zbioru danych wszystkich badanych genów *pet* wykorzystanych do kalibracji osi czasu. Czas ewolucji wynoszący 1,0 odpowiada korzeniu drzewa, a wartość 0 jego wierzchołkom

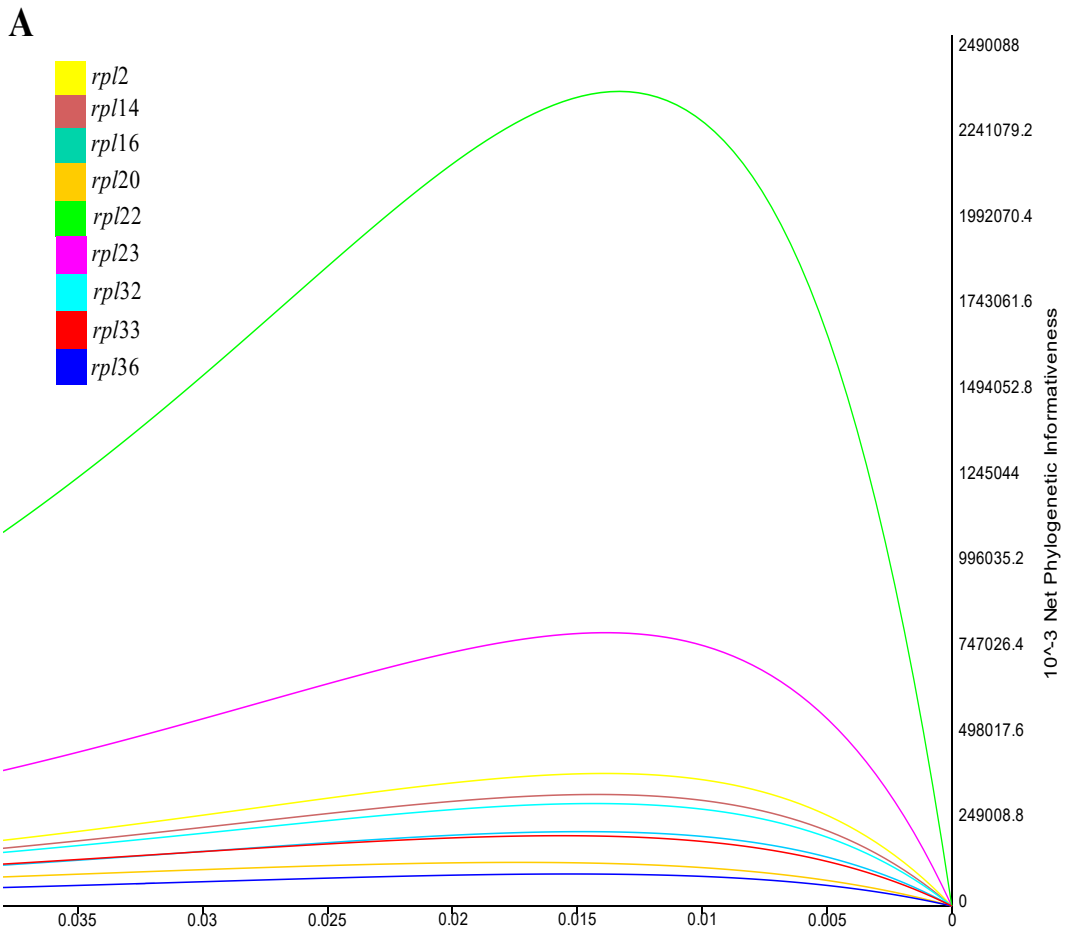
Geny kodujące podjednostki dehydrogenazy NADH

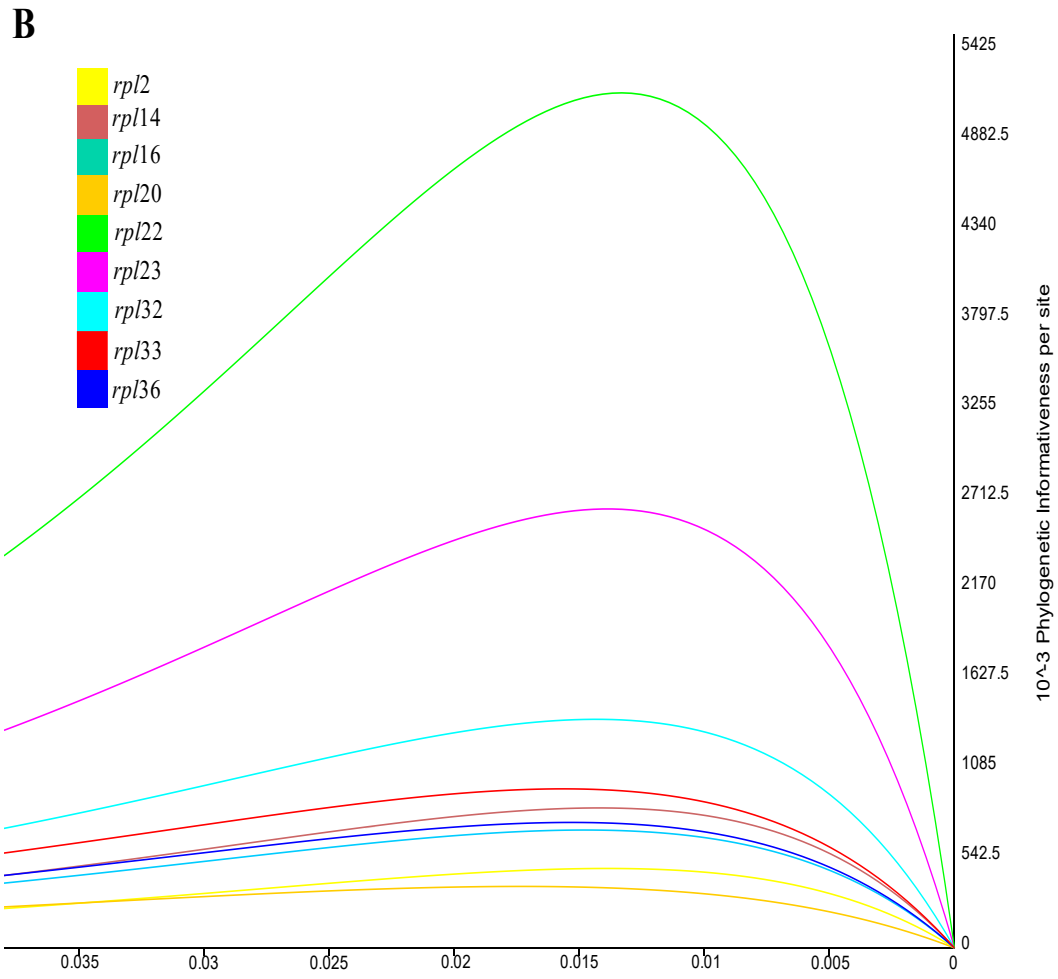




Rycina 44. Informatywność filogenetyczna. Profil pierwszy (A) ogólny sieciowy, drugi (B) powstał w przeliczeniu na miejsca filogenetycznej informatywności dla wybranych regionów w obrębie genów *ndh*. Oś czasu została skalibrowana do drzewka ML o ultrametryzowanym obliczeniu wyprowadzony ze zbioru danych wszystkich badanych genów *ndh* wykorzystanych do kalibracji osi czasu. Czas ewolucji wynoszący 1,0 odpowiada korzeniu drzewa, a wartość 0 jego wierzchołkom

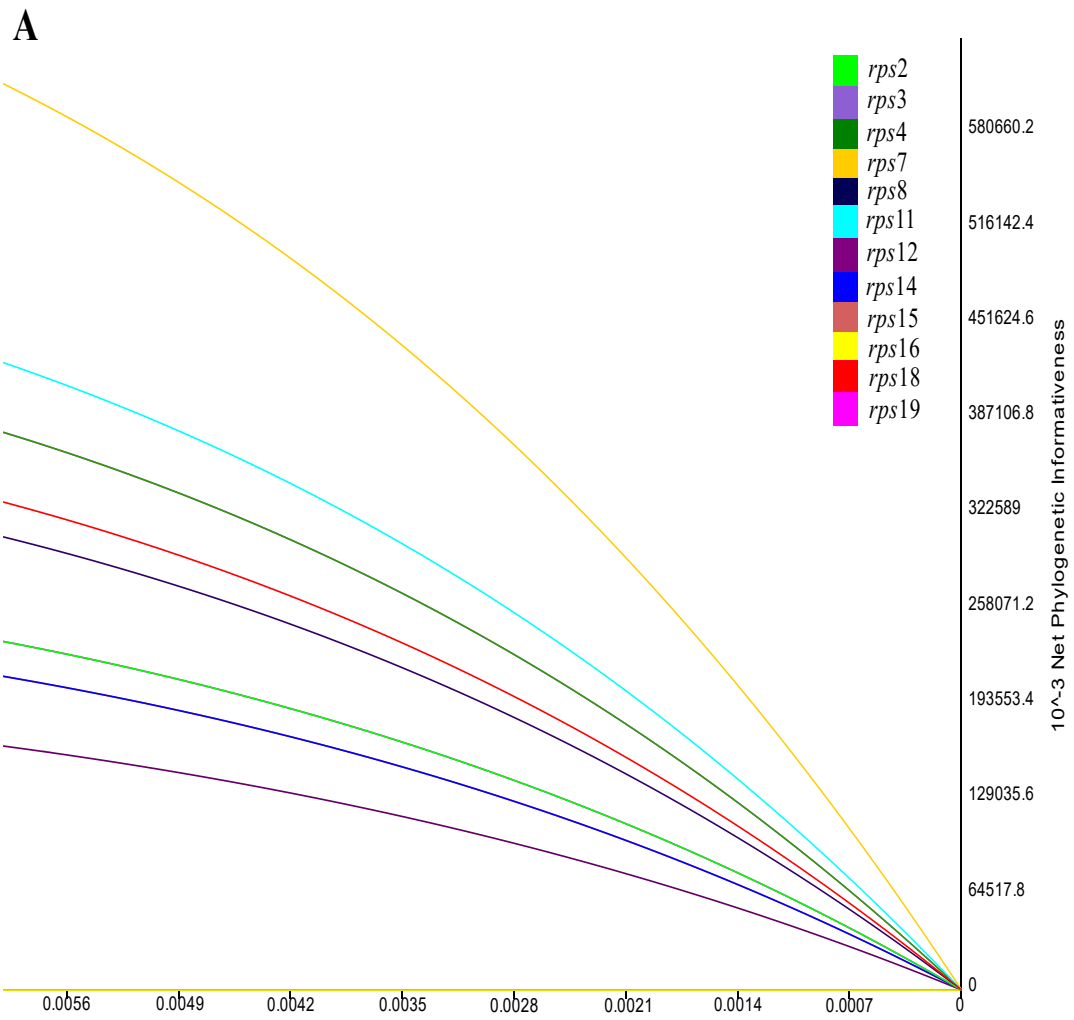
Geny kodujące białka dużej podjednostki rybosomu

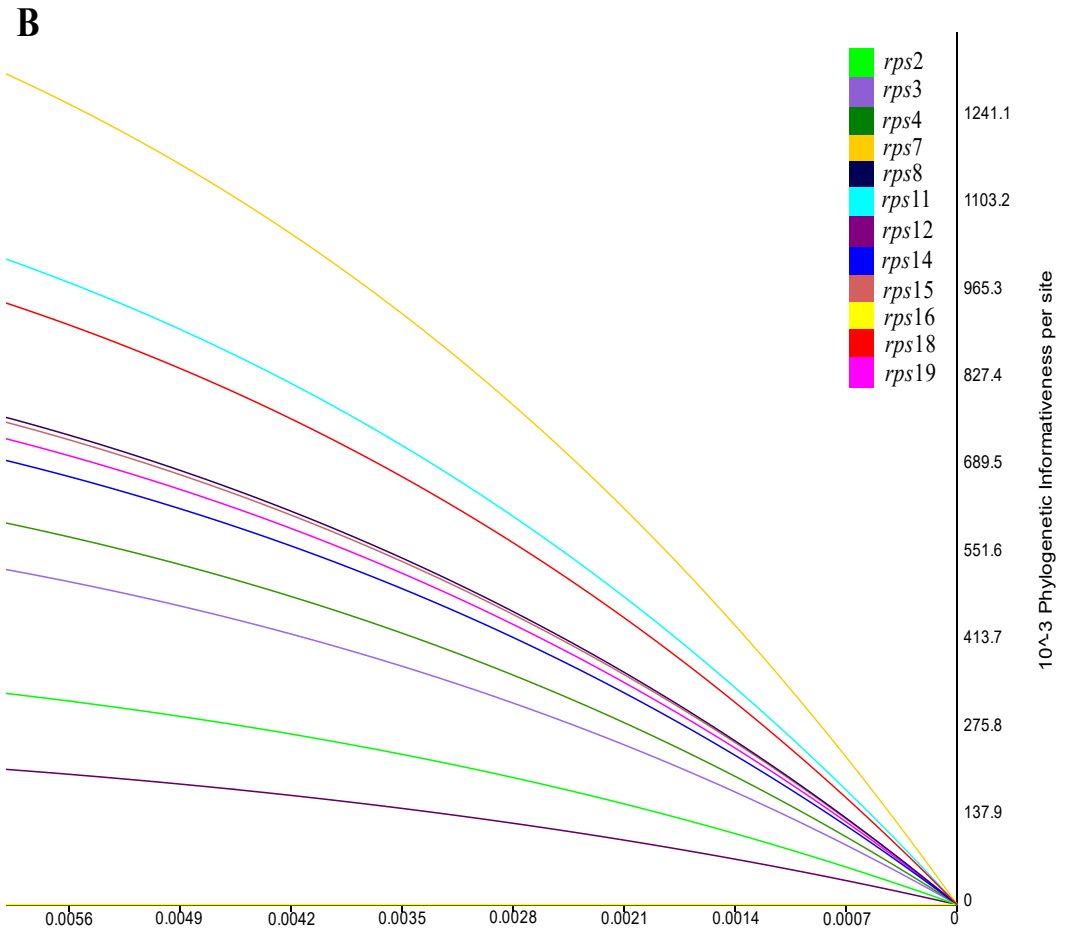




Rycina 45. Informatywność filogenetyczna. Profil pierwszy (A) ogólny sieciowy, drugi (B) powstał w przeliczeniu na miejsca filogenetycznej informatywności dla wybranych regionów w obrębie genów *rpl*. Oś czasu została skalibrowana do drzewka ML o ultrametryzowanym obliczeniu wyprowadzony ze zbioru danych wszystkich badanych genów *rpl* wykorzystanych do kalibracji osi czasu. Czas ewolucji wynoszący 1,0 odpowiada korzeniu drzewa, a wartość 0 jego wierzchołkom

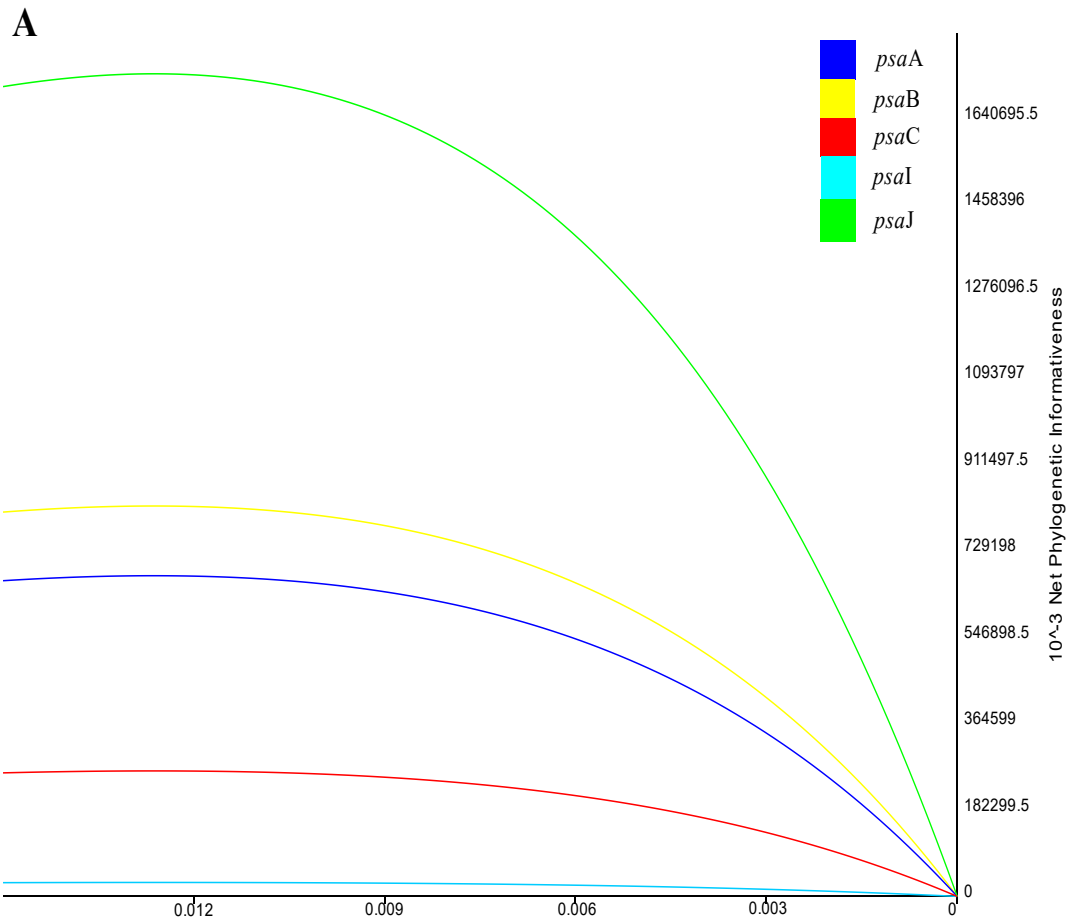
Geny kodujące białka małej podjednostki rybosomu

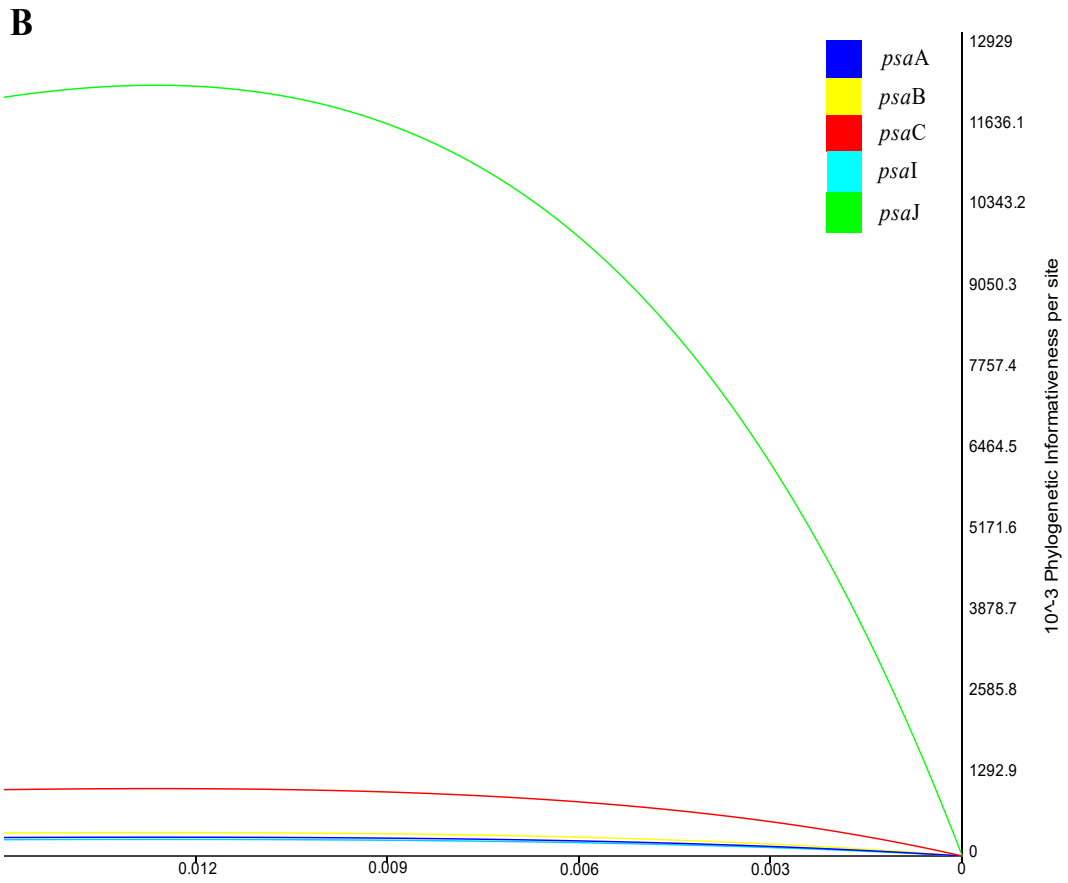




Rycina 46. Informatywność filogenetyczna. Profil pierwszy (A) ogólny sieciowy, drugi (B) powstał w przeliczeniu na miejsca filogenetycznej informatywności dla wybranych regionów w obrębie genów *rps*. Oś czasu została skalibrowana do drzewka ML o ultrametryzowanym obliczeniu wyprowadzony ze zbioru danych wszystkich badanych genów *rps* wykorzystanych do kalibracji osi czasu. Czas ewolucji wynoszący 1,0 odpowiada korzeniu drzewa, a wartość 0 jego wierzchołkom

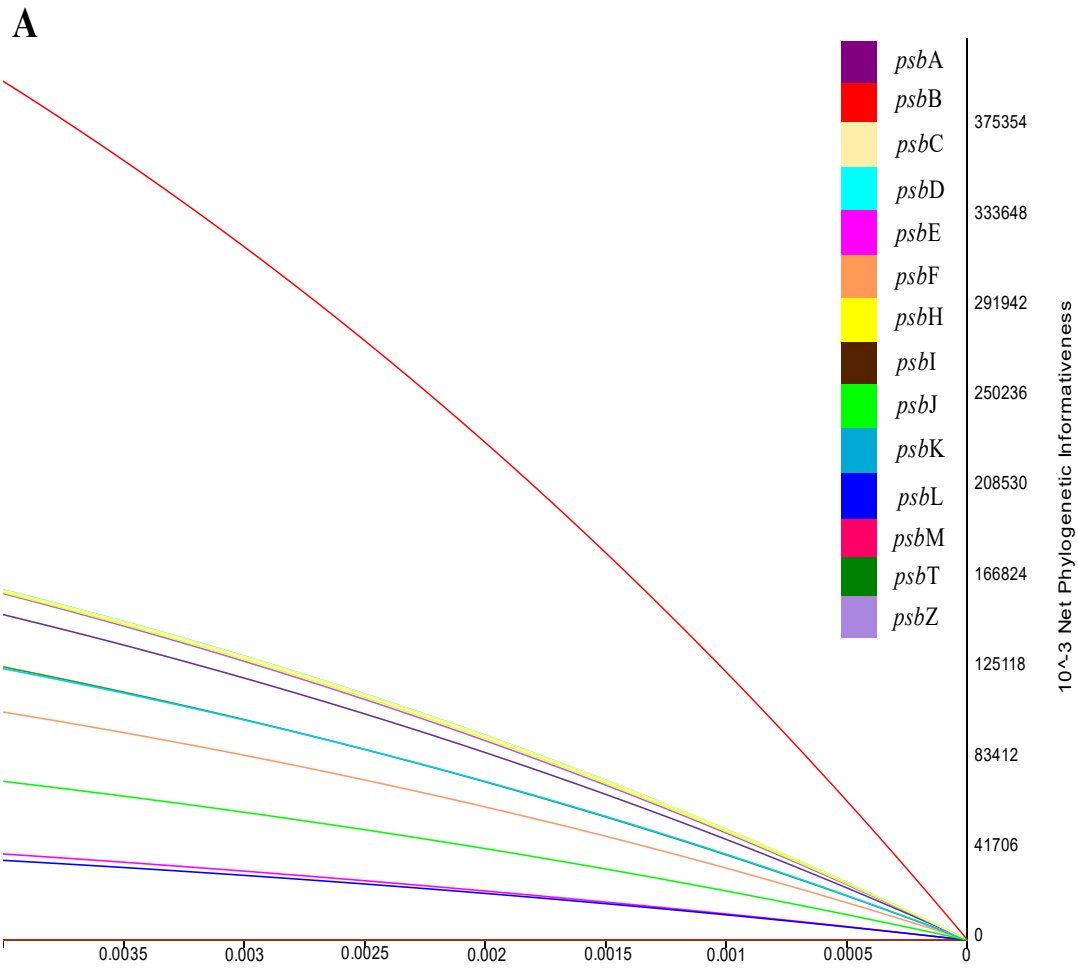
Geny kodujące podjednostki fotosystemu I

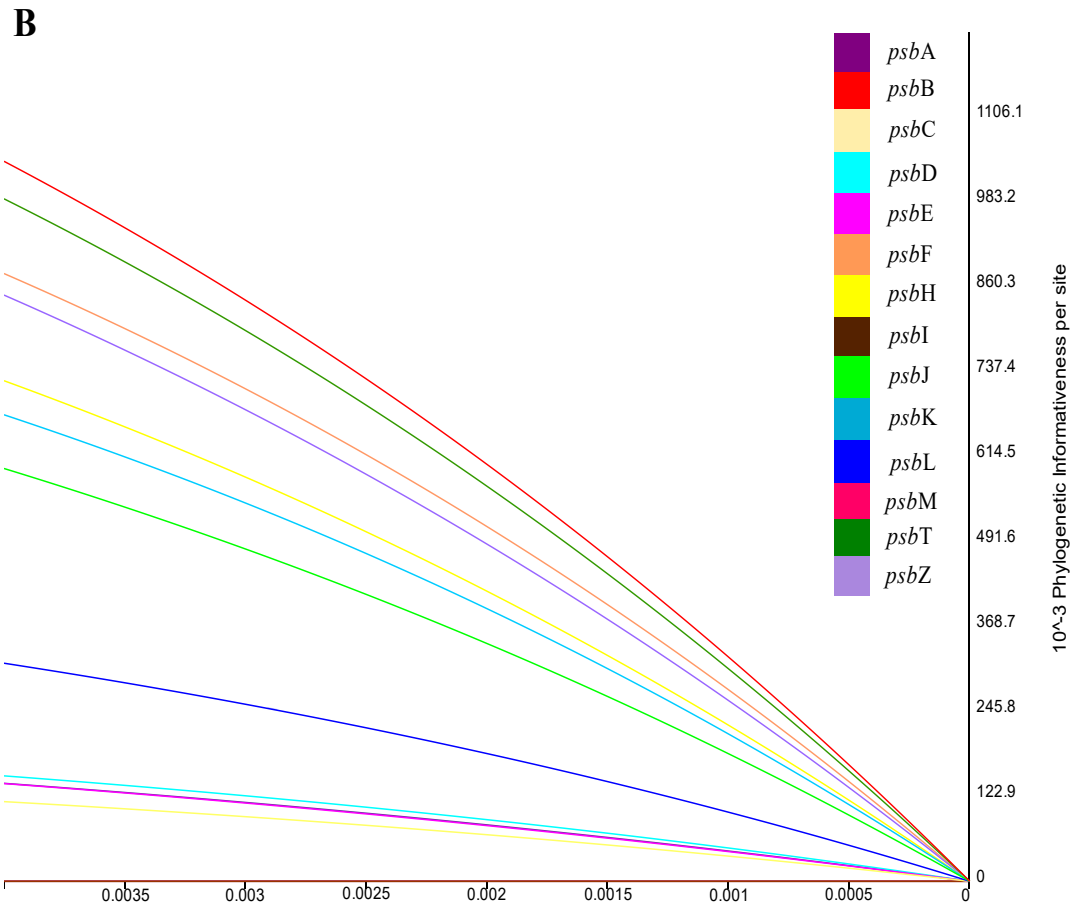




Rycina 47. Informatywność filogenetyczna. Profil pierwszy (A) ogólny sieciowy, drugi (B) powstał w przeliczeniu na miejsca filogenetycznej informatywności dla wybranych regionów w obrębie genów *psa*. Oś czasu została skalibrowana do drzewka ML o ultrametryzowanym obliczeniu wyprowadzony ze zbioru danych wszystkich badanych genów *psa* wykorzystanych do kalibracji osi czasu. Czas ewolucji wynoszący 1,0 odpowiada korzeniu drzewa, a wartość 0 jego wierzchołkom

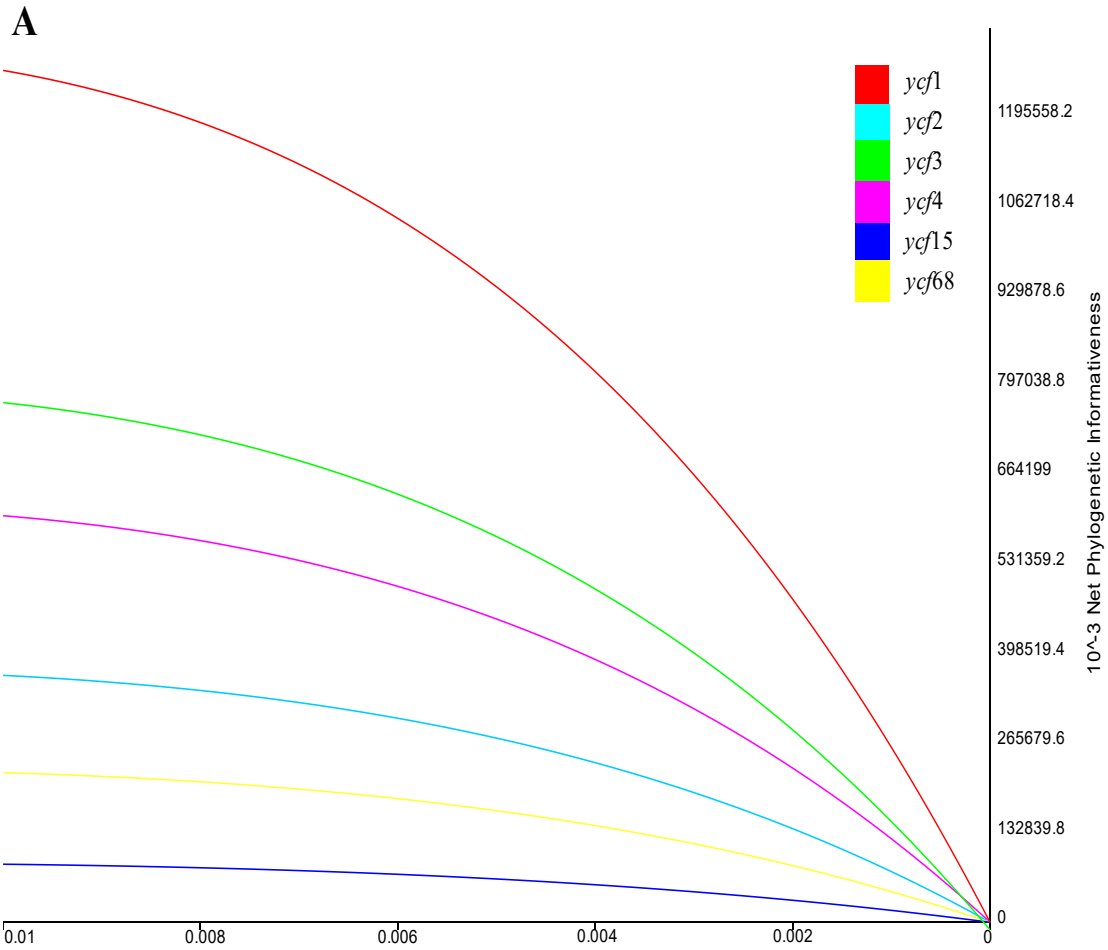
Geny kodujące podjednostki fotosystemu II

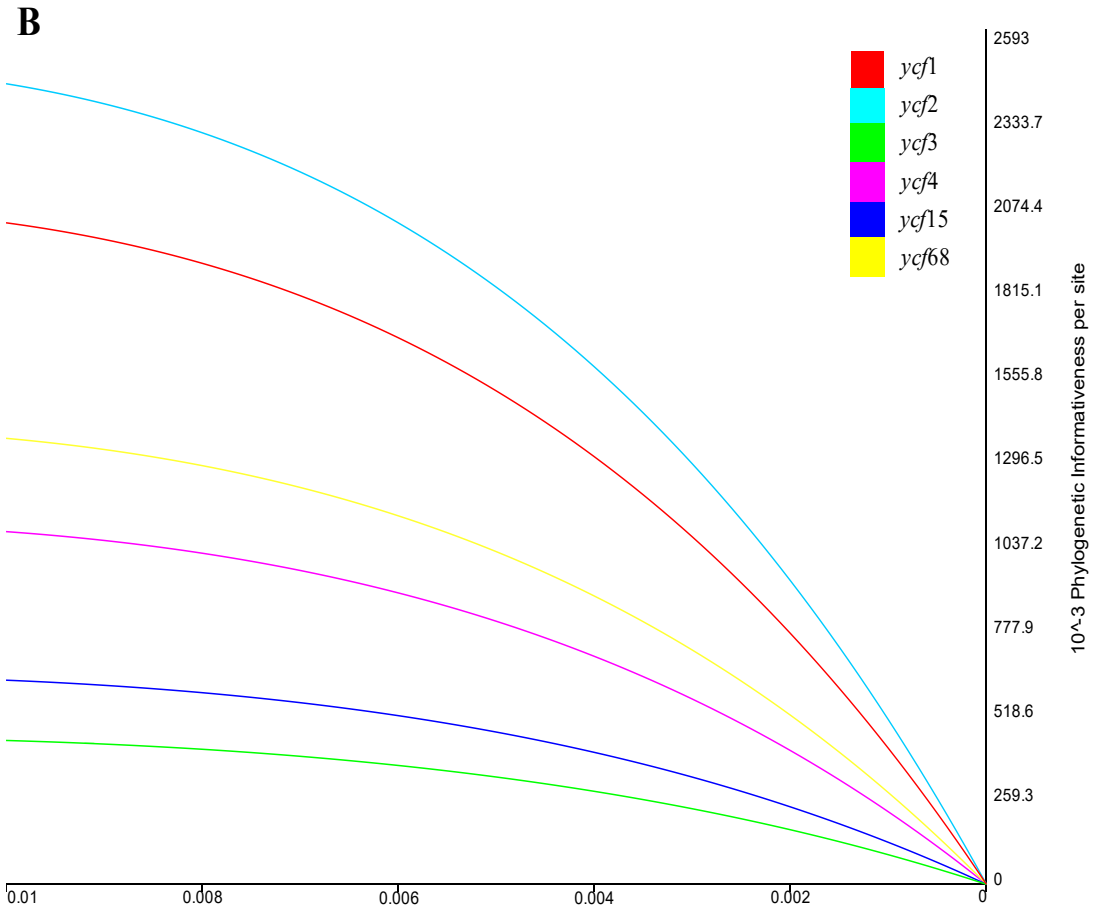




Rycina 48. Informatywność filogenetyczna. Profil pierwszy (A) ogólny sieciowy, drugi (B) powstał w przeliczeniu na miejsca filogenetycznej informatywności dla wybranych regionów w obrębie genów *psb*. Oś czasu została skalibrowana do drzewka ML o ultrametryzowanym obliczeniu wyprowadzony ze zbioru danych wszystkich badanych genów *psb* wykorzystanych do kalibracji osi czasu. Czas ewolucji wynoszący 1,0 odpowiada korzeniu drzewa, a wartość 0 jego wierzchołkom

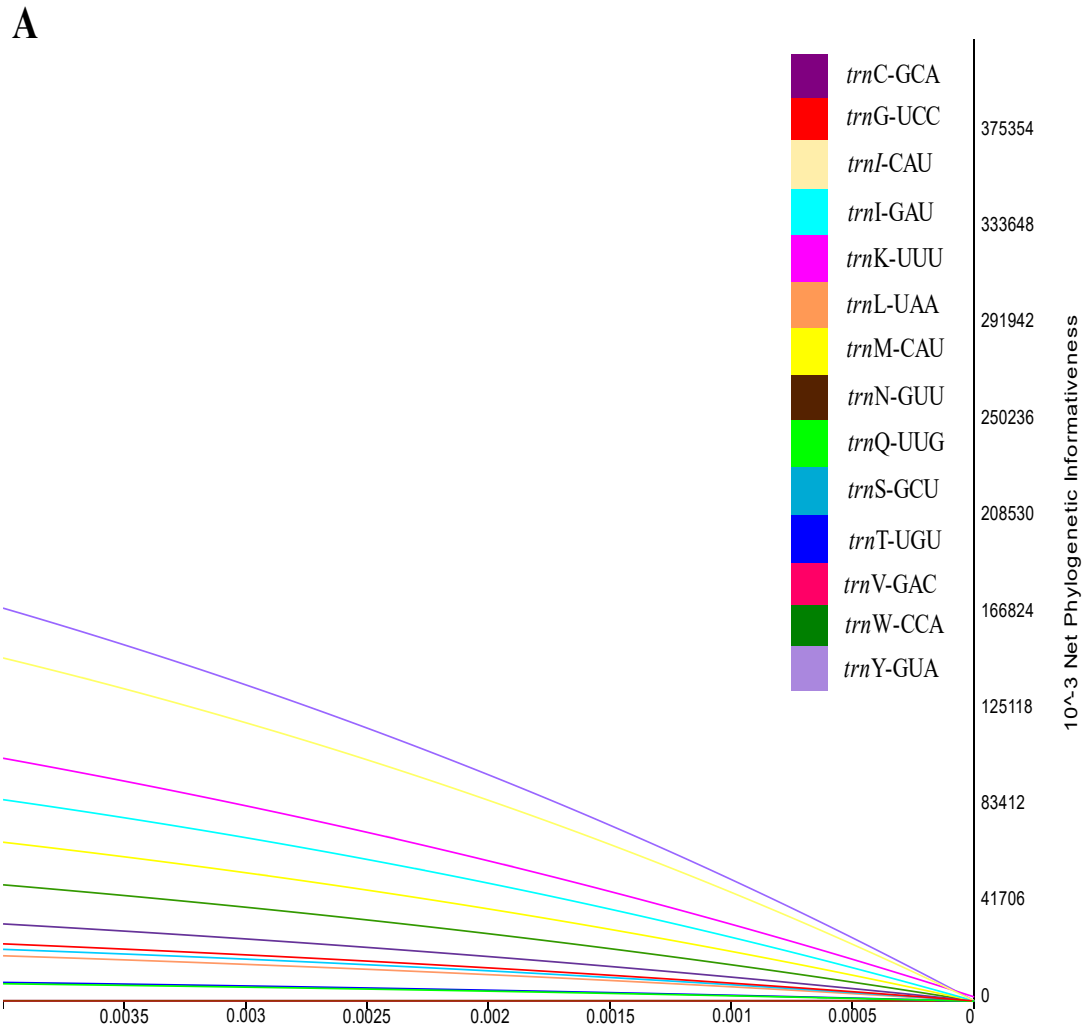
Hipotetyczne geny

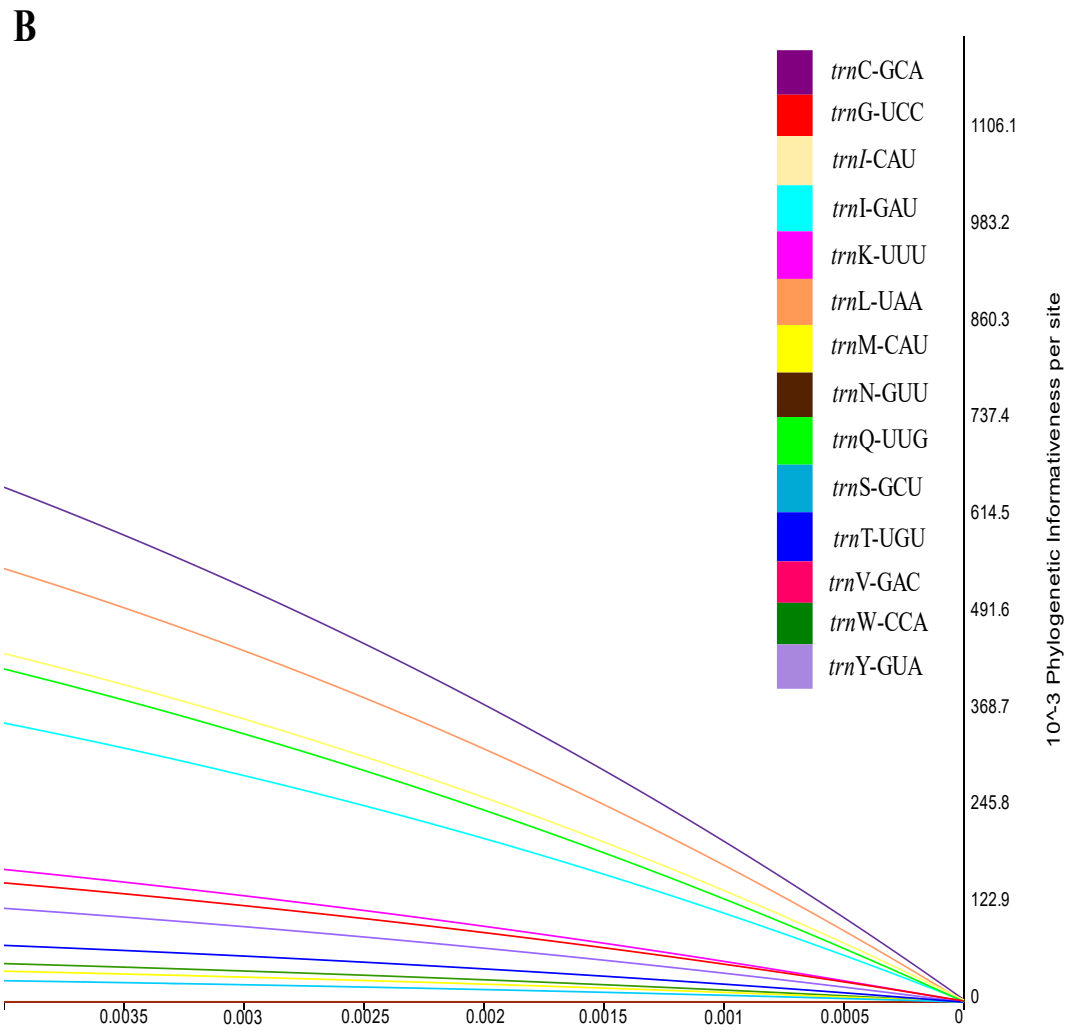




Rycina 49. Informatywność filogenetyczna. Profil pierwszy (A) ogólny sieciowy, drugi (B) powstał w przeliczeniu na miejsca filogenetycznej informatywności dla wybranych regionów w obrębie genów *ycf*. Oś czasu została skalibrowana do drzewka ML o ultrametryzowanym obliczeniu wyprowadzony ze zbioru danych wszystkich badanych genów *ycf* wykorzystanych do kalibracji osi czasu. Czas ewolucji wynoszący 1,0 odpowiada korzeniu drzewa, a wartość 0 jego wierzchołkom

Geny kodujące transportujące RNA





Rycina 50. Informatywność filogenetyczna. Profil pierwszy (A) ogólny sieciowy, drugi (B) powstał w przeliczeniu na miejsca filogenetycznej informatywności dla wybranych regionów w obrębie genów *trn*. Oś czasu została skalibrowana do drzewka ML o ultrametryzowanym obliczeniu wyprowadzony ze zbioru danych wszystkich badanych genów *ycf* wykorzystanych do kalibracji osi czasu. Czas ewolucji wynoszący 1,0 odpowiada korzeniu drzewa, a wartość 0 jego wierzchołkom

Profil badanej informatywności filogenetycznej sieciowej dla wybranych genów w obrębie badanych gatunków, obejmujący kompleksową informatywność regionów kodujących białka, wykazał bardzo niski sygnał filogenetyczny podtrzymywany przez geny *trn*, z których duża część nie została nawet uwzględniona w obszarze grafów. Niewystarczająco wysoki stopień PI ogólnego ujawniono również w regionie *rrn*4,5. Piki tego profilu nakładały się na ramy osi kalibracyjnej. Najwyższy zakres PI netto uzyskano dla genów *rpo*C2, *ndh*F, *rpl*22, *rps*16, *rrn*16, *psa*J, *ycf*1, *rpl*22, *atp*H, *atp*B, *pet*A oraz *psb*B z dużą różnicą w stosunku do poziomu następnego regionu w wartości sieciowej PI (Ryciny A 40-49).

Rozdzielczość profilu PI w przeliczeniu na miejsce potwierdza najniższy poziom informatywności dla genu *rrn*4,5 i dla wielu pojedynczych genów z poszczególnych grup. Potwierdza także najwyższy poziom informatywności odpowiednio dla genów *rpo*C2, *ndh*F, *rpl*22, *rps*16, *rrn*16, *psa*J i *psb*B (Ryciny B 40-41, 44-48). W związku z tym, jeśli chodzi o wynik informatywności filogenetycznej w przeliczeniu na miejsce i ogólnej, najbardziej obiecującymi markerami filogenetycznym dla rodzaju *Erica* okazały się właśnie geny *rpo*C2, *ndh*F, *rpl*22, *rps*16, *rrn*16, *psa*J, *psb*B oraz *atp*F.

Gen *rpo*C2 okazał się ponownie obiecującym markerem genetycznym, potwierdzając tym samym wyniki badań z roku 2017, gdy analizie poddano 50 przedstawicieli z rodzaju *Erica*, a dochodzenia filogenetyczne prowadzono na fragmentach 4 genów z grupy *rpo* (Gomolińska i in. 2017).

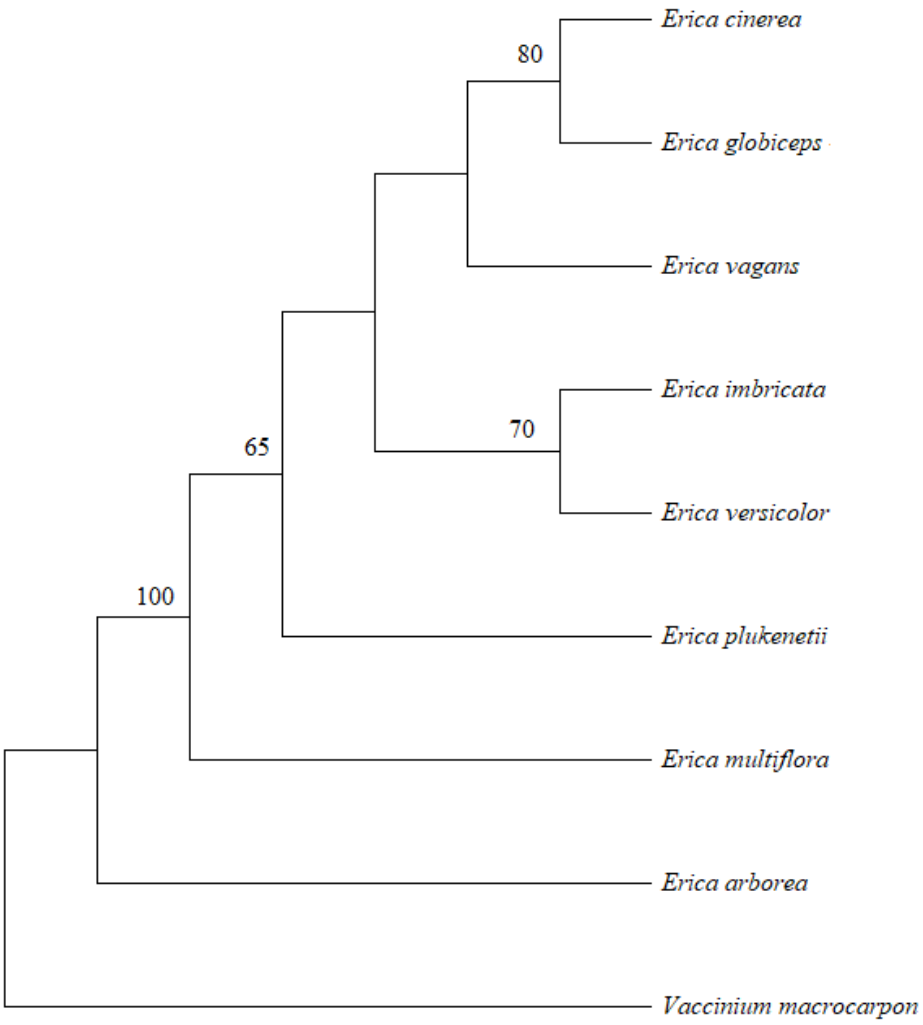
Porównanie filogenetyczne grup genów chloroplastowych wykazuje wysoki stopień zachowania genów *mat*K oraz *ccs*A, szczególnie *ccs*A w przeliczeniu miejsc filogenetycznej informatywności, co sugeruje, że geny te mogą poprzedzać dywergencję genów *rbc*L, *acc*D czy *cem*A (Rycina 39) oraz wzbogacają zasób kandydatów do zadowalających markerów filogenetycznych dla rodzaju *Erica*.

Geny z rodziny *pet* mogą ewoluować i rozszerzać się szybciej niż rodzina genów *trn* (Rycina 43 oraz Rycina 50). Przeprowadzona analiza wykazała, że podczas ekspansji rodziny genów *atp, pet* ewoluowało wiele ich grup, co skutkowało wysokim poziomem funkcjonalnej dywergencji wśród wszystkich genów *pet*, uniemożliwiając wyłonienie pretendenta najwyżeszej informatywności filogenetycznej.

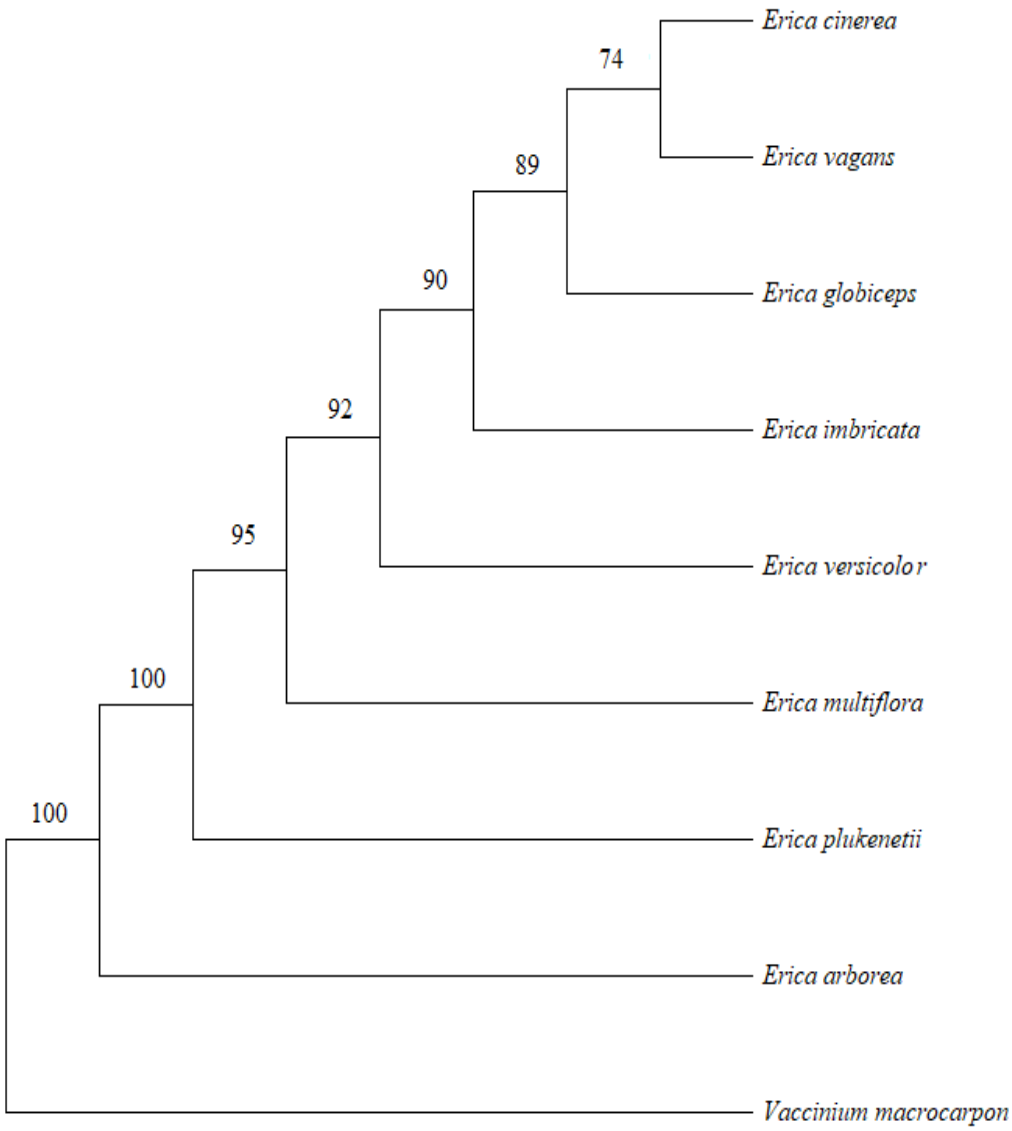
Drzewa filogenetyczne skonstruowane na pojedycznych genach (Ryciny 23-27), z relewantnym wsparciem dla kładów, wskazały gatunek *Erica arborea* jako prawdopodobny gatunek ancestralny. Analizy przeprowadzone na poszczególnych grupach genów chloroplastowych potwierdziły tę hipotezę w analizach opartych na genach z grup: *pet, psa, psb* oraz *trn*., statystycznie jest to zatem najczęściej wyłaniany gatunek, będący w najwcześniejszej linii ewolucyjnej. Celem głębszej weryfikacji tej hipotezy przeprowadzono jeszcze analizę filogenetyczną na podstawie danych z genów wytypowanych w badaniu informatywności filogenetycznej jako najlepsze markery genetyczne, dla przedstawicieli z rodzaju *Erica.*

Najwyższe wsparcie bootstarp otrzymano dla drzew skonstruowanych dla genów *atp*F, *ndh*F, *psb*B oraz *rpo*C2, które zaprezentowano na Rycinach 51-54. We wszystkich 4 przypadkach najstarszym gatunkiem ewolucyjnie okazał się wrzosiec drzewiasty (*Erica arborea*). Wnikliwa analiza porównawcza wszystkich otrzymanych dendrogramów konsensusowych oddala możliwość pełnego i przekonującego odtworzenia powiązań między analizowanymi gatunkami, co może wynikać m.in. ze stosunkowo ograniczonej liczby gatunków poddanych tym analizom, pomimo obfitych zasobów informacji genetycznej, które sobą niosą. Odtworzenie historii ewolucyjnej jest zatem wciąż niewykonalne. Gdy najbardziej odległy gatunek to gatunek bikontynentalny, możemy jedynie przychylić się do teorii rozejścia się linii ewolucyjnej na granicy kontynentu Afryki i Europy, jednak wciąż niewyjaśnionym pozostaje skąd gatunki z tego rodzaju rozpoczęły swoją ekspansję (Kron i in. 2002).

Zaprezentowane w tej sekcji rezultaty mają szansę poskutkować jeszcze lepszym zrozumieniem ewolucji roślin z rodzaju *Erica* i charakterystyki funkcjonalnej poszczególnych rodzin/grup genów genomu chloroplastowego w przyszłości.



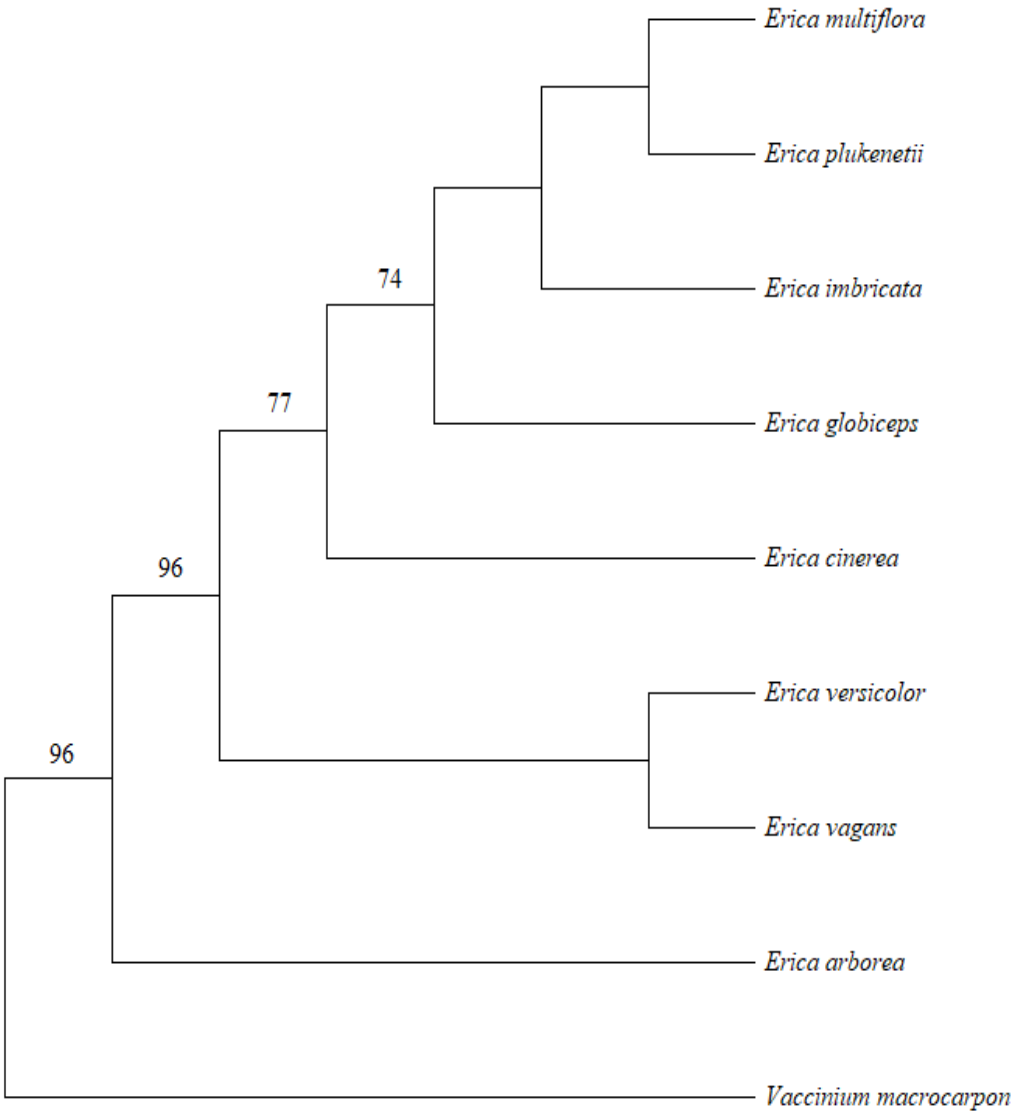
Rycina 51. Konsensusowe drzewo filogenetyczne przedstawicieli z rodzaju *Erica* skontruowane analizą bayesowską dla kompletnych genów *ndh*F. Wartości bootstrap poniżej 50% są ukryte. Liczby nad gałęziami reprezentują wsparcie bootstrapowe uzyskane przez maksymalne prawdopodobieństwo (wartość ML). Grupę zewnętrzną stanowił gatunek *Vaccinium macrocarpon*.



Rycina 52. Konsensusowe drzewo filogenetyczne przedstawicieli z rodzaju *Erica* skontruowane analizą bayesowską dla kompletnych genów *rpo*C2. Wartości bootstrap poniżej 50% są ukryte. Liczby nad gałęziami reprezentują wsparcie bootstrapowe uzyskane przez maksymalne prawdopodobieństwo (wartość ML). Grupę zewnętrzną stanowił gatunek *Vaccinium macrocarpon*



Rycina 53. Konsensusowe drzewo filogenetyczne przedstawicieli z rodzaju *Erica* skontruowane analizą bayesowską dla kompletnych genów *psb*B. Wartości bootstrap poniżej 50% są ukryte. Liczby nad gałęziami reprezentują wsparcie bootstrapowe uzyskane przez maksymalne prawdopodobieństwo (wartość ML). Grupę zewnętrzną stanowił gatunek *Chamaedaphne calyculata*



Rycina 54. Konsensusowe drzewo filogenetyczne przedstawicieli z rodzaju *Erica* skontruowane analizą bayesowską dla kompletnych genów *atp*F. Wartości bootstrap poniżej 50% są ukryte. Liczby nad gałęziami reprezentują wsparcie bootstrapowe uzyskane przez maksymalne prawdopodobieństwo (wartość ML). Grupę zewnętrzną stanowił gatunek *Vaccinium macrocarpon*

5. Dyskusja

5.1. Genom chloroplastowy *Erica* a inne genomy chloroplastowe

Aktualnie według opublikowanych danych literaturowych dostępnych jest około ośmiu tysięcy kompletnych, opracowanych genomów chloroplastowych, w tym dotychczas żadnego całego genomu chloroplastowego z rodzaju *Erica*. Wiąże się to z ogromnym trudem złożenia kompletnej sekwencji plastomu w tym rodzaju. Składały się na to wielokrotne próby, nielinearność w pokryciu danych oraz segmenty DNA z motywem licznych powtórzeń tandemowych znaczącej długości. W bazie danych są osiągalne jedynie poszczególne regiony chloroplastowe (głównie *atp*-*rbc*L, *psb*A-*trn*H) lub pojedyncze geny (głównie *rbc*L i *mat*K, *rpo*C1, *rpo*B, *rpl*16) dla wielu gatunków z rodzaju *Erica* (Benson i in. 2013).

Przez dłuższy czas dostępne były jedynie pojedyncze kompletne genomy chloroplastowe z rodziny *Ericaceae*, głównie rodzajów z grupy niefotosyntetyzujących i miksotroficznych. Na przestrzeni ostatnich lat opracowano ich stosunkowo więcej, prowadząc intensywne prace genomiczne, w głownej mierze pośród rodzaju *Vaccinium* oraz *Rhododendron* (Lin i in. 2020, Kim i in. 2020, Ma i in. 2021).

Badania z genomiki, ewolucyjne i filogenetyczne, w rodzinie *Ericaceae* nie obejmowały dotychczas przedstawicieli z rodzaju *Erica*, nie uwzględniały także genomu *Chamaedaphne calyculata*. Zwiększająca się liczba kompletnych genomów plastydowych może być szczególnie korzystna dla dalszych działań genetyki populacyjnej, analiz molekularnych, ewolucyjnych i genetyki konserwatorskiej.

Zbadano tutaj tylko dwa kompletne genomy *Erica*, oraz jedynie osiem gatunków z rodzaju, który liczy 856 gatunków, w związku z czym może to nie w pełni pokazać rzeczywisty stan filogenezy rodzaju *Erica*.

Jednak badanie to było pierwszą tego typu kompleksową analizą filogenetyczną *Erica* na poziomie całego genomu chloroplastowego, czy sekwencji NOR a wynikające topologie oparte na całym genomie chloroplastowym lub zestawach danych CDS ostatecznie pokazują, że *Erica arborea* może być kandydatem na ancestora *Erica*.

Międzygatunkowy przepływ genów w odpowiedzi na selekcję ekologiczną mógł doprowadzić do trudności w analazie filogenetyki molekularnej. Biorąc pod uwagę, że genomy chloroplastowe i jądrowe ewoluują niezależnie, sama filogeneza genomów chloroplastowych może być niewystarczająca do podejmowania decyzji taksonomicznych dotyczących całego rodzaju *Erica*. Dlatego potrzeba więcej metod i zdecydowanie większej liczby gatunków w dalszych badaniach, aby poprawić naszą zdolność do lepszego zrozumienia filogenezy i ewolucji rodzaju *Erica*. Wyniki pokazały pewne różnice i ewolucję między kompletnymi genomami chloroplastów *Erica,* w tym wielkość, strukturę i różnorodność nukleotydów, co dostarczyło cennych informacji do ekspandacji wiedzy z zakresu klasyfikacji gatunków, ich charakterystyki i ewolucji.

5.1.1. Analiza porównawcza struktury i zawartości genów genomów chloroplastowych gatunków fotosyntetyzujących z rodziny *Ericaceae*

Genomy chloroplastowe porównywanych reprezentantów z rodziny *Ericaceae* to koliste cząsteczki o wielkości plasującej się w zakresie przekraczającym długość występującą u większości roślin kwiatowych, ze względu na dodatkową proporcję niekodującego DNA reprezentowaną przez powtórzenia i pseudogeny, za wyjątkiem wielkości plastomu *Arbutus unedo*, który osiągnął jedynie ~ 150 tys. pz (Martinez-Arberola i in. 2013).

Plastomy te ustrukturyzowane są w typową czterofragmentową konstrukcję składającą się z dwóch odwróconych powtórzeń (IRA, IRB), oddzielonych dużą pojedynczą kopią (LSC) i małą pojedynczą kopią (SSC). Wyjątek stanowi genom gatunku *Rhododendron pulchrum*, który nie zawierał regionów odwróconych powtórzeń (IR) (Shen i in. 2020). Utrata regionu IR jest rzadkim zjawiskiem u roślin okrytonasiennych, jednak tego typu wyjątki odnotowywano już wcześniej. Brak regionów odwróconych powtórzeń wykryto m.in. po raz pierwszy w rodzaju *Passiflora* (Męczennica) w gatunkach *Passiflora capsularis* i *Passiflora costaricensis*, który jest gatunkiem wykazującym sporadyczne dziedziczenie chloroplastów po obojgu rodzicach.(Cauz-Santos i in. 2020).

Zawartość procentowa GC w analizowanych plastomach jest zbliżona u wszystkich gatunków *Ericaceae* (uplasowuje się w granicach 35-37%) i jest porównywalna do zakresu odnotowanego w genomach chloroplastowych roślin z kladu *Asteridae* (Astrowych) i *Rosidae* (Różowe) (Wang i in. 2019; Zhang i in. 2021). Wyższe wartości zawartości procentowej GC odnotowano m.in. w rodzinie *Apocynaceae* (Toinowatych), gdzie odnotowano wartość totalną 38,3% (Zheng i in. 2022).

Rozmiary małego regionu pojedynczej kopii obserwowane u analizowanych przedstawicieli fotosyntetyzujących z rodziny *Ericaceae* u większości gatunków rozkłada się w zbliżonych granicach (2,249 - 3,518 pz). Na ogół mały obszar pojedynczej kopii jest u wrzosowatych znacznie zmniejszony, przy wielkości w granicach ~ 3 tys. pz, za wyjątkiem kontrastującego wielkością regionu SSC u gatunku *Pyrola rotundifolia*, gdzie osiąga wielkość 11,945 pz. (Tabela 11). Gatunek ten wykazał także obecność nowego genu, powstałego w wyniku duplikacji genu *rps*4. Redukcja regionu SSC przekracza tę zanotowaną u niefotosyntetyzujących roślin pasożytniczych, wyróżniających się wybitną redukcją większości genów chloroplastowych, np. w rodzajach *Cuscuta* (Kianianka) czy *Epifagus* (monotypowy rodzaj z jedynym przedstawicielem *Epifagus virginiana* znanym pod nazwą „krople bukowe") (Miao i in. 2022). Ekstremalne skrócenie regionu SSC łączy się z powieleniem i włączeniem pewnych operonów (np. *ndh*H-D u *Vaccinium macrocarpon*) w każdy z dwóch regionów IR i jednoczesne rozszerzenie regionów powtórzeniowych. Porównywane genomy zawierają dużą liczbę regionów powtórzeniowych i translokacji. Analiza porównawcza rozszerzonego zestawu plastomów *Ericaceae* sugeruje, że geny *clp*P i *acc*D, obecne w genomach plastydowych prawie wszystkich roślin, nie zostały utracone w tej rodzinie, mimo że we wcześniejszych badaniach stawiano tego typu hipotezy (Fajardo i in. 2013, Martinez-Arberola i in. 2013, Logacheva i in. 2016).

Uliniowanie całych genomów chloroplastowych z innymi reprezentantami *Asteridae* wykazuje wysoki stopień konserwatywności wielu regionów kodujących wraz z unikatowymi rearanżacjami charakterystycznymi dla poszczególnych gatunków. Pierwszy całkowicie zsekwencjonowany plastyd *Nicotiana tabacum* (Tytoń szlachetny), stanowi punkt odniesienia nowo badanych cpDNA jako gatunek uznany za posiadający porządek i zawartość genów przodków roślin okrytozalążkowych. Jest to wzmocnione przez podobieństwo cpDNA tytoniu do m.in. podstawowych okrytozalążkowych, *Amborella*, *Nymphaea* (Grzybienie), *Calycanthus* (Kielichowiec), *Drimys* (Zacierp), *Liriodendron* (Tulipanowiec) czy *Piper* (Pieprz) (Sahoo i Rakshit, 2022). W niektórych pochodnych liniach okrytozalążkowych ten stan przodków został nieco lub silnie zmodyfikowany przez inwersje, utratę genów, zawartość bądź brak ORF czy zmiany w obrębie IR (Fonseca i in. 2022).

Jednak ogólny wzór podobny do *Nicotiana tabacum* jest szeroko rozpowszechniony np. w przypadku rodzajów *Eucalyptus* (Eukaliptus), *Gossypium* (Bawełna) czy *Panax* (Wszechlek/Panaks) (Steane, 2005; Lee i in. 2006). Około 70 % *Asteridae*, których genom chloroplastowy został całkowicie zsekwencjonowany przedstawiał ogólną zawartość genów tego „odnośnikowego" gatunku (Ruhlman i Jansen, 2021).

Gatunki *Ericaceae* odbiegają od porządku przyjętego we wzorcu *Nicotiana tabacum* w większym zakresie niż inne gatunki astrowatych, z uwagi na występujące szeroko zakrojone zmiany. Za wyjątkiem plastomu *Arbutus unedo*, wielkość genomów chloroplastowych przedstawicieli fotosyntetyzujących *Ericaceae* jest dłuższa niż większości roślin kwitnących ze względu na zwiększony udział niekodującego DNA reprezentowane przez różnego typu powtórzenia i pseudogeny. Wynika to m.in. z mechanizmu generującego transfer genów z regionu SSC do regionów odwróconych powtórzeń (IR) co wiąże się z duplikacją i relokacją genów, a także dwukrotną ekspansją IR oraz znaczącym zminiaturyzowaniem regionu SSC. Niektóre z przegrupowań strukturalnych w plastomach obejmują przesunięcia granic IR (rozszerzenia i skurczenia), które przyczyniły się do rozległych przegrupowań. Na przykład rozmiary IR są niezwykle zmienne ze względu na przesunięcia granic IR, wahając się od 7 000 pz w gatunku południowoafrykańskiej monsoni (*Monsonia speciosa*), do 88 000 pz w gatunku afrykańskiej pelargonii (*Pelargonium transvaalense*) (Turudić i in. 2022). Gatunek ten ma największy znany do tej pory plastom z IR większym niż 87,4 pz. Rodzina *Geraniaceae* (Bodziszkowate), do której należy *Pelargonium transvaalense* słynie również z ekstremalnych rekonfiguracji genomów plastydów, wykazujących wiele unikalnych układów, relokacji i powtórzeń (Weng i in. 2017). Pomimo gigantycznych wielkości regionów IR, zmienności strukturalnej i różnic w wielkościach plastomów przedstawiciele *Geraniaceae*, w tym reprezentanci *Pelargonium* posiadają regularną liczbę genów kodujących białka i zwykle 29 genów kodujących tRNA. Dane dostępne dla *Geraniaceae* i innych linii roślin kwitnących sugerują, że te przyrosty wielkości i rekonfiguracji plastomu niekoniecznie są związane z istotnymi zmianami w ekspresji genów lub ogólnej funkcji organelli (Guisinger i in. 2011; Turudić i in. 2022).

Rozległe ekspansje IR są rzadkie i zwykle wiążą się ze zmianami w składzie genów i rearanżacjami strukturalnymi. Przy czym średni rozmiar IR okrytozalążkowych wynosi zwykle 25 000 pz (Weng i in. 2017). Rozszerzenia i kurczenia IR powodują również zmiany w zawartości genów z powodu przeniesienia genów z regionów SSC do IR lub w inny sposób. Zaproponowane w literaturze metody to konwersje genów, pęknięcia dwuniciowe i delecje genomowe jako możliwe mechanizmy przesunięć granic IR (Cauz-Santos i in. 2020). Chociaż obecność IR może odgrywać ważną rolę w stabilności plastomu wśród fotosyntezujących okrytozalążkowych, zaobserwowano także proces utraty IR m.in. u niektórych gatunków roślin strączkowych , u sukulenta saguaro (karnegia olbrzymia, *Carnegiea gigantea*) i niektórych gatunków iglicy (rodzaj *Erodium*) (Blazier i in. 2016). Przesunięcia i inwersje granic IR mogą być źródłem duplikacji genów i spadku poziomu ich zmienności. Geny relokowane; przeniesione z regionów SSC do IR mają tendencję do zmniejszania szybkości podstawień poprzez aktywność naprawczą zależną od kopii. Duplikacje genów znalezionych w plastomach wynikają często z połączenia inwersji i przesunięć IR w procesie tzw. duplikacji tandemowej. (Guisinger i in. 2011; Ruhlman i Jansen, 2021; Qin i in. 2022).

Największa dywergencja sekwencji badanych plastomów z rodziny *Ericaceae* obserwowana jest w pobliżu pozycji ok. 90 000 pz, w okolicy sekwencji obejmujących regiony IR oraz SSC, co wiąże się z faktem, iż region IR-SSC uznany jest nierzadko za region hipermutacyjny i porównywane pomiędzy gatunkami granice IRb-SSC wykazywały wielokrotnie większe różnice niż region IRa-LSC. Sam region SSC w wielu badaniach wykazał także największą różnorodność nukleotydów. Nie jest to jednak normą, w przypadku badań prowadzonych np. na przedstawicielach rodziny *Apiaceae* granice między regionami IR i LSC były dość zmienne, połączenia IRB-SSC i SSC-IRA natomiast wykazywały wysoki stopień konserwatywności (Kang i in. 2019).

Region LSC wrzosowatych obrazuje kilka przeniesień genów w plastomie jednego z reprezentantów wrzosowatych. Zakłada się, że np. u *Arbutus unedo* nastąpiły minimum dwie główne inwersje segmentów w plastomie, przez co porządek genów w tym gatunku odbiega nawet od ogólnego schematu innych gatunków wrzosowatych. Wśród podzbiorów plastomu region LSC ma największy rozmiar, stanowiąc na ogół ponad połowę całej wielkości plastomu.

Ze względu na zalety wypływające ze swojej długości, region LSC zawiera większość miejsc informatywnych całego genomu, dzięki czemu może stać się alternatywą dla całego plastomu. Wiele badań wykazało, iż wydajność, przydatność w identyfikacji gatunków i analizy filogenetyczne mają prawie taką samą wydajność w odniesieniu do regionu LSC co do całego genomu (Schmitz-Linneweber i in. 2002, Yukawa i in. 2006, Ku i in. 2013, Yang i in. 2018).

Plastomy badanych przedstawicieli fotosyntetyzujących *Ericaceae* zawierają łącznie od 73 (*Rhododendron pulchrum*) do 147 genów (*Vaccinium macrocarpon*), biorąc pod uwagę zduplikowane geny jako jeden gen (Tabela 11). Obecne są w ich składzie wszystkie geny związane z procesem fotosyntezy, zwykle znajdujące się w plastomach roślin okrytonasiennych fotosyntetyzujących (komponenty kodujące fotosystemy I i II, kompleks cytochromu b6 / f, *rbc*L, kompleks syntazy ATP).

Podobieństwa w obrębie tego samego rodzaju w rodzinie *Ericaceae* przedstawiają się w różny sposób dla każdego z trzech rodzajów, wśród których porównano więcej niż jednego przedstawiciela. W przypadku rodzaju *Vaccinium* zauważyć można rozbieżność w zawartości genów. Brak genu *trn*G-GCC u *Vaccinium oldhamii* i *Vaccinium macrocarpon*, gdzie z kolei wystąpiły geny *pet*B, *pet*D, *rps*7, *rps*12, *rps*16 i *rpo*A, które nie zostały odnotowane u pozostałych przedstawicieli tego rodzaju, poddanych niniejszemu zestawieniu (Kim i in. 2019). Zazwyczaj wypadnięcie jednego genu *trn* powoduje tzw. codon usage bias, który odnosi się do różnic w częstości występowania kodonów synonimicznych w kodującym DNA. W genomie chloroplastowym gatunku *Vaccinium macrocarpon*, jako w jedynym z swoim rodzaju, wystąpiły geny *ycf*2, *ycf*15b, *ycf*68 oraz *inf*A, natomiast – również jako w jedynym z rodzaju – nie wystąpiły geny *psb*Z oraz *rpl*20 (Fajardo i in. 2013). U *Vaccinium bracteatum* z kolei nie odnotowano genów *rpl*2, *trn*D-GUC, *trn*fM-CAU, *trn*K-UUU, *trn*R-UCU, *trn*V-UAC oraz *trn*Y-GUA (Kim i in. 2020).

Porównując kompletne genomy chloroplastowe rodzaju *Vaccinium*, zauważyć można także dysonanse w liczbach intronów i kopii niektórych genów. Egzemplifikacją rozbieżności względem liczby kopii genów jest chociażby przykład genu *rps*12, który wystąpił w dwóch kopiach u gatunków *Vaccinium oldhamii* i *Vaccinium macrocarpon*, czy genów *rps*3, *rps*18, *rpl*22 i *rpo*A, które wystąpiły w dwóch kopiach jedynie u

Vaccinium macrocarpon. W kwestii liczby intronów przykładem są m.in. geny *rps*16 i *trn*G-UCC, które miały po jednym intronie u wszystkich przedstawicieli oprócz *Vaccinium macrocarpon*; gen *rps*3, który posiadał jeden intron u trzech przedstawicieli *Vaccinium*, za wyjątkiem *Vaccinium oldhamii* i *Vaccinium macrocarpon*; gen *trn*V-UAC nie zawierał intronów jedynie w *Vaccinium bracteatum*, a gen *trn*K-UUU nie zawierał intronów zarówno w *Vaccinium bracteatum*, jak i w *Vaccinium macrocarpon*; natomiast geny *pet*B, *pet*D oraz *psb*A zawierały po jednym intronie tylko w *Vaccinium oldhamii* i *Vaccinium macrocarpon* (Schlautman i in. 2017; Kim i in. 2020). Wspólne i charakterystyczne dla wszystkich plastomów *Vaccinium* okazało się wystąpienie dwóch intronów genu *ycf*3 (Fajardo i in. 2013; Kim i in. 2019).

W przypadku rodzaju *Rhododendron* różnice obrazują się już na poziomie samej wielkości plastomu. *Rhododendron pulchrum* posiada zdecydowanie najmniejszy genom, z najmniejszą liczbą genów, natomiast *Rhododendron platypodum* to posesor największego genomu chloroplastowego w całej rodzinie (Shen i in. 2020; Ma i in. 2021). Zawiera on 16 genów z dwiema kopiami, 6 genów tRNA z dwiema kopiami, jeden z trzema, oraz jeden z aż pięcioma kopiami, wśród genów rRNA 4 mają po dwie kopie. Geny *rps*12 oraz *ycf*3 posiadają po dwa introny, a geny *trn*I-GAU, *trn*A-UGC, *ndh*A, *rps*16, *ndh*B, *pet*D, *pet*B, *rpl*16, *trn*G-UCC, *atp*F, *trn*V-UAC, *trn*L-UAA, *trn*K-UUU mają po jednym intronie. *Rhododendron pulchrum* wykazał w genie *ycf*3 obecność dwóch intronów. Z kolei w genach *ndh*A, *trn*V-UAC, *trn*L-UAA, *trn*A-UGC, *trn*I-GAU oraz *trn*G-UCC tego gatunku zanotowano po jednym intronie (Shen i in. 2020). *Rhododendron delavayi* posiada 12 genów z dwiema kopiami posród genów kodujących białka. Wśród genów tRNA ma sześć genów z dwiema kopiami oraz jeden gen z trzema kopiami (Li i in. 2020). Region IR *Rhododendron delavayi* uległ znacznemu rozszerzeniu do długości 46,968 pz, przy czym u *Vaccinium macrocarpon* przyjmuje on długość 34,232 pz, u *Arbutus unedo* 27,924 pz, u *Pyrola rotundifolia* 23,938 pz, a u gatunków *Erica* – odpowiednio 35,290 pz (*Erica arborea*) oraz 35,276 pz (*Erica versicolor*). W plastomie *Rhododendron delavayi* niektóre z genów normalnie występujących w LSC konserwatywnych genomów plastydowych, takie jak *cem*A, *rps*16, *psa*I, *ycf*4, *pet*A zostały przeniesione do regionu IR.

Inne geny normalnie występujące w SSC, takie jak *ccs*A, *ndh*A, *ndh*D, *ndh*E, *ndh*G, *ndh*H, *ndh*I, *rps*15, *psa*C, *rpl*32 również przeniosły się do IR. Z drugiej strony geny takie jak np. *rpl*2, *rpl*23, *ycf*2, *ndh*B czy *rps*7 zostały przeniesione z regionu IR do LSC. Generalnie u tego gatunku więcej genów uległo transferowi do regionu IR niż z regionu IR w genomie plastydowym, co w konsekwencji wpłynęło na zwiększenie jego totalnej długości (Li i in. 2020).

Geny kodujące składniki kompleksu dehydrogenazy NADH często są obcinane, zawierają przesunięcia ramki, kodony stop w ramce, etc., co może wskazywać na wczesne stadia pseudogenizacji lub edycji RNA – tak jest u *Pyrola rotundifolia, Vaccinium macrocarpon* czy *Chamaedaphne calyculata* (Fajardo i in. 2013; Szczecińska i in. 2014; Logacheva i in. 2016). Kompleks NDH bierze udział w chlororespiracji i cyklicznym transporcie elektronów w fotosystemie I, przypuszczalnie równoważy reaktywny tlen gatunków (ROS) w celu złagodzenia stresu wywołanego przez czynnik indukujący ROS16. Geny NDH są nieobecne w starszych liniach *Archaeplastida*, obecne natomiast u większości reprezentantów *Streptophyta*, co wraz z ich funkcją sugeruje, że ich obecność w plastomach wiąże się z adaptacją do lądowego trybu życia (Sabater, 2021). Jednak wtórna utrata genów z grupy *ndh* z cpDNA był odnotowany w kilku liniach roślin. Tego typu zjawisko zauważono m. in. u *Plantaginaceae* (Babkowate), gdzie plastom brzeżycy (*Litorella*) nie posiada funkcjonalnych kopii genów *ndh*, co przypisuje się amfibijnemu trybowi życia i częściowym poleganiu na fotosyntezie CAM (Mower i in. 2021). Ponadto utratę genów *ndh* zauważono także w rodzaju *Impatiens* (Niecierpek) u *Impatiens morsei*, niektórych i niektórych przedstawicieli rodzin *Geraniaceae* (Bodziszkowate), *Pinaceae* (Sosnowate) i *Orchidaceae* (Storczykowate) (Blazier i in. 2016; Qiu i in. 2023).

Ogólnofilogenetyczne badania transkryptomu ukazało, że jest to całkowita utrata kompleksu genów *ndh*, a nie zastępowanie genów plastydów przez geny jądrowe i/lub mitochondrialne. Rośliny tytoniu z usuniętymi genami *ndh* są żywotne w optymalnych warunkach wzrostu, ale podatne na stres wodny, natomiast mutacje w genie *ndh*F okazały się wpływać na wydajność fotosyntezy przy zmieniających się natężeniach światła (Sabater, 2021; Mower i in. 2021; Qiu i in. 2023).

Tego typu wyniki sugerują, że kompleks NDH odgrywa rolę w interakcji roślina-środowisko i że jest zależny od czynników ekologicznych. Gen *acc*D jest nienaruszony u wszystkich gatunków *Ericaceae*. Sekwencja genu *ycf*1 zawiera skrócony ORF (~ 200 aminokwasów) i dlatego jest mało prawdopodobne, aby działał prawidłowo i sprawnie u fotosyntetyzujących *Ericaceae* (Fajardo i in. 2013, Martinez-Arberola i in. 2013, Szczecińska i in. 2014, Logacheva i in. 2016).

Filogeneza kompletnych sekwencji genomu cpDNA trzynastu gatunków fotosytentyzujących z rodziny *Ericaceae* została przeprowadzona na podstawie danych uzyskanych z sekwencjonowania oraz pobranych z GenBank NCBI. Sekwencje zostały wyrównane w programie ClustalW. Drzewo ML zostało skonstruowane przy użyciu MEGA11 (Tamura i in. 2021). Skonstruowano także drzewo BI na podstawie wnioskowania bayesowskiego w MrBayes. ML i analiza BI wygenerowały taką samą topologię drzewa. Zgodność drzew filogenetycznych skonstruowanych tymi metodami ma potwierdzenie w literaturze. Podobna sytuacja miała miejsce przy klasyfikacji krasnorostów, gdzie drzewo ML było zgodne z drzewem bayesowskim, z wyjątkiem niektórych słabo obsługiwanych regionów, a test statystyczny pokazał, że drzewo BI nie jest znacząco mniej prawdopodobne niż drzewo ML (Verbruggen i in. 2010; Torres i in. 2021). Zbieżność topologii drzew ML i BI miała też miejsce w analizie przedstawicieli rodzaju *Erica* (Gomolińska i in. 2017). Ocena wpływu decyzji metodologicznych na powstałe hipotezy ma kluczowe znaczenie w filogenetyce. Warto poddać ocenie, w jaki sposób wybór modelu, definicja ortologii i czynniki zakłócające wpływają na wyniki filogenomiczne. W przeprowadzonych przez zespół Torresa badaniach porównawczych ze 157 filogenomicznych zestawów danych otrzymano wyniki z trzech połączonych metod filogenetycznych (maksymalne podobieństwo, maksymalna parsymonia i metoda Bayesowska). Uzyskane w ten sposób drzewa były względem siebie bardzo podobne, 96,7% podobieństwa dla wszystkich węzłów pomiędzy metodami BI i ML, 90,6% dla relacji ML i MP oraz 89,1% dla porównania relacji BI i MP. Główne wnioski większości badań były zgodne co do trzech metod filogenetycznych, a niezgodność dotyczyła tych węzłów, które były dotychczas uważane za wyjątkowo oporne i problematyczne we wnioskowaniach systematycznych.

Różnice między metodami były proporcjonalnie większe w zestawach danych analizujących relacje na wyższych poziomach taksonomicznych (w szczególności na poziomach typu i królestw) i niezależne od liczby znaków zawartych w zestawach danych (Verbruggen i in. 2010; Gomolińska i in. 2017; Torres i in. 2021).

Analiza bootstrap z 2000 powtórzeń została zastosowana do potwierdzenia stabilności każdego z węzłów drzewa. Rekonstrukcja filogenetyczna rodziny *Ericaceae* przeprowadzona na bazie kompletnych plastomów pokryła się z tą, która została przeprowadzona na podstawie regionów odwróconych powtórzeń. Jest to pierwsza tego typu analiza, która uwzględnia nie tylko rodzaj *Erica*, ale także rodzaj *Chamaedaphne,* pomijany dotychczas w zestawieniach wrzosowatych przez innych badaczy. W jednym z wcześniejszych badaniach *Rhododendron pulchrum* był bliższy rodzajowi *Vaccinium* (Shen i in. 2019), jednak w późniejszych latach ogólna rewizja rodziny zdaje się być zbieżna z otrzymaną w niniejszych analizach (Ma i in. 2021).

5.1.2. Analiza porównawcza regionów powtórzeniowych oraz struktury i zawartości genów w plastomach okrytonasiennych

Porównawcze badania plastomów wykazują, iż niektóre regiony plastomów są predysponowane do mutacji z powodu nielosowego rozkładu lokalizacji sekwencji powtórzeniowych, tandemów i mutacji. Analizy powtórzeń SSR, powtórzeń oligonukleotydowych i powtórzeń tandemowych wykazują podobieństwa dwóch analizowanych kompletnych genomów *Erica*. Powtórzenia SSR uchodzą często za niechciany produkt replikacji genomu, bądź błąd w procesie kopiowania genomu, mogą stanowić potężne narzędzie w zrozumieniu procesów ewolucyjnych (Schlautman i in. 2017).

Przeważnie SSR zidentyfikowane u *Erica* były mononukleotydami w porównaniu z dinukleotydami i trinukleotydami, podczas gdy poziom A/T SSR w przypadku mononuklotydów i AT/TA SSR w przypadku dinukleotydu wykazaywał obfitość. Podobny trend dystrybucji SSR odnotowany został również w genomach chloroplastów innych rodzin okrytonasiennych (Wang i in. 2023). Lokalizacja SSR zidentyfikowana w tych analizach mogą pomóc w zrozumieniu struktury genetycznej gatunków *Erica*. Ich liczba na tle innych gatunków jest stosunkowo wysoka – w badaniach nad czterema genomami chloroplastowymi z rodzaju *Pinellia* (Pinelia) zlokalizowano ich od 68 do 111 (Cui i in. 2022), u przedstawicieli rodzaju *Uncaria* (Czepota) oszacowano liczbę SSR w przedziale od 31 do 49 (Chen i in. 2022), w 5 analizowanych plastomach *Lepidium* (Pieprzyca) liczba SSR wyniosła od 87 do 98 (Zhou i in. 2022), w rodzaju *Gaultheria* (Golteria) liczba SSR wahała się od 57 do 75 (Xu i in. 2022), w rodzaju *Carpesium* od 37 do 41 (Shi i in. 2022), a w plastomach z rodzaju *Avena* (Owies) było ich 276 (Liu i in. 2020), podczas gdy liczba powtórzeń SSR w rodzaju *Erica* wyniosła odpowiednio 874 u *Erica versicolor* i 814 u *Erica arborea.*

W rodzinie *Ericaceae* dotychczas najwięcej powtórzeń SSR zanotowano w gatunku *Rhododendron pulchrum*, gdzie stanowiły liczbę łącznie 221 powtórzeń SSR oraz *Rhododendron delavayi*, gdzie zanotowano ich 371, co tłumaczono skomplikowanym podłożem genetycznym w rodzaju rododendronów (Shen i in. 2020; Li i in. 2020).

W niniejszej pracy określono także powtórzenia oligonukleotydowe w genomach chloroplastów *Erica arborea* i *Erica versicolor*. Postuluje się, że powtórzenia tego typu generują substytucje i indele w genomach, mogą stanowić punkt wyjścia do określenia mutacyjnych hotspotów. Te z kolei często są sugerowane do wykorzystania w opracowaniu efektywnych markerów molekularnych do analizy filogenetycznej, genetyki populacyjnej i barkodów gatunkowych. Liczba powtórzeń oligonukleotydowych wyniosła po 50 w obydwu gatunkach, ich długość osiągnęła od 187 pz do 704 pz w przypadku *Erica arborea* oraz od 193 pz do 706 pz u *Erica versicolor*. Podobną liczbę powtórzeń oligonukleotydowych rozpoznano w analizie pięciu plastomów rodzaju *Uncaria*, gdzie było ich od 49 do 51, jednak były one zdecydowanie krótszymi sekwencjami osiągając na ogół 30 pz - maksymalnie do 40 pz. Ponadto w rodzaju tym wykryte zostały powtórzenia palindromowe, których nie wykazano w przedstawicielach *Erica* (Chen i in. 2022). Zbieżne wyniki otrzymano także dla trzech analizowanych gatunków w rodzaju *Carpesium*, gdzie powtórzeń tego typu odkryto 39-40, o długości nieprzekraczającej 30-40 pz i jednoczesnym występowaniu powtórzeń palindromowych (Shi i in. 2022), u sześciu reprezentantów z rodzaju *Adiantum* (Niekropień/Złotowłos) powtórzeń tego typu zanotowano do 53, w przedziale długości sekwencji od 30 pz do 123 pz (Gu i in. 2022), natomiast w rodzajach *Polygonatum* (Kokoryczka), *Disporopsis*, *Disporum* (Parnik) oraz *Maianthemum* (Konwalijka) badając 17 reprezentantów liczba zdetektowanych powtórzeń nie przekroczyła 70 pz, a najwyższa liczba powtórzeń to 86, z powtórzeniami palindromowymi włącznie (Wang i in. 2022).

Liczbowa przewaga powtórzeń oligonukleotydowych miała miejsce np. w 11 analizowanych genomach chloroplastowych z rodzaju *Avena*, gdzie wystąpiło ich aż do 757, jednak ich długość była zdecydowanie dużo krótsza – sekwencje powtórzeniowe nie przekraczały 95 pz (Liu i in. 2020).

Wśród przedstawicieli rodziny *Ericaceae* tendencja do długich sekwencji powtórzeniowych w niektórych rodzajach zdecydowanie wzrasta. Wśród przedstawicieli z rodzaju *Gaultheria* zanotowano powtórzenia o długości nawet do 1722 pz w liczbie od 182 do 259 (Xu i in. 2022). W gatunku *Rhododendron delavayi* długość takich powtórzeń w zakresie od 15 pz do 100 pz odnaleziono w liczbie 1754, a 114 powtórzeń mieściło się w zakresie od 101 pz do 1000 pz (Li i in. 2020), z kolei w gatunku *Rhododendron pulchrum* najdłuższe powtórzenie oligonukleotydowe osiągęło 951 pz, razem zidentyfikowano ich 576 (Shen i in. 2020).

Powtórzenia tandemowe (TR) to wszechobecne, niestabilne elementy genomowe, które w przeszłości były określane jako niefunkcjonalne DNA. Przeprowadzona analiza powtórzeń tandemowych wykazała występowanie TR o rozmiarach jednostek zdecydowanie dłuższym niż 6 pz, zlokalizowanych niemal całkowicie w regionach niekodujących. Wśród 92 znalezionych powtórzeń tandemowych u *Erica versicolor* ich wielkość plasowała się od 2 do 330 pz. Natomiast liczba kopii tych powtórzeń charakteryzowała się bardzo znacznym rozrzutem, który dla niektórych z tych powtórzeń wyniósł od 2 do 6 kopii, podczas gdy w innych powtórzeniach liczba kopii osiągnęła wyniki kilkudziesięciu kopii. Z kolei w 99 wykrytych powtórzeniach tandemowych u *Erica arborea* wielkość okresu powtórzeń tandemowych wyniosła od 3 do 330 pz, z liczbą kopii od 2 do 63. Podobne wyniki powtórzeń tandemowych wykryto podczas analizy genomów chloroplastowych *Salvia plebeia*, gdzie wielkość okresu powtórzeń tandemowych wśród sześciu analizowanych gatunków wyniosła od 9 do 410 pz. Jednak w ich przypadku liczba kopii powtórzeń była zdecydowanie niższa i wynosiła na ogół 3 kopie (Cui i in. 2020).

W bardzo wielu przypadkach liczba powtórzeń tandemowych, mimo że zbliżona do wyników uzyskanych dla wrzośców, koreluje z dużo mniejszymi długościami sekwencji. W jedenastu genomach z rodzaju *Ficus* (Figowiec) zdetektowano od 69 do 82 powtórzeń tandemowych, których długość nie przekraczała 20 pz (Huang i in. 2022). W 5 genomach chloroplastowych z rodzaju *Lepidium* zanotowano od 21 do 30 powtórzeń tandemowych (Zhou i in. 2022), w rodzaju *Panax* powtórzeń tandemowych wykryto 24, o długości od 9 do 27 pz (Dinh i in. 2023), w rodzaju *Uncaria* znalezionych od 20 do 32 powtórzeń tandemowych miało długość od 25 do 98 pz (Chen i in. 2022),

od 1 do 12 powtórzeń tandemowych długośći w przedziale 15 do 102 pz zidentyfikowano u przedstawicieli z rodzaju *Adiantum* (Gu i in. 2022). Bardzo duże liczby powtórzeń tandemowych wykryto w 11 badanych osobnikach z rodzaju *Avena*: od 702 do 712, jednak były one na ogół krótkie (7pz), najdłuższe nie przekroczyły 95 pz (Liu i in. 2020).

Zgłębiając dane dotyczące rodziny *Ericaceae* w badaniach gatunku *Arbutus unedo* stwierdzono, że liczba powtórzeń tandemowych koreluje z wielkością genomu, jednak pomimo niewielkiego plastomu znaleziono tam 53 powtórzenia tandemowe, co stanowiło w zestawieniu z innymi astrowymi jeden z najwyższych wyników. W badaniu tym analizie poddano 36 gatunków, z czego tylko 5 przekroczyło liczbę powtórzeń tandemowych wykrytą u, w większości średnia TR wynosiła 30, a ich wielkość na ogół nie przekraczała 50 pz, największe tandemy wykryto u *Vaccinium macrocarpon* i *Arbutus unedo* (219 i 213 pz) (Martinez-Alberola i in. 2013). Tendencję gromadzenia się powtórzeń tandemowych w rodzinie *Ericaceae* potwierdziły badania prowadzone m.in. nad *Pyrola rotundifolia* (Logacheva i in. 2016). W *Rhododendron delavayi* znaleziono łącznie 96 powtórzeń tandemowych, w tym powtórzenie tandemowe o długości 126 pz i dwa o długości 505 pz (Li i in. 2020).

Swoiste nagromadzenie dużych powtórzeń tandemowych w regionach niekodujących genomów chloroplastowych może mieć znaczenie funkcjonalne w odniesieniu do genomiki porównawczej i ewolucyjnej (Shi i in. 2022). Powtórzenia oligonukleotydowe zidentyfikowane w tych badaniach z jednej strony są pomocne w identyfikacji regionów silnie polimorficzncych, z drugiej wraz z dużymi powtórzeniami tandemowymi stanowiły potężną barierę w procesie składania genomów. Z tego też powodu udało się złożyć dwa kompletne genomy z całkowitego zasobu genomowego ośmiu gatunków. W rodzinie *Ericaceae* dużo niższa liczba SSR i krótsze sekwencje powtórzeniowe wielu przedstawicieli w rodzaju *Vaccinium* sprawiają, że podobieństwo międzygatunkowe jest dużo wyższe niż w rodzaju *Rhododendron,* gdzie powtórzeń jest zdecydowanie więcej i ich sekwencje są zdecydowanie dłuższe (Shen i in. 2020; Li i in. 2020; Fahrenkrog i in. 2022). Zazwyczaj głębokie zsekwencjonowanie jednego genomu z konkretnego rodzaju jest wystarczające dla stworzenia wiarygodnej referencji dla pozostałych, w przypadku niniejszej pracy wiarygodny alignment dotyczy jedynie sekwencji kodujących, bowiem większość powtórzeń skupia się w obszarach sekwencji

niekodujących – w przypadku SSR jest to przewaga rzędu 70%, natomiast powtórzenia oligonukleotydowe oraz tandemowe skupiają się w regionach niekodujących niemal całkowicie. Skumulowanie długich sekwencji powtórzeniowych wpływa na zachodzenie rearanżacji, rekombinacji i w kosekwencji odpowiada za destabilizację genomów. Sąsiednie relacje sekwencji powtórzeniowych skłaniają do wyłonienia tezy, że powtórzenia tandemowe odgrywają istotną rolę w powstawaniu mutacji plastomicznych (Long i in. 2023). Ponadto dystrybucja tandemowych powtórzeń zbudowanych z dużych motywów, szeroko rozpowczechniona w plastomie, może wpływać na rearanżacje genomu przez zakłócanie replikacji DNA. Źródłem zmienności w sekwencjach powtórzeniowych jest „poślizg" polimerazy DNA podczas procesu replikacji. W wyniku tego zdarzenia zachodzi do wypadnięcia lub wstawienia znaków powtórzonych (Feng i in. 2023).

Za pomocą wielu programów bioinformatycznych dokonano porównania otrzymanych kompletnych sekwencji genomów chloroplastowych *Erica versicolor* oraz *Erica arborea*. Podobieństwa między sekwencjami mogą wynikać z analogii strukturalnej tudzież funkcjonalnej bądź też stanowią wypadkową losową. Zgodność sekwencji można określać liczbowo np. przez procent identyczności. Zjawisko różnicowania sekwencji w toku ewolucji prowadzi do zmian w sekwencjach poszczególnych genów i np. do wyłaniania się zjawiska dywergencji. Różnicowanie się sekwencji i zachodzące w nich modyfikacje przejawiają się w występowaniu insercji i delecji – te wywołują przerwy w wyrównaniu sekwencji oraz w zamianach nukleotydów – tranzycjach i transwersjach. W ujęciu całego genomu oraz porównania wielu genomów wpisuje się to w kategorię analiz filogenetycznych, które rozpoczynają się od zestawienia porównywanych sekwencji, poprzez wykonanie uliniowania. Poza narzędziami bioinformatycznymi, dorównuje się poszczególne sekwencje także ręcznie, czego dokonano w niniejszej pracy zestawiając w tabelach porównania sekwencji kodujących i niekodujących w regionach LSC, SC i IRA pomiędzy porównywanymi plastomami. Dopiero w kolejnych procedurach można rozpocząć jakiekolwiek wnioskowanie filogenetyczne lub konstruowanie i wyznaczenie drzew filogenetycznych. Dokonanie dopasowania sekwencji dwóch kompletnych genomów chloroplastowych z rodzaju *Erica* wykazało znaczące różnice w obrębie regionów niekodujących, a także polimorfizm regionów kodujących białka. Pomimo wysoce konserwatywnego charakteru plastomów,

u blisko spokrewnionych gatunków zanotowuje się zauważalne różnice – pomimo strukturalnego podobieństwa, dopasowanie kompletnych sekwencji chloroplastowych ujawnia dużą liczbę polimorfizmów pojedynczych nukleotydów (SNP) i insercji / delecji (InDels). W regionach kodujących ustalono występowanie 454 miejsc polimorficznych. Stosunek mutacji synonimicznych wobec niesynonimicznych wyniósł 1,35. Wśród genów funkcjonalnych mutacje synonimiczne są częstsze i dlatego odpowiadają za większość sygnału filogenetycznego (Wang i in. 2022).

Powszechne zjawisko polegającego na tym, że liczba niesynonimicznych SNP jest mniejsza niż synonimicznych SNP w genomie chloroplastu innej rośliny miewa czasem dysonans, co miało miejsce w badaniach nad sześcioma kompletnymi plastomami z rodzaju *Chenopodium* (Komosa) z rodziny *Amaranthaceae* (Szarłatowatych) (Park i in. 2021). Niesynonimiczny i synonimiczny wzorzec substytucji nukleotydów jest ważnym markerem ewolucji. Stosunek mutacji niesynonimicznych do synonimicznych wskazuje na presję selekcyjną wywieraną na geny. Kiedy wartość tego stosunku jest mniejsza od jedności wskazuje na selekcję oczyszczającą, gdy przyjmuje wartość 1 dowodzi o ewolucji neutralną, podczas gdy wartość powyżej jedności demonstruje selekcję pozytywną (Lawrie i in. 2013). Geny, które wykazywały wyższą wartość mutacji niesynonimicznych w stosunku do mutacji synonimicznych u niektórych gatunków, mogą podlegać niskiej presji selekcyjnej i większej selekcji pozytywnej z powodu pewnych warunków środowiskowych, takich jak gen *clp*P w rodzinie *Malvaceae* (Malwowate): *Firmiana Major* (7.4748), *Heritiera parvifolia* (5.5191) i *Althaea officinalis* (Prawoślaz lekarski; 1.1626). Wiązać się to może z różnym stopniem stresów biotycznych lub abiotycznych, z jakimi borykają się różne gatunki w swoich niszach ekologicznych (Abdullah i in. 2020). Niektóre geny wyróżniły się wysokim poziomem mutacji niesynonimicznych w niniejszym badaniu, przy czym gen *rps*12 wykazał zaistnienie jedynie 6 mutacji niesynonimicznych. Gen *mat*K wykazał stosunek mutacji niesynonimiczny wobec synonimicznych: 1,1428; *psa*B (8), *rpo*C2 (0,5357) czy *ndh*F (0,6923). Wysokie poziomy mutacji niesynonimicznych *mat*K odnotowyano dotychczas w wielu liniach roślin (Han i in. 2022). W rezultacie region ten jest cenną sekwencją kodu kreskowego w badaniach filogenetycznych.

Poziomy niesynonimicznych mutacji w sekwencji genu *mat*K mogą sugerować, że gen taki przyzbliża się do pseudogenizacji, ale konserwatywna wymiana aminokwasów w głębokich podziałach filogenetycznych sprawia, że podstawowe funkcje zostają zachowane (Barthet i Hilu, 2008). Podobnie jak gen *mat*K, gen *ndh*F posiada specyficzne miejsca, które wydają się być wysoce konserwatywne dla podstawowych funkcji a niesynonimiczne zmiany poza tymi miejscami nie zaburzają ich funkcji ani struktury (Park i in. 2021). Porównanie częstotliwości występowania mutacji synonimicznych i niesynonimicznych celem określenia zmian funkcjonalnych w sekwencjach kodujących, może nie ująć należycie wpływu zmian aminokwasowych na funkcje komórek i wpływu późniejszej selekcji na te zmiany. W innych przypadkach selekcja może działać w celu promowania różnorodności aminokwasowej i funkcjonalnej w niektórych genach, aby umożliwić szybką adaptację do zmian środowiskowych, takich jak np. odmienny poziom natężenia światła (Bludau i Aebersold, 2020), w zestawieniu danych z badanych genomów *Erica*, gen związany z procesem fotosyntezy *psa*B posiada także wysoki poziom zmian niesynonimicznych.

W zestawieniu dwóch plastomów *Erica* mutacji SNP zanotowano w regionach kodujących 405, podczas gdy w regionach niekodujących wystąpiło ich aż 960. Jest to znaczący polimorfizm, w zestawieniu dwóch gatunków tego samego rodzaju. W przypadku np. trzciny cukrowej wiele porównań nie wykazywało żadnych zmian, a jedno z nich zlokalizowało występowanie zaledwie 24 mutacji SNP (Xu i in. 2019). Lokalna dystrybucja i separacja ekologiczna mogą być przyczyną znacznych różnic genetycznych między gatunkami tego samego rodzaju i mogą wyjaśnić wysoką różnorodność nukleotydową gatunków w tym badaniu. Stosunek zróżnicowania nukleotydowego rejonów niekodujących wobec rejonów kodujących był u *Erica* ponad czterokrotnie wyższy. Polimorfizmy pojedynczego nukleotydu, po znormalizowaniu na pary zasad, zgodnie z oczekiwaniami były zlokalizowane w większej liczbie w regionach międzygenowych, ponieważ regiony kodujące są na ogół bardziej konserwatywne niż regiony niekodujące, a te ewoluują szybciej niż sekwencje kodujące (Shaw i in. 2005). Podobne trendy kreują się w wielu tego typu badaniach. Analiza dziewięciu odmian pszenicy wykazała występowanie w sumie 564 mutacji SNP i 160 indeli, z których 230 SNO i 4 InDel zlokalizowano odpowiednio w regionach kodujących białka (Hassoubah i in. 2020).

Wyższy poziom dywergencji w regionach niekodujących względem sekwencji kodujących oszacowano także w rodzaju *Avena* , *Chenopodium* czy rodzinie *Malvaceae* (Liu i in. 2020, Abdullah i in. 2020; Park i in. 2021). Nie jest to jednak normą, rozkład SNP i InDel bywa także odnotowany ze stosunkowo niewielką przewagą w regionach niekodujących, jak miało to miejsce w badaniach porównujących 13 sekwencji plastomów ziemniaka, gdzie jedynie ponad 50% SNP jest obecnych w regionach międzygenowych i intronowych i tylko w jednym przypadku statystyka SNP w regionach niekodujących przewyższyła o ponad 100 tę zlokalizowaną w CDS (Achakkagari i in. 2020).

Średnia odległość genetyczna kompletnych genomów *Erica* wyniosła 0,01385. Odległość genetyczna pomiędzy genomami *Vaccinium* plasowała się w obrębie wartości zdecydowanie niższych – zmienność dla pary gatunków *Vaccinium bracteatum* a *Vaccinium oldhamii* wyniosła p = 0.0087. Natomiast zdecydowanie wyższą wartość p = 0,8838 otrzymano dla pary gatunków z rodzaju *Rhododendron*: między *Rhododendron platypodum* a *Rhododendron pulchrum* (najwyższy poziom zmienności). Odległości genetyczne w rodzinie *Ericaceae* plasują się w wartościach stosunkowo wysokich. W rodzinie *Solanaceae* (Psiankowate) na poziomie genomu odległości genetyczne między pięcioma gatunkami z rodzaju *Physalis* (Miechunka) wahały się od 0,0007 do 0,0048, a średnia odległość genetyczna wynosiła zaledwie 0,0029 (Feng i in. 2020). Jednak nie są to wartości nie występujące w literaturze. W rodzinie *Rosaceae* (Różowate) międzygatunkowe odległości genetyczne wahały się od 0,0285 dla *Eriobotryae japonica* (Nieśplik japoński) i *Eriobotryae deflexa* do 0,8665 dla *Magnolia grandiflora* (Magnolia wielkokwiatowa) i *Dillenia turbinata,* gdzie występowały także wartości zmienności 0,7495 dla *Magnolia grandiflora* i *Eriobotryae fragrans* czy 0,1195 dla *Magnolia grandiflora* i *Dillenia indica* (Li i in. 2021).

Analizy porównawcze regionów kodujących ujawniły szerokie spektrum zmienności, od mutacji punktowych do małych/średnich insercji/delecji (InDels), przy czym najbardziej zmiennymi genami okazały się *rpl*22, *rpl*23, *rpo*B, *rpo*C1, *rpo*C2, *ndh*A, *ndh*B, *ycf*1 oraz *ycf*2. Ogólnie rzecz biorąc, łącznie 60 genów ma SNP, a kilka genów wykazuje wyższą frekwencję SNP od pozostałych.

Największą liczbę SNP zidentyfikowano w genach *clp*P (47 SNP), *rpo*C2 (43 SNP), *ycf*1 (40 SNP) i *rpl*22 (21 SNP). Gen *clp*P wykazał także najwyższą liczbę InDel wśród regionów kodujących – 13 indeli. Podobnie sytuacja przedstawiała się w analizie plastomów ziemniaka, gdzie gen clpP wykazał 17 indeli i było to najwięcej w zestawie genów. Jednak w tamtej analizie wysoką liczbę indeli wykazały także geny *mat*K i *ycf*1, które w niniejszej analizie nie wykazały ani jednej zmiany typu InDel (Achakkagari i in. 2020).

Spośród opisanych genów kodujących białka w rodzaju *Erica* 20 wykazało charakter doskonale zachowanych sekwencji nukleotydowych. Stwierdzono powolne tempo ewolucji genów fotosyntezy, co jest powszechne w komórkach roślinnych (Menezes i in. 2018). Globalne dopasowania parami specyficzne dla gatunku wykazały, że średnia zmienność nukleotydowa mieści się w zakresie od 0,015 dla regionów kodujących do 0,067 dla regionów niekodujących. Jest to wartość wysoka, dla 6 plastomów z rodzaju *Chenopodium* średnia wartość różnorodności nukleotydów wyniosła 0,0000625 (Park i in. 2021). W badaniach nad rodziną *Orchidaceae* dochodzenia Bi i in. (2018) różnorodność nukleotydów ośmiu gatunków z rodzaju *Fritillaria* (Szachownica) ustaliły w granicach od 0 do 0,02583, w badaniach z 2019 Somaratne i in. wartość π wynosiła od 0 do 0,05872 dla dziewięciu genomów chloroplastowych gatunków *Eragrostis* (Miłka), natomiast w 2020 wartości π były zdecydowanie wyższe, wahały się od 0 do 0,34889 w rodzaju *Paphiopedilium* (Sabotek) (Vu i in. 2020).

Ogólnie wyższa zmienność nukleotydów występuje w regionach niekodujących i wpływa na odcinki A/T, liczbę jednostek powtórzeń tandemowych i regiony DNA o niskiej złożoności. Ponadto w sekwencjach niekodujących zidentyfikowano bardzo duże InDele. Niektóre z nich mogą być bardziej artefaktami niż realnymi insercjami/delecjami (Hassoulbah i in. 2020). Ponad 90% zgromadzonych indeli obecnych jest w regionach międzygenowych, co może mieć wpływ na regulację ekspresji genów. Jedynie niewielka liczba InDel odnotowana została w genach.

W oparciu o wyniki informatywności filogenetycznej w przeliczeniu na miejsce oraz ogólnej, sieciowej najbardziej obiecującymi markerami filogenetycznym dla rodzaju *Erica* okazały się geny *rpo*C2, *ndh*F, *rpl*22, *rps*16, *rrn*16, *psa*J, *psb*B, *mat*K, *ccs*A oraz *atp*F. Dwa z nich wykazały stuprocentową identyczność w porównaniu między genomami *Erica arborea* a *Erica versicolor* – były to *rps*16 oraz *psa*J. Pozostałe natomiast wykazały duży stopień polimorfizmu. Wartości π tych genów wyniosły odpowiednio: *rpo*C2 (0,0115), *ndh*F (0,0232), *rpl*22 (0,1873), *rrn*16 (0,0322), *psb*B (0,0026), *atp*F (0,0047), *mat*K (0,0098), *ccs*A (0,0041). Geny *ndh*F, *rpo*C2 i *rpl*22 w wysoce konserwatywnych plastomach wykazały zmienność w badaniach nad rodziną *Apocynaceae* (Toinowate) (Zheng i in. 2022). Ponadto gen rpl22 w wielu badaniach wykazuje wysoką zmienność na tle zestawu genów chloroplastowych np. w rodzinie *Caryophyllaceae* (Goździkowate), gdzie wykazał najszybszą ewolucję (Lin i in. 2022). Wytypowanie genu *rpo*C2 na skuteczny marker filogenetyczny w rodzaju *Erica* miało już miejsce we wcześniejszych badaniach (Gomolińska i in. 2017). Ze względu na wysoką informatywność filogenetyczną i najniższą niezgodność drzewa genowego, *rpo*C2 jest zalecany także jako obiecujący kod kreskowy plastydowego DNA do badań taksonomicznych w rodzinie *Gentianeae* (Goryczkowate) wraz z wielokrotnie przewijającym się pretendentem genem *mat*K (Zhang i in. 2020). Prowadzone badania wykazują na ogół przydatność genu *mat*K w rozwiązywaniu międzyrodzajowych lub międzygatunkowych relacji rośliny kwitnących oraz traw (Han i in. 2022). Wyrażenie „na ogół" nie jest tu przypadkowe, albowiem w niektórych rodzinach roślin gen *mat*K umożliwia identyfikacje jedynie części gatunków. W rodzinie *Myristicaceae* (Muszkatołowce) nawet poniżej 50%, w rodzinie *Orchidaceae* (Storczykowate) ponad 90% przedstawicieli (Li i in. 2015). Z kolei eksperyment polegający na przeegzaminowaniu 92 gatunków z 32 rodzajów różnych grup roślin zaprezentował wynik 56% skuteczności genu *mat*K w identyfikacji gatunkowej (Fazekas i in. 2008). Zatem limity, które dostarczała sekwencja tego genu w kategorii barkodu sprawiły, iż sugerowano połączenie go wraz z genem *rbc*L, jako swoistym „uzupełnieniem" między innymi z uwagi na łatwość jego amplifikacji. Wolne tempo ewolucji *rbc*L spośród genów chloroplastowych roślin okrytonasiennych ogranicza jego przydatność identyfikacyjną jako samodzielnej jednostki, dlatego do indentyfikacji roślin proponuje się zestaw obu tych genów (*rbc*L wraz z *mat*K).

Subsydiarnym faktem jest, iż na trzeciej konferencji kodów kreskowych DNA, która odbyła się w Meksyku w roku 2009, członkowie grupy roboczej Consortium for the Barcode of Life (CBOL) Plant Working Group zaleciła kombinację rdzeń-kod kreskowy składającą się z części dwóch plastydowych regionów kodujących, *mat*K i *rbc*L, uzupełnionymi dodatkowymi markerami zgodnie z konkretnymi wymaganiami, do których najczęściej należy ITS bądź wybrany region międzygenowy (Park i in. 2023).

Gen małej podjednostki białka rybosomalnego 16 (*rps*16) był używany do pomyślnego rozwiązywania powiązań między rodzajami w wielu rodzinach i podrodzinach, jak np. u *Fabaceae* (Bobowate) czy *Apiaceae* (Selerowate) (Lee i in. 2001; Downie i Katz-Downie, 1999). Jednak stwierdzono, że wykazuje on zbyt dużą zmienność, by rozstrzygać relacje na poziomie rodzajów (Ingram i Doyle, 2003). Chociaż pierwotnie sugerowano, że *rps*16 jest cennym narzędziem do badań na poziomie rodziny i niższym, sugeruje się, że nie zapewnia on wystarczającego wsparcia dla analiz. Ponadto region ten nie może być używany w niektórych taksonach, ze względu na absencję całości bądź części genu *rps*16 w sekwencjach niektórych przedstawicieli *Linaceae* (Lnowate), *Polygalaceae* (Krzyżownicowate), *Turneraceae*, *Violaceae* (Fiołkowate), *Fabaceae* (Bobowate), *Pinus* (Sosna) (Shaw i in. 2005).

Gen kodujący podjednostkę F dehydrogenazy NADP (*ndh*F) obecny w regionie SSC, wykorzystywany jest często z powodzeniem do rekonstrukcji relacji taksonomicznych wnosząc większy wkład znaków informacyjnych niż inne geny będące markerami filogenetycznymi. Zadowalające wyniki z zastosowaniem tego genu osiągnięto zarówno u jednoliściennych jak i u roślin kwitnących z rodziny *Asteraceae*, gdzie oszacowano że wnosi trzykrotnie więcej informacji filogenetycznych niż *rbc*L, ponieważ jest dłuższy i ewoluuje dwukrotnie szybciej oraz przy opracowaniu systematyki endemitów z rodziny *Brassicaceae* (Kapustowate) (Kim i in. 1995; Givnish i in. 2006; Fuentes-Soriano i Kellogg, 2021).

Plastydowy gen *psb*B, który koduje białko rdzeniowe Fotosystemu II i posiada najwyższy poziom translacji wśród genów chloroplastowych, potwierdził użyteczność filogenetyczną swojej sekwencji podczas przetestowania w wielu rodzajach roślin z rodziny *Lamiaceae* (Jasnowate) (Yasodha i in. 2018).

Gen *atp*F kodujący jedną z podjednostek syntazy H+-ATP, która jest niezbędna do transportu elektronów i fotofosforylacji podczas fotosyntezy również wykazuje zastosowanie w filogenetyce – prawdopodobnie odegrał znaczącą rolę w dywergencji gatunków z rodzaju *Quercus* (Dąb) (Yin i in. 2018). Ponadto większość wolniejszych genów ewolucyjnych była związana z fotosyntezą i samoreplikacją. Ze względu na swoje podstawowe funkcje geny te są bardziej konserwatywne i nie wykazują tendencji do zmian, podobnie jak cechy wykazujące wiele innych roślin (Cauz-Santos i in. 2020).

Spośród regionów niekodujących najczęściej proponowanym regionem używanym jako marker jest region *psb*A-*trn*H, który w niniejszych badaniach wykazał stuprocentową konserwatywność między dwoma porównywanymi genomami chloroplastowymi *Erica arborea* i *Erica versicolor*. Region międzygenowy *psb*A-*trn*H jest jednym z najbardziej zmiennych regionów w genomie chloroplastów u nagonasiennych. Jest często proponowany jako odpowiedni do badań nad kodami kreskowymi DNA i jest przydatny w filogenetyce na poziomie gatunku. Jest on bardzo często stosowany w badaniach filogenetycznych i bywa dołączany do kompleksu markerów opartych na genach chloroplastowych (Park i in. 2023). Został także uznany za cenny marker identyfikacji gatunków w rodzinie *Ericaceae* na przykładzie rodzaju *Rhododendron*, gdzie w 94 próbach z 53 gatunków rodzaju wykazał skuteczność plasującą się na poziomie 93,6% (Liu i in. 2012).

Regiony międzygenowe, które w niniejszych badaniach wykazały najwyższy poziom zmienności to m.in. *rpl*32-*ndh*F, *ndh*I-*psa*I, *clp*P-*yc*f3 oraz *rps*16-*rrn*16. Zmienność długości, spowodowana stosunkowo dużymi indelami, odnotowano w opisanych powyżej regionach. W rodzinie *Asteraceae* (Astrowate) w badaniach nad 19 plastomami z rodzaju *Mikania,* region *rpl*32-*ndh*F również okazał się jednym z najbardziej zmiennych regionów w genomie chloroplastowym , tuż obok *rpl*16-*rps*3, który w badaniach nad *Erica* także wykazał zmienność, jednak na niższym poziomie (wartość $\pi = 0, 016$) (Thode i in. 2021). W rodzinie *Lauraceae* (Laurowate) region ten stanowił jeden z kompleksu jedenastu bardziej rozbieżnych sekwencji międzygenowych, jednak nie zaklasyfikował się w kategorii hiperzmiennych (Bai i in. 2022).

Znaczne zróżnicowanie wskazujące na potencjał potencjalnego markera molekularnego wykazał ten region także podczas badań nad dziewięcioma genomami chloroplastowymi banana (rodzaj *Musa*) (Song i in. 2022). Pozostałe regiony stanowią unikalne sekwencje dla tego konkretnego rodzaju *Erica*, natomiast wysoce zmienne regiony międzygenowe stosowane powszechnie w badaniach np. *atp*B-*rbc*L, *rps*16-*trn*R, *apt*F-*apt*I, *rpo*B-*psb*D, *rps*4-*trn*L, *trn*P-*psb*B, *trn*S-*rps*14, *rpl*16-r*pl*2; albo wykazywały zmienność jednak nie na najwyższym poziomie, albo nie wystąpiły w stricte takiej aranżacji sekwencji (Lin i in. 2022).

Wszystkie zidentyfikowane sekwencje genów z poszczególnych grup/rodzin zostały poddane przyrównaniu wielu sekwencji, który został wykonany w celu przeprowadzenia analiz filogenetycznych oraz skonstruowania drzew filogenetycznych, które nie rozstrzygnęły w pełni relacji filogenetycznych w rodzaju *Erica*. Brak rozstrzygnięć związków filogenetycznych i ewolucji genów miał miejsce także w rodzaju *Pinellia*, gdzie znaczące rozbieżności pomiędzy sekwencjami plastomów zaprzeczyły wysokiemu stopniowi konserwatywności w obrębie tego samego rodzaju (Cui i in. 2022).

5.2. Analiza NOR

Porządek sekwencji międzygenowego regionu dystansowego rDNA 26S-18S scharakteryzowano po raz pierwszy w rodzaju *Erica*. Ponadto ujawniono zmienność długości tych regionów u ośmiu przedstawicieli jednego z głównych rodzajów *Ericaceae*.

W filogenetycznych konkluzjach dotyczących różnych poziomów organizacji taksonomicznej dane jądrowej sekwencji rDNA są szeroko i często stosowane. W królestwie roślin najbardziej znaczącym filogenetycznym markerem jądrowym jest region ITS. Region ITS został z powodzeniem wykorzystany w wielu badaniach rekonstrukcji filogenetycznej w różnych kategoriach systematycznych oraz w oszacowaniu związku między bliskimi poznanymi gatunkami w obrębie rodzajów. Przegląd literatury ukazuje nierozwiązane problemy dotyczące szczegółowej analizy filogenetycznej i otrzymanych drzew filogenetycznych w obrębie kladów obejmujących rodzaj *Erica* (Kron i in. 2002).

Badanie przeprowadzone w ramach tej pracy ujawniło niski poziom zmienności ITS u gatunków *Erica* w opracowaniu ich filogenetycznego połączenia. Podobnie profil filogenetycznej informatywności i filogenetyczne drzewo otrzymane z analizy ML o niskim poziomie rozdzielczości ujawniły stosunkowo niski wskaźnik ewolucji sekwencji ITS. Koreluje to z wynikami badań prowadzonych na rodzaju *Stipa*, gdzie najniższy poziom zróżnicowania genetycznego przypisano również regionom ITS oraz ITS2 (Krawczyk i in. 2017). Nie stanowi to normy, bowiem regiony ITS stanowią jedne z bardziej prominentnych markerów filogenetycznych u roślin. Były one z sukcesem stosowane m.in. w wielu badaniach rekonstrukcji filogenetycznych na poziomie rodziny czy rodzaju w rodzinie *Poaceae* (Wiechlinowate) oraz są jednym z powszechniej stosowanych markerów służących określeniu relacji filogenetycznych w rodzinie *Brassiceccae* (Kapustowate) (Edger i in. 2014, Gao i in. 2015).

Zdecydowanie bardziej pomocne w rozdzielczości filogenetycznej z wyższym poziomem niesionej informacji w niniejszych badaniach okazały się wyniki oparte na informacjach dostarczonych przez analizę regionu IGS.

Wiele przeprowadzonych badań wykazało, że region IGS jest cennym markerem w badaniach filogenetycznych roślin kwitnących i doskonałym dla definiowania zmienności wewnątrzgatunkowej, czego dowiodły badania prowadzone na przestrzeni ostatnich lat. Pomiar różnorodności biologicznej jest kluczowym, ale trudnym przedsięwzięciem, czego przykładem są dęby, gdzie złożone wzorce zróżnicowania morfologicznego, ekologicznego, biogeograficznego i genetycznego kolidują z tradycyjną taksonomią, która mierzy różnorodność biologiczną pod względem liczby gatunków (lub wyższych taksonów). Analizy wykorzystujące markery IGS nad rodzajem *Quercus* (Dąb) dowiodły jego wysoki potencjał (Piredda i in. 2020).

IGS charakteryzuje się wysokim tempem ewolucji, co pozwala na ujawnienie wewnątrzgatunkowego polimorfizmu, a także charakteryzuje się odmiennym, niejednorodnym tempem ewolucji, co ilustrują różnorodne funkcje jego elementów. Ogólnie w regionie IGS fragment 5'ETS wskazuje na wyższy poziom konserwatywności, który pokazuje poziom zmienności zbliżony do ITS lub nieznacznie wyższy i bywa także z powodzeniem stosowany jako marker genetyczny w badaniach powiązań filogenetycznych, np. w rodzaju *Rosa* (Róża) (Vozárová i in. 2021).

W gatunkach *Erica* liczba miejsc Pi w regionie 5'ETS nie przekroczyła liczby miejsc Pi wykrytych dla ITS1. Większy poziom informatywności rozpoznany został w regionie IGS, indukowany przez dużą zmienność NTS z jego powtarzającymi się fragmentami. Podobnie kladogram jak i filogenetyczny wykres informatywności dają bardziej wiarygodne dowody na rekonstrukcję filogenetyczną opartą o dane IGS, niż te same informacje otrzymane z sekwencji regionu ITS. Sygnał filogenetyczny uzyskany z 3’ETS + NTS jest dość silny, aby na jego podstawie konstruować wnioski i omawiać związki wewnątrzgatunkowe w rodzaju *Erica*. Zostało to potwierdzone przez analizę ML przeprowadzoną na grupie przedstawicieli *Erica*. Kladogram oparty na sekwencjach 3’ETS + NTS w zestawieniu z kladogramem opartym na regionie ITS daje więcej informacji pomocnych w filogenetycznym wnioskowaniu w studium przypadku analizowanych gatunków. Jednostka 3’ETS + NTS w wewnętrznym regionie dystansowym jest bardziej pomocna w analizie filogenetycznej z rodzaju *Erica* niż inne markery cpDNA stosowane w innych badaniach filogenetycznych.

Na przykład badania filogenetyczne obejmujące szybko ewoluujące sekwencje cpDNA nie potwierdziły analizowanych genów jako wiarygodnych markerów pomocnych w zrozumieniu zależności filogenetycznych między gatunkami należącymi do rodzaju *Erica*. Kladogram oparty na regionie 3'ETS + NTS jest bardziej przydatny niż kladogram oparty na genach *rpo* (Gomolińska i in. 2017). Badanie genów *rpo* oddzieliło *Erica multiflora, Erica arborea* i *Erica cinerea*, które należą do jednego podrodzaju *Euerica*. W kladogramie uzyskanym z partycji IGS zlokalizowano te gatunki w bliższej relacji. Szybka ewolucja sekwencji NTS implikuje występowanie mutacji, homoplazji, nasycenia substytucjami, co powoduje coś w rodzaju szumu informacyjnego. Występowanie tego jest przedstawione w profilu PI, gdzie wierzchołek jest przesunięty w krótkim odstępie czasu od właściwego miejsca na wykresie. Analizy schematu PI ilustrują kongresje mutacji, które nie niosąc ważnych informacji - po lewej stronie szczytu - przeszły przez dużą część badanej osi czasu. Z kolei prawa strona piku wykazuje wysoki poziom informatywności badanych sekwencji dla najwcześniejszego przejścia ewolucyjnego. Sekwencje 3'ETS + NTS są wiarygodnym markerem do powierzchownych badań filogenetycznych. W badaniach opartych na wyższym poziomie ewolucji region 3'ETS + NTS nie jest właściwy i nie może pokazać historii ewolucji i może sugerować niewłaściwe proporcje w renderowaniu linii ewolucyjnej. Z drugiej strony zastosowanie sekwencji dostosowanych w tej pracy do szerokiej skali rozdzielczości filogenetycznej powoduje poważne problemy z uzyskaniem niezawodnego dopasowania, co jest spowodowane dużą rozbieżnością w zmienności poszczególnych fragmentów. Do przeprowadzenia szczegółowego i właściwego badania konieczne było przypisanie powtarzających się motywów regionu NTS i regionów TIS. Przypisanie tych regionów pomaga przygotować wyrównanie wszystkich sekwencji. Region NTS jest konstruowany i kształtowany przez różne delecje i multiplikacje całej tej sekwencji. Przypisanie tych regionów pomaga przygotować wyrównanie wszystkich sekwencji. Aby otrzymać właściwy reprezentatywny obraz ewolucji, należy logicznie komponować powtarzalne fragmenty NTS. Problem z prawidłową interpretacją leży w mnożeniu różnych motywów i dwóch typach mutacji (podstawienie nukleotydów i indeli). Zestaw najlepszego wyboru układu sekwencji powtarzalnych był bardzo trudny, również ze względu na duże rozbieżności długości NTS u badanych gatunków.

Do wyznaczenia motywów powtórzeń w obrębie IGS konieczny jest wybór algorytmu i metodyki najlepiej dopasowanej do badanych sekwencji, na co wskazują badania prowadzone nad gatunkami z rodzajów *Stipa* (Ostnica), *Fagus* (Buk) czy *Quercus* (Dąb). Liczba i rozmieszczenie otrzymanych motywów zależy od liczby analizowanych gatunków. Jedna sekwencja jednego gatunku może mieć inny rozkład powtórzeń niż pełny zestaw studium przypadku. W niniejszej pracy metodologia badań zakłada długość pojedynczego powtórzenia jako stałą wartość, z rozbieżnościami opartymi jedynie na polimorfizmie pojedynczego nukleotydu. Ta sama metoda została wdrożona we wspomnianych badaniach implikacji filogenetycznych w *Stipa* oraz w analizie porównawczej pomiędzy *Fagus sylvatica* (Buk zwyczajny) i *Quercus suber* (Dąb korkowy). (Inácio i in. 2014, Krawczyk i in. 2017).

To rozwiązanie nie jest uniwersalne i zawsze praktykowane. Inni naukowcy wybierali wcześniej metody, która uwzględniały różnice w długości pojedynczych powtórzeń oraz poszukiwanie motywów metodą wyszukiwania wzrokowego (Bailey i in. 2015). Różnorodność zastosowanej metody powoduje dysonans w interpretacji otrzymanych wyników, zwłaszcza gdy są one ze sobą połączone. Jedynym przypadkiem, w którym możliwe jest uzyskanie niemal identycznych wyników, jest sytuacja analizy występowania regularnych i nieskomplikowanych modeli iteracji. Dzięki temu wielu badaczy może dyskutować z różnymi wynikami, porównywać własne wyniki z danymi literaturowymi i znajdować wspólne wzorce. Analiza zmienności IGS w *Stipa* pokazuje wiele wspólnych implikacji w IGS. Zaczynając od odpowiednich powtórzeń, zgodnych wzorców lub wspólnej płaszczyzny do dogłębnej analizy, jej organizacja i struktura są w ogólnym ujęciu podobne do większości innych badanych przypadków. W wielu przypadkach w analizie IGS oznaczono kilka – na ogół od dwóch do pięciu – powtórzeń z regularną częstotliwością pomiędzy poddanymi badaniom. Na przykład w analizie porównawczej *Fagus sylvatica* i *Quercus suber* wykazuje 2 podpowtórzenia u *Fagus sylvatica* (R1 i R2) w sumie w wielu kopiach oraz trzy typy podpowtórzeń u *Quercus suber* (A, B i C) w kilku niedoskonałych kopiach. Charakterystyka molekularna *Punica granatum* (Granat właściwy) wskazała również tylko trzy typy powtarzających się sekwencji (A, B i C) (Parvaresh i in. 2014).

Podobieństwa między długością regionu IGS a jego sekwencjami mają tendencję do zwiększania się pomiędzy blisko spokrewnionymi taksonami. U roślin długość regionu IGS waha się od 1,1 Kb w *Vicia tetrasperma* (Wyka czteronasienna) do 23 Kb w rodzaju *Pinus*. Wiele badań dotyczących niektórych wariantów długości IGS, rozpoznaje dwa warianty długości IGS, jak w przypadku dębu. Sekwencje IGS z unikalnymi podpowtórzeniami mają wspólną historię ewolucyjną. Pomimo opublikowanych badań dotyczących szczegółowych badań IGS nie wykazano takiego stopnia zróżnicowania i wielokrotności powtórzeń w analizowanym rodzaju lub grupie z jednoczesnymi rozbieżnościami między poszczególnymi osobnikami (Inácio i in. 2014, Rosato i in. 2016, Krawczyk i in. 2017).

Rozbieżność IGS w kierunku indywidualizacji gatunkowej w rodzinie *Ericaceae* charakteryzuje się złożonymi podpowtórzeniami oraz różnicami w długości IGS. Sekwencja i budowa strukturalna *Erica* IGS zostały już zdekodowane, zamiast tego, dzięki przedstawionym tu badaniom, można wywnioskować część tendencji ewolucyjnych tego regionu zarówno w rodzaju, jak i w rodzinie *Ericaceae*. Regiony międzygenowe *Erica* i *Chamaedaphne calyculata*, podobnie jak u dębów, wykazywały ogólną organizację typową dla większości IGS złożone z pięciu różnych regionów: sub-powtórzenia (SR), bogate w AT, promotor i sekwencje transkrybowane z zewnątrz (3′ i 5). ′-ETS) - o cechach strukturalnych roślinnych sekwencji IGS i wszystkich elementów funkcjonalnych potrzebnych do aktywności genu rRNA. Zarejestrowano niezgodność długości IGS między testowanymi gatunkami. W *Chamaedaphne calyculata* długość IGS osiągnęła 5041 pz i wyodrębniono 9 motywów nieskorelowanych z motywami dostępnymi w publicznych repozytoriach danych. W rodzaju *Erica* długość IGS sięgała od 2890 u *Erica vagans* do 8620 pz u *Erica arborea*. Motywy rozkładu i konfiguracji w IGS *Erica* były zupełnie inne niż u *Chamaedaphne* i istotnie różniły się pomiędzy sobą. Nie można stworzyć wspólnych bloków motywów dla wszystkich gatunków *Erica*, jak miało to miejsce pośród przedstawicieli innych rodzajów. Początkowy powtarzający się region oflankowany przez jeden unikalny motyw (ER10) otwierał wszystkie IGS-y rodzaju *Erica*. Kolejnym wspólnym wzorem jest część końca 5'ETS, gdzie wspólne dla wszystkich gatunków *Erica* są trzy motywy (kolejno: ER8, ER7 i ER8).

W przeciwnym razie ogólny test składu motywów wykazuje znacznie większe rozbieżności w NTS niż w 5'ETS i brak wspólnych powtórzeń dostępnych i opublikowanych danych. Zestaw motywów nie jest klarowny i łatwy jak w – już skomplikowanym systemie – w powtarzalnych regionach *Stipa* czy u badanych dębów. Różnice między gatunkami odzwierciedlają ilość mutacji, które można określić na podstawie poziomu indywidualnej dywergencji rDNA w genomie gatunku. Wysokie podobieństwo obserwuje się między regionami kodującymi rDNA, ale nie między poszczególnymi podpowtórzeniami IGS, może to być spowodowane tym, że regiony kodujące i niekodujące nie rozwijają się razem i wykazują odmienne tempo zmienności. Spacery mają tendencję do bycia bardziej zmiennymi niż regiony kodujące ze względu na dobór naturalny i zapobieganie rozprzestrzenianiu się mutacji w genach. Każdy z badanych gatunków reprezentuje inny region geograficzny, posiada różnice morfologiczne i ekologiczne – prawdopodobnie brak wspólnego schematu kolejności motywów odpowiada określonej reakcji na zmiany środowiskowe, rozróżnienie i historię ewolucji (Krawczyk i in. 2017, Countihno i in. 2016).

Podsumowując prowadzoną pracę, region IGS, a w szczególności region w pobliżu 26S rDNA reprezentują sekwencje, które mogą być wykorzystane jako marker molekularny w analizie filogenetycznej rodzaju *Erica*. Porównanie jednostek powtórzeń w obrębie NOR z *Chamaedaphne calyculata* i przedstawicieli *Erica* pokazuje, że organizacja jednostek nDNA *Ericaceae* nie odbiega od uniwersalnej kombinacji. Tempo ewolucji w powtórzeniach jest niezwykle wysokie i wstępnie wyznaczyło testowany region do przyszłej dogłębnej rekonstrukcji filogenezy i szczegółowych analiz w rodzaju *Erica*. Odkrycia te dodatkowo potwierdzają intensywną inkoherencję rodzaju *Erica* i poszerzają wiedzę na temat różnorodności, rozmieszczenia i ewolucji rodziny *Ericacea*.

5.3. Potencjał superbarkordingowy plastomów oraz regionów NOR

Kody kreskowe DNA roślin często wykazywały swoje ograniczenia w rozróżnianiu gatunków, zwłaszcza w przypadku blisko spokrewnionych taksonów, czyniąc nierealistyczną możliwość wykorzystania uniwersalnego zestawu znaczników do identyfikacji gatunków roślin wyższych. To ograniczenie jest dodatkowo zaostrzone w przypadku zastosowania na poziomie wewnątrzgatunkowym. Podczas gdy tradycyjne kody kreskowe często mają trudności z niezawodnym rozróżnieniem w obrębie gatunków, pełne sekwencje plastomów okazują się skuteczne dzięki globalnym wzorcom zmian sekwencji. Rzeczywiście, dzięki zastosowaniu pełnego kodu kreskowego plastomu jako metody z wyboru do identyfikacji roślin, przewiduje się rosnące wykorzystanie plastomów pełnej długości w identyfikacji i identyfikowalności organizmów (Yu i in. 2021).

Aktualnie za sprawą intensywnego rozwoju metod sekwencjonowania DNA sugeruje się użycie kompletnej sekwencji genomu chloroplastowego jako roślinnego barkodu. Wynika to m.in. z wysokiego stopnia konserwatyzmu w strukturze genomu, niskiego poziomu substytucji nukleotydowych oraz braku rekombinacji. Zastosowanie pełnych sekwencji plastomów wyklucza zjawisko subiektywnego rekomendowania konkretnych regionów w genomie chloroplastowym przez badaczy. Barkoding DNA jest obecnie szeroko stosowanym i skutecznym narzędziem, które umożliwia szybką i dokładną identyfikację gatunków roślin; jednak żadne z dostępnych loci nie działa we wszystkich gatunkach. Ponieważ kody kreskowe DNA pojedynczego locus nie mają odpowiednich zmienności u blisko spokrewnionych taksonów, ostatnie badania nad barkodami położyły duży nacisk na zastosowanie sekwencji całego genomu chloroplastowego, które są teraz łatwiej dostępne w wyniku udoskonalenia technologii sekwencjonowania. Podczas gdy sekwencjonowanie genomu chloroplastów może już dostarczyć wiarygodnych barkodów do dokładnej identyfikacji roślin, nie jest jeszcze zasobooszczędne i nie oferuje jeszcze szybkości analizy zapewnianej przez barkody z jednym locus niewyspecjalizowanym obiektom laboratoryjnym (Zhang i in. 2021).

Wykorzystanie genomu chloroplastowego jako barkodu w literaturze widnieje pod terminem tzw. „superbarkodu". Takie podejście do barkodu DNA, które dla wybranych grup taksonów łączy najlepsze wykorzystanie barkodów pojedynczego miejsca i super barkodów w celu skutecznej identyfikacji roślin. Specyficzne barkody mogą zwiększyć zdolność do rozróżniania blisko spokrewnionych roślin zarówno na poziomie gatunku jak i populacji (Pouchon i in. 2022).

Superbarkoding z powodzeniem został wykorzystany do szczegółowych badań nad wieloma rodzinami roślin, między innymi u roślin leczniczych, które były stosunkowo trudne do identyfikacji. Barkoding DNA jako technologia identyfikacji i klasyfikacji molekularnej jest stosowana w roślinach leczniczych od roku 2008. Zastosowanie tej techniki w znacznym stopniu zapewniło bezpieczeństwo i skuteczność materiałów leczniczych, uwierzytelniając gatunki roślin leczniczych (Yu i in. 2021). Gatunki o najbliższych relacjach filogenetycznych są również podobne pod względem składników chemicznych i efektów leczniczych. Wraz z rozwojem sekwencjonowania nowej generacji (NGS) wykazano również, że technologia barkodowania skutecznie identyfikuje gatunki w próbkach mieszanych. W rodzinie goryczkowatych (*Gentianeceae*) rodzaj goryczka (*Gentiana*) stanowi grupę roślin leczniczych wykazujących właśnie komplikację identyfikacyjną. Badania przeprowadzone nad reprezentantami rodzaju goryczki zasiedlających prowincję Yunnan zaimplementowały identyfikację przy pomocy superbarkodingu DNA w oparciu o genomy chloroplastowe. Otrzymane wyniki uskazały, że superkodowanie DNA ma oczywiste zalety w rozróżnianiu gatunków w przypadku skomplikowanych rodzajów roślin, analizy te dostarczyły nie tylko naukowej podstawy do identyfikacji ale i ochrony poszczególnych przedstawicieli z rodzaju *Gentiana* (HS i DQ, 2021). Podobne podejście zastosowano w rodzinie makowatych (*Papaveraceae*). Przeanalizowano struktury molekularne i powiązania filogenetyczne, używając kompletnych genomów chloroplastów gatunków maku polnego (*Papaver rhoeas*) i maku wschodniego (*Papaver orientale*), dochodząc do wniosku, że genom chloroplastowy może być użyty jako potężne narzędzie do rozwiązywania pozycji filogenetycznych i powiązań w bardzo istotnej medycznie rodzinie *Papaveraceae* (Zhou i in. 2018).

Analiza rozbieżności sekwencji czterech gatunków *Papaveraceae* podobnie jak w niniejszej analizie przedstawicieli z rodzaju *Erica* wykazała, że najbardziej rozbieżne regiony znajdują się w niekodujących sekwencjach z minimalnymi różnicami, które charakteryzują się hiperzmiennością o cechach sugerujących wykorzystanie w barkodingu. W rodzaju szachownicy (inaczej korona; *Fritillaria*) przeanalizowano 26 kompletnych genomów chloroplastowych i ukazano, że z pomocą superbarkodingu można dokonać skutecznych rozróżnień na poziomie gatunkowym, co zapewnia bezprecedensową głębię i skalę w strategiach identyfikacji na niższych poziomach taksonomicznych. Ustalono, iż cpDNA posiada kwalifikacje do pełnienia funkcji uniwersalnego markera w zestawieniu z tradycyjnymi markerami identyfikacji molekularnej, zarówno z uwagi na wspomnianą haploidalność, nierekombinowanie, wysoce konserwatywne sekwencje które mogą funkcjonować jako miejsca pojedyncze. Jednak w przeciwieństwie do pojedynczych sekwencji genów mają większą zmienność i potencjał do identyfikacji gatunków blisko spokrewnionych na niższych poziomach taksonomicznych. Superbarkoding w rodzaju *Fritillaria* wykazał 100% skuteczność identyfikacyjną i również podobnie jak w rodzaju *Erica* otrzymane wyniki kładły duży nacisk na impakt regionów niekodujących i kumulujące się w nich zmiany (Wu i in. 2021). Stwierdzono, że na jego podstawie można z powodzeniem odróżnić wszystkie zarejestrowane w encyklopedii farmakologicznej gatunki z rodzaju, ponadto badania te wykazały iż superbarkod oparty na pełnej długości sekwencji genomu chloroplastowego może nie tylko pomyślnie identyfikować gatunki z rodzaju, ale także analiza relacji filogenetycznych dowiodła iż superbarkoding odzwierciedla parametry biogeograficzne (Wu i in. 2021; Zhang i in. 2021).

Podejście z wykorzystaniem superkodów kreskowych z wykorzystaniem całych plastomów oferuje rozwiązanie ograniczeń związanych z używaniem krótkich regionów kodów kreskowych w celu wyraźnego rozróżnienia różnorodności międzygatunkowej i wewnątrzgatunkowej. Ponieważ w wielu roślinach plastom jest dziedziczony po matce, brak rekombinacji pozwala zachować rozmiar genomu, liczbę genów i kolejność genów w większości roślin. Jednak między gatunkami kumulują się wystarczające różnice, aby umożliwić oszacowanie ich ścieżki ewolucyjnej. Markery plastydowe wykazują nieraz

fałszywe wykrywanie spowodowane złożonością genomu wynikającą chociażby ze zróżnicowania wewnątrzgatunkowego. Zbiorcze zastosowanie plastomów i sekwencji rDNA może zmniejszyć błędy wykrywania i zwiększyć moc superbarkodingu (Park i in. 2023). Obecnie coraz częściej stosuje się metodę łączenia analizy chloroplastowego DNA wraz z rDNA, celem skuteczności wyklarowania wniosków, co miało miejsce m.in. w badaniu ewolucji i różnorodności *Euonymus hamiltonianus* (Trzmielina Hamiltona) z rodziny *Celastraceae* (Dławiszowate) (Park i in. 2022).

W badaniach nad rozróżnieniem morfologicznie podobnych gatunków roślin zielnych z rodziny Asparagaceae (Szparagowate): *Liriope platyphylla* oraz *Ophiopogon japonicus* (Konwalnik japoński), posłużono się kompletnymi sekwencjami plastomu oraz regionem 45S rDNA, gdzie zidentyfikowano polimorfizmy między- oraz wewnątrzgatunkowe. Drzewa oparte na jednej i drugiej partycji danych wykazały identyczną typologię na poziomie międzygatunkowym, a prowadzone analizy potwierdziły potencjał superbarkodingu (Lee i in. 2022). Jednakże nie jest to rozwiązanie absolutnie doskonałe i pozbawione rys, nie ma bowiem do tej pory żadnej metody stuprocentowej, która sprosta wszystkim wyzwaniom labilności warunków środowiska i mechanizmów ewolucyjnych. Dowodzą temu niezadowalające wyniki badań nad rodzajem *Schima* z rodziny *Theaceae* (Herbatowate), gdzie zastosowanie macierzy danych kompletnych plastomów i jądrowego rybosomalnego DNA (rDNA) jako superbarkodu żadna z analizowanych prób nie została prawidłowo przypisana do gatunku przy użyciu standardowych kodów kreskowych DNA i macierzy nrDNA, natomiast jedynie 27,27% gatunków z wieloma rozszerzeniami zostało rozróżnionych przy użyciu genomu plastydowego i jego podzielonych zestawów danych. Osiągnięto tymi badaniami najniższy szacowany wskaźnik sukcesu superkodu kreskowego w literaturze do tej pory (Yu i in. 2022).

6. Wnioski

1. Porównanie pełnych sekwencji genomów chloroplastowych *Erica arborea* i *Erica cinerea* wykazało istniejące różnice pomiędzy analizowanymi genomami. Genom chloroplastowy *Erica versicolor* jest dłuższy od *Erica arborea* o 4 197 pz; rejon LSC jest dłuższy o 4 224 pz, rejony SSC są niemal identyczne w długości (odpowiednio 2 571 pz i 2 570 pz), natomiast rejony odwróconych powtórzeń są dłuższe u *Erica arborea* o 14 pz.

   Porównując genomy, zgodnie z zestawieniem wykonanym za pomocą Geneious Alignment, zidentyfikowano 131,517 miejsc identycznych pomiędzy sekwencjami analizowanych genomów o poziomie identyczności pairwise identity 91,6%. Frekwencja nukleotydów po wyrównaniu genomów rozkłada się następująco: A (32,4%), T (32%), C (17,8%) i G (17,8%). Natomiast średnia zawartość GC 98,552 pz (35,6%).

2. Po zestawieniu porównawczym kompletnych genomów z rodzaju *Erica* względem siebie, wyłączając drugi region odwróconego powtórzenia IRB, łącznie zidentyfikowano 5 570 miejsc polimorficznych. W regionach kodujących zidentyfikowano 454 miejsc polimorficznych, w tym 49 insercji/delecji, 233 mutacje synonimiczne i 172 mutacje niesynonimiczne.

   Stosunek mutacji synonimicznych do niesynonimicznych (S/N) wynosi 1,35. W regionach niekodujących zostało zidentyfikowanych 5116 miejsc polimorficznych, w tym 3556 insercji/delecji oraz 960 mutacji typu SNP. Średnia odległość genetyczna dla całych, porównywanych genomów wyniosła 0,01385.

3. Najbardziej zmiennymi rejonami w regionach kodujących okazały się głównie geny *rpl*22, *rpo*C2, *rpo*C1, *ndh*A, *ndh*B i *rpl*23. Natomiast wśród regionów międzygenowych najbardziej zmienne obszary to *clp*P-*ycf*3, *ndh*I-*psa*I, *rps*16-*rrn*16 i *rpl*32-*ndh*F. Najwyższą wartość zróżnicowania nukleotydowego w regionach kodujących osiągnął gen *rpo*C1 ($\pi = 0{,}3502$).

Natomiast najwyższą wartość zróżnicowania nukleotydowego w regionach niekodujących wykazał region *rps*16-*rrn*16 (π = 0,6918). Średnia zróżnicowania nukleotydowego dla regionów kodujących wyniosła π = 0,015, z kolei dla regionów niekodujących π = 0,067.

4. Charakter plastomów wskazuje na tempo ewolucji występujące pomiędzy gatunkiem południowoafrykańskim (*Erica versicolor)* a rozpierzchniętym gatunkiem bikontynentalnym (*Erica arborea).* Poziom konserwatywności obu plastomów po ich wzajemnym porównaniu uplasował się jedynie na poziomie 63%.

5. Najbardziej prawdopodobnym gatunkiem ancestralnym na aktualnym poziomie badań w rodzaju *Erica* wydaje się być *Erica arborea*, co potwierdziły analizy filogenetyczne z wysokim wsparciem dla kladów, zarówno dla analiz dotyczących pojedynczych genów, jak i całych grup. Potwierdziły to konsensusowe kladogramy otrzymane dla genów/grup genów: *cem*A, *ccs*A, *acc*D, *mat*K, *rbc*L, *ndh*F, *pet*, *psa*, *psb*, *psb*B, *rpo*C2 oraz dla zestawu genów *trn*.

6. Przeprowadzone analizy filogenetyczne wytypowujące informatywność filogenetyczną netto oraz w przeliczeniu na pojedyncze miejsca, wskazały na najbardziej obiecujące markery filogenetyczne w obrębie rodzaju *Erica*. Okazały się nimi geny: *rpo*C2, *ndh*F, *rpl*22, *rps*16, *rrn*16, *psa*J, *psb*B, *atp*F, *mat*K oraz *ccs*A które mogą stanowić asumpt do opracowania sekwencji barkodowych dla badanego rodzaju.

7. Przeprowadzona analiza sekwencji powtórzeniowych wykazała wystąpienie bardzo dużej liczby powtórzeń SSR: u 874 *Erica versicolor* oraz u 814 *Erica arborea*, z czego znaczny procent (w granicach 70%) zlokalizowany był w regionach niekodujących. Zidentyfikowano także 50 powtórzeń oligonukleotydowcych o bardzo okazałych rozmiarach od 187 do 704 pz dla *Erica arborea* oraz od 193 do 706 pz dla *Erica versicolor*.

Dodatkowo znaleziono powtórzenia tandemowe w liczbie 92 dla *Erica versicolor* oraz 99 dla *Erica arborea*, które osiągały do 330 pz. Powtórzenia o dużych rozmiarach, zatem zarówno oligonukleotydowe jak i tandemowe zidentyfikowano w przeważającej liczbie w sekwencjach niekodujących genomów, co może wpływać na rearanżacje genomu przez zakłócanie replikacji DNA i wpłynęło na skomplikowany proces składania genomów.

8. W odniesieniu do analiz regionów NOR, najbardziej informatywnym filogenetycznie obszarem okazał się 3'ETS + NTS, jednak rekonstrukcja filogenetyczna bazująca na tej sekwencji nie wytypowała zadowalającego gatunku ancestralnego, z uwagi na rozłączenie na dwa klady.

9. Wzorce zróżnicowania sekwencji powtórzonych w obrębie IGS odzwierciedlają w pewnym stopniu procesy ewolucyjne zachodzące w rodzaju *Erica*. Sekwencje IGS posiadające unikalne powtórzenia reprezentują wspólną historię ewolucyjną, natomiast przedstawiciele z największą liczbą zróżnicowanych powtórzeń (zdetektowanych motywów) obrazują największą dywergencję i trudności w określeniu filogenezy – jak w przypadku bikontynentalnego gatunku *Erica arborea*.

10. Rekonstrukcja filogenetyczna rodziny *Ericaceae* w oparciu o całe genomy oraz skonstruowana na podstawie regionów odwróconych powtórzeń po raz pierwszy uwzględniła nie tylko gatunki z rodzaju *Erica*, ale także gatunek *Chamaedaphne calyculata*, ponadto nie wykazała żadnych aberracji wobec badań prowadzonych w latach ubiegłych.

11. Analiza regionów odwróconych powtórzeń pozwala w znaczącym stopniu odróżnić sekwencje należące do rodzaju *Erica* od innych przedstawicieli z rodziny *Ericaceae*, unikatowa architektonika regionów IR świadczy o silnym rozwoju ewolucyjnym linii badanego rodzaju.

Praca ta uwzględnia także po raz pierwszy budowę regionów IR dla rodzaju *Chamaedaphne* (dla gatunku *Chamaedaphne calyculata*).

12. Zestawienie poszczególnych drzew oraz analiz filogenetycznych uniemożliwiło zadowalające przeprowadzenie rekonstrukcji filogenetycznej w obrębie rodzaju *Erica*. Niemożliwym pozostaje także ustalenie szczegółowych powiązań pomiędzy analizowanymi gatunkami, ze względu na znacząco odbiegające układy wzajemnych powiązań międzygatunkowych.

Spis literatury

Abdullah, Mehmood, F., Shahzadi, I., Ali, Z., Islam, M., Naeem, M., Mirza, B., Lockhart, P.J., Ahmed, I., Waheed, M.T. (2020). Correlations among oligonucleotide repeats, nucleotide substitutions, and insertion–deletion mutations in chloroplast genomes of plant family *Malvaceae*. *Journal of Systematics and Evolution*, *59* (2), 388-402.

Achakkagari, S.R., Kyriakidou, M., Tai, H.H., Anglin, N.L.. Ellis, D., Strömvik, M.V. (2020). Complete plastome assemblies from a panel of 13 diverse potato taxa. *PloS One*, *10*, doi: 10.1371/journal.pone.0240124.

Adeoyo, O.R. (2021). A Review on Mycorrhizae and Related Endophytic Fungi as Potential Sources of Enzymes for the BioEconomy. *Journal of Advances in Microbiology*, *21*(8), 35-49.

Aguilera, P.M., Gabriele, M., Debat, H.J., Bubillo, R.E., Marti, D.A. (2015). The 18S-25S ribosomal RNA unit of yerba mate (*Ilex paraguariensis* A. St.-Hil.), *Plant Biosystems – An International Journal Dealing with all Aspects of Plant Biology*, *150*: 6, 1240-1248.

Aird, D., Ross, M.G., Chen, W.S., i in. (2011). Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biology*, *12*, 18-21.

Amiryousefi, A., Hyvönen, J., Poczai, P. (2018). IRscope: an online program to visualize the junction sites of chloroplast genomes. *Bioinformatics*, *34* (17), 3030-3031.

Amroun, D., Hamoudi, M., Khennouf, S., Boutefnouchet, S., Harzallah, D., Amrane, M., Dahamna, S. (2021). In-vivo anti-inflammatory activity and safety assessment of the aqueous extract of Algerian *Erica arborea* L. (*Ericaceae*) aerial parts. *Journal of Ethnopharmacology*, Vol. *271*, 113881.

Anastassakis, K. (2022). A Few More and Recently Reported Herbs, w: *Androgenetic Alopecia From A to Z. Vol.2 Drugs, Herbs, Nutrituion and Supplements*. Springer. Cham, 507-519.

Arrigoni, R., Vacherie, B., Benzoni, F., Stefani, F., Karsenti, E., Jaillon, O., Not, F., Nunes, F., Payri, C., Wincker, P., Barbe, V. (2017). A new sequence data set of SSU rRNA gene for *Scleractinia* and its phylogenetic and ecological applications. *Molecular Ecology Resources*, *17*(5), 1054-1071.

Asaf, S., Khan, A.L., Lubna, Khan, A., Lee, I., Al-Harrasi, A. (2020). Expanded inverted repeat region with large scale inversion in the first complete plastid genome sequence of *Plantago ovata*. *Scientific Reports*, *10* **,** 3881.

Avise, J.C. i in. (red.) (1996). *Conservation Genetics: Case Histories from Nature*; Springer Publisher, 2-510.

Azuma, Y., Tsuru, S., Habuchi, M., Takami R., Takano S., Yamamoto K., Hosoda K. (2023). Synthetic symbiosis between a cyanobacterium and a ciliate toward novel chloroplast-like endosymbiosis. *Scientific Reports*, *13*, 6104.

Baba, T., Hirose, D., Noma, S., Ban, T. (2021). Inoculation with two *Oidiodendron maius* strains differentially alters the morphological characteristics of fibrous and pioneer roots of *Vaccinium virgatum* 'Tifblue' cuttings. *Scientia Horticulturae,* vol. *281*, 109948.

Bai, X., Peng, J., Yang, Y., Xiong, B. (2022), The Complete Chloroplast Genome Sequence of *Machilus chuanchienensis* (Lauraceae): Genome Structure and Phylogenetic Analysis. *Genes*, *13*, 2402.

Bailey, T. L., Bodén, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W.W., Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching", *Nucleic Acids Research*, *37*, 202-208.

Barthet, M.M., Hilu, K.W. (2008). Evaluating Evolutionary Constraint on the Rapidly Evolving Gene matK Using Protein Composition. *Journal of Molecular Evolution*, *66*, 85–97.

Beier, S., Thiel, T., Münch, T., Scholz, U., Mascher, M. (2017). MISA-web: a web server for microsatellite prediction. *Bioinformatics*, *33*, 2583–2585.

Bekkai, D., Majdoub, Y.O., Bekkai, H., Cacciola, F., Miceli, N., Taviano, M.F., Cavò, O., Errabii, T., Vinci, R.L., Mondello, L., El Kbiach, M.B. (2022). Determination of the Phenolic Profile by Liquid Chromatography, Evaluation of Antioxidant Activity and Toxicity of Moroccan *Erica multiflora*, *Erica scoparia*, and *Calluna vulgaris* (*Ericaceae*). *Molecules*, *27* (13), 3979.

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research*, *27* (2), 573-580.

Benson, D.A. Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Sayers, E.W. (2013). GenBank. *Nucleic Acids Research*, *41*(D1), 36-42.

Blazier, J.C., Jansen, R.K., Mower, J.P., Govindu, M., Zhang, J., Weng, M.L., Ruhlman, T.A. (2016). Variable presence of the inverted repeat and plastome stability in *Erodium*. *Annals of Botany*, *117* (7), 1209–1220.

Bludau, I., Aebersold, R. (2020). Proteomic and interactomic insights into the molecular basis of cell functional diversity. *Nature Reviews Molecular Cell Biology*, *21*, 327–340.

Booyens, J. (2012). Sunburst : colourful associations between sunbirds and flowers at Kirstenbosch. *Veld* & *Flora*, *98* (2), 56-59.

Borowska-Zuchowska, N., Kwasniewski, M., Hasterok, R. (2016). Cytomolecular analysis of ribosomal DNA evolution in a natural allotetraploid *Brachypodium hybridum* and its putative ancestors – dissecting complex repetitive structure of intergenic spacers. *Frontiers in Plant Science*, *7*: 1499, 1-15.

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., Buckler, E.S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, *23*, 2633-2635.

Bradley, R., Burt, A.J., Read, D.J. (1982). The biology of mycorrhiza in the *Ericaceae*. *New Phytologist*, *91* (2), 197-209.

Brickell, C. (2003). *RHS A-Z Encyclopedia of Garden Plants. Third edition*. Dorling Kindersley, London.

Brown, N., Duncan, G. (2006). *Grow fynbos plants*. Kirstenbosch Gardening Series. South African National Biodiversity Institute, Cape Town, 1-240.

Casarrubia, S., Martino, E., Daghino, S., Kohler, A., Morin, E., Khouja, H.R., Murat, C., Barry, K.W., Lindquist, E.A., Martin, F.M., Perotto, S. (2020). Modulation of Plant and Fungal Gene Expression Upon Cd Exposure and Symbiosis in Ericoid Mycorrhizal *Vaccinium myrtillus*. *Frontiers in Microbiology*, *11*, doi: 10.3389/fmicb.2020.00341.

Cairney, J.W.G., Meharg, A.A. (2003). Ericoid mycorrhiza: a partnership that exploits harsh edaphic conditions. *European Journal of Soil Science*, *54*, 735-740.

Campbell-Martinez, G.E., Steppe, C., Wilson, S.B., Ball, M., Peterson, C. (2022). Effects of temperature, seed provenance, and substrate on germination of the endemic and threatened Paper nailwort (*Paronychia chartacea ssp. chartacea*). *Florida Scientist*, *85* (1), 1-9.

Castro, E.B., Carro, C.B. (2022). Reseña etnobotánica de la Maragatería Alta (León). *Ethnobotanical review of Maragatería Alta* (province of León, Spain), *83*, 105-126.

Cauz-Santos, L.A., Costa, Z.P., Callot, C., Cauet, S., Zucchi, M.I., Berges, H., Berg, C., Vieira, M.L.C. (2020). A Repertory of Rearrangements and the Loss of an Inverted Repeat Region in *Passiflora* Chloroplast Genomes. *Genome Biology and Evolution*, *12* (10), 1841–1857.

Ceballos, G., Ehrlich, P.R., Baranosky, A.D., García, A., Pringle, R.M., Palmer, T.M. (2015). Accelerated human-in duced species loss: Entering the sixth mass extinction. *Science Advances*, *1* (5): e1400253, 1-5.

Chan, C.X., Gross, J., Yoon, H.S., Bhattacharya, B. (2011). Plastid Origin and Evolution: New Models Provide Insights into Old Problems. *Plant Physiology*, *155* (4), 1552-1660.

Chang, C.C., Lin, H.C., Lin, I.P., Chow, T.Y., Chen, H.H., Chen, W.H., Cheng, C.H., Lin, C.Y., Liu, S.M., Chang, C.C., Chaw, S.M. (2006). The chloroplast genome of *Phalaenopsis aphrodite* (Orchidaceae): comparative analysis of evolutionary rate with

that of grasses and its phylogenetic implications. *Molecular Biology Evolution*, *23* (2), 279-291.

Chauss, D., Freiwald, T., McGregor, R., Yan, B., Wang, L., Nova-Lamperti, E., Kumar, D., Zhang, Z., Teague, H., West, E.E., Vannella, K.M., Ramos-Benitez, M.J., Bibby, J., Kelly, A., Malik, A., Freeman, A.F., Schwartz, D.M., Portilla, D., Chertow, D.S., John, S., Lavender, P., Kemper, C., Lombardi, G., Mehta, N.N., Cooper, N., Lionakis, M.S., Laurence, A., Kazemian, M., Afzali, B. (2022). Autocrine vitamin D signaling switches off pro-inflammatory programs of T_H1 cells. *Nature Immunology*, *23*, 62–74.

Chautá, A., Kumar, A., Mejia, J., Stashenko, E.E., Kessler, A. (2022). Defensive functions and potential ecological conflicts of floral stickiness. *Scientific Reports*, *12*, 19848.

Chen, M.M., Zhang, M., Liang, Z.S., He, Q.L. (2022). Characterization and Comparative Analysis of Chloroplast Genomes in Five *Uncaria* Species Endemic to China. *International Journal of Molecular Science*, *23* (19), 11617.

Chen, L., Liu, X., Wang, Z., Wu, X., Hong, K., Xie, C. (2023). Comparative Chloroplast Genome Analyses of Six Hemlock Trees in East Asia: Insights into Their Genomic Characterization and Phylogenetic Relationship. *Forests*, *14* (11), 2136.

Coutinho, P.J., Carvahlo, A., Martín, A., Ribeioro, T., Moraris-Cecílio, L., Lima-Brito, J. (2016). Oak ribosomal DNA: characterization by FISH and polymorphism assessed by IGS PCR-RFLP. *Plant Systematic and Evolution*, *302*, 527-544.

Cowie, R.H., Bouchet, P., Fontaine, B. (2022). The Sixth Mass Extinction: fact, fiction or speculation? *Biological Reviews*, *97* (2), 640-663.

Crayn, D.M., Hislop, M., Puente-Lelièvre, C. (2020). A phylogenetic recircumscription of *Styphelia* (Ericaceae, Epacridoideae, Styphelieae). *Australian Systematic Botany*, *33* (2), 137-168.

Cui, N., Liao, B.S., Liang, C.L., Li, S.F., Zhang, H., Xu, J., Li, X.W., Chen, S.L. (2020). Complete chloroplast genome of *Salvia plebeia*: organization, specific barcode and phylogenetic analysis. *Chinese Journal of Natural Medicines*, *18* (8), 563-572.

Cui, N., Chen, W., Li, X., Wang, P. (2022). Comparative chloroplast genomes and phylogenetic analyses of *Pinellia*. *Molecular Biology Reports*, *49*, 7873–7885.

Daniell, H., Jin, S., Zhu, X.G., Gitzendanner, M.A., Soltis, D.E., Soltis, P.S. (2021). Green giant—a tiny chloroplast genome with mighty power to produce high-value proteins: history and phylogeny. *Plant Biotechnology Journal*, *19* (3), 430-447.

Darling, A.E., Mau, B., Perna, N.T. (2010). Progressive Mauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. *PloS One*, *5* (6), 1-17.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, *9* (8), 772.

Désamoré, A., Laenen, B., Devos, N., Popp, M., González-Mancebo, J.M., Carine, M.A., Vanderpoorten, A. (2010). Out of Africa: north-westwards Pleistocene expansions of the heather *Erica arborea*. *Journal of Biogeography*, *38* (1), 164-176.

Dinerstein, E., Vynne, C., Sala, E., Joshi, A.R., Fernando, S., Lovejoy, T.E., Mayorga, J., Olson, D., Asner, G.P., Baillie, J.E.M., Burgess, N.D., Burkart, K., Noss, R.F., Zhang, Y.P., Baccini, A., Birch, T., Hahn, N., Joppa, L.N., Wikramanayake, E. (2019). A global deal for nature: Guiding principles, milestones, and targets. *Science Advances*, *5* (4), eaaw2869. https://doi. org/10.1126/sciadv.aaw2869.

Dinh, X.T., Mai, X.C., Nguyen, M.L. (2023). Development of short tandem repeat (STR) and derived cleaved amplified polymorphic (dCAPS) markers for distinguishing species and varieties of the genus *Panax* in Vietnam. *Genetic Resources and Crop Evolution*, *70*, 2047–2054.

Dobrogojski, J., Adamiec, M., Luciński, R. (2020). The chloroplast genome: a review. *Acta Physiologiae Plantarum*, *42* (98), 39-45.

Downie, S., Katz-Downie, R.D.S. (1999). Phylogenetic analysis of chloroplast rps16 intron sequences reveals relationships within the woody southern African Apiaceae subfamily *Apioideae*. *Canadian Journal of Botany*, *77*, 1120– 1135.

Durán Gómez, J.A., Rodriguez, R.R. (2022). Los brezales litorales sobre dunas en Cantabria y su flora asociada. *Flora Montiberica*, *84*, 94-97.

Dutta, U. (2022). Essentials of Cytogenetic and Molecular Cytogenetic Laboratory Testing. Cambridge Scholars Publishing. United Kingdom.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32* (5), 1792-97.

Edger, P.P., Tang, M., Bird, K.A., Mayfield, D.R., Conant, G., Mummenhoff, K., Koch, M.A., Pires, C.J. (2014). Secondary Structure Analyses of the Nuclear rRNA Internal Transcribed Spacers and Assessment of Its Phylogenetic Utility across the Brassicaceae (Mustards). *PLoS ONE*, *9* (7), 41-47.

Erickson, M.F., Pessoa, D.M.A. (2022). Determining factors of flower coloration. *Acta Botanica Brasilica*, *36*, e2021abb0299.

Fadaei, S., Khan, S., Young, M., Sherr, I., Zwiazek, J.J. (2021). Impact of soil stockpiling on ericoid mycorrhizal colonization and growth of velvetleaf blueberry (*Vaccinium myrtilloides*) and Labrador tea (*Ledum groenlandicum*). *The Journal of the Society of Ecological Restoration*, *29* (1), e13276.

Fahrenkrog, A.M., Matsumoto, G., Toth, K., Jokipii-Lukkari, S., Salo, H.M., Häggman, H., Benevenuto, J., Munoz, P. (2022). Chloroplast genome assemblies and comparative analyses of major *Vaccinium* berry crops. *Scientific Reports*, *12*, 21600.

Fajardo, D., Senalik, D., Ames, M., Zhu, H., Steffan, S.A., Harbut, R., Palashock, J., Vorsa, N., Gillspie, E., Kron, K., Zalapa, J.E. (2013). Complete plastid genome sequence of *Vaccinium macrocarpon*: structure, gene content, and rearragements revealed by next generation sequencing. *Tree Genetics & Genomes*, *9*, 489-498.

Fazekas, A.J., Burgess, K.S., Kesanakurti, P.R., Graham, S.W., Newmaster, S.G., Husband, B.C., Percy, D.M., Hajibabaei, M., Barrett, S.C.H. (2008). Multiple miltilocus DNA barcodes from the plastid genome discriminate plant species equally well. *PLoS ONE*, *3* (7), e2808.

Feknous, N., Boumendjel, M. (2022). Natural bioactive compounds of honey and their antimicrobial activity. *Czech Journal of Food Sciences*, *40*, 163-178.

Feng, H., Chen, Y., Xu, X., Luo, H., Wu, Y., He, Ch. (2020). The complete chloroplast genome of *Musa acuminate* var. *chinensis*. *Mitochondrial DNA part B, 5* (3), 2691-2692.

Feng, S., Zheng, K., Jiao, K., Cai, Y., Chen, C., Mao, Y., Wang, L., Zhan, X., Ying, Q., Wang, H. (2020). Complete chloroplast genomes of four *Physalis* species (Solanaceae): lights into genome structure, comparative analysis, and phylogenetic relationships. *BMC Plant Biology*, *20*, 242.

Feng, J., Xiong, Y., Su, X., Liu, T., Xiong, Y., Zhao, J., Lei, X., Yan, L., Gou, W., Ma, X. (2023). Analysis of Complete Chloroplast Genome: Structure, Phylogenetic Relationships of *Galega orientalis* and Evolutionary Inference of Galegeae. *Genes*, *14* (1), 176.

Fernandes, L., Ribeiro, H., Oliveira, A., Silva, A.S., Freitas, A., Henriques, M., Rodrigues, M.E. (2021). Portuguese honeys as antimicrobial agents against *Candida* species. *Journal of Traditional and Complementary Medicine*, *11* (2), 130-136.

Foden, W., Potter, L. (2005). *Erica imbricata* L. National Assessment: Red List of South African Plants version 2020. Dostęp 2022/12/13.

Fonseca, L.H.M., Nazareno, A.G., Thode, V.A., Zuntini, A.R., Lohmann, L.G. (2022). Putting small and big pieces together: a genome assembly approach reveals the largest Lamiid plastome in a woody vine. *PeerJ Plant Biology*, *10*, e13207.

Frankham, R., Ballou, D.J., Briscoe, D.B. (2010). *Introduction to Conservation Genetics*. Cambridge University Press, 1-25.

Frazer, K.A., Pachter, L., Poliakov, A., Rubin, E.M., Dubchak, I. (2004). VISTA: computational tools for comparative genomics. *Nucleic Acid Research, 1* (32), 273-279.

Fu Nan, C., Mo, Z.Q., Yang, J.B., Cai, J., Ye, L.J., Zou, J.Y., Qin, H.T., Zheng, W., Hollingsworth, P.M., Li, D.Z., Gao, L.M. (2022). Testing genome skimming for species discrimination in the large and taxonomically difficult genus *Rhododendron*. *Molecular Ecology Resources*, *22*, 404-414.

Fuentes-Soriano, S., Kellogg, E.A. (2021). Molecular Systematics of Tribe Physarieae (*Brassicaceae*) Based on Nuclear ITS, LUMINIDEPENDENS, and Chloroplast *ndh*F. *Systematic Botany*, *46* (3), 611-627.

Gao, G., Deng, J., Gou, X., Wang, Q., Ding, C., Zhang, L., Zhou, Y., Yang, R. (2015). Phylogenetic relationships among *Elysmus* and related diploid genera (Triticeae: Poaceae) based on nuclear rDNA ITS sequences. *Biologia*, *70* (2), 183-189.

Garcia, S., Lim, K.Y., Chester, M., Garnatje, T., Pellicer, J., Vallès, J., Leitch, A.R., Kovařik, A. (2009). Linkage of 35S and 5S rRNA genes in *Artemisia* (family Asteraceae): first evidence from angiosperms. *Chromosoma*, *118*, 85-97.

Geer, L.Y., i in. (2010). The NCBI BioSystems database. *Nucleic Acids Research*, *38*, 492-496.

Geneious 9.0 https://www.geneious.com

Genre, A., Lanfranco, L., Perotto, S., Bonfante, P. (2020). Unique and common traits in mycorrhizal symbioses. *Nature Reviews Microbiology*, *18*, 649-660.

Gillespie, E., Kron, K. (2010). Molecular phylogenic relationships and a revised classification of the subfamily *Ericoidae* (*Ericaceae*). *Molecular Phylogenetics and Evolution*, *56*, 343-354.

Girbig, M., Misiaszek, A.D., Müller, C.W. (2022). Structural insights into nuclear transcription by eukaryotic DNA-dependent RNA polymerases. *Nature Reviews Molecular Cell Biology*, *23*, 603–622.

Givnish, T.J., Pires, J.C., Graham, S.W., McPherson, M.A., Prince, L.M. (2006). Phylogenetic relationships of monocots based on the highly infotmative plastid gene ndhF: Evidence for widespread concerted convergence. *Aliso, 22*, 28-51.

Glen, H.F. (2002). *Cultivated Plants of Southern Africa*. Jacana, Johannesburg.

Goldblatt, P., Manning, J.C. (2000). *Cape Plants: A conspectus of the Cape Flora of South Africa.* Strelitzia 9. Cape Town: National Botanical Institute.

Gomolińska, A.M., Szczecińska, M., Sawicki, J., Krawczyk, K., Szkudlarz, P. (2017). Phylogenetic analysis of selected representatives of the genus Erica based on the genes encoding the DNA-dependent RNA polymerase I. *Biodiversity Research and Conservation*, *46*, 1-18.

Green, B.R. (2011). Chloroplast genomes of photosynthetic eukaryotes. The Plant Journal. Special Issue: *The Plant Genome: An Evolutionary View on Structure and Function*, *66* (1), 34-44.

Greiner, S., Lehwark, P., Bock, R. (2019). OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Research*, *47*, W59-W64.

Gu, X., Zhu, M., Su, Y., Wang, T. (2022). A Large Intergenic Spacer Leads to the Increase in Genome Size and Sequential Gene Movement around IR/SC Boundaries in the Chloroplast Genome of *Adiantum malesianum* (Pteridaceae). *International Journal of Molecular Science*, *23* (24), 15616.

Guisinger, M.M., Kuehl, J.V., Boore, J.L., Jansen, R.K. (2011). Extreme Reconfiguration of Plastid Genomes in the Angiosperm Family *Geraniaceae*: Rearrangements, Repeats, and Codon Usage. *Molecular Biology and Evolution*, *28* (1), 583–600.

Gunn, M., Codd, L.E.W. (1981). Botanical exploration of southern Africa : an illustrated history of early botanical literature on the Cape flora : biographical accounts of the leading plant collectors and their activities in southern Africa from the days of the East India Company until modern times. Cape Town: Published for the Botanical Research Institute by A.A. Balkema.

Guo, L., Guo, S., Xu, J., He, L., Carlson, J.E., Hou, X. (2020). Phylogenetic analysis based on chloroplast genome uncover evolutionary relationship of all the nine species and six cultivars of tree peony. *Industrial Crops and Products*, *153*, 112567.

Hall, T.A. (1999). BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nuclear Acids Symposium Series*, *41*, 95–98.

Hamm, C.A., Rademacher, V., Landis, D.A., Williams, B.L. (2013). Conservation Genetics and the Implication for Recovery of the Endagered Mitchell's Satyr Butterfly, *Neonympha mitchellii mitchellii*. *Journal of Heredity*, *105* (1), 19-27.

Han, T., Li, M., Li, J., Lv, H., Ren, B., Chen, J., Li, W. (2019). Comparison of chloroplast genomes of *Gynura* species: sequence variation, genome rearrangement and divergence studies. *BMC Genomics*, *20* (791), doi.org/10.1186/s12864-019-6196-x.

Han, Y., Gao, Y., Li, Y., Zhai, X., Zhou, H., Ding, Q., Ma, L. (2022). Chloroplast Genes Are Involved in The Male-Sterility of K-Type CMS in Wheat. *Genes*, *13* (2), 310.

Hassoubah, S.A., Farsi, R.M., Alrahimi, J.S., Nass, N.M., Bahieldin, A. (2020). Comparison of Plastome SNPs/INDELs among Different Wheat (*Triticum* sp.) Cultivars. *Research Biosciences Biotechnology Asia*, *17* (1), 2807.

Hilário, S., Santos, L., Phillips, A.L.J., Alves, A. (2022). Caveats of the internal transcribed spacer region as a barcode to resolve species boundaries in *Diaporthe*. *Fungal Biology*, *126* (1), 54-74.

Hitchcock, A. (2016). The history of the *Erica* collection at Kirstenbosch : part 2. *Veld & Flora*, *102* (1), 8-13.

Hitchcock, A. (2020). *Erica verticillata*, from Extinction to Restoration. w: *Science and Actions for Species Protection*. Noah's Arks for the 21st Century. Red. Braun von J., Kauffels T., Raven P., Vogel J., Sorondo M.S. Pontificiae Academiae Scientarvum Scripta Varia 146, 153-176.

Hitchcock, A., Williams, J., Cowell, C. (2020). Lessons learned as *Erica turgida* is returned. *Journal for Nature Conservation, 56*, 58-68.

Holz, H., Segar, J., Veldez, J., Staude, I.R. (2022). Assessing extinction risk across the geographic ranges of plant species in Europe. *Plants People Planet, 4* (3), 303-311.

HS, H., DQ, Z. (2021). DNA super-barcoding of several medicinal species in *Gentiana* from Yunnan province. *China Journal of Chinese Materia Medica, 46* (20), 5260-5269.

Huang, Y., Li, J., Yang, Z., An, W., Xie, C., Liu, S., Zheng, X. (2022). Comprehensive analysis of complete chloroplast genome and phylogenetic aspects of ten *Ficus* species. *BMC Plant Biology, 22* (253), doi.org/10.1186/s12870-022-03643

Huson, D.H., Bryant, D. (2006). Application of Phylogenetic Networks in Evolutionary Studies. *Molecular Biology and Evolution, 23* (2), 254-267.

Inácio, V., Rocheta, M., Morais-Cecílio, L. (2014). Molecular organization of the 25S-18S rDNA IGS of *Fagus sylvatica* and *Quercus suber*. A comparative analysis. *PLoS ONE, 9* (6), 1-12.

Ingram, A. L., Doyle, J. J. (2003). The origin and evolution of *Eragrostis* tef (Poaceae) and related polyploids: evidence from nuclear waxy and plastid rps16. *American Journal of Botany, 90*, 116– 122.

Izza Ab Ghani, N., Arifin, W., Ismail, A. (2022). Conservation Genetics for Managing Biodiversity. *IntechOpen*. doi: 10.5772/intechopen.101872.

Jarosz, Z., Michałojć, Z., Pitura, K., Dzida, K., Koter, M. (2021). Influence of Fertilization and Mycorrhizae on the Nutritional Status of Rhododendron (*Rhododendron hybridum*) in a Nursery. *Agriculture, 11* (6), 538.

Jarrous, N., Mani, D., Ramanathan, A. (2022). Coordination of transcription and processing of tRNA. *The FERBS Journal, 289* (13), 3630-3641.

Jhuraney, A., Velkova, A., Johnson, R.C., Kessing, B., Carvalho, R.S., Whiley, P., Spurdle, A.B., Vreeswijk, M.P.G., Caputo, S.M., Millot, G.A., Vega, A., Coquelle, N., Galli, A., Eccles, D., Blok, M.J., Pal, T., Luijt, R.B., Pena, M. S., Neuhausen, S.L., Donenberg, T., Machackova, E., Thomas, S., Vellee, M., Couch, F.J., Tavtigian, S.V., Glover, J.N.M., Carvalho, M.A., Brody, L.C., Sharan, S.K., Monteiro, A.N. (2015). BRCA1 Circos: a visualisation resource for functional analysis of missense variants. *Journal of Medical Genetics*, *52*, 224-230.

Kang, L., Xie, D., Xiao, Q., Peng, C., Yu, Y., He, X. (2019). Sequencing and analyses on chloroplast genomes of *Tetrataenium candicans* and two allies give new insights on structural variants, DNA barcoding and phylogeny in *Apiaceae* subfamily *Apioideae*. *PeerJ, 7*, e8063, doi.org/10.7717/peerj.8063.

Kardos, M., Armstrong, E.E., Fitzpatrick, S.W., Funk, C.W. (2021). The crucial role of genome-wide genetic variation in conservation. *PNAS*, *118* (48), doi.org/10.1073/pnas.2104642118.

Kaźmierczakowa, R., Bloch-Orłowska, J., Celka, Z., Cwener, A., Dajdok, Z., Michalska-Hejduk, D., Pawlikowski, P., Szczęśniak, E., Ziarnek, K. (2016). *Polska czerwona lista paprotników i roślin kwiatowych. Polish red list of pteridophytes and flowering plants*. Kraków: Instytut Ochrony Przyrody Polskiej Akademii Nauk.

Kershaw, F., Bruford, M.W., Funk, W.C., Grueber, C.E., Hoban, S., Hunter, M.E., Laikre, L., MacDonald, A.J., Meek, M.H., Mittan, C., O'Brien, D., Ogden, R., Shaw, R.E., Vernesi, C., Segelbacher, G. (2022). The Coalition for Conservation Genetics: Working across organizations to build capacity and achieve change in policy and practice. *Conservation Science and Practice A journal of the Society for Conservation Biology*, e12635 doi.org/10.1111/csp2.12635.

Kim, K.J., Jansen, R.K. (1995). ndhF sequence evolution and the major clades in the sunflower family. *Proceedings of the National Academy of Sciences*, *92* (22), 10379-10383.

Kim, S.C., Baek, S.H., Lee, J.W., Hyun, H.J. (2019). Complete chloroplast genome of *Vaccinium oldhamii* and phylogenetic analysis. *Mitochondrial DNA Part B, 4* (1), 902-903.

Kim, Y., Shin, J., Oh, D.R., Kim, D.W., Lee, H.S., Choi, C. (2020). Complete chloroplast genome sequences of *Vaccinium bacteatum* Thunb., *V. vitis-idaea* L., and *V. uliginosum* L. (Ericaceae). *Mitochondrial DNA Part B, 5* (2), 1843-1844.

Kim, Y., Shin, J., Oh, D.R., Kim, D.W., Choi, C. (2020). Comperative Analysis of Complete Chloroplast Genome Sequences and Insertion-Deletion (Indel) Polymorphism to Distinguish Five *Vaccinium* Species. *Forests, 11* (927), 1-13.

Khela, S. (2013). *Erica cinerea*. The IUCN Red List of Threatened Species 2013: e.T203006A2758531. doi.: 10.2305/IUCN.UK.2013-2.RLTS.T203006A2758531.en. Dostęp 13 grudnia 2022.

Kobayashi, N., Naktsuka, A., Ohta, H., Kurashige, Y., Handa, T., Scariot, V., Caser, M., Demasi, S., Riek, J.D., Keyser, E.D., Huylenbroeck, J. (2021). Contribution of the *Rhododendron ripense* Makino Chloroplast Genome to the Development of Evergreen Azalea Cultivars. *The Horticular Journal, 90* (2), 223-231.

Köroğlu, A., Hürkul, M.M., Kendir, G., Küçükboyaci, N. (2019). In vitro antioxidant capacities and phenolic contents of four *Erica* L. (*Ericaceae*) taxa native to Turkey. *Journal of Research in Pharmacy, 23* (1), 93-100.

Krawczyk, K., Nobis, M., Nowak, A., Szczecińska, M., Sawicki, J. (2017). Phylogenetic implications of nuclear rRNA IGS variation in Stipa L. (Poaceae). *Scientific Reports, 7*, 1-11.

Kron, K.A., Judd, W.S., Stevens, P.F., Crayn, D.M., Anderberg, A.A., Gadek, P.A., Quinn, C.J., Luteyn, J.L. (2002). Phylogenetic Classification of *Ericaceae*: Molecular and Morphological Evidence. *The Botanical Review*, vol.*68* (3), 335-423.

Kron, K.A., Luteyn, J.L. (2005). Origin and biogeographic patterns in *Ericaceae*: New insight from recent phylogenetic analyses. W: (red.) Friis, I., Balslev, H. *Plant Diversity and Complexity Pattern–Local, Regional and Global Dimensions*. Biology Skrifter 55. Copenhagen: The Roy. Academic Science Letter, 479–500.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., Marra, M.A.. (2009). Circos: an information aesthetic for comparative genomics. *Genome Research*, *19*, 1639-1645.

Ku, C., Hu, J.M., Kuo, C.H. (2013). Complete plastid genome sequence of the basal *Asterid Ardisia polysticta* Miq. and comparative analyses of asterid plastid genomes. *PLoS ONE*, *8*, e62548.

Kumaran, N., Padmalal, D. (2022). *Holocene Climate Change and Environment*. Elsevier, 629-650.

Kurtz, S., Choudhuri, J.V., Ohlebusch, E., Schleiermacher, C., Stoye, J., Giegerich, R. (2001). REPuter: The Manifold Applications of Repeat Analysis on a Genomic Scale. *Nucleic Acids Research*, *29* (22), 4633-4642.

Lan, T., Albert, V.A. (2011). Dynamic distribution patterns of ribosomal DNA and chromosomal evolution in *Paphiopedilum*, a lady's slipper orchid. *BMC Plant Biology*, *11*, 126.

Lanfranco, L., Fiorilli, V. (2022). Life of Microbes Inside the Plant: Beneficial Fungal Endophytes and Mycorrhizal Fungi w: de Bruijn, F.J., Smidt, H., Cocolin, L.S., Sauer, M., Dowling, D., Thomashow, L. (red): *Good Microbes in Medicine, Food Production, Biotechnology, Bioremediation, and Agriculture*. Wiley Online Library.

Lawrie, D.S., Messer, P.W., Hershberg, R., Petrov, D.A. (2013). Strong Purifying Selection at Synonymous Sites in *D. melanogaster*. *PLOS Genetics*, *9* (5), e1003527.

Le Maitre, N.C. (2017). *Phylogenetics of the genus Erica and anthocyanin synthesis gene expression in Erica plukenetii*. PhD Thesis. Stellenbosch University.

Le Maitre, N.C., Pirie, M.D., Bellstedt, D.U. (2019). Floral Color, Anthocyanin Synthesis Gene Expression and Control in Cape Erica Species. *Frontiers in Plant Science*, *10*, 1565, 1-9.

Lee, J., Hymowitz, T. (2001). A molecular phylogenetic study of the subtribe *Glycininae* (Leguminosae) derived from the chloroplast DNA rps16 intron sequences. *American Journal of Botany*, *88*, 2064– 2073.

Lee, S.B., Kaittanis, C., Jansen, R.K., Hostetler, J.B., Tallon, L.J., Town, C.D., Daniell, H. (2006). The complete chloroplast genome sequence of *Gossypium hirsutum* : organization and phylogenetic relationships to other angiosperms. *BMC Genomics*, *7*, 61-74.

Lee, Y., Park, H.S., Jeon, J.H., Park, J.Y., Kim, S.H., Huh, J., Woo, S., Jeong, D.W., Yang, D.W. (2022). Complete plastid and 45S rDNA sequences allow authentication of *Liriope platyphylla* and *Ophiopogon japonicus*. *Current Plant Biology*, *30*, 100244.

Leopold, D.R., Peay, K.G., Vitousek, P.M., Fukami, T. (2021). Diversity of putative ericoid mycorrhizal fungi increases with soil age and progressive phosphorus limitation across a 4.1-million-year chronosequence. *FEMS Microbiology Ecology*, *97* (3), fiab016.

Li, X., Yang, Y., Henry, R.J., Rossetto, M., Wang, Y., Chen, S. (2015). Plant DNA barcoding: from gene to genome. *Biology Reviews of the Cambridge Philosophical Society*, *90* (1), 157-166.

Li, H., Guo, Q., Li, Q., Yang, L. (2020). Long-reads reveal that *Rhododendron delavayi* plastid genome contains extensive repeat sequences, and recombination exist among plastid genomes of photosynthetic *Ericaceae*. *Peer J*, *8*, e9048.

Li, F., Xie, X., Huang, R., Tian, E., Li, C., Chao, Z. (2021). Chloroplast genome sequencing based on genome skimming for identification of Eriobotryae Folium. *BMC Biotechnology*, *21*, 69.

Liang, H., Zhang, Y., Deng, J., Gao, G., Ding, C., Zhang, L., Yang, R. (2020). The Complete Chloroplast Genome Sequences of 14 *Curcuma* Species: Insights Into Genome Evolution and Phylogenetic Relationships Within Zingiberales. *Frontiers in Genetics*, *11*, 3389-3397.

Lin, S., Liu, J., He, X., Wang, J., Wang, Z., Zhang, X., Bao, M., Fu, X. (2022). Comprehensive Comparative Analysis and Development of Molecular Markers for *Dianthus* Species Based on Complete Chloroplast Genome Sequences. *International Journal of Molecular Science*, *23*, 12567.

Linder, H.P. (2003). The radiation of the Cape flora, southern Africa. *Biological Review*, *78*, 597-638.

Liu, C., Shi, L.C., Zhu, Y.J., Chen, H.M., Zhang, J.H., Lin, X.H., Guan, X.J. (2012). Cp GAVAS, an integrated web server for the annotation, visualization, analysis, and Gen Bank submission of completely sequenced chloroplast genome sequences. *BMC Genomics*, *13* (1), 715.

Liu, Y., Zhang, L., Liu, Z., Luo, K., Chen, S., Chen, K. (2012). Species identification of *Rhododendron* (Ericaceae) using the chloroplast deoxyribonucleic acid *PsbA-trnH* genetic marker. *Pharmacognozy magazine*, *8* (29), 29-36.

Liu J., Chen T., Zhang Y., Li Y., Gong J., Yi Y. 2020. The complete chloroplast genome of *Rhododendron delavayi* (Ericaceae). Mitochondrial DNA Part B 5(1): 37-38.

Liu, Q., Li, X., Li, M., Xu, W., Schwarzacher, T., Heslop-Harrison, J.S. (2020). Comparative chloroplast genome analyses of *Avena*: insights into evolutionary dynamics and phylogeny. *BMC Blant Biology*, *20*, 406.

Liu, W., Zhang, B. (2022). The landscape of genome sequencing and assembling in plants. *Functional & Integrative Genomics*, *22*, 1147-1152.

Llorens, L., Penuelas, J., Estiarte, M. (2003). Ecophysiological responses of two Mediterranean shrubs, *Erica multiflora* and *Globularia alypum*, to experimentally drier and warmer conditions. *Physiologia Plantarum*, *119* (2), 231-243.

Logacheva, M.D., Schelkunov, M.I., Shtranikova, V.Y., Matveeva, M.V., Penin, A.A. (2016). Comparative analysis of plastid genomes of non-photosynthetic *Ericaceae* and their photosynthetic relatives. *Scientific Reports, 6,* 300-312.

Long, L., Li, Y., Wang, S., Liu, Z., Wang, J., Yang, M. (2023). Complete chloroplast genomes and comparative analysis of *Ligustrum* species. *Scientific Reports, 13,* 212.

Lovejoy, T.E., Hanah, L. (2019). *Biodiversity and Climate Change*. New Haven: Yale University Press.

Lopez-Giraldez, F., Townsend, J.P. (2011). PhyDesign: an online application for profiling phylogenetic informativeness. *BMC Evolutionary Biology, 11,* 152.

Ma, L.H., Zhu, H.X., Wang, C.Y., Li, M.Y., Wang, H.Y. (2021). The complete chloroplast genome of *Rhododendron platypodium* (Ericaceae): an endemic and endangered species from China. *Mitochondrial DNA Part B, 6* (1), 196-197.

Maciszewski, K., Dabbagh, N., Preisfeld, A., Karnkowska, A. (2022). Maturyoshka: A maturase inside a maturase, and other peculiarities of the novel chloroplast genomes of marine euglenophytes. *Molecular Phylogenetics and Evolution, 170,* 107441

Mahadani, P., Dasgupta, M., Vijayan, J., Kar, C.S., Ray, S. (2022). DNA Barcoding in Plants: Past, Present, and Future, w: Singh, R.L., Mondal, S., Parihar, A., Singh, P.K. (red.) *Plant Genomics for Sustainable Agriculture*. Singapore: Springer.

Martinez-Alberola, F., del Campo, E.M., Lázaro-Gimeno, D., Mezquita-Claramonte, S., Molins, A., Mateu-Andrés, I., Pedrola-Monfort, J., Casano, L.M., Barreno, E. (2013). Balanced Gene Losses, Duplications and Intensive Rearrangements Led to an Unusual Regularly Sized Genome in *Arbutus unedo* Chloroplasts. *PLoS ONE, 8* (11), 79-85.

Mayor, C., Brudno, M., Schwartz, J.R., Poliakov, A., Rubin, E.M., Frazer, K.A., Pachter, L.S., Dubchak, I. (2000). VISTA : visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics, 16* (11), 1046-1047.

Mayrose, I., Graur, D., Ben-Tal, N., Pupko, T. (2004) Comparison of Site-Specific Rate-Inference Methods for Protein Sequences: Empirical Bayesian Methods Are Superior. *Molecular Biology and Evoloution, 21*, 1781-1791.

McCarren, S., Coetzee, A., Midgley, J. (2021). Corolla stickiness prevents nectar robbing in *Erica*. *Journal of Plant Research, 134* (5), 963-970.

McClintock, B. (1934). The relation of a particular chromosomal element to the development of the nucleoli in *Zea mays*. *Zeitschrift fur Zellforschung und mikroskopische Anatomie, 21* (2), 294–328.

McGuire, A.F., Kron, K.A. (2005). Phylogenetic relationships of European and African *Ericas*. *International Journal of Plant Science, 166*, 311-318.

Miao, Y., Chen, H., Xu, W., Yang, Q., Liu, C., Huang, L. (2022). Structural mutations of small single copy (SSC) region in the plastid genomes of five *Cistanche* species and inter-species identification. *BMC Plant Biology, 22*, 412.

Mirek, Z., Piękoś-Mirkowa, H., Zając, A., Zając, M. (2002). *Flowering plants and pteridophytes of Poland. A checklist. Biodiversity of Poland. Vol. 1*. Kraków: W. Szafer Institute of Botany, Polish Academy of Sciences.

Mirek, Z., Zarzycki, K., Wojewoda, W., Szeląg, Z. (2006). *Red list of plants and fungi in Poland. Czerwona lista roślin i grzybów Polski.* Kraków: Instytut Botaniki im. W. Szafera PAN.

Mohammed, T., Firoz, A., Ramadan, A.M. (2022). RNA Editing in Chloroplast: Advancements and Opportunities. *Issues of Molecular Biology, 44* (11), 5593-5604.

Molnár, A.P., Demeter, L., Biró, M., Chytrý, M., Bartha, S., Gantuya, B., Molnár, Z. (2023). Is there a massive glacial–Holocene flora continuity in Central Europe? *Biological Reviews, 98* (6), 2307-2319.

Mower, J.P., Guo, W., Partha, R., Fan, W., Levsen, N., Wolff, K., Nugent, J.M., Pabón-Mora, N., González, F. (2021). Plastomes from tribe Plantagineae (Plantaginaceae) reveal infrageneric structural synapormorphies and localized hypermutation for *Plantago* and functional loss of *ndh* genes from *Littorella*. *Molecular Phylogenetics and Evolution*, *162*, 107217.

Mugrabi De Kuppler, A.L., Fagúndez, J., Bellstedt, D.U., Oliver, E.G.H., Léon, J., Pirie, M.D. (2015). Testing reticulate versus coalescent origins of *Erica lusitanica* using a species phylogeny of the northern heathers (*Ericeae, Ericaceae*). *Molecular Phylogenetic Evolution*, *88*, 121–131.

Nakahama, N. (2021). Museum specimens: An overlooked and valuable material for conservation genetics. *Ecological Research, 36* (1), 13-23.

Nazemiyeh, H., Bahadori, F., Delazar, A., Ay, M., Topçu, G., Nahar, L., Majinda, R.R.T., Sarker, S.D. (2008). Antioxidant phenolic compounds from the leaves of *Erica arborea* (*Ericaceae*). *Natural Product Research*, *22* (16), 1385-1392.

Nazemiyeh, H., Bahadori, F., Delazar, A., Ay, M., Topçu, G., Kolak, U., Nahar, L., Auzie, A.A., Sarker, S.D. (2008). Tricetin 4'-α-Lrhamnopyranoside: a new flavonoid from the aerial parts of *Erica arborea*. *Chemistry of Natural Compounds*, *44* (2), 174-177.

Nelson, E.C., Pirie, M.D. (2022). Where have all the heathers gone? *Sibbaldia: The International Journal of Botanic Garden Horticulture*, *21*, 75-92.

Nusrat, S., Harbig, T., Gehlenborg, N. (2019). Tasks, Techniques, and Tools for Genomic Data Visualization. *Computer Graphics Forum: journal of the European Association for Computer Graphics, 38* (3), 781-805.

Oborská-Oplová, M., Gerhardy, S., Panse, V.G. (2022). Orchestrating ribosomal RNA folding during ribosome assembly. *BioEssays Ideas that Push the Boundaries*, *44* (8), 1-15.

Ojeda, F., Budde, K.B., Heuertz, M. (2016). Biogeography and evolution of seeder and resprouter forms of *Erica coccinea* (*Ericaceae*) in the fire-prone Cape fynbos. *Plant Ecology*, *217* (6), 751-761.

Olejniczak, S.A., Łojewska, E., Kowalczyk, T., Sakowicz, T. (2016). Chloroplasts: state of research and practical applications of plastome sequencing. *Planta*, *244* (3), 1-19.

Oliver, E.G.H. (1993). Studies in the *Ericoideae* (*Ericaceae*). XI. The generic relationship between *Erica* and *Blaeria*. *Kew Bulletin*, *48*, 771–780.

Oliver, E.G.H. (1993). Studies in the *Ericoideae* (*Ericaceae*). XII. The placing of the genus *Blaeria* into synonymy under *Erica*; nomenclatural and taxonomic changes for the southern African region. *Bothalia*, *23*, 1–7.

Oliver, E.G.H., Oliver, I.M. (2003). *Ericaceae*. w: *Plants of southern Africa: an annotated checklist*. Red: Germishuizen, G., Meyer, N.L. Strelitzia: 424-451. Pretoria: National Botanical Institute.

Palmer, J.D. (1985). Comparative organization of chloroplast genomes. *Annual Reviews of Genetic*, *19*, 325-354.

Palomar, V.M., Jaksich, S., Fujii, S., Kuciński, J., Wierzbicki, A.T. (2022). High-resolution map of plastid-encoded RNA polymerase binding patterns demonstrates a major role of transcription in chloroplast gene expression. *The Plant Journal*, *111* (4), 1139-1151.

Panagopoulos, A. (2022). Study and evaluation of the characteristics of saline wastewater (brine) produced by desalination and industrial plants. *Environmental Science and Pollution Research*, *29*, 23736–23749.

Park, J., Min, J., Kim, Y., Chung, Y. (2021). The Comparative Analyses of Six Complete Chloroplast Genomes of Morphologically Diverse *Chenopodium album* L. (Amaranthaceae) Collected in Korea. *International Journal of Genomics*, 6643444.

Park, Y.S., Kang, J.S., Park, J.Y., Shim, H., Yang, H.O., Kang, J.H., Yang, T.J. (2022). Analysis of the complete plastomes and nuclear ribosomal DNAs from *Euonymus hamiltonianus* and its relatives sheds light on their diversity and evolution. *PLoS ONE*, *17* (10), e0275590.

Park, H.S., Mohanan, P., Yang, T.Y. (2023). Super-Barcoding of Medicinal Materials Based on Complete Plastid Genomes and Nuclear Ribosomal DNA w*: Authentication Of Chinese Medicinal Materials By DNA Technology: Techniques and Applications (2nd Edition)* pod red. Shaw, P., Cao, H., Lo, Y. World Scientific, 363-390.

Parvaresh, M., Talebi, M., Badraldin-Ebrahim Sayed-Tabatabaei. (2014). Molecular characterization of ribosomal DNA intergenic spacer (IGS) region in pomegranate (*Punica granatum* L.). *Plant Systematics and Evolution, 300*, 899-908.

Piredda, R., Grimm, G.W., Schulze, E.D., Denk, T., Simeone, M.C. (2020). High-throughput sequencing of 5S-IGS in oaks: Exploring intragenomic variation and algorithms to recognize target species in pure and mixed samples. *Molecular Ecology Resources*, *21* (2), 495-510.

Pirie, M.D., Oliver, E.G.H., Bellstedt, D.U. (2011). A densely sampled ITS phylogeny of the Cape flagship genus *Erica* L. suggest numerous shifts in floral macro-morphology. *Molecular Phylogenetics and Evolution*, *61*, 593-601.

Pirie, M.D., Oliver, E., De Kuppler, A.M., Gehrke, B., Le Maitre, N., Kandziora, M., Bellstedt, D. (2016). The biodiversity hotspot as evolutionary hot-bed: spectacular radiation of *Erica* in the Cape Floristic Region. *BMC Evolutionary Biology*, *16*, 190.

Pirie, M.D., Kandziora, M., Nürk, N.M., Le Maitre, N.C., Mugrabi de Kuppler, A., Gehrke, B., Oliver, E.G.H., Bellstedt, D. (2019). Leaps and bounds: geographical and ecological distance constrained the colonisation of the Afrotemperate by *Erica*. *BMC Evolutionary Biology, 19*, 222, doi.org/10.1186/s12862-019-1545-6.

Pirie, M.D., Blackhall-Miles, R., Bourke, G., Crowley, D., Ebrahim, I., Forest, F., Knaack, M., Koopman, R., Lansdowne, A., Nürk, N.M., Osborne, J., Pearce, T.R., Rohrauer, D., Smit, M., Wilman, V. (2022). Preventing species extinctions: A global conservation consortium for *Erica*. *Plants People Planet*, *4* (4), 335-344.

Posada, D. (2008). jModelTest: Phylogenetic Model Averaging. *Molecular Biology and Evolution*, *25*, 1253–1256.

Pouchon, C., Boyer, F., Roquet, C., Denoeud, F., Chave, J., Coissac, E., Alsos, I.G., Lavergne, S. (2022). ORTHOSKIM: In silico sequence capture from genomic and transcriptomic libraries for phylogenomic and barcoding applications. *Molecular Ecology Resources, 22* (5), 2018-2037.

Pond, S.L.K., Frost, S.D.W., Muse, S.V. (2005). HyPhy: hypothesis testing using phylogenies. *Bioinformatics*, *21* (5), 676–679.

Qiu, H., Zhang, Z.H., Wang, M.Z., Jin, X.J., Lin, J.D., Comes, H.P., Chen, J.X., Cui, R.N., Duan, R.Q., Li, P. (2023). Plastome evolution and phylogenomics of *Impatiens* (Balsaminaceae). *Planta*, *257*, 45.

Raimondo, D., von Staden, L., Foden, W., Victor, J.E., Helme, N.A., Turner, R.C., Kamundi, D.A., Manyama, P.A. (2009). *Red List of South African Plants*. Strelitzia 25. Pretoria: South African National Biodiversity Institute.

Rambaut, A. (2010). *FigTree v1.3.1*. Institute of Evolutionary Biology, University of Edinburgh, Edinburgh. http://tree.bio.ed.ac.uk/software/figtree/.

Raven, P. (2020) Plant Conservation: Challenges and Opportunities. w: Science and Actions for Species Protection. Noah's Arks for the 21st Century. Red. Braun von, J., Kauffels, T., Raven, P., Vogel, J., Sorondo, M.S. *Pontificiae Academiae Scientarvum Scripta Varia, 146*, 113-124.

Ray, J., Bordolui, S.K. (2021). Role of Seed Banks in the Conservation of Plant Diversity and Ecological Restoration. *Research and Reviews: Journal of Environmental Sciences*, *3* (2), 1-16.

Reich, D., Berger, A., Balthazar von, M., Chartier, M., Sherafati, M., Schönenberg, J., Manafzadeh, S., Staedler, Y.M. (2020). Modularity and evolution of flower shape: the role of function, development, and spandlers in *Erica*. *New Phytologist*, *226*, 267-280.

Rice, P., Longden, I. (2000). EMBOSS: the European Molecular Open Software Suite.*Trends in Genetics*, *16* (6), 276-277.

Ritchie, P.D.L., Clarke, J.J., Cox, P.M., Huntingford, C. (2021). Overshooting tipping point thresholds in a changing climate. *Nature*, *592*, 517-523.

Rosato, M., Kovařík, A., Garilleti, R., Rosselló, J.A. (2016). Conserved organisation of 45S rDNA sites and rDNA gene copy number among major clades of early land plants. *PLoS ONE, 11* (9), 17-28.

Ruhlman, T.A., Jansen, R.K. (2021). Plastid Genomes of Flowering Plants: Essential Principles. w: Maliga, P. (red.) *Chloroplast Biotechnology. Methods in Molecular Biology*, vol *2317*. New York: Humana.

Sabater, B. (2021). On the Edge of Dispensability, the Chloroplast *ndh* Genes. *International Journal of Molecular Science*, *22* (22), 12505.

Sadki, C., Hacht, B., Souliman, A., Atmani, F. (2010). Acute diuretic activity of aqueous *Erica multiflora* flowers and *Cynodon dactylon* rhizomes extracts in rats. *Journal of Ethnopharmacology*, *128* (2), 352-356.

Saha, A., Das, S., Moin, M., Dutta, M., Bakshi, A., Madhav, M.S., Kirti, P.B. (2017). Genome-Wide Identification and Comprehensive Expression Profiling of Ribosomal Protein Small Subunit (RPS) Genes and their Comparative Analysis with the Large Subunit (RPL) Genes in Rice. *Frontiers in Plant Science, 8*, 1553.

Sahoo, S., Rakshit, R. (2022). The pattern of coding sequences in the chloroplast genome of *Atropa belladonna* and a comparative analysis with other related genomes in the nightshade family. *Genomics & Informatics, 20* (4), e43.

Sato, N. (2021). Are Cyanobacteria an Ancestor of Chloroplasts or Just One of the Gene Donors for Plants and Algae? *Genes*, *12* (6), 823.

Sargin, S.A. (2021). Plants used against obesity in Turkish folk medicine: A review. *Journal of Ethnopharmacology*, *270*, 113841.

Sarwar Golam, A.K.M. (2011). *Pollen morphology of Ericaceae and its systematic significance*. Germany: LAP Lam Academy Press.

Saunders, C.J., Miller, N.A., Soden, S.E., Dinwiddie, D.L., Noll, A. Abu Alnadi, N., Andraws, N., Patterson, M.L., Krivohlavek, L.A., Fellis, J., Humphray, S., Saffrey, P., Kingsbury, Z., Weir, J.C., Betley, J., Grocock, R.J., Margulies, E.H., Farrow, E.G., Artman, M., Safina, N.P., Petrikin, J.E., Hall, K.P., Kingsmore, S.F. (2012). Rapid whole-genome sequencing for genetic disease diagnosis in neonatal intensive care units. *Scintific Translational Medicine*, *4* (154), 154-155.

Schlautman, B., Covarrubias-Pazaran, G., Fajardo, D., Steffan, S., Zalapa, J. (2017). Discriminating power of microsatellites in cranberry organelles for taxonomic studies in *Vaccinium* and Ericaceae. *Genetic Resources and Crop Evolution*, *64*, 451-466.

Schmitz-Linneweber, C., Regel, R., Du, T.G., Hupfer, H., Herrmann, R.G. (2002). The plastid chromosome of *Atropa belladonna* and its comparison with that of *Nicotiana tabacum*: the role of RNA editing in generating divergence in the process of plant speciation. *Molecular Biology & Evolution*, *19*, 1602-1612.

Schumann, D., Kirsten, G., Oliver, E.G.H. (1992). *Ericas of South Africa*. Cape Town: Fernwood Press.

Shao, Z.Q., Xue, J.Y., Wang, Q., Wang, B., Chen, J.Q. (2019). Revisiting the Origin of Plant NBS-LRR Genes. *Trends in Plant Science*, *24* (1), 9-12.

Shaw, J., Lickey, E.B., Schilling, E., Small, R.L. (2007). Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: The tortoise and the hare III. *American Journal of Botany*, *94*, 275–288.

Shaver, J.M., Oldenburg, D.J., Bendich, A.J. (2008). The Structure of Chloroplast DNA Molecules and the Effects of Light on the Amount of Chloroplast DNA during Development in *Medicago truncatula*. *Plant Physiology*, *146* (3), 1064-1074.

Shen, J., Li, X., Zhu, X., Huang, X., Jin, S. (2020). The Complete Plastid Genome of *Rhododendron pulchrum* and Comparative Genetic Analysis of *Ericaceae* Species. *Forests*, *11* (158), 1-14.

Shi, X., Xu, W., Wan, M., Sun, Q., Chen, Q., Zhao, C., Sun, K., Shu, K. (2022). Comparative analysis of chloroplast genomes of three medicinal *Carpesium* species: Genome structures and phylogenetic relationships. *PLoS ONE*, *10*, doi.org/10.1371/journal.pone.0272563

Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi-Shinozaki, K., Ohto, C., Torazawa, K., Meng, B.Y., Sugita, M., Deno, H., Kamogashira, T., Yamada, K., Kusuda, J., Takaiwa, F., Kato, A., Tohdoh, N., Shimada, H., Sugiura, M. (1986). The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *The EMBO Journal*, *5* (9), 2043-2049.

Singh, M., Yadaf, P., Yadaf, A.K. (2021). Cloning and Sequence Analysis of 5S Ribosomal RNA Gene(s) and Associated Intergenic Spacer Regions in *Carica* Species. *International Journal of Biology*, *13* (1), 1-13.

Singhal, R., Pal, R., Dutta, S. (2022). Chloroplast Engineering: Fundamental Insights and Its Application in Amelioration of Environmental Stress. *Applied Biochemistry and Biotechnology*, doi.org/10.1007/s12010-022-03930-8.

Şirin, E., Ertürk, Y., Kazankaya, A. (2022). Effects of PGPR, AMF and Trichoderma Applications on Adaptation Abilities to Different Biotic and Abiotic Conditions in Medicinal and Aromatic Plants. *Turkish Journal of Agriculture - Food Science and Technology*, *10* (2), 166-173.

Slatko, B.E., Gardner, A.F., Ausubel, F.M. (2018). Overview of Next-Generation Sequencing Technologies. *Current Protocols in Molecular Biology*, doi.org/10.1002/cpmb.59.

Smith, D.R., Lee, R.W., Cushman, J.C., Magnuson, J.K., Tran, D., Polle, J.E. (2010). The *Dunaliella salina* organelle genomes: large sequences, inflated with intronic and intergenic DNA. *BMC Plant Biology*, *10*, 83-85.

Soltis, P.S., Soltis, D.E. (2021). Plant genomes: Markers of evolutionary history and drivers of evolutionary change. *Plants People Planet, 3* (1), 74-82.

Song, B., Buckler, E.S., Wang, H., Wu, Y., Rees, E., Kellogg, E.A., Gates, D.J., Khaipho-Burch, M., Bradbury, P.J., Ross-Ibarra, J., Hufford, M.B., Romay, M.C. (2021). Conserved noncoding sequences provide insights into regulatory sequence and loss of gene expression in maize. *Genome Research, 31*, 1245-1257.

Song, W., Ji, C., Chen, Z., Cai, H., Wu, X., Shi, C., Wang, S. (2022). Comparative Analysis the Complete Chloroplast Genomes of Nine *Musa* Species: Genomic Features, Comparative Analysis, and Phylogenetic Implications. *Frontiers in Plant Sciences, 13*, doi.org/10.3389/fpls.2022.832884.

Stamatakis, A. (2014). RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics, 30* (9), 1312-1313.

Steane, D.A. (2005). Complete nucleotide sequence of the chloroplast genome from the Tasmanian blue gum, *Eucalyptus globulus* (Myrtaceae). *DNA Research, 12*, 215-220.

Stevens, P.F. (1970). *Calluna*, *Cassiope* and *Harrimanella*: a taxonomic and evolutionary problem. *New Phytology, 69*, 1133-1150.

Stevens, P.F. (2008). A classification of the *Ericaceae*: subfamilies and tribes. *Botanical Journal of Linnean Society, 64*, 1-53.

Stevens, P.F., Luteyn, J., Oliver, E.G.H., Bell, T.L., Brown, E.A., Crowden, R.K., George, A.S., Jordan, G.J., Ladd, P., Lemson, K., McClean, C.B., Menadue, Y., Pate, J.S., Stace, H.M., Weiller, C.M. (2004). *Ericaceae*. w: Kubitzki, K. (red.) *The Families and Genera of Vascular Plants.* Berlin-Heidelberg- New York: Springer-Verlad, 145-194.

Stribley, D.P., Read, D.J. (2006). The biology of mycorrhiza in the *Ericaceae*. *New Phytologist, 73* (4), 731-741.

Sultanov, D., Hochwagen, A. (2022). Varying strength of selection contributes to the intragenomic diversity of rRNA genes. *Nature Communications*, *13*, 7245.

Suna, S., Özcan-Sinir, G., Tamer, C.E., İncedayi, B., Çopur, Ö.U. (2018). Antioxidant Capacity and Physicochemical Characteristics of Carbonated *Erica Arborea* Tea Beverage. *Beverages*, *4* (3), 50.

Szczecińska, M., Gomolińska, A., Szkudlarz, P., Sawicki, J. (2014). Plastid and nuclear genomic resources of a relict and endangered plant species: *Chamaedaphne calyculata* (L.) Moench (*Ericaceae*). *Turkish Journal of Botany*, *38*, 1229-1238.

Tampakaki, P.A., Fotiadis, T.C., Ntatsi, G., Savvas, D. (2017). A novel sumbiovar (aegeanense) of the genus *Ensife*r nodulates *Vigna unguiculata*. *Journal of Scintific Food Agriculture*, *97*, 4314-4325.

Tamura, K., Stecher, G., Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis version 11. *Molecular Biology and Evolution*, *38*, 3022-3027.

Thaenkham, U., Chaisiri, K., Hui En Chan, A. (2022). DNA Sequence Alignment and Phylogenetic Inferences: Guidelines for Analysis and the Selection of Appropriate Methods for Molecular Systematics, w: *Molecular Systematics of Parasitic Helminths*. Singapur: Springer, 201-223.

Thode, V.A., Oliveira, C.T., Loeuille, B., Siniscalchi, C.M., Pirani, J.R. (2021). Comparative analyses of *Mikania* (Asteraceae: Eupatorieae) plastomes and impact of data partitioning and inference methods on phylogenetic relationships. *Scientific Reports*, *11*, 13267.

Thompson, J.D., Higgins, D.G., Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, *22* (22), 4673-4680.

Tlas, L.T., Chamaa, I., Fandi, J., Haddad, E. (2021). In-vitro antibacterial activity and chemical composition of Syrian *Erica manupuliflora* essential oil. *Bulletin of Pharmaceutical Sciences Assiut University*, *44* (1), 63-71.

Tong, Z., Zhou, J., Xiu, Z., Jiao, F., Hu, Y., Zheng, F., Chen, X., Li, Y., Fang, D., Li, S., Wu, X., Zeng, J., Zhao, S., Jian, J., Xiao, B. (2020). Construction of a high-density genetic map with whole genome sequencing in *Nicotiana tabacum* L. *Genomics*, *112* (2), 2028-2033.

Torres, A., Goloboff, P.A., Catalano, S.A. (2021). Assessing topological congruence among concatenation-based phylogenomic approaches in empirical datasets. *Molecular Phylogenetics and Evolution*, *161*, 935-942.

Townsend, J.P., Su, Z., Tekle, Y.I. (2012). Phylogenetic signal and noise: predicting the power of a data set to resolve phylogeny. *Systematic Biology*, *61*, 835-849.

Turudić, A., Liber, Z., Grdiša, M., Jakše, J., Varga, F., Šatović, Z. (2022). Chloroplast Genome Annotation Tools: Prolegomena to the Identification of Inverted Repeats. *International Journal of Molecular Science*, *23* (18), 10804.

Tuzlacı, E., Eryaşar Aymaz, P. (2001). Turkish natural medicinal plants, Part IV: Gönen (Balıkesir). *Fitoterapia*, *72* (4), 323-343.

Qin, Q., Li, J., Zeng, S., Xu, Y., Han, F., Yu, J. (2022). The complete plastomes of red fleshed pitaya (*Selenicereus monacanthus*) and three related *Selenicereus* species: insights into gene losses, inverted repeat expansions and phylogenomic implications. *Physiology and Molecular Biology of Plants*, *28*, 123–137.

Valgaz, A.P.F., Barcos-Arias, M., Naranjo-Morán, J., Tapia, D.P., Moreira-Gómez, R. (2022). Ericaceous Plants: A Review for the Bioprospecting of Ericoid Mycorrhizae from Ecuador. *Diversity*, *14* (8), 648.

Van der Niet, T., Pirie, M.D., Shuttleworth, A., Johnson, S.D., Midgley, J.J. (2014). Do pollinator distributions underlie the evolution of pollination ecotypes in the Cape shrub *Erica plukenetii*? *Annals of Botany*, *113* (2), 301-316.

Van der Niet, T., Pires, K., Steenhuisen, S-L. (2020). Flower constancy of the Cape honey bee pollinator of two co-flowering Erica species from the Cape Floristic Region (South Africa). *South African Journal of Botany*, *132*, 371-377.

Vera, M.L., Alonso, M.C., Obeso, J.R. (1995). Rebrote después de incendio e inversión reproductiva en *Erica cinerea* y *E. vagans*. *Orsis*, *10*, 7-15.

Verbruggen, H., Maggs, C.A., Saunders, G.W., Gall, L.L., Yoon, H.S., Clerck, O. (2010). Data mining approach identifies research priorities and data requirements for resolving the red algal tree of life. *BMC Evolutionary Biology*, *10*, 10-16.

Vohník, M., Réblová, M. (2023). Fungi in hair roots of *Vaccinium* spp. (Ericaceae) growing on decomposing wood: colonization patterns, identity and *in vitro* symbiotic potential. Mycorrhiza. *Mycorrhiza*, *33*, 69-86.

Vozárová, R., Herklotz, V., Kovařik, A., Tynkevich, Y.O., Volkov, R.A., Ritz, C.M., Lunerová, J. (2021). Ancient Origin of Two 5S rDNA Families Dominating in the Genus *Rosa* and Their Behavior in the Canina-Type Meiosis. *Frontiers in Plant Science*, *12*, 643548

Wang, B., Li, M., Yuan, Y. (2019). The complete chloroplast genome of *Pericallis hybrida* (Asteridae). *Mitochondrial DNA Part B, 4* (1), 233-234.

Wang, M., Wang, X., Sun, J, Wang, Y., Ge, Y., Dong, W., Yuan, O., Huang, L. (2021). Phylogenomic and evolutionary dynamics of inverted repeats across *Angelica* plastomes. *BMC Plant Biology*, *21*, 26-29.

Wang, J., Qian, J., Jiang, Y., Chen, X., Zheng, B., Chen, S., Yang, F., Xu, Z., Duan, B. (2022). Comparative Analysis of Chloroplast Genome and New Insights Into Phylogenetic Relationships of *Polygonatum* and Tribe Polygonateae. *Frontiers in Plant Science*, *13*, 992189-94.

Wang, F., Zhao, W., Dong, Z., Ma, L., Li, W., Li, Z., Xin, P. (2023). Analysis of the Chloroplast Genome Characteristics of 6 Species of *Yucca*. *Bulletin of Botanical Research*, *43* (1), 109-119.

Weng, M.L., Ruhlman, T.A., Jansen, R.K. (2017). Expansion of inverted repeat does not decrease substitution rates in *Pelargonium* plastid genomes. *New Phytologist*, *214* (2), 842-851.

Westwood, M., Cavender, N., Meyer, A., Smith, P. (2021). Botanic garden solutions to the plant extinction crisis. *Plants People Planet, 3* (1), 22-32.

Wicke, S., Schneeweiss, G., de Pamphilis, C., Müller, K.F., Quandt, D. (2011). The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Molecular Biology, 76*, 273-297.

Wicke, S., Costa, A., Munoz, J., Quand, D. (2011). Restless 5 S: the re-arrangement(s) and evolution of the nuclear ribosomal DNA in land plants. *Molecular Phylogenetic Evolution, 61* (2), 321-332.

Willi, Y., Kristensen, T.N., Sgrò, C.M., Hoffmann, A.A. (2021). Conservation genetics as a management tool: The five best-supported paradigms to assist the management of threatened species. *PNAS, 119* (1), e2105076119.

Wrońska-Pilarek, D., Szkudlarz, P., Bocianowski, J. (2018). Systematic importance of morphological features of pollen grains of species from *Erica* (*Ericaceae*) genus. *PLoS ONE, 13* (10), e0204557.

Wu, W., Li, J., Liu, Y., Jiang, M., Lan, M., Liu, C. (2021). Peculiarities of the inverted repeats in the complete chloroplast genome of *Strobilanthes bantonensis* Lindau. *Mitochondrial DNA Part B, 6*: 4, 1440-1447.

Wu, L., Wu, M., Cui, N., Xiang, L., Li, Y., Li, X., Chen, S. (2021). Plant super-barcode: a case study on genome-based identification for closely related species of *Fritillaria*. *Chinese Medicine, 16* (52), 2-11.

Wyman, S.K., Jansen, R.K., Boore, J.L. (2004). Automatic annotation of organellar genomes with DOGMA. DOGMA citation. *Bioinformatics, 20* (17), 3252-3255.

Xiong, Q., Hu, Y., Lv, W., Wang, Q., Liu, G., Hu, Z. (2021). Chloroplast genomes of five *Oedogonium* species: genome structure, phylogenetic analysis and adaptive evolution. *BMC Genomics, 22* (707), doi.org/10.1186/s12864-021-08006-1

Xu, X., Rudin, C., Seltzer, M. (2019). Optimal sparse decision trees. *Advances in Neural Information Processing Systems*, *32*, 2-9.

Xu, F., He, L., Gao, S., Su, Y., Li, F., Xu, L. (2019). Comparative Analysis of two Sugarcane Ancestors *Saccharum officinarum* and *S. spontaneum* based on Complete Chloroplast Genome Sequences and Photosynthetic Ability in Cold Stress. *International Journal of Molecular Science*, 20 (15), 3828.

Xu, Y.L., Shen, H.H., Du, X.Y., Lu, L. (2022). Plastome characteristics and species identification of Chinese medicinal wintergreens (*Gaultheria*, Ericaceae). *Plant diversity*, *44* (6), 519-529.

Yaici, K., Dahamna, S., Moualek, I., Belhadj, H., Houali, K. (2021). Évaluation de la teneur des composés phénoliques, des propriétés antioxydantes et antimicrobiennes de l'espèce *Erica arborea* L. (Ericaceae) dans la médecine traditionnelle du Tell sétifien dans l'Est Algérien. *Phytothérapie*, *19* (4), 226-234.

Yang, Z., Zhao, T., Ma, Q., Liang, L., Wang, G. (2018). Comparative Genomics and Phylogenetic Analysis Revealed the Chloroplast Genome Variation and Interspecific Relationships of *Corylus* (Betulaceae) Species. *Frontiers in Plant Science*, *9*, 920-927.

Yasodha, R., Vasudeva, R., Balakrishnan, S., Sakthi, A.R., Abel, N., Binai, N., Rajashekar, B., Bachpai, V.K.W., Pillai, C., Dev, S.A. (2018). Draft genome of a high value tropical timber tree, Teak (*Tectona grandis* L. f): insights into SSR diversity, phylogeny and conservation. *DNA Research*, *25* (4), 409–419.

Yin, K., Zhang, Y., Li, Y., Du, F.K. (2018). Different Natural Selection Pressures on the atpF Gene in Evergreen Sclerophyllous and Deciduous Oak Species: Evidence from Comparative Analysis of the Complete Chloroplast Genome of *Quercus aquifolioides* with Other Oak Species. *International Journal of Molecular Science*, *19* (4), 1042.

Yoichi, W., Tomaru, N. (2014). Patterns of geographic distribution have a considerable influence on population genetic structure in one common and two rare species of *Rhododendron* (*Ericaceae*). *Tree Genetics & Genomes*, *10*, 827-837.

Yu, J., Wu, X., Liu, C., Newmaster, S., Ragupathy, S., Kress, W.J. (2021). Progress in the use of DNA barcodes in the identification and classification of medicinal plants. *Ecotoxicology and Environmental Safety*, Vol. *208,* 911-927.

Yu, X.Q., Jiang, Y.Z., Folk, R.A., Zhao, J.L., Fu, C.N., Fang, L., Peng, H., Yang, Y.B., Yang, S.H. (2022). Species discrimination in *Schima* (Theaceae): Next-generation super-barcodes meet evolutionary complexity. *Molecular Ecology Resources, 22* (8), 3161-3175.

Yukawa, M., Tsudzuki, T., Sugiura, M. (2006). The chloroplast genome of *Nicotiana sylvestris* and *Nicotiana tomentosiformis*: complete sequencing confirms that the *Nicotiana sylvestris* progenitor is the maternal genome donor of *Nicotiana tabacum*. *Molecular Genetetics & Genomics, 275,* 367-373.

Zaveri, S.H., Dasgupta, S. (2022). Transplastomic Plants: Development, Innovation and Application for Way Towards Eco Friendly Agruculture, w: Satyannarayana, B., Barwant, M. (red): *Recent Trends of Innovations in Chemical and Biological Scence Volume* I, 148-154.

Zhang, H., Gou, X., Li, S., Wang, C., Peng, C., Hu, C., Zhang, M., Yu, L., Su, H. (2021). Complete mitochondrial genome of Black Drongo *Dicrurus macrocercus* (Passeriformes: Dicruridae) and phylogenetic analyses. *Mitochondrial DNA Part B Resources*, *6* (8), 2442-2444.

Zhang, W., Sun, Y., Liu, J., Xu, Ch., Zou, X., Chen, X., Liu, Y., Wu, P., Yang, W., Zhou, S. (2021). DNA barcoding of *Oryza*: conventional, specific, and super barcodes. *Plant Molecular Biology, 105,* 215-228.

Zhang, F.X., Landis, J.B., Wang, H.X., Zhu, Z.X., Wang, H.F. (2021). Comparative analysis of chloroplast genome structure and molecular dating in Myrtales. *BMC Plant Biology, 21,* 219.

Zhang, X., Sun, Y., Landis, J.B., Lv, Z., Shen, J., Lin, N., Sun, J., Deng, T., Sun, H., Wang, H. (2020). Plastome phylogenomic study of Gentianeae (*Gentianaceae*): widespread gene tree discordance and its association with evolutionary rate heterogeneity of plastid genes. *BMC Plant B*iology, *20*, 340.

Zheng, C., Fan, J., Caraballo-Ortiz, M.A., Liu, Y., Liu, T., Fu, G., Zhang, Y., Yang, P., Su, X. (2022). The complete chloroplast genome and phylogenetic relationship of *Apocynum pictum* (Apocynaceae), a Central Asian shrub and second-class national protected species of western China. *Gene*, *830*, 146517.

Zhou, J., Cui, Y., Chen, X., Li, Y., Xu, Z., Duan, B., Li, Y., Song, J., Yao, H. (2018). Complete chloroplast genomes of *Papaver rhoeas* and *Papaver orientale*: molecular structures, comparative analysis, and phylogenetic analysis. *Molecules*, *23*, 437-442.

Zhou, Q., Chen, Y., Wang, J., Li, M., Zeng, W., Wang, Y., Li, Y., Zhao, H. (2022). A comparative study of the chloroplast genomes of five *Lepidium* species with high medicinal value. *Functional Plant Biology*, *50* (1), 29-45.

Zuo, R., Jiang, P., Sun, C., Chen, C., Lou, X. (2020). Analysis of the chloroplast genome characteristics of *Rhus chinensis* by de novo sequencing. *Chinese Journal of Biotechnology*, *36* (4), 772-781.

Zydlik, Z., Zydlik, P., Kleiber, T. (2019). The effect of the mycorrhization on the content of macroelements in the soil and leaves of blueberry cultivated after replantation. *Zemdirbyste-Agriculture*, *106* (4), 345–350.

Spis rycin

Rycina 1. Rozmieszczenie geograficzne badanych populacji……..……54

Rycina 2. Schemat przedstawiający poszczególne etapy prowadzonych prac…..…….55

Rycina 3. Schemat składania genomu *Erica versicolor* in silico i jego adnotacja…….74

Rycina 4. Mapa genomu chloroplastowego *Erica versicolor*. Geny umieszczone po wewnętrznej stronie okręgu są transkrybowane w kierunku przeciwnym wobec ruchu wskazówek zegara, natomiast kierunek transkrybowania genów ulokowanych po zewnętrznej stronie okręgu zachodzi zgodnie z ruchem wskazówek zegara. Strzałki wskazują kierunek transkrypcji. Wewnętrzny okrąg wskazuje duży region pojedynczych kopii (LSC), regiony odwróconych powtórzeń (IRA, IRB) i mały region pojedynczych kopii (SSC). Zwizualizowano na nim zawartość G/C (odcień ciemnoszary) i A/T (odcień jasnoszary). Geny są oznaczone kolorami na podstawie ich funkcji, przynależności do odpowiednich rodzin przytoczonych w legendzie poniżej genomu……87

Rycina 5. Mapa genomu chloroplastowego *Erica arborea*. Geny umieszczone po wewnętrznej stronie okręgu są transkrybowane w kierunku przeciwnym wobec ruchu wskazówek zegara, natomiast kierunek transkrybowania genów ulokowanych po zewnętrznej stronie okręgu zachodzi zgodnie z ruchem wskazówek zegara. Strzałki wskazują kierunek transkrypcji. Wewnętrzny okrąg wskazuje duży region pojedynczych kopii (LSC), regiony odwróconych powtórzeń (IRA, IRB) i mały region pojedynczych kopii (SSC).. Zwizualizowano na nim zawartość G/C (odcień ciemnoszary) i A/T (odcień jasnoszary). Geny są oznaczone kolorami na podstawie ich funkcji, przynależności do odpowiednich rodzin przytoczonych w legendzie poniżej genomu……92

Rycina 6. Wykres Circos oparty na całkowitym dopasowaniu genomów chloroplastowych między dwoma gatunkami rodzaju *Erica* przedstawionymi w pracy. Mapa połączeń dotyczy wizualizacji zdetektowanych powtórzeń tandemowych w genomach chloroplastowych, których wartość dopasowania wyniosła minimum 90%.....…..…….99

Rycina 7. Zestawienie liczbowe sekwencji powtórzeniowych mikrosatelitarnych w genomach chloroplastowych rodzaju *Erica*……100

Rycina 8. Drzewo zbudowane na regionach IR gatunków z rodziny *Ericaceae*. Grupę zewnętrzną stanowił gatunek *Actinidia arguta*. Wartości bootstrap poniżej 50% zostały ukryte…….....103

Rycina 9. Odległość między sąsiednimi genami a połączeniami dużej pojedynczej kopii (LSC), małej pojedynczej kopii (SSC) i dwóch odwróconych powtórzeń (IR) wśród genomów cp z dwunastu gatunków z rodziny *Ericaceae*. Każdy gatunek i odpowiadająca mu długość sekwencji genomu chloroplastów są przedstawione po lewej stronie każdej ścieżki. Geny transkrybowane na nici dodatnie i ujemne są przedstawione powyżej i poniżej odpowiadających im ścieżek odpowiednio w kierunkach od prawej do lewej i od lewej do prawej. Strzałki pokazują odległość pz początkowej lub końcowej współrzędnej danego genu od odpowiedniego miejsca połączenia. W przypadku genów rozciągających się z jednego regionu do drugiego, pasek T na górze lub poniżej pokazuje zasięg ich części wraz z odpowiadającymi im wartościami. Geny w pobliżu skrzyżowań to realistycznie wyskalowane projekcje odległości pz dla każdego miejsca. …….……...105-107

Rycina 10. Porównanie sekwencji regionów odwróconych IR trzynastu gatunków z rodziny *Ericaceae*. Graficzna informacja o podobieństwie oparta na mVISTA przedstawia identyczność sekwencji IR z *Actinidia arguta* jako regionem odniesienia. Dla wykresów zastosowano punkt odcięcia 50% identyczności. Na każdym wykresie oś w skali Y reprezentuje identyczność procentową (50 do 100%). Przerywane przestrzenie wskazują bardzo rozbieżne regiony w porównaniu z regionem odniesienia. Cyfry odpowiadają kolejno następującym gatunkom: 1 – *Erica versicolor*, 2 – *Erica arborea*, 3 – *Arbutus unedo*, 4 – *Chamaedaphne calyculata*, 5 – *Pyrola rotundifolia*, 6 –

Rhododendron delavai, 7 – *Rhododendron platypodum*, 8 – *Rhododendron pulchrum*, 9 – *Vaccinium bracteatum*, 10 – *Vaccinium macrocarpon*, 11 – *Vaccinium oldhamii*, 12 – *Vaccinium uliginosum*, 13 – *Vaccinium vitis-idaea*……108

Rycina 11. Syntenie i rearanżacje wykryte w kompletnych regionach odwróconych powtórzeń genomów chloroplastowych sześciu gatunków: przedstawicieli z rodziny *Ericaceae* oraz *Actinidia arguta*. Jednokolorowe pola połączone liniami wskazują lokalne bloki współliniowe i reprezentują regiony synteniczne. Profile podobieństwa tożsamości sekwencji są reprezentowane przez histogramy w każdym bloku, a sekwencje transkrybowane w odwrotnych kierunkach są oznaczone kolorowymi blokami powyżej i poniżej linii środkowej odwróconych powtórzeń badanych gatunków. ……..……..109

Rycina 12. Porównanie sekwencji genomów chloroplastowych z rodzaju *Erica*. Graficzna informacja o podobieństwie oparta na mVISTA przedstawia poziom identyczności sekwencji plastomu Erica versicolor z Erica arborea jako odniesieniem. Dla wykresów zastosowano punkt odcięcia 70% identyczności.Oś Y reprezentuje procent identyczności w zakresie 50%-100% Przerywane przestrzenie wskazują rozbieżne regiony w zestawieniu z regionem odniesienia…….115

Rycina 13. Porównanie genomu chloroplastowego *Erica arborea* z genomem chloroplastowym *Erica versicolor*. Dane uzyskane w programie Mauve 2.3.1. Jednokolorowe pola połączone liniami wskazują lokalne bloki współliniowe i reprezentują regiony synteniczne. Widoczny profil podobieństwa tożsamości sekwencji reprezentowany jest przez histogramy wewnątrz bloków……..……116

Rycina 14. Porównanie sekwencji genomów chloroplastowych *Erica arborea* i *Erica versicolor*. Kolor intensywnie zielony oznacza 100% poziom identyczności, kolor zielonkawo-brunatny to podobieństwo niższe niż 100%, ale powyżej 30%, czerwony przypisany jest dla poziomu podobieństwa poniżej 30%. Histogramy poniżej schematu dopasowania sekwencji to zawartość procentowa AT (kolor zielony) oraz GC (kolor niebieski) w poszczególnych genomach……117

Rycina 15. Przykładowe logo sekwencji konsensusowej; wykonane dla fragmentu genu *rpl*22 (11 243-11 295 pz). Kolor intensywnie zielony odpowiada dopasowaniu sekwencji genu między genomami *Erica arborea* a *Erica versicolor* na poziomie 100%. Kolor zielonkawo-brunatny to miejsca wystąpienia mutacji……118

Rycina 16. Graficzne zestawienie metodą dot plot dwóch kompletnych sekwencji plastomów *Erica arborea* i *Erica versicolor* celem identyfikacji regionów bliskiego podobieństwa między nimi ……...……..…129-130

Rycina 17. Konsensusowe drzewo filogenetyczne przedstawicieli z rodziny *Ericeae* skontruowane analizą bayesowską dla kompletnych genomów chloroplastowych. Wartości bootstrap poniżej 50% są ukryte. Liczby nad gałęziami reprezentują wsparcie bootstrapowe uzyskane przez maksymalne prawdopodobieństwo (wartość ML)……...…….144

Rycina 18. Schemat organizacji strukturalnej otrzymanego klastera genów rDNA rodzaju *Erica*……..146

Rycina 19. Procentowy rozkład zmienności w obrębie poszczególnych fragmentów funkcjonalnych regionów NOR…….148

Rycina 20. Informatywność filogenetyczna. Profil (A) ogólny, sieciowy i (B) w przeliczeniu na miejsca filogenetycznej informatywności dla wybranych regionów w obrębie rDNA. Oś czasu została skalibrowana do drzewka ML o ultrametryzowanym obliczeniu dla pełnej jednostki rDNA. Czas ewolucji wynoszący 1,0 odpowiada korzeniu drzewa, a wartość 0 jego wierzchołkom…….….156-157

Rycina 21. Zestawienie dla regionów ITS i 3'ETS + NTS. Kladogramy konsensusowe otrzymane z analizy ML całego ITS (A) i regionu 3'ETS + NTS (B). Wartości bootstrap poniżej 55% nie zostały ujęte…….159

Rycina 22. Konsensusowa sieć filogenetyczna wygenerowana w Splits Tree, zrekonstruowana przy użyciu inkluzywnego zestawu danych sekwencji kompletnych genów cpDNA przedstawicieli z rodzaju *Erica*…….……167

Rycina 23. Konsensusowe drzewo filogenetyczne przedstawicieli z rodzaju *Erica* skontruowane analizą bayesowską dla kompletnych genów *rbc*L. Wartości bootstrap poniżej 50% są ukryte. Liczby nad gałęziami reprezentują wsparcie bootstrapowe uzyskane przez maksymalne prawdopodobieństwo (wartość ML).....169

Rycina 24. Konsensusowe drzewo filogenetyczne przedstawicieli z rodzaju *Erica* skontruowane analizą bayesowską dla kompletnych genów *mat*K. Wartości bootstrap poniżej 50% są ukryte. Liczby nad gałęziami reprezentują wsparcie bootstrapowe uzyskane przez maksymalne prawdopodobieństwo (wartość ML)……170

Rycina 25. Konsensusowe drzewo filogenetyczne przedstawicieli z rodzaju *Erica* skontruowane analizą bayesowską dla kompletnych genów *ccs*A. Wartości bootstrap poniżej 50% są ukryte. Liczby nad gałęziami reprezentują wsparcie bootstrapowe uzyskane przez maksymalne prawdopodobieństwo (wartość ML)……171

Rycina 26. Konsensusowe drzewo filogenetyczne przedstawicieli z rodzaju *Erica* skontruowane analizą bayesowską dla kompletnych genów *cem*A. Wartości bootstrap poniżej 50% są ukryte. Liczby nad gałęziami reprezentują wsparcie bootstrapowe uzyskane przez maksymalne prawdopodobieństwo (wartość ML) ….....172

Rycina 27. Konsensusowe drzewo filogenetyczne przedstawicieli z rodzaju *Erica* skontruowane analizą bayesowską dla kompletnych genów *acc*D. Wartości bootstrap poniżej 50% są ukryte. Liczby nad gałęziami reprezentują wsparcie bootstrapowe uzyskane przez maksymalne prawdopodobieństwo (wartość ML)……173

Rycina 28. Konsensusowe drzewo filogenetyczne przedstawicieli z rodzaju *Erica* skontruowane analizą bayesowską dla kompletnych genów *atp*. Wartości bootstrap poniżej 50% są ukryte. Liczby nad gałęziami reprezentują wsparcie bootstrapowe uzyskane przez maksymalne prawdopodobieństwo (wartość ML)……174

Rycina 29. Konsensusowe drzewo filogenetyczne przedstawicieli z rodzaju *Erica* skontruowane analizą bayesowską dla kompletnych genów *ndh*. Wartości bootstrap poniżej 50% są ukryte. Liczby nad gałęziami reprezentują wsparcie bootstrapowe uzyskane przez maksymalne prawdopodobieństwo (wartość ML)……175

Rycina 30. Konsensusowe drzewo filogenetyczne przedstawicieli z rodzaju *Erica* skontruowane analizą bayesowską dla kompletnych genów *pet*. Wartości bootstrap poniżej 50% są ukryte. Liczby nad gałęziami reprezentują wsparcie bootstrapowe uzyskane przez maksymalne prawdopodobieństwo (wartość ML)……...….176

Rycina 31. Konsensusowe drzewo filogenetyczne przedstawicieli z rodzaju *Erica* skontruowane analizą bayesowską dla kompletnych genów *rpo*. Wartości bootstrap poniżej 50% są ukryte. Liczby nad gałęziami reprezentują wsparcie bootstrapowe uzyskane przez maksymalne prawdopodobieństwo (wartość ML)……177

Rycina 32. Konsensusowe drzewo filogenetyczne przedstawicieli z rodzaju *Erica* skontruowane analizą bayesowską dla kompletnych genów *r*RNA. Wartości bootstrap poniżej 50% są ukryte. Liczby nad gałęziami reprezentują wsparcie bootstrapowe uzyskane przez maksymalne prawdopodobieństwo (wartość ML)……178

Rycina 33. Konsensusowe drzewo filogenetyczne przedstawicieli z rodzaju *Erica* skontruowane analizą bayesowską dla kompletnych genów *ycf*. Wartości bootstrap poniżej 50% są ukryte. Liczby nad gałęziami reprezentują wsparcie bootstrapowe uzyskane przez maksymalne prawdopodobieństwo (wartość ML)……179

Rycina 34. Konsensusowe drzewo filogenetyczne przedstawicieli z rodzaju *Erica* skontruowane analizą bayesowską dla kompletnych genów *psa*. Wartości bootstrap poniżej 50% są ukryte. Liczby nad gałęziami reprezentują wsparcie bootstrapowe uzyskane przez maksymalne prawdopodobieństwo (wartość ML)……180

Rycina 35. Konsensusowe drzewo filogenetyczne przedstawicieli z rodzaju *Erica* skontruowane analizą bayesowską dla kompletnych genów *rpl*. Wartości bootstrap poniżej 50% są ukryte. Liczby nad gałęziami reprezentują wsparcie bootstrapowe uzyskane przez maksymalne prawdopodobieństwo (wartość ML)……181

Rycina 36. Konsensusowe drzewo filogenetyczne przedstawicieli z rodzaju *Erica* skontruowane analizą bayesowską dla kompletnych genów *rps*. Wartości bootstrap poniżej 50% są ukryte. Liczby nad gałęziami reprezentują wsparcie bootstrapowe uzyskane przez maksymalne prawdopodobieństwo (wartość ML)……..……..182

Rycina 37. Konsensusowe drzewo filogenetyczne przedstawicieli z rodzaju *Erica* skontruowane analizą bayesowską dla kompletnych genów *psb*. Wartości bootstrap poniżej 50% są ukryte. Liczby nad gałęziami reprezentują wsparcie bootstrapowe uzyskane przez maksymalne prawdopodobieństwo (wartość ML)……183

Rycina 38. Konsensusowe drzewo filogenetyczne przedstawicieli z rodzaju *Erica* skontruowane analizą bayesowską dla kompletnych genów *trn*. Wartości bootstrap poniżej 50% są ukryte. Liczby nad gałęziami reprezentują wsparcie bootstrapowe uzyskane przez maksymalne prawdopodobieństwo (wartość ML)……....184

Rycina 39. Informatywność filogenetyczna. Profil pierwszy (A) ogólny sieciowy, drugi (B) powstał w przeliczeniu na miejsca filogenetycznej informatywności dla wybranych regionów w obrębie genów *acc*D, *cem*A, *ccs*A, *mat*K, *rbc*L. Oś czasu została skalibrowana do drzewka ML o ultrametryzowanym obliczeniu wyprowadzony ze zbioru danych wszystkich badanych genów wykorzystanych do kalibracji osi czasu. Czas ewolucji wynoszący 1,0 odpowiada korzeniu drzewa, a wartość 0 jego wierzchołkom…….….....187-188

Rycina 40. Informatywność filogenetyczna. Profil pierwszy (A) ogólny sieciowy, drugi (B) powstał w przeliczeniu na miejsca filogenetycznej informatywności dla wybranych regionów w obrębie genów *rrn*. Oś czasu została skalibrowana do drzewka ML o ultrametryzowanym obliczeniu wyprowadzony ze zbioru danych wszystkich badanych genów wykorzystanych do kalibracji osi czasu. Czas ewolucji wynoszący 1,0 odpowiada korzeniu drzewa, a wartość 0 jego wierzchołkom…….....189-190

Rycina 41. Informatywność filogenetyczna. Profil pierwszy (A) ogólny sieciowy, drugi (B) powstał w przeliczeniu na miejsca filogenetycznej informatywności dla wybranych regionów w obrębie genów *rpo*. Oś czasu została skalibrowana do drzewka ML o ultrametryzowanym obliczeniu wyprowadzony ze zbioru danych wszystkich badanych genów *rpo* wykorzystanych do kalibracji osi czasu. Czas ewolucji wynoszący 1,0 odpowiada korzeniu drzewa, a wartość 0 jego wierzchołkom…….....191-192

Rycina 42. Informatywność filogenetyczna. Profil pierwszy (A) ogólny sieciowy, drugi (B) powstał w przeliczeniu na miejsca filogenetycznej informatywności dla wybranych regionów w obrębie genów *atp*. Oś czasu została skalibrowana do drzewka ML o ultrametryzowanym obliczeniu wyprowadzony ze zbioru danych wszystkich badanych genów *atp* wykorzystanych do kalibracji osi czasu. Czas ewolucji wynoszący 1,0 odpowiada korzeniu drzewa, a wartość 0 jego wierzchołkom…….....193-194

Rycina 43. Informatywność filogenetyczna. Profil pierwszy (A) ogólny sieciowy, drugi (B) powstał w przeliczeniu na miejsca filogenetycznej informatywności dla wybranych regionów w obrębie genów *pet*. Oś czasu została skalibrowana do drzewka ML o ultrametryzowanym obliczeniu wyprowadzony ze zbioru danych wszystkich badanych genów *pet* wykorzystanych do kalibracji osi czasu. Czas ewolucji wynoszący 1,0 odpowiada korzeniu drzewa, a wartość 0 jego wierzchołkom……....…195-196

Rycina 44. Informatywność filogenetyczna. Profil pierwszy (A) ogólny sieciowy, drugi (B) powstał w przeliczeniu na miejsca filogenetycznej informatywności dla wybranych regionów w obrębie genów *ndh*. Oś czasu została skalibrowana do drzewka ML o ultrametryzowanym obliczeniu wyprowadzony ze zbioru danych wszystkich badanych genów *ndh* wykorzystanych do kalibracji osi czasu. Czas ewolucji wynoszący 1,0 odpowiada korzeniu drzewa, a wartość 0 jego wierzchołkom……....….197-198

Rycina 45. Informatywność filogenetyczna. Profil pierwszy (A) ogólny sieciowy, drugi (B) powstał w przeliczeniu na miejsca filogenetycznej informatywności dla wybranych regionów w obrębie genów *rpl*. Oś czasu została skalibrowana do drzewka ML o ultrametryzowanym obliczeniu wyprowadzony ze zbioru danych wszystkich badanych genów *rpl* wykorzystanych do kalibracji osi czasu. Czas ewolucji wynoszący 1,0 odpowiada korzeniu drzewa, a wartość 0 jego wierzchołkom……...……..199-200

Rycina 46. Informatywność filogenetyczna. Profil pierwszy (A) ogólny sieciowy, drugi (B) powstał w przeliczeniu na miejsca filogenetycznej informatywności dla wybranych regionów w obrębie genów *rps*. Oś czasu została skalibrowana do drzewka ML o ultrametryzowanym obliczeniu wyprowadzony ze zbioru danych wszystkich badanych genów *rps* wykorzystanych do kalibracji osi czasu. Czas ewolucji wynoszący 1,0 odpowiada korzeniu drzewa, a wartość 0 jego wierzchołkom……....….201-202

Rycina 47. Informatywność filogenetyczna. Profil pierwszy (A) ogólny sieciowy, drugi (B) powstał w przeliczeniu na miejsca filogenetycznej informatywności dla wybranych regionów w obrębie genów *psa*. Oś czasu została skalibrowana do drzewka ML o ultrametryzowanym obliczeniu wyprowadzony ze zbioru danych wszystkich badanych genów *psa* wykorzystanych do kalibracji osi czasu. Czas ewolucji wynoszący 1,0 odpowiada korzeniu drzewa, a wartość 0 jego wierzchołkom……..…..….203-204

Rycina 48. Informatywność filogenetyczna. Profil pierwszy (A) ogólny sieciowy, drugi (B) powstał w przeliczeniu na miejsca filogenetycznej informatywności dla wybranych regionów w obrębie genów *psb*. Oś czasu została skalibrowana do drzewka ML o ultrametryzowanym obliczeniu wyprowadzony ze zbioru danych wszystkich badanych genów *psb* wykorzystanych do kalibracji osi czasu. Czas ewolucji wynoszący 1,0 odpowiada korzeniu drzewa, a wartość 0 jego wierzchołkom……..……...205-206

Rycina 49. Informatywność filogenetyczna. Profil pierwszy (A) ogólny sieciowy, drugi (B) powstał w przeliczeniu na miejsca filogenetycznej informatywności dla wybranych regionów w obrębie genów *ycf*. Oś czasu została skalibrowana do drzewka ML o ultrametryzowanym obliczeniu wyprowadzony ze zbioru danych wszystkich badanych genów *ycf* wykorzystanych do kalibracji osi czasu. Czas ewolucji wynoszący 1,0 odpowiada korzeniu drzewa, a wartość 0 jego wierzchołkom……..……..….207-208

Rycina 50. Informatywność filogenetyczna. Profil pierwszy (A) ogólny sieciowy, drugi (B) powstał w przeliczeniu na miejsca filogenetycznej informatywności dla wybranych regionów w obrębie genów *trn*. Oś czasu została skalibrowana do drzewka ML o ultrametryzowanym obliczeniu wyprowadzony ze zbioru danych wszystkich badanych genów *ycf* wykorzystanych do kalibracji osi czasu. Czas ewolucji wynoszący 1,0 odpowiada korzeniu drzewa, a wartość 0 jego wierzchołkom……..……..….209-210

Rycina 51. Konsensusowe drzewo filogenetyczne przedstawicieli z rodzaju *Erica* skontruowane analizą bayesowską dla kompletnych genów *ndh*F. Wartości bootstrap poniżej 50% są ukryte. Liczby nad gałęziami reprezentują wsparcie bootstrapowe uzyskane przez maksymalne prawdopodobieństwo (wartość ML). Grupę zewnętrzną stanowił gatunek *Vaccinium macrocarpon*…….213

Rycina 52. Konsensusowe drzewo filogenetyczne przedstawicieli z rodzaju *Erica* skontruowane analizą bayesowską dla kompletnych genów *rpo*C2. Wartości bootstrap poniżej 50% są ukryte. Liczby nad gałęziami reprezentują wsparcie bootstrapowe uzyskane przez maksymalne prawdopodobieństwo (wartość ML). Grupę zewnętrzną stanowił gatunek *Vaccinium macrocarpon.* …..……...…214

Rycina 53. Konsensusowe drzewo filogenetyczne przedstawicieli z rodzaju *Erica* skontruowane analizą bayesowską dla kompletnych genów *psb*B. Wartości bootstrap poniżej 50% są ukryte. Liczby nad gałęziami reprezentują wsparcie bootstrapowe uzyskane przez maksymalne prawdopodobieństwo (wartość ML). Grupę zewnętrzną stanowił gatunek *Chamaedaphne calyculata* ….....……..…….215

Rycina 54. Konsensusowe drzewo filogenetyczne przedstawicieli z rodzaju *Erica* skontruowane analizą bayesowską dla kompletnych genów *atp*F. Wartości bootstrap poniżej 50% są ukryte. Liczby nad gałęziami reprezentują wsparcie bootstrapowe uzyskane przez maksymalne prawdopodobieństwo (wartość ML). Grupę zewnętrzną stanowił gatunek *Vaccinium macrocarpon*……216

Spis tabel

Tabela 1. Lista gatunków wykorzystanych do analiz z ich przynależnością do podrodzaju i sekcji w obrębie rodzaju *Erica* wraz z miejscem zebrania materiału i kolekcjonerem próby……...53

Tabela 2. Długość i liczba odczytów dla poszczególnych osobników, uzyskanych z sekwencjonowania NGS (całkowity otrzymany zasób genomowy)……85

Tabela 3. Dystrybucja genów w regionie plastydowym u *Erica versicolor*…….……88

Tabela 4. Geny ustalone w genomie chloroplastowym *Erica versicolor* ([a]geny zlokalizowane w regionach pojedynczych kopii. [b]pseudogeny, [c]geny zlokalizowane w regionie odwróconych powtórzeń,[d]2 kopie, 1 funkcjonalna i 2 pseudogen)……...88-90

Tabela 5. Dystrybucja genów w regionie plastydowym u *Erica arborea*…….93

Tabela 6. Geny ustalone w genomie chloroplastowym *Erica arborea* ([a]geny zlokalizowane w regionach pojedynczych kopii. [b]pseudogeny, [c]geny zlokalizowane w regionie odwróconych powtórzeń, [d]2 kopie, 1 funkcjonalna i 2 pseudogen)…….93-95

Tabela 7. Zestawienie polimorfizmów w sekwencjach kodujących pomiędzy badanymi genomami *Erica arborea* a *Erica versicolor*. Długość sekwencji w genomie (pz), liczba mutacji oraz insercji w genach i rejonach kodujących (CDS) u *Erica arborea* względem *Erica versicolor*. SNP – liczba substytucji, InDel – liczba insercji/delecji w CDS, π – wartości zróżnicowania nukleotydowego, p – odległość genetyczna, NS – mutacje niesynonimiczne, kategorie genu wg Changa i in. (2006): GE – geny ekspresji genów, GA – geny aparatu fotosyntetycznego, GM – geny metabolizmu fotosyntetycznego, R – różne. Kolejność genów zgodna z wynikiem porównania otrzymanego za pomocą Geneious Alignment v 9. Wykluczono z tabeli geny identyczne pomiędzy genomami oraz drugie powtórzenie odwrócone IRB..……...……..121-124

Tabela 8. Zestawienie polimorfizmów w sekwencjach niekodujących pomiędzy genomem chloroplastowym *Erica arborea* a genomem chloroplastowym *Erica versicolor*. Długość sekwencji w genomie [pz], liczba mutacji oraz insercji w sekwencjach międzygenowych u *Erica arborea* względem *Erica versicolor*. SNP – liczba substytucji, InDel – liczba insercji/delecji, p – odległość genetyczna, π – wartości zróżnicowania nukleotydowego. Wykluczono z tabeli sekwencje identyczne pomiędzy genomami oraz drugie powtórzenie odwrócone IRB……...……...124-127

Tabela 9. Charakterystyka wyrównanych sekwencji genomowych w odniesieniu do obu kompletnych sekwencji z rodzaju *Erica* oraz wszystkich kompletnych sekwencji plastomów fotosyntetyzujących przedstawicieli rodziny *Ericaceae*……..…….….133-134

Tabela 10. Odległość genetyczna (p) między badanymi gatunkami w plastomach fotosyntetyzujących *Ericaceae* ……..……135-137

Tabela 11. Porównanie szczegółów budowy genomów chloroplastowych *Erica arborea*, *Erica vesricolor* oraz plastomów fotosyntetyzujących przedstawicieli rodziny *Ericaceae*……139-143

Tabela 12. Całkowita długość otrzymanych sekwencji klasterów genów rDNA wraz z nadanymi numerami akcesyjnymi GenBank ……..…147

Tabela 13. Charakterystyka wybranych elementów funkcjonalnych i domen w ramach rDNA w rodzaju *Erica*. Jeśli w gatunku wykryto więcej niż jeden przypuszczalny TIS, rozważano opcję z krótszym 5'ETS……150-151

Tabela 14. Zestawienie przypuszczalnego TTS u przedstawicieli rodzaju *Erica*. Pozycja pierwszego nukleotydu TTS dla przypuszczalnej sekwencji TTS w odniesieniu do końca 3’ 26S rDNA…….....153

Tabela 15. Zestawienie przypuszczalnego TIS u przedstawicieli rodzaju *Erica*. Pozycja pierwszego nukleotydu TIS dla przypuszczalnej sekwencji TIS w odniesieniu do końca 5’ 18S rDNA……..154

Tabela 16. Zestawienie regionów wewnętrznych powtórzeń (motywów) określonych dla gatunków poddanych analizom……..162-163

Tabela 17. Wykres wykrytych motywów w rodzaju *Erica* wraz z ich wartością Pi……164-165

Spis fotografii

Opublikowane fotografie ośmiu gatunków *Erica* poddanych badaniom w niniejszej pracy oraz jedynego reprezentanta flory polskiej - wrzośca bagiennego (*Erica tetralix*).

Zdjęcia gatunków na następujących licencjach: CC BY-SA 2.0, CC BY-SA 3.0 oraz CC BY-SA 4.0, które pozwalają na udostępnianie, w tym kopiowanie i rozpowszechnianie materiału na dowolnym nośniku i w dowolnym formacie, w dowolnym celu, nawet komercyjnym. Z warunkiem uznania autorstwa, tj. podaniem odpowiedniego autora przy konkretnym zdjęciu oraz podaniem linku do strony (tu witryny: commons.wikipedia.org oraz casabio.org).

Zdjęcia gatunków, które zostały opublikowane za zasadach domeny publicznej przez właścicieli ich praw autorskich do danych dzieł, udostępnione są jako własność publiczna z zapewnieniem każdemu prawa do użycia tych zdjęć w dowolnym celu, bez żadnych ograniczeń. W obrębie fotografii nie wprowadzono żadnych zmian.

Fotografia 1. *Erica tetralix*, fot. Krzysztof Ziarnek.....38

https://commons.wikimedia.org/wiki/Category:Erica_tetralix#/media/File:Erica_tetralix_kz27.jpg

Fotografia 2. *Erica arborea*, fot. Javier Martin…….....43

https://commons.wikimedia.org/wiki/Erica_arborea#/media/File:Erica_arborea_Flowers Closeup_19March2009_SierraMadrona.jpg

Fotografia 3. *Erica versicolor*, fot. Stan Shebs…….....44

https://commons.wikimedia.org/wiki/Erica_versicolor#/media/File:Erica_versicolor_3.jpg

Fotografia 4. *Erica cinerea,* fot. Balles2601.....45

https://commons.wikimedia.org/wiki/Erica_cinerea#/media/File:Erica_cinerea_2601.JPG

Fotografia 5. *Erica vagans,* fot. Patrice78500…….....46

https://commons.wikimedia.org/wiki/Erica_vagans#/media/File:Bruyère_vagabonde.JPG

Fotografia 6. *Erica globiceps,*fot. David Gwynne-Evans (Cameron McMaster).....47

https://casabio.org/observations/147483/Erica globiceps/Boskloof 10Aug08_079.jpg

Fotografia 7. *Erica plukenetii,* fot. Joachim Huber…….....49

https://commons.wikimedia.org/wiki/Erica_plukenetii#/media/File:Western_Cape_Fynbos,_South_Africa_(3217278809).jpg

Fotografia 8. *Erica imbricata,* fot. Andrew Massyn..….....50

https://commons.wikimedia.org/wiki/File:Erica_imbricata.JPG#/media/File:Erica_imbricata.JPG

Fotografia 9. *Erica multiflora,* fot. Thomas Schoch.....51

https://commons.wikimedia.org/wiki/Erica_multiflora#/media/File:Erica_multiflora_Mallorca.jpg

Suplement

Suplement 1.

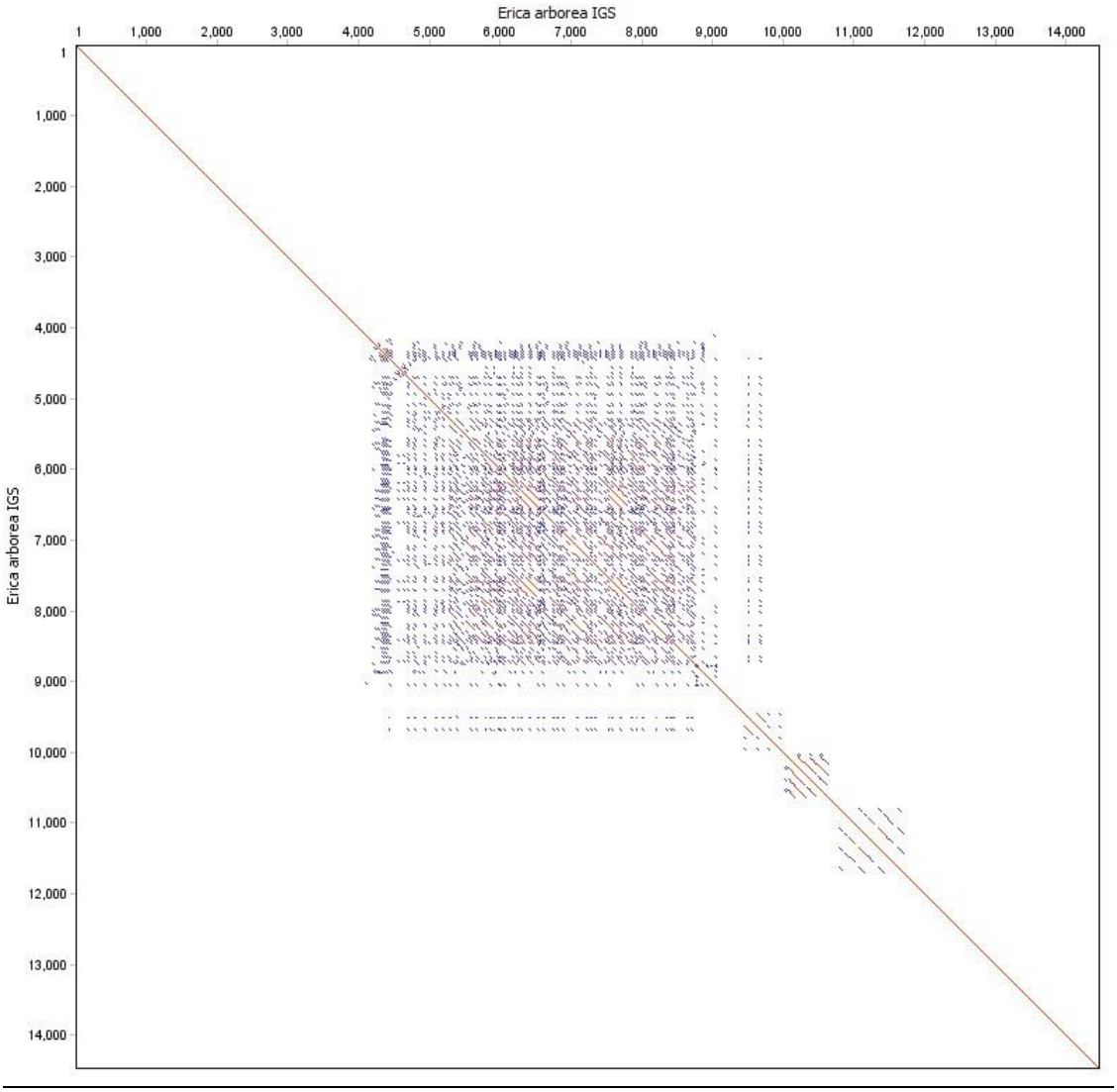
Sekwencje genów chloroplastowych otrzymanych dla gatunków *Erica cinerea*, *Erica globiceps*, *Erica imbricata*, *Erica multiflora*, *Erica plukenetii* oraz *Erica vagans*, które nie zostały zdeponowane w bazie danych sekwencji genetycznych GenBank, zestawiono w wersji dodatku do pracy w formie elektronicznej na osobnym pliku Word oraz na odrębnym nośniku elektronicznym (tu: dysk CD).

Sekwencje zapisane zostały w formacie FASTA osobno dla pojedynczych genów chloroplastowych i osobno dla grup genów chloroplastowych. W przypadku zapisu sekwencji dla całej grupy genów podano wyraźny zapis z dokładną lokalizacją poszczególnych genów w danej grupie.

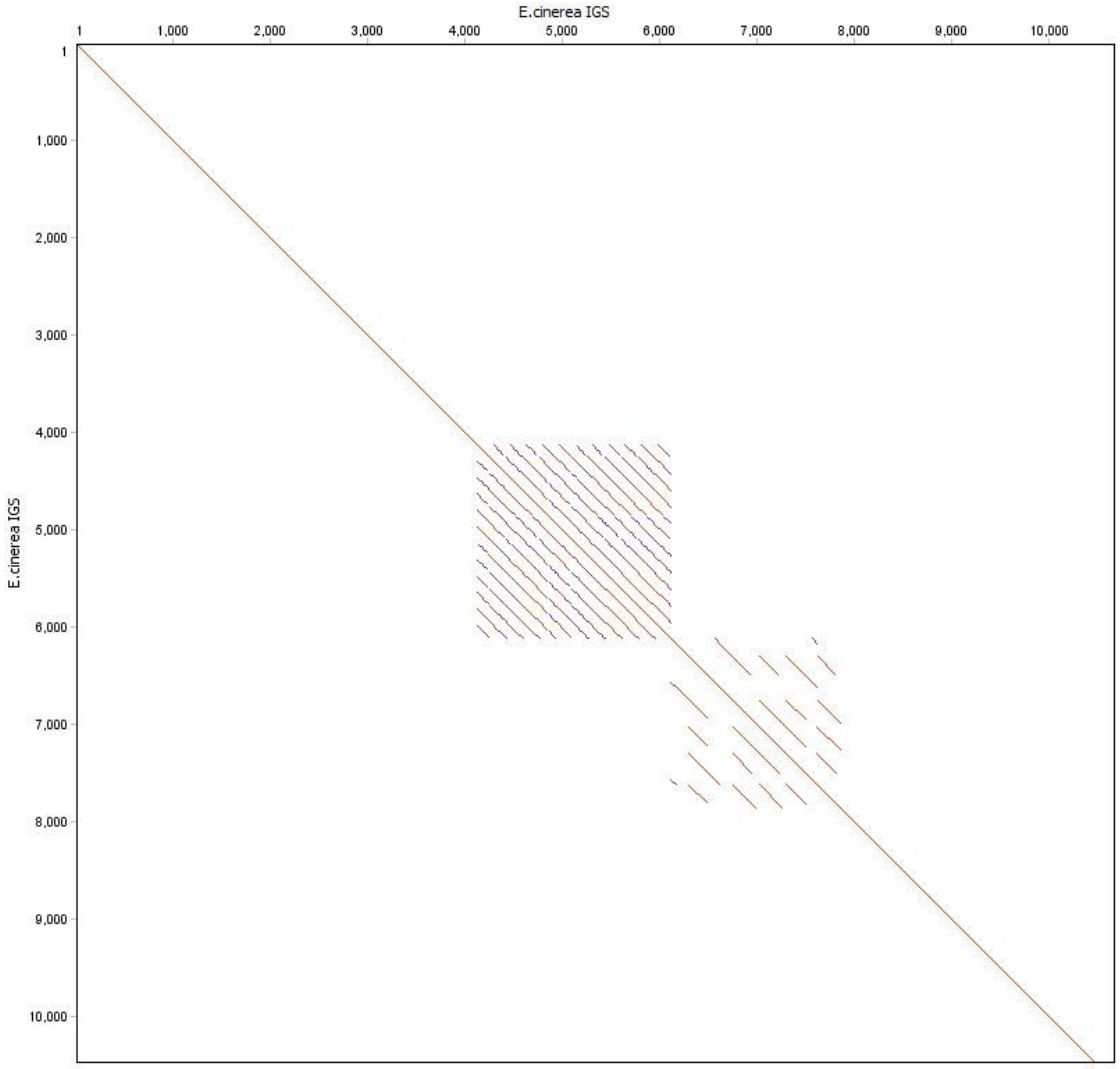
Do dokumentu zawierającego zestawienia sekwencji niezdeponowanych w bazie danych GenBank, dołączono także kompletne sekwencje genów chloroplastowych gatunków, dla których otrzymano całe genomy chloroplastowe (*Erica arborea* i *Erica versicolor*), tworząc skompilowane pliki, które posłużyły do analiz filogenetycznych w niniejszej pracy. Sekwencje genów chloroplastowych gatunku *Erica versicolor* stanowiły referencję do otrzymania sekwencji poszczególnych genów dla pozostałych badanych gatunków z rodzaju *Erica*.

Suplement 2.

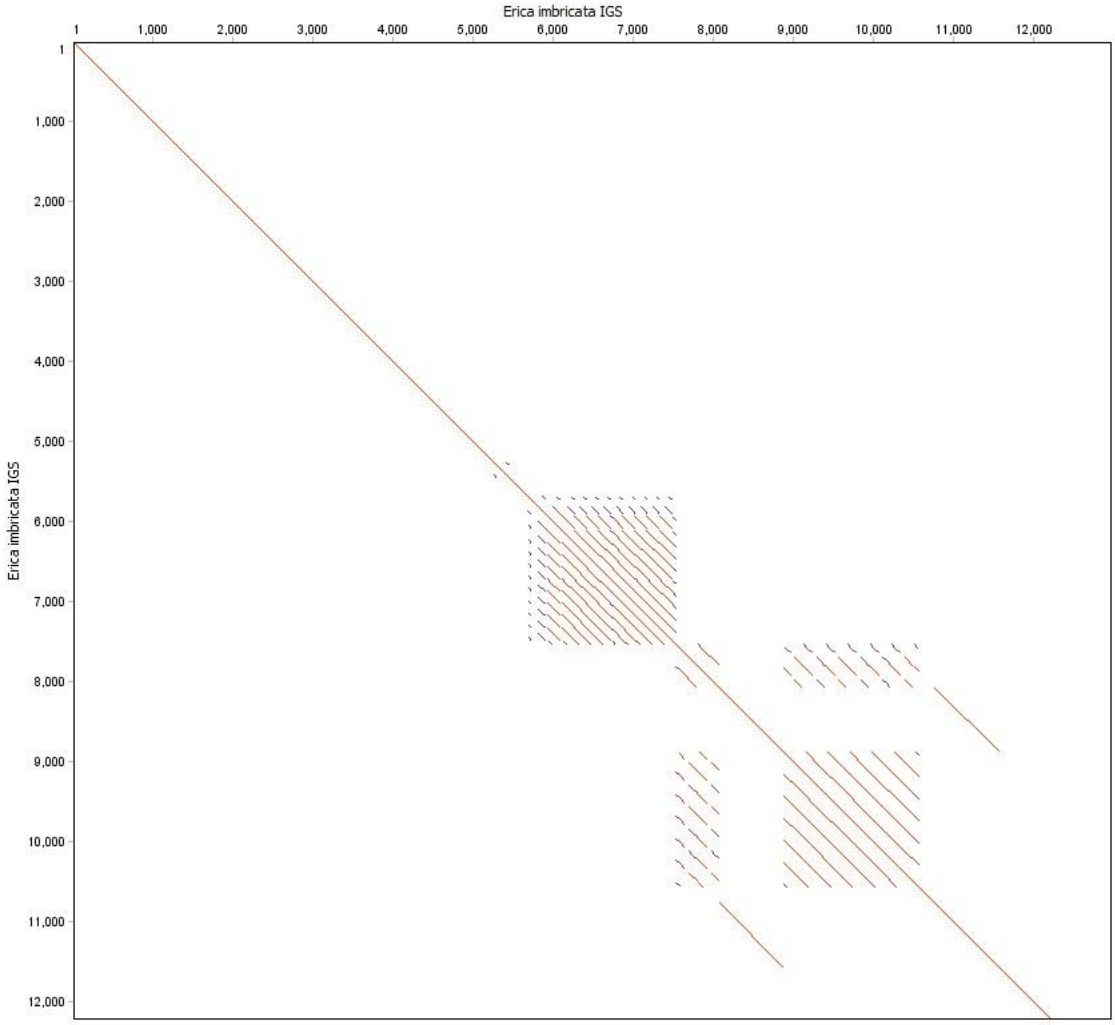
Wykresy dot plot regionów IGS poszczególnych przedstawicieli z rodzaju *Erica*.



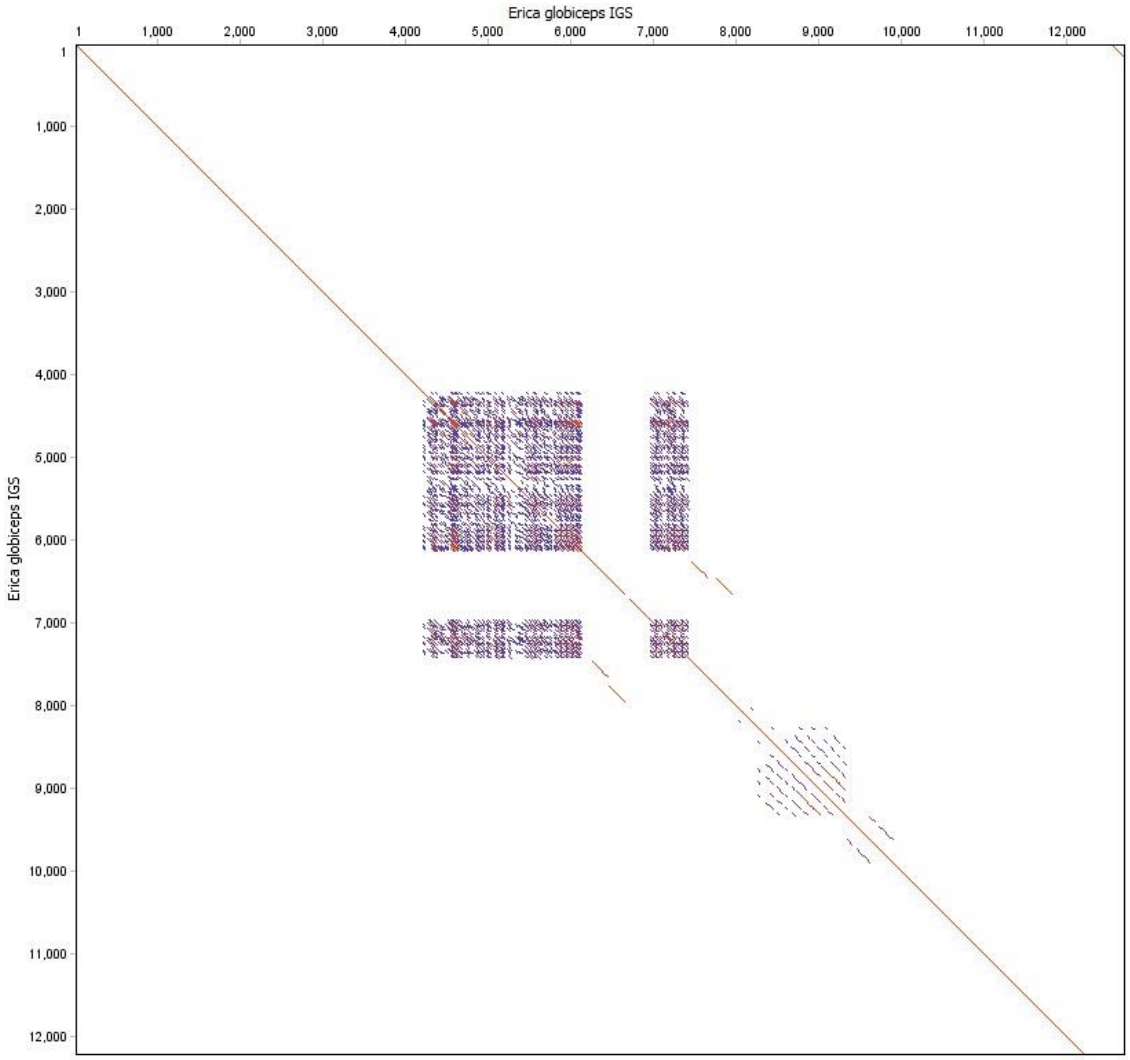
Wykres dot plot odstępów międzygenowych. Samoporównanie *Erica arborea* IGS. Rozmiar ramki zawierającej blok między powtórzeniami odpowiada długości między powtórzeniami



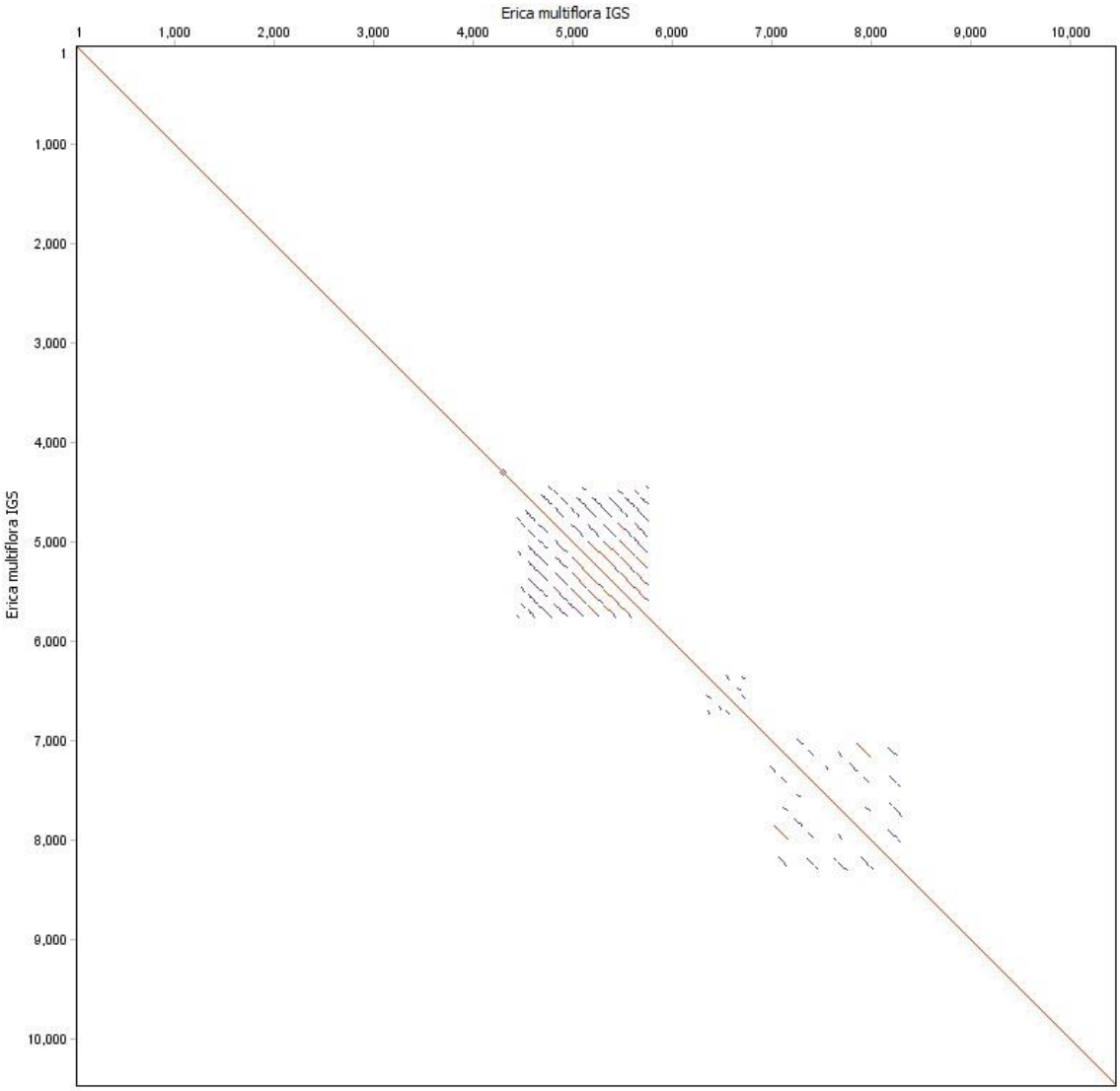
Wykres dot plot odstępów międzygenowych. Samoporównanie *Erica cinerea* IGS. Rozmiar ramki zawierającej blok między powtórzeniami odpowiada długości między powtórzeniami



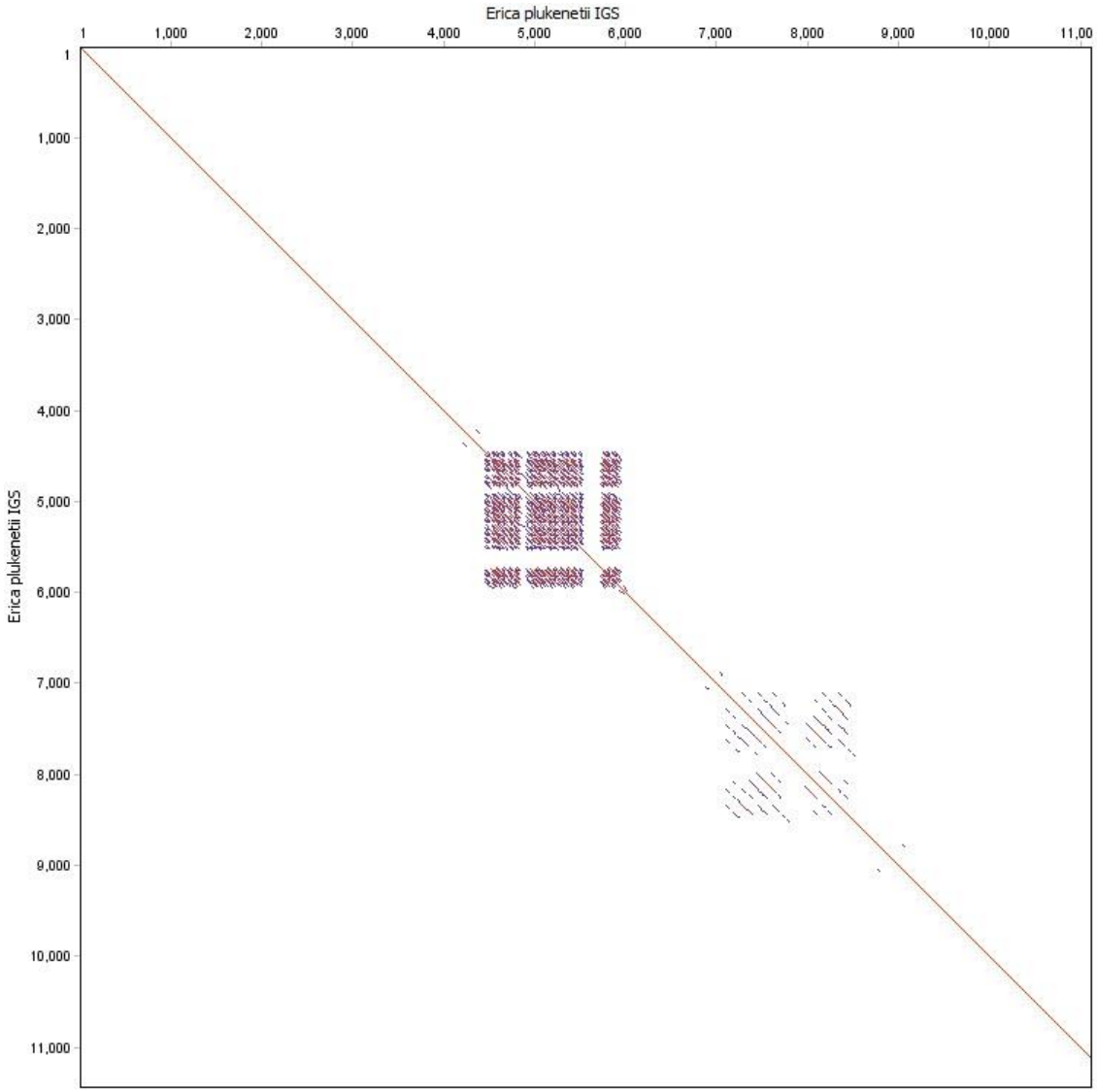
Wykres dot plot odstępów międzygenowych. Samoporównanie *Erica imbricata* IGS. Rozmiar ramki zawierającej blok między powtórzeniami odpowiada długości między powtórzeniami



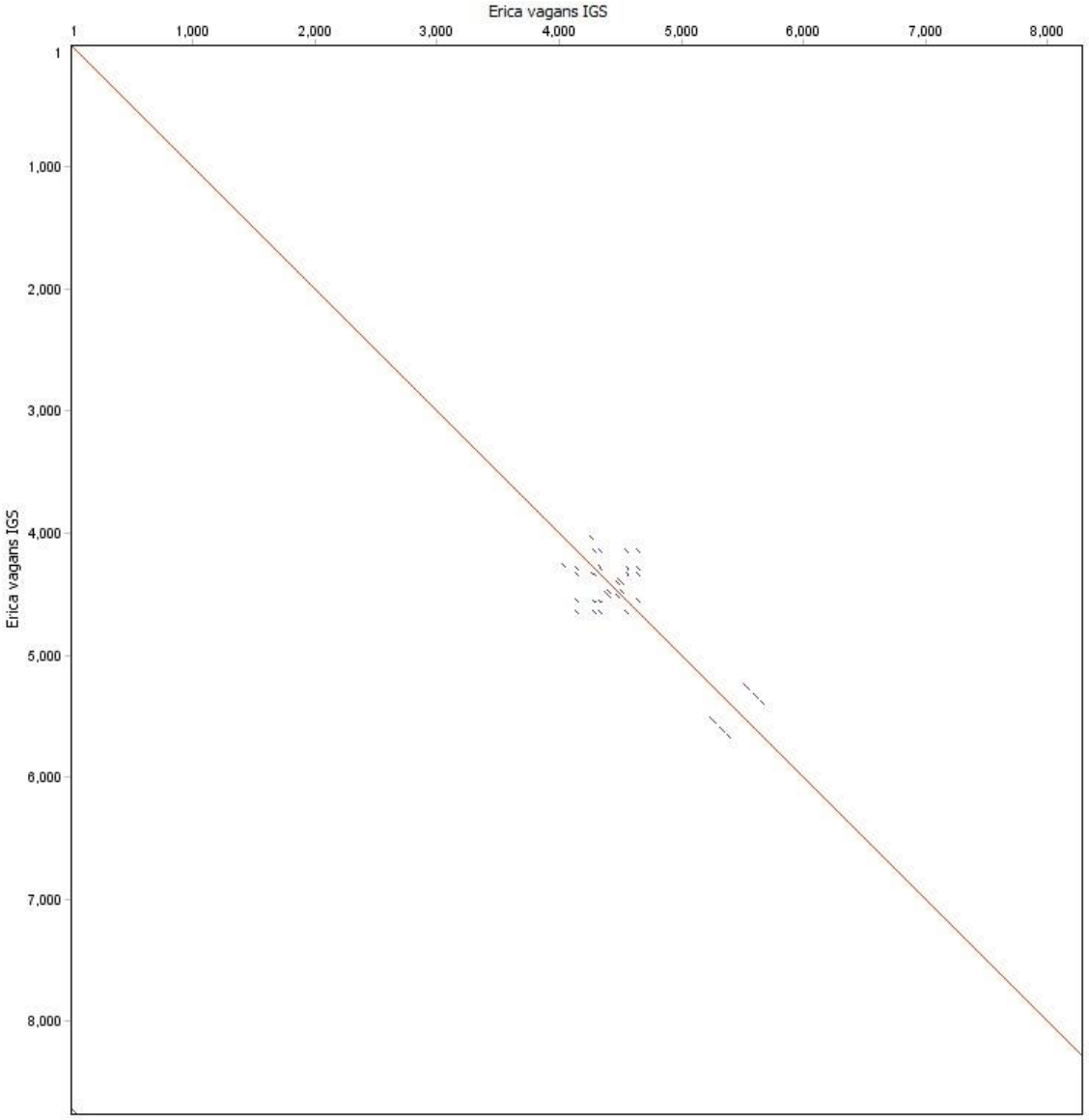
Wykres dot plot odstępów międzygenowych. Samoporównanie *Erica globiceps* IGS. Rozmiar ramki zawierającej blok między powtórzeniami odpowiada długości między powtórzeniami



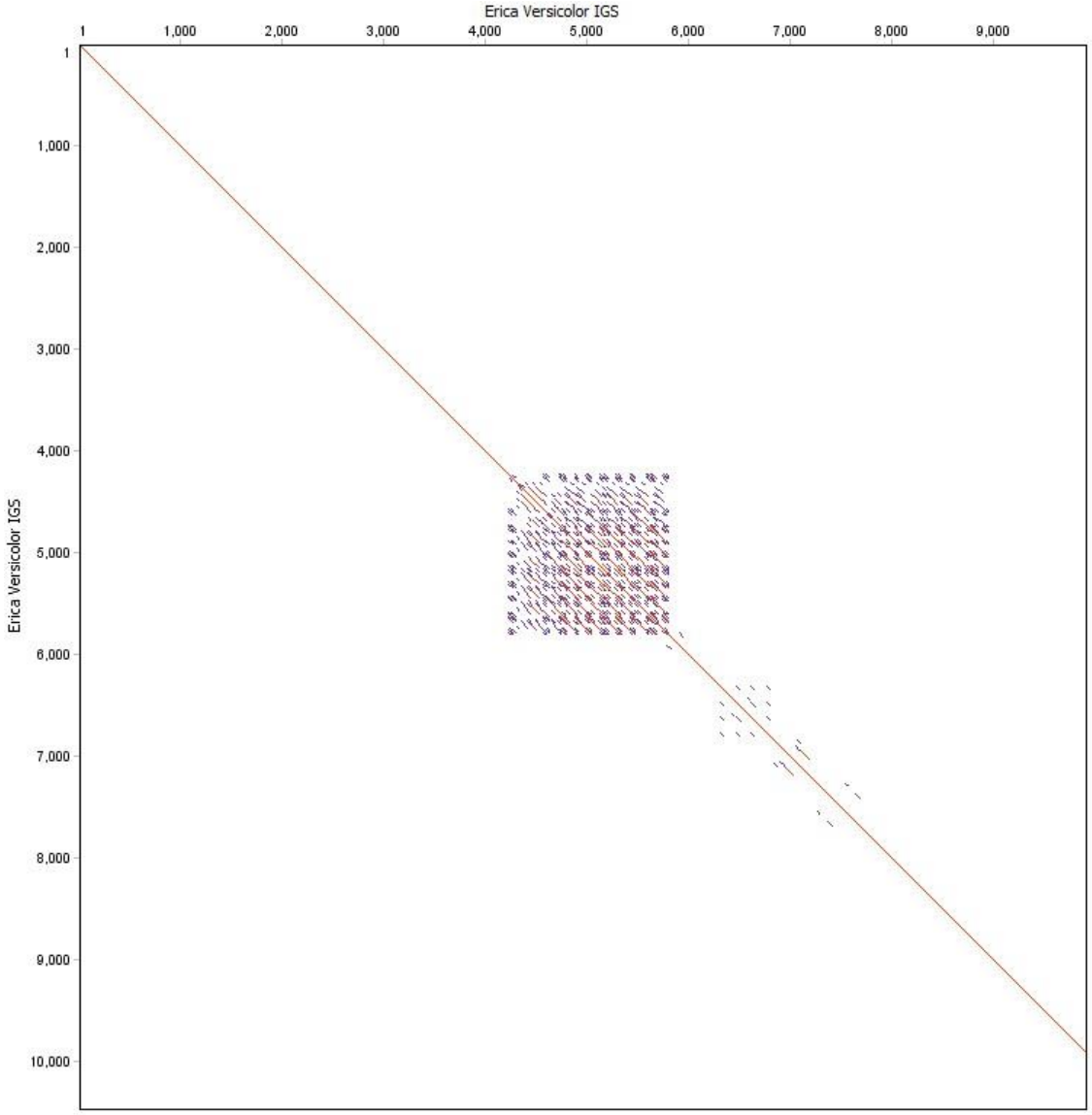
Wykres dot plot odstępów międzygenowych. Samoporównanie *Erica multiflora* IGS. Rozmiar ramki zawierającej blok między powtórzeniami odpowiada długości między powtórzeniami



Wykres dot plot odstępów międzygenowych. Samoporównanie *Erica plukenetii* IGS. Rozmiar ramki zawierającej blok między powtórzeniami odpowiada długości między powtórzeniami



Wykres dot plot odstępów międzygenowych. Samoporównanie *Erica vagans* IGS. Rozmiar ramki zawierającej blok między powtórzeniami odpowiada długości między powtórzeniami



Wykres dot plot odstępów międzygenowych. Samoporównanie *Erica vesricolor* IGS. Rozmiar ramki zawierającej blok między powtórzeniami odpowiada długości między powtórzeniami

Suplement 3.

Zestawienie wykrytych motywów w obrębie IGS u przedstawicieli rodzaju *Erica* wraz z liczbą ich powtórzeń.

Motyw	Gatunek i liczba powtórzeń								Całkowita liczba
	E.arborea	*E. cinerea*	*E.imbricata*	*E.multiflora*	*E.plukenetii*	*E.vagans*	*E.versicolor*	*E.globiceps*	
ER1	4	5	8	5	2	3	2	2	**31**
ER2	5	-	12	-	9	-	2	7	**35**
ER3	-	17	1	8	-	-	1	-	**27**
ER4	4	2	7	5	2	3	1	-	**24**
ER5	26	-	1	-	3	3	1	1	**35**
ER6	9	-	-	1	10	1	10	15	**46**
ER7	1	1	2	1	1	2	1	1	**10**
ER8	3	2	4	3	3	2	3	3	**23**
ER9	3	-	3	3	3	-	4	3	**19**

Motyw	Gatunek i liczba powtórzeń								Całkowita liczba
	E.arborea	*E. cinerea*	*E.imbricata*	*E.multiflora*	*E.plukenetii*	*E.vagans*	*E.versicolor*	*E.globiceps*	
ER10	1	1	1	1	1	1	1	1	**8**
ER11	3	1	-	1	2	1	1	2	**11**
ER12	1	1	1	-	-	1	1	1	**6**
ER13	-	3	2	-	-	-	-	2	**7**
ER14	-	2	1	2	1	1	2	2	**11**
Razem:	60	35	43	30	37	18	30	40	**293**

Suplement 4.

Zestawienie wykrytych motywów w *Chamaedaphne calyculata* z ich wartością Pi.

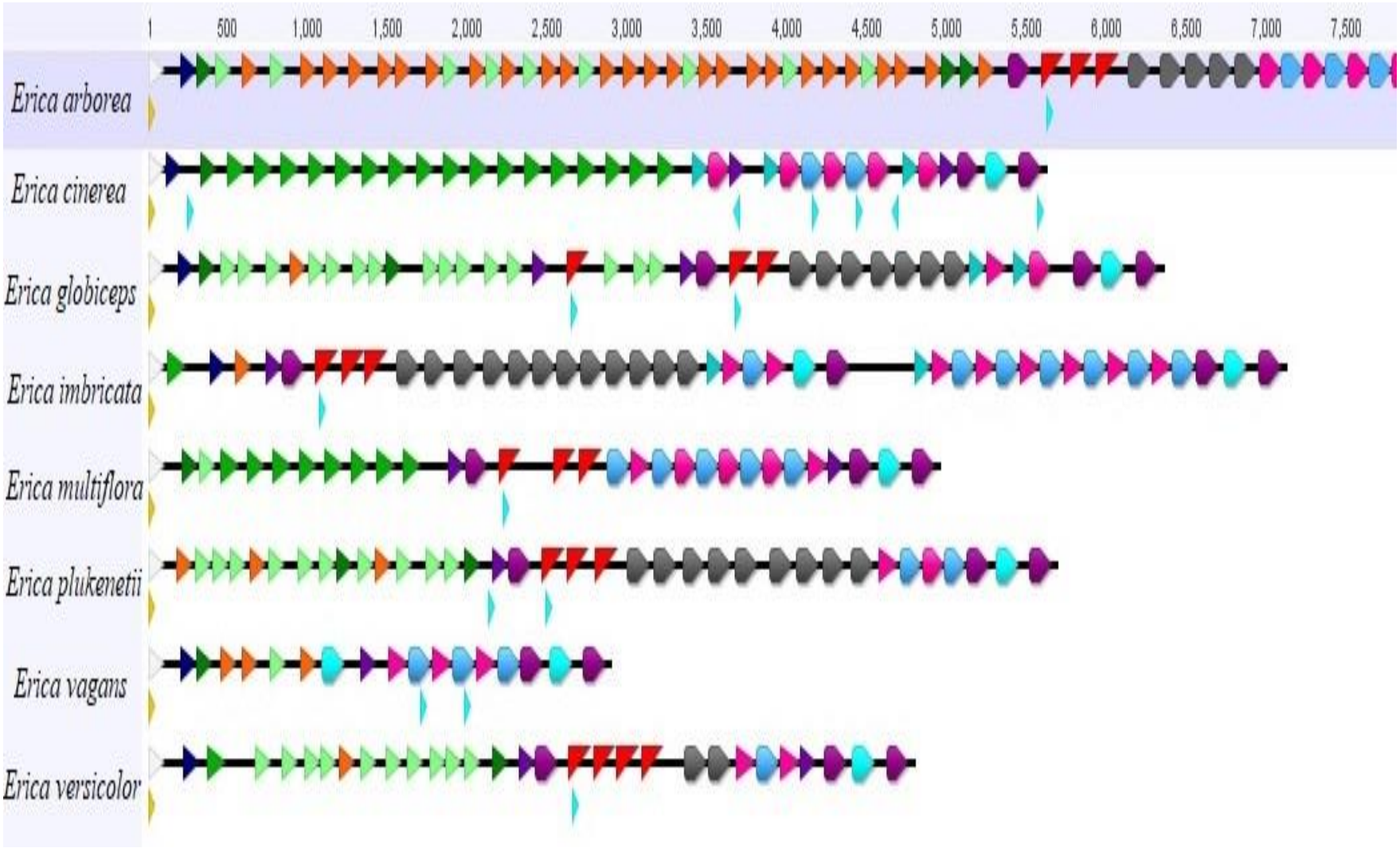
Motywy *Chamaedaphne calyculata*	Liczby powtórzeń	Wartość zróżnicowania nukleotydowego (Pi)
Motif 1	7	0,14182
Motif 2	4	0,08937
Motif 3	7	0,09895
Motif 4	5	0,07455
Motif 5	2	0,17949
Motif 6	4	0,38117
Motif 7	5	0,41455
Motif 8	6	0,14000
Motif 9	2	0,21429
Razem:	**42**	

Suplement 5.

Powtórzone regiony zdefiniowane dla rodzaju *Erica*.

Motywy (kod nazwy wraz z odpowiadającym mu kolorem na rycinie):

- **ER1** (błękitny),
- **ER2** (szary),
- **ER3** (zielony),
- **ER4** (fuksja),
- **ER5** (pomarańczowy),
- **ER6** (seledynowy),
- **ER7** (cyjanowy),
- **ER8** (fioletowy),
- **ER9** (czerwony),
- **ER10** (biały),
- **ER11** (ciemny zielony),
- **ER12** (granatowy),
- **ER13** (różowy),
- **ER14** (ciemny fiolet)


1
500
1,000
1,500
2,000
2,500
3,000
3,500
4,000
4,500
5,000
5,500
6,000
6,500
7,000
7,500
Erica arborea
Erica cinerea
Erica globiceps
Erica imbricata
Erica multiflora
Erica plukenetii
Erica vagans
Erica versicolor

Suplement 6.

Powtórzone regiony zdefiniowane dla gatunku *Chamaedaphne calyculata*.

Motywy (kod nazwy wraz z odpowiadającym mu kolorem na rycinie):

- **MOTIF 1** (pomarańczowy),
- **MOTIF 2** (niebieski),
- **MOTIF 3** (zielony),
- **MOTIF 4** (szary),
- **MOTIF 5** (czerwony),
- **MOTIF 6** (granatowy),
- **MOTIF 7** (biały),
- **MOTIF 8** (błękitny),
- **MOTIF 9** (róż indyjski)

TTS
MOTIF 7
MOTIF 7
MOTIF 6
MOTIF 7
MOTIF 7
MOTIF 2
MOTIF 5
MOTIF 3
MOTIF 2
MOTIF 3
MOTIF 2
MOTIF 3
MOTIF 3
MOTIF 5
TTS
MOTIF 6
MOTIF 6
MOTIF 3
MOTIF 6
MOTIF 2
MOTIF 3
MOTIF 1
MOTIF 4
MOTIF 8
MOTIF 1
MOTIF 4
MOTIF 8
MOTIF 9
MOTIF 4
MOTIF 1
MOTIF 9
MOTIF 8
MOTIF 9
MOTIF 8
MOTIF 4
MOTIF 1
MOTIF 8
MOTIF 1
MOTIF 4
MOTIF 8
MOTIF 9
MOTIF 8

Suplement 7.

Frekwencja zidentyfikowanych motywów SSR u *Erica versicolor*.

Powtórzenia	3	4	5	6	7	8	9	10	11	12	13	14
	15	16	17	18	19	20	21	22	23	RAZEM		
A	-	-	-	-	128	51	23	16	14	1	3	2
	-	-	1	-	-	-	-	-	1	**240**		
C	-	-	-	-	9	1						
										10		
G	-	-	-	-	9	1			1			
										11		
T	-	-	-	-	108	42	20	16	9	5	3	2
	1	2	2	2	1	-	1	-	-	**214**		
AG	2									**2**		
AT	23	1	1							**25**		
CT	2									**2**		
GA	2									**2**		
GT	1									**1**		
TA	22	7	1	1	-	-	1			**32**		
TC	2									**2**		
AAC	2									**2**		
AAG	30	-	20							**50**		
AAT	4									**4**		
ACG	3									**3**		

Powtórzenia	3	4	5	6	7	8	9	10	11	RAZEM
AGA	4									**4**
AGC	1									**1**
ATA	5	-	1							**6**
ATC	1									**1**
ATG	3									**3**
ATT	8									**8**
CAA	1									**1**
CAG	1									**1**
CAT	1									**1**
CCT	1									**1**
CGT	1									**1**
CTA	2									**2**
CTT	2									**2**
GAA	2	-	61	4						**67**
GCA	1									**1**
GCT	1									**1**
GGA	1									**1**
TAA	7									**7**
TAG	1									**1**
TAT	7									**7**
TCA	2									**2**
TCT	11	-	7							**18**
TTA	6									**6**

Powtórzenia	3	4	5	6	7	8	9	10	11	RAZEM
TTC	30	81	4							**115**
TTG	1									**1**
AAAT	1									**1**
AAGA	1									**1**
AATA	1									**1**
AATC	1									**1**
AATT	1									**1**
AGAA	1									**1**
GATT	1									**1**
TATT	1									**1**
TTTA	1									**1**
TTTC	3									**3**
TTTG	1									**1**
TAAAA	1									**1**

Suplement 8.

Frekwencja zidentyfikowanych motywów SSR u *Erica arborea*.

Powtórzenia	3	4	5	6	7	8	9	10	11	12	13	14
	15	16	17	18	19	20	21	22	23	RAZEM		
A	-	-	-	-	115	40	35	20	4	1	3	-
	-	-	1	-	-	-	-	-	1	**219**		
C	-	-	-	-	9	1						
										10		
G	-	-	-	-	7	1			1			
										9		
T	-	-	-	-	98	42	26	19	7	4	3	1
						-	1	-	-	**201**		
AG	2									**2**		
AT	14	1								**15**		
CT	2									**2**		
GA	2									**2**		
GT	1									**1**		
TA	28	5								**33**		
TC	1									**1**		
AAC	2									**2**		
AAG	32	14	2							**48**		
AAT	5	1								**6**		
ACG	3									**3**		

Powtórzenia	3	4	5	6	7	8	9	10	11	RAZEM
AGA	3	1								**4**
AGC	1									**1**
ATA	5									**5**
ATC	1									**1**
ATG	3									**3**
ATT	8									**8**
CAA	1									**1**
CAG	1									**1**
CAT	1									**1**
CCT	1									**1**
CGT	1									**1**
CTA	1									**1**
CTT	2									**2**
GAA	3		57	1						**61**
GCA	1									**1**
GCT	1									**1**
GGA	1									**1**
TAA	7	1	-	1						**9**
TAG	1									**1**
TAT	5									**5**
TCA	2									**2**
TCT	11	-	6							**17**
TTA	8									**8**

Powtórzenia	3	4	5	6	7	8	9	10	11	RAZEM
TTC	34	72	4							**110**
TTG	1									**1**
AAAT	1									**1**
AAGA	1									**1**
AATA	1									**1**
AATC	1									**1**
AATT	1									**1**
AGAA	1									**1**
GATT	1									**1**
TATT	1									**1**
TTTA	1									**1**
TTTC	3									**3**
TTTG	1									**1**
TAAAA	1									**1**

Suplement 9.

Frekwencja zaklasyfikowanych typów powtórzeń (biorąc pod uwagę komplementarność sekwencji) u *Erica versicolor*.

Powtórzenia	3	4	5	6	7	8	9	10	11	12	13	14
	15	16	17	18	19	20	21	22	23	RAZEM		
A/T	-	-	-	-	236	93	43	32	23	6	6	2
	3	2	2	3	1	-	-	1	1	**454**		
C/G	-	-	-	-	18	2	-	-	1	**21**		
AC/GT	1									**1**		
AG/CT	8									**8**		
AT/TA	45	8	2	1	-	1				**57**		
AAC/GTT	4									**4**		
AAG/CTT	79	169	8							**256**		
AAT/ATT	37	1								**38**		
ACG/CGT	4									**4**		
ACT/AGT	3									**3**		
AGC/CTG	5									**5**		
AGG/CCT	2									**2**		

Powtórzenia	3	4	5	6	7	8	9	10	11	RAZEM
ATC/ATG	7									**7**
AAAC/GTTT	1									**1**
AAAG/CTTT	5									**5**
AAAT/ATTT	4									**4**
AATC/ATTG	2									**2**
AATT/AATT	1									**1**
AAAAT/ATTTT	1									**1**

Suplement 10.

Frekwencja zaklasyfikowanych typów powtórzeń (biorąc pod uwagę komplementarność sekwencji) u *Erica arborea*.

Powtórzenia	3	4	5	6	7	8	9	10	11	12	13	14
	15	RAZEM										
A/T	-	-	-	-	213	82	61	39	11	5	6	1
	2									420		
C/G	-	-	-	-	16	2	1			19		
AC/GT	1									1		
AG/CT	7									7		
AT/TA	42	5	1							48		
AAC/GTT	4									4		
AAG/CTT	85	149	8							242		
AAT/ATT	38	2								41		
ACG/CGT	4									4		
ACT/AGT	2									2		
AGC/CTG	5									5		
AGG/CCT	2									2		

Powtórzenia	3	4	5	6	7	8	9	10	11	RAZEM
ATC/ATG	7									**7**
AAAC/GTTT	1									**1**
AAAG/CTTT	5									**5**
AAAT/ATTT	2									**2**
AATC/ATTG	2									**2**
AATT/AATT	1									**1**
AAAAT/ATTTT	1									**1**

Suplement 11.

Dystrybucja powtórzeń oligonukleotydów w genomach chloroplastowych *Erica arborea*.

	Długość powtórzenia [pz]	Start	Typ powtórzenia	Długość drugiego powtórzenia	Początek drugiego powtórzenia
Erica arborea	27751	2	**P**	27751	140081
	3321	29286	**P**	3321	135270
	1029	92157	**P**	1029	138945
	724	93591	**F**	724	95094
	690	32281	**F**	690	82276
	402	94003	**P**	402	138423
	393	6724	**P**	393	57178
	393	57178	**F**	393	160717
	362	6430	**F**	362	70736
	362	70736	**P**	362	161042
	326	82276	**P**	326	135270
	312	95506	**P**	312	138513
	304	71927	**F**	304	130444
	286	93462	**F**	286	94416
	283	7131	**F**	283	13900
	283	7131	**P**	283	153651
	283	13900	**P**	283	160420
	283	153651	**F**	283	160420

	Długość powtórzenia [pz]	Start	Typ powtórzenia	Długość drugiego powtórzenia	Początek drugiego powtórzenia
	271	57818	**P**	271	70504
	251	6248	**P**	251	52192
	251	52192	**F**	251	161335
	245	12333	**P**	245	12698
	245	12333	**F**	245	154891
	245	12698	**F**	245	155256
	245	154891	**P**	245	155256
	238	12798	**F**	238	13056
	238	12798	**P**	238	154540
	238	13056	**P**	238	154798
	238	154540	**F**	238	154798
	222	90707	**F**	222	91037
	208	12603	**F**	208	12993
	208	12603	**P**	208	154633
	208	12993	**P**	208	155023
	208	154633	**F**	208	155023
	194	11630	**P**	194	12366
	194	11630	**F**	194	12716
	194	11630	**P**	194	154924
	194	11630	**F**	194	155274
	194	12366	**F**	194	156010
	194	12716	**P**	194	156010
	194	154924	**F**	194	156010

	Długość powtórzenia [pz]	Start	Typ powtórzenia	Długość drugiego powtórzenia	Początek drugiego powtórzenia
	194	155274	**P**	194	156010
	187	11974	**F**	187	12352
	187	11974	**P**	187	12737
	187	11974	**F**	187	154910
	187	11974	**P**	187	155295
	187	12352	**P**	187	155673
	187	12737	**F**	187	155673
	187	154910	**P**	187	155673
	187	155295	**F**	187	155673

Suplement 12.

Dystrybucja powtórzeń oligonukleotydów w genomach chloroplastowych *Erica versicolor*.

	Długość powtórzenia [pz]	Start	Typ powtórzenia	Długość drugiego powtórzenia	Początek drugiego powtórzenia
Erica versicolor	35276	101199	**P**	35276	139046
	706	105304	**F**	706	10763
	706	105304	**P**	706	167177
	706	107638	**P**	706	169511
	706	167177	**F**	706	169511
	701	25121	**P**	701	38257
	445	105818	**F**	445	107991
	445	167085	**P**	445	167085
	445	107991	**P**	445	169258
	445	167085	**F**	445	169258
	397	50192	**F**	397	174190

	Długość powtórzenia [pz]	Start	Typ powtórzenia	Długość drugiego powtórzenia	Początek drugiego powtórzenia
	353	105657	**F**	353	105818
	353	105657	**P**	353	169350
	353	105818	**P**	353	169511
	353	169350	**F**	353	169511
	298	3124	**F**	298	3685
	283	122259	**F**	283	129059
	283	122259	**P**	283	146179
	283	129059	**P**	283	152979
	283	146179	**F**	283	152979
	278	123193	**P**	278	124053
	278	123193	**F**	278	151190
	278	124053	**F**	278	152050
	278	151190	**P**	278	152050
	276	123489	**F**	276	124584

Długość powtórzenia [pz]	Start	Typ powtórzenia	Długość drugiego powtórzenia	Początek drugiego powtórzenia
276	123489	**P**	276	150661
276	124584	**P**	276	151756
276	150661	**F**	276	151756
247	58554	**F**	247	58884
213	19908	**P**	213	145295
213	19908	**F**	213	130013
211	123105	**F**	211	123666
211	123105	**P**	211	151644
211	123666	**P**	211	152205
211	151644	**F**	211	152206
208	24912	**P**	208	38959
201	124501	**F**	201	124885
201	124501	**P**	201	150435
201	124885	**P**	201	150819

	Długość powtórzenia [pz]	Start	Typ powtórzenia	Długość drugiego powtórzenia	Początek drugiego powtórzenia
	201	150435	**F**	201	150819
	195	123239	**F**	195	124496
	195	123239	**P**	195	150830
	195	124090	**P**	195	124496
	195	124090	**F**	195	150830
	195	124496	**F**	195	151236
	195	124496	**P**	195	152087
	195	150830	**P**	195	151236
	195	150830	**F**	195	152087
	193	123924	**F**	193	124311
	193	123924	**P**	193	151017

Suplement 13.

Sekwencje powtórzeń tandemowych w genomie chloroplastowym *Erica arborea*.

Indeksy powtórzeń względem początku sekwencji	Rozmiar okresu powtórzenia [pz]	Liczba kopii dopasowana do wzorca konsensusu	Wielkość formacji konsensusu (może nieznacznie różnić się od wielkości okresu)	Skład procentowy dla każdego z nukleotydów [%]				Region występowania powtórzenia w genomie i jego typ [K-kodujący NK – niekodujący]
				A	C	G	T	
986--1034	20	2.8	17	26	12	6	55	IR: *rpl*32-*trn*L-UAG; NK
10956--10989	17	2.1	17	35	17	23	23	IR: *psa*I-*ndh*I; NK
11071--11114	22	2.0	23	2	22	9	65	IR: *psa*I-*ndh*I; NK
11076--11123	22	2.2	22	0	22	10	66	IR: *psa*I-*ndh*I; NK
11404--11496	21	4.4	21	8	25	10	54	IR: *psa*I-*ndh*I; NK
11382--11502	36	3.6	36	8	24	12	53	IR: *psa*I-*ndh*I; NK
11446--11477	12	2.7	12	0	25	6	68	IR: *psa*I-*ndh*I; NK
11568--11959	21	20.5	21	8	25	10	55	IR: *psa*I-*ndh*I; NK

Indeksy powtórzeń względem początku sekwencji	Rozmiar okresu powtórzenia [pz]	Liczba kopii dopasowana do wzorca konsensusu	Wielkość formacji konsensusu (może nieznacznie różnić się od wielkości okresu)	Skład procentowy dla każdego z nukleotydów [%]				Region występowania powtórzenia w genomie i jego typ [K-kodujący NK – niekodujący]
				A	C	G	T	
11546--11974	36	13.0	36	8	25	11	54	IR: *psa*I-*ndh*I; NK
11334--11636	164	1.9	163	12	21	16	49	IR: *psa*I-*ndh*I; NK
11610--11636	12	2.2	12	0	25	7	66	IR: *psa*I-*ndh*I; NK
11778--11939	9	18.0	9	6	25	9	57	IR: *psa*I-*ndh*I; NK
11619--11966	63	5.3	64	8	25	10	55	IR: *psa*I-*ndh*I; NK
11649--11924	99	2.8	99	7	25	10	55	IR: *psa*I-*ndh*I; NK
11736--11939	6	34.0	6	6	25	10	57	IR: *psa*I-*ndh*I; NK
11975--12578	33	18.8	33	55	11	25	8	IR: *psa*I-*ndh*I; NK
12046--12077	12	2.7	12	68	6	25	0	IR: *psa*I-*ndh*I; NK
11975--12344	36	10.4	35	55	10	25	7	IR: *psa*I-*ndh*I; NK
12009--12077	3	23.0	3	60	8	26	4	IR: *psa*I-*ndh*I; NK

Indeksy powtórzeń względem początku sekwencji	Rozmiar okresu powtórzenia [pz]	Liczba kopii dopasowana do wzorca konsensusu	Wielkość formacji konsensusu (może nieznacznie różnić się od wielkości okresu)	Skład procentowy dla każdego z nukleotydów [%]				Region występowania powtórzenia w genomie i jego typ [K-kodujący NK – niekodujący]
				A	C	G	T	
12058--12593	84	6.5	84	55	10	25	8	IR: *psa*I-*ndh*I; NK
12172--12578	36	12.3	33	55	10	25	8	IR: *psa*I-*ndh*I; NK
11975--12587	132	4.8	131	55	11	25	7	IR: *psa*I-*ndh*I; NK
12559--12602	9	4.9	9	68	6	25	0	IR: *psa*I-*ndh*I; NK
12618--12677	18	3.5	18	10	26	10	53	IR: *psa*I-*ndh*I; NK
12604--13363	33	23.3	33	8	25	10	55	IR: *psa*I-*ndh*I; NK
12690--12721	12	2.7	12	0	25	6	68	IR: *psa*I-*ndh*I; NK
12604--13330	132	5.6	132	8	25	10	55	IR: *psa*I-*ndh*I; NK
12822--12853	12	2.7	12	0	25	6	68	IR: *psa*I-*ndh*I; NK
12690--12890	6	33.5	6	6	25	10	57	IR: *psa*I-*ndh*I; NK
12711--13330	258	2.4	258	8	25	10	55	IR: *psa*I-*ndh*I; NK

Indeksy powtórzeń względem początku sekwencji	Rozmiar okresu powtórzenia [pz]	Liczba kopii dopasowana do wzorca konsensusu	Wielkość formacji konsensusu (może nieznacznie różnić się od wielkości okresu)	Skład procentowy dla każdego z nukleotydów [%]				Region występowania powtórzenia w genomie i jego typ [K-kodujący NK – niekodujący]
				A	C	G	T	
12987--13310	6	54.0	6	7	26	9	56	IR: *psa*I-*ndh*I; NK
13302--13384	18	4.6	18	12	26	10	50	IR: *psa*I-*ndh*I; NK
13248--13364	36	3.3	36	7	27	8	56	IR: *psa*I-*ndh*I; NK
16001--16038	14	2.7	14	42	5	21	31	IR: *ndh*A; K
32579--32636	28	2.2	28	43	3	10	43	IR: *trn*T-UGU-*ycf*15; NK
32587--32644	27	2.2	28	34	3	13	48	IR: *trn*T-UGU-*ycf*15; NK
32973--32998	13	2.0	13	69	0	15	15	IR: *trn*T-UGU-*ycf*15; NK
43908--43943	12	3.0	12	69	8	5	16	LSC: *rps*3; K
50080--50198	21	5.7	21	46	14	21	17	LSC: *clp*P; K
60035--60145	53	2.1	53	55	6	23	14	LSC: *psb*E-*pet*L; NK
72248--72276	15	1.9	15	6	20	10	62	LSC: *clp*P; K

Indeksy powtórzeń względem początku sekwencji	Rozmiar okresu powtórzenia [pz]	Liczba kopii dopasowana do wzorca konsensusu	Wielkość formacji konsensusu (może nieznacznie różnić się od wielkości okresu)	Skład procentowy dla każdego z nukleotydów [%]				Region występowania powtórzenia w genomie i jego typ [K-kodujący NK – niekodujący]
				A	C	G	T	
72250--72287	15	2.5	15	10	18	10	60	LSC: *clp*P; K
72272--72326	21	2.6	21	14	21	9	54	LSC: *clp*P; K
72750--72787	18	2.2	17	2	10	2	84	LSC: *clp*P; K
82574--82631	28	2.2	28	43	3	10	43	LSC: *trn*T-UGU-*trn*H-GAG; NK
82582--82639	27	2.2	28	34	3	13	48	LSC: *trn*T-UGU-*trn*H-GAG; NK
89546--89587	15	2.8	15	16	4	9	69	LSC: *mat*K-*acc*D; NK
90625--90819	102	1.9	102	21	22	16	38	LSC: *mat*K-*acc*D; NK
90787--90843	24	2.4	24	17	19	14	49	LSC: *mat*K-*acc*D; NK
90685--90952	126	2.1	126	20	23	14	42	LSC: *mat*K-*acc*D; NK
90872--91018	39	4.2	38	21	23	15	40	LSC: *acc*D-*trn*T-GGU; NK

Indeksy powtórzeń względem początku sekwencji	Rozmiar okresu powtórzenia [pz]	Liczba kopii dopasowana do wzorca konsensusu	Wielkość formacji konsensusu (może nieznacznie różnić się od wielkości okresu)	Skład procentowy dla każdego z nukleotydów [%]				Region występowania powtórzenia w genomie i jego typ [K-kodujący NK – niekodujący]
				A	C	G	T	
90811--91149	102	3.3	102	21	22	15	40	LSC: *acc*D-*trn*T-GGU; NK
90625--91280	228	2.9	234	21	22	15	39	LSC: *acc*D-*trn*T-GGU; NK
91015--91275	126	2.1	126	21	22	14	41	LSC: *acc*D-*trn*T-GGU; NK
91117--91173	24	2.4	24	17	19	14	49	LSC: *acc*D-*trn*T-GGU; NK
90625--91297	330	2.0	330	21	22	15	39	LSC: *acc*D-*trn*T-GGU; NK
91644--91694	12	4.2	12	9	25	0	64	LSC: *acc*D-*trn*T-GGU; NK
91644--91706	24	2.6	24	11	23	1	63	LSC: *acc*D-*trn*T-GGU; NK
93195--93224	15	2.0	15	66	0	13	20	LSC: *acc*D-*trn*T-GGU; NK

Indeksy powtórzeń względem początku sekwencji	Rozmiar okresu powtórzenia [pz]	Liczba kopii dopasowana do wzorca konsensusu	Wielkość formacji konsensusu (może nieznacznie różnić się od wielkości okresu)	Skład procentowy dla każdego z nukleotydów [%]				Region występowania powtórzenia w genomie i jego typ [K-kodujący NK – niekodujący]
				A	C	G	T	
109506--109543	19	2.0	19	57	5	15	21	LSC: *psa*A-*psb*K; NK
110085--110158	34	2.2	34	33	5	25	35	LSC: *psa*A-*psb*K; NK
111513--111553	20	2.0	20	46	12	9	31	LSC: *trn*S-GGA-*trn*G-UCC; NK
114585--114652	35	1.9	35	32	11	22	33	LSC: *trn*S-GGA-*trn*G-UCC; NK
130325--130404	39	2.1	39	21	5	12	61	LSC: *clp*P; K
130765--130793	15	1.9	15	6	20	10	62	LSC: *clp*P-*ycf*3; NK
130767--130804	15	2.5	15	10	18	10	60	LSC: *clp*P-*ycf*3; NK
130797--130844	21	2.3	21	14	25	10	50	LSC: *clp*P-*ycf*3; NK
138887--138912	3	8.7	3	65	0	0	34	IR: *rps*16-*rrn*16; NK
151797--151834	14	2.7	14	31	21	5	42	IR: *ndh*A; K

Indeksy powtórzeń względem początku sekwencji	Rozmiar okresu powtórzenia [pz]	Liczba kopii dopasowana do wzorca konsensusu	Wielkość formacji konsensusu (może nieznacznie różnić się od wielkości okresu)	Skład procentowy dla każdego z nukleotydów [%]				Region występowania powtórzenia w genomie i jego typ [K-kodujący NK – niekodujący]
				A	C	G	T	
154724--154755	12	2.7	12	68	6	25	0	IR: *ndh*I-*psa*I; NK
154525--154848	6	54.0	6	56	9	26	7	IR: *ndh*I-*psa*I; NK
154687--154848	3	54.0	3	57	9	26	6	IR: *ndh*I-*psa*I; NK
154505--155231	258	2.8	258	55	10	25	8	IR: *ndh*I-*psa*I; NK
154982--155013	12	2.7	12	68	6	25	0	IR: *ndh*I-*psa*I; NK
155114--155145	12	2.7	12	68	6	25	0	IR: *ndh*I-*psa*I; NK
154945--155145	6	33.5	6	57	10	25	6	IR: *ndh*I-*psa*I; NK
155158--155217	18	3.5	18	53	10	26	10	IR: *ndh*I-*psa*I; NK
155077--155145	3	23.0	3	60	8	26	4	IR: *ndh*I-*psa*I; NK
155233--155610	6	63.0	6	7	25	10	56	IR: *ndh*I-*psa*I; NK
155233--155610	9	42.0	9	7	25	10	56	IR: *ndh*I-*psa*I; NK

Indeksy powtórzeń względem początku sekwencji	Rozmiar okresu powtórzenia [pz]	Liczba kopii dopasowana do wzorca konsensusu	Wielkość formacji konsensusu (może nieznacznie różnić się od wielkości okresu)	Skład procentowy dla każdego z nukleotydów [%]				Region występowania powtórzenia w genomie i jego typ [K-kodujący NK – niekodujący]
				A	C	G	T	
155254--155866	33	17.9	36	8	25	11	55	IR: *ndh*I-*psa*I; NK
155230--155860	132	5.0	128	7	25	11	55	IR: *ndh*I-*psa*I; NK
155260--155777	84	6.2	85	8	25	11	55	IR: *ndh*I-*psa*I; NK
155758--155789	12	2.7	12	0	25	6	68	IR: *ndh*I-*psa*I; NK
155516--155727	72	2.7	83	8	25	10	55	IR: *ndh*I-*psa*I; NK
155758--155826	3	23.0	3	4	26	8	60	IR: *ndh*I-*psa*I; NK
155876--156267	21	20.5	21	55	10	25	8	IR: *ndh*I-*psa*I; NK
155909--156289	63	5.9	63	55	11	24	8	IR: *ndh*I-*psa*I; NK
155861--156246	36	11.4	35	55	10	25	8	IR: *ndh*I-*psa*I; NK
155911--156186	99	2.8	99	55	10	25	7	IR: *ndh*I-*psa*I; NK
155896--156099	6	34.0	6	57	10	25	6	IR: *ndh*I-*psa*I; NK

Indeksy powtórzeń względem początku sekwencji	Rozmiar okresu powtórzenia [pz]	Liczba kopii dopasowana do wzorca konsensusu	Wielkość formacji konsensusu (może nieznacznie różnić się od wielkości okresu)	Skład procentowy dla każdego z nukleotydów [%]				Region występowania powtórzenia w genomie i jego typ [K-kodujący NK – niekodujący]
				A	C	G	T	
156199--156225	12	2.2	12	66	7	25	0	IR: *ndh*I-*psa*I; NK
156339--156431	21	4.4	21	54	10	25	8	IR: *ndh*I-*psa*I; NK
156358--156389	12	2.7	12	68	6	25	0	IR: *ndh*I-*psa*I; NK
156199--156501	164	1.9	163	49	16	21	12	IR: *ndh*I-*psa*I; NK
156721--156764	22	2.0	23	65	9	22	2	IR: *ndh*I-*psa*I; NK
156846--156879	17	2.1	17	23	23	17	35	IR: *ndh*I-*psa*I; NK
166801--166838	20	1.9	19	55	5	13	26	IR: *trn*L-UAG-*rpl*32; NK

Suplement 14.

Sekwencje powtórzeń tandemowych w genomie chloroplastowym *Erica versicolor*.

Indeksy powtórzeń względem początku sekwencji	Rozmiar okresu powtórzenia [pz]	Liczba kopii dopasowana do wzorca konsensusu	Wielkość formacji konsensusu (może nieznacznie różnić się od wielkości okresu)	Skład procentowy dla każdego z nukleotydów [%]				Region występowania powtórzenia w genomie i jego typ [K- kodujący NK – niekodujący]
				A	C	G	T	
161--186	13	2.0	13	69	0	15	15	LSC: *trn*T-*ycf*15; NK
11573--11608	12	3.0	12	69	8	5	16	LSC*: rps*3; K
17755--17873	21	5.7	21	47	14	22	15	LSC: *clp*P; K
24614--24719	34	3.1	34	32	2	14	50	LSC: *psb*B-*ycf*1; NK
24830--24857	10	2.8	10	75	10	0	14	LSC: *psb*B-*ycf*1; NK
27773--27883	53	2.1	53	56	5	23	14	LSC: *psb*E-*pet*L; NK
34835--34864	2	15.0	2	46	0	0	53	LSC: *trn*C-GCA-*pet*N; NK
40029--40075	21	2.2	20	12	23	12	51	LSC: *clp*P-*atp*E; NK
40115--40165	9	5.7	9	9	13	3	72	LSC: *clp*P-*atp*E; NK
40115--40166	18	2.9	18	9	15	3	71	LSC: *clp*P-*atp*E; NK

Indeksy powtórzeń względem początku sekwencji	Rozmiar okresu powtórzenia [pz]	Liczba kopii dopasowana do wzorca konsensusu	Wielkość formacji konsensusu (może nieznacznie różnić się od wielkości okresu)	Skład procentowy dla każdego z nukleotydów [%]				Region występowania powtórzenia w genomie i jego typ [K- kodujący NK – niekodujący]
				A	C	G	T	
50362--50412	27	1.9	28	35	5	9	49	LSC: *trn*T-UGU-*trn*H-GUG; NK
50364--50420	27	2.1	28	33	5	12	49	LSC: *trn*T-UGU-*trn*H-GUG; NK
58492--58621	63	2.1	63	20	24	13	41	LSC: *acc*D-*trn*T-GGU; NK
58504--58635	63	2.1	63	19	27	14	38	LSC: *acc*D-*trn*T-GGU; NK
58557--58824	126	2.1	126	20	24	14	41	LSC: *acc*D-*trn*T-GGU; NK
58659--58722	24	2.7	24	15	23	10	50	LSC: *acc*D-*trn*T-GGU; NK
58683--59028	102	3.4	102	20	24	14	39	LSC: *acc*D-*trn*T-GGU; NK
58887--59154	126	2.1	126	20	24	14	40	LSC: *acc*D-*trn*T-GGU; NK
58557--59161	228	2.7	228	21	24	15	39	LSC: *acc*D-*trn*T-GGU; NK
58989--59052	24	2.7	24	15	23	10	50	LSC: *acc*D-*trn*T-GGU; NK
58555--59161	330	1.8	330	21	24	15	39	LSC: *acc*D-*trn*T-GGU; NK

Indeksy powtórzeń względem początku sekwencji	Rozmiar okresu powtórzenia [pz]	Liczba kopii dopasowana do wzorca konsensusu	Wielkość formacji konsensusu (może nieznacznie różnić się od wielkości okresu)	Skład procentowy dla każdego z nukleotydów [%]				Region występowania powtórzenia w genomie i jego typ [K- kodujący NK – niekodujący]
				A	C	G	T	
59516--59672	12	13.1	12	8	24	0	66	LSC: *acc*D-*trn*T-GGU; NK
61508--61532	10	2.5	10	52	0	0	48	LSC: *trn*T-GGU-*trn*T-GGU; NK
74873--74911	19	2.1	19	58	0	15	25	LSC: *psa*A-*trn*Q-UUG; NK
75454--75527	34	2.2	34	33	5	25	35	LSC: *trn*Q-UUG – *psb*K; NK
79971--80038	35	1.9	35	33	11	22	32	LSC: *atp*A-*atp*F; NK
96447--96493	24	2.0	24	25	6	10	57	LSC: *clp*P; K
96884--96939	21	2.7	20	16	25	7	51	LSC: *clp*P-*ycf*3; NK
97657--97750	31	3.0	31	29	24	7	38	LSC: *clp*P-*ycf*3; NK
102701--102731	15	2.1	15	64	0	0	35	IR: *trn*M-CAU-*rps*16; NK
105087--105181	39	2.4	39	51	13	7	27	IR: *rps*16-*rrn*16; NK
105853--106161	161	1.9	161	31	11	26	30	IR: *rps*16-*rrn*16; NK

Indeksy powtórzeń względem początku sekwencji	Rozmiar okresu powtórzenia [pz]	Liczba kopii dopasowana do wzorca konsensusu	Wielkość formacji konsensusu (może nieznacznie różnić się od wielkości okresu)	Skład procentowy dla każdego z nukleotydów [%]				Region występowania powtórzenia w genomie i jego typ [K- kodujący NK – niekodujący]
				A	C	G	T	
107982--108334	161	2.2	161	33	11	26	28	IR: *rps*16-*rrn*16; NK
120399--120436	14	2.7	14	31	21	5	42	IR: *ndh*A; K
123050--123193	36	4.2	36	55	10	25	8	IR: *ndh*I-*psa*I; NK
123171--123202	12	2.7	12	68	6	25	0	IR: *ndh*I-*psa*I; NK
123050--123866	33	24.6	35	54	11	25	8	IR: *ndh*I-*psa*I; NK
123050--123361	105	3.1	98	55	10	25	8	IR: *ndh*I-*psa*I; NK
123068--123850	99	7.9	100	55	10	25	8	IR: *ndh*I-*psa*I; NK
123050--123866	132	6.2	131	54	11	25	8	IR: *ndh*I-*psa*I; NK
123732--123763	12	2.7	12	68	6	25	0	IR: *ndh*I-*psa*I; NK
123776--123835	18	3.5	18	53	10	26	10	IR: *ndh*I-*psa*I; NK
123695 --123763	3	23.0	3	60	8	26	4	IR: *ndh*I-*psa*I; NK

Indeksy powtórzeń względem początku sekwencji	Rozmiar okresu powtórzenia [pz]	Liczba kopii dopasowana do wzorca konsensusu	Wielkość formacji konsensusu (może nieznacznie różnić się od wielkości okresu)	Skład procentowy dla każdego z nukleotydów [%]				Region występowania powtórzenia w genomie i jego typ [K- kodujący NK – niekodujący]
				A	C	G	T	
123859--124493	36	18.2	36	8	25	11	54	IR: *ndh*I-*psa*I; NK
123911--124488	99	6.0	96	8	25	10	55	IR: *ndh*I-*psa*I; NK
124040--124237	9	22.0	9	7	25	10	57	IR: *ndh*I-*psa*I; NK
123875--124493	135	4.7	135	8	25	11	55	IR: *ndh*I-*psa*I; NK
123998--124237	6	40.0	6	7	25	10	57	IR: *ndh*I-*psa*I; NK
124280--124486	33	6.4	33	7	25	10	56	IR: *ndh*I-*psa*I; NK
124385--124416	12	2.7	12	0	25	6	68	IR: *ndh*I-*psa*I; NK
124112--124488	84	4.5	84	7	25	10	55	IR: *ndh*I-*psa*I; NK
124486--125064	36	16.8	35	55	11	25	8	IR: *ndh*I-*psa*I; NK
124538--125075	99	5.6	99	55	11	25	8	IR: *ndh*I-*psa*I; NK
124523--125011	6	81.5	6	56	11	25	7	IR: *ndh*I-*psa*I; NK

Indeksy powtórzeń względem początku sekwencji	Rozmiar okresu powtórzenia [pz]	Liczba kopii dopasowana do wzorca konsensusu	Wielkość formacji konsensusu (może nieznacznie różnić się od wielkości okresu)	Skład procentowy dla każdego z nukleotydów [%]				Region występowania powtórzenia w genomie i jego typ [K- kodujący NK – niekodujący]
				A	C	G	T	
125343--125386	22	2.0	23	65	9	22	2	IR: *ndh*I-*psa*I; NK
125468--125501	17	2.1	17	23	23	17	35	IR: *ndh*I-*psa*I; NK
135430--135467	20	1.9	19	55	5	13	26	IR: *trn*L-UAG-*rpl*32; NK
136476--136516	17	2.5	16	26	9	0	63	SSC: *rpl*32-*ndh*F; NK
140022--140070	20	2.8	17	26	12	6	55	IR: *rpl*32-*trn*L-UAG; NK
149999--150032	17	2.1	17	35	17	23	23	IR: *psa*I-*ndh*I; NK
150114--150157	22	2.0	23	2	22	9	65	IR: *psa*I-*ndh*I; NK
150119--150166	22	2.2	22	0	22	10	66	IR: *psa*I-*ndh*I; NK
150425--151014	36	17.1	36	8	24	11	55	IR: *psa*I-*ndh*I; NK
150425--150962	99	5.6	99	8	25	11	55	IR: *psa*I-*ndh*I; NK
150531--150977	9	49.7	9	7	25	10	56	IR: *psa*I-*ndh*I; NK

Indeksy powtórzeń względem początku sekwencji	Rozmiar okresu powtórzenia [pz]	Liczba kopii dopasowana do wzorca konsensusu	Wielkość formacji konsensusu (może nieznacznie różnić się od wielkości okresu)	Skład procentowy dla każdego z nukleotydów [%]				Region występowania powtórzenia w genomie i jego typ [K- kodujący NK – niekodujący]
				A	C	G	T	
150468--150892	117	3.5	117	7	25	11	56	IR: *psa*I-*ndh*I; NK
150489--150977	6	81.5	6	7	25	11	56	IR: *psa*I-*ndh*I; NK
151084--151115	12	2.7	12	68	6	25	0	IR: *psa*I-*ndh*I; NK
151007--151641	36	18.2	35	54	11	25	8	IR: *psa*I-*ndh*I; NK
151098--151141	21	2.1	21	56	9	27	6	IR: *psa*I-*ndh*I; NK
151011--151115	3	35.0	3	59	9	25	5	IR: *psa*I-*ndh*I; NK
151179--151641	99	4.6	98	55	11	25	8	IR: *psa*I-*ndh*I; NK
151083--151424	84	4.1	84	55	10	25	7	IR: *psa*I-*ndh*I; NK
151263--151502	6	40.0	6	57	10	25	7	IR: *psa*I-*ndh*I; NK
151012--151628	84	7.2	83	55	10	25	8	IR: *psa*I-*ndh*I; NK
151007--151625	135	4.7	133	55	11	25	8	IR: *psa*I-*ndh*I; NK

Indeksy powtórzeń względem początku sekwencji	Rozmiar okresu powtórzenia [pz]	Liczba kopii dopasowana do wzorca konsensusu	Wielkość formacji konsensusu (może nieznacznie różnić się od wielkości okresu)	Skład procentowy dla każdego z nukleotydów [%]				Region występowania powtórzenia w genomie i jego typ [K- kodujący NK – niekodujący]
				A	C	G	T	
151634--152450	36	24.6	36	8	25	11	54	IR: *psa*I-*ndh*I; NK
151737--151768	12	2.7	12	0	25	6	68	IR: *psa*I-*ndh*I; NK
151683--152419	63	11.7	63	8	25	10	55	IR: *psa*I-*ndh*I; NK
151650--152419	99	7.8	99	8	25	10	55	IR: *psa*I-*ndh*I; NK
151737--151805	3	23.0	3	4	26	8	60	IR: *psa*I-*ndh*I; NK
151634--152450	132	6.2	134	8	25	11	54	IR: *psa*I-*ndh*I; NK
152073--152366	9	32.7	9	7	25	10	56	IR: *psa*I-*ndh*I; NK
152298--152329	12	2.7	12	0	25	6	68	IR: *psa*I-*ndh*I; NK
152139--152450	105	3.1	102	8	25	10	55	IR: *psa*I-*ndh*I; NK
155064--155101	14	2.7	14	42	5	21	31	IR: *ndh*A; K
167166--167518	161	2.2	161	28	26	11	33	IR: *rrn*16-*rps*16; NK

Indeksy powtórzeń względem początku sekwencji	Rozmiar okresu powtórzenia [pz]	Liczba kopii dopasowana do wzorca konsensusu	Wielkość formacji konsensusu (może nieznacznie różnić się od wielkości okresu)	Skład procentowy dla każdego z nukleotydów [%]				Region występowania powtórzenia w genomie i jego typ [K- kodujący NK – niekodujący]
				A	C	G	T	
169339--169852	161	3.2	161	29	26	11	32	IR: *rrn*16-*rps*16; NK
170319--170413	39	2.4	39	27	7	13	51	IR: *rrn*16-*rps*16; NK
172769--172799	15	2.1	15	35	0	0	64	IR: *rps*16-*trn*M-CAU; NK
174341--174391	27	1.9	28	35	5	9	49	LSC:*trn*T-UGU-*ycf*15; NK
174343--174399	27	2.1	28	33	5	12	49	LSC: *trn*T-UGU-*ycf*15; NK

Suplement 15.

Zestawienie polimorfizmów w sekwencjach odpowiadających dojrzałym cząsteczkom tRNA i rRNA pomiędzy badanymi genomami *Erica arborea* a *Erica versicolor*.

Długość sekwencji w genomie [pz], SNP – liczba substytucji, InDel – liczba insercji/delecji, p – odległość genetyczna.

Wykluczono z tabeli geny identyczne pomiędzy genomami.

Nazwa genu	***Erica arborea* [pz]**	***Erica vesricolor* [pz]**	**Liczba SNP**	**Liczba InDel**	**Odległość genetyczna (p)**
tRNA					
trnI-CAU	74	74	1	0	0,01388
trnY-GUA	84	84	1	0	0,01198
trnV-UAC	88	88	2	0	0,00303
trnL-UAA	75	75	5	0	0,08051
trnT-UGU	73	73	4	0	0,05769
trnfM-CAU	67	67	1	0	0,01539
trnG-UCC	71	71	4	0	0,05032
trnT-GGU	73	71	4	2	0,02009
rRNA					
*rrn*16	1493	1492	3	1	0,03223
*rrn*23	2803	2803	5	0	0,00179

Suplement 16.

Analiza zmiennych miejsc w genomach chloroplastów *Erica*.

	Liczba miejsc	**Miejsca zmienne**		**Miejsca informatywne**		**Zróżnicowanie nukleotydowe**
		Liczba	**%**	**Liczba**	**%**	
LSC	104 460	1394	1,33%	1394	1,33%	0,01493
SSC	2 571	46	1,78%	46	1,78%	0,00209
IR	35 596	3857	10,9%	3857	10,9%	0,11962
Kompletny genomom chloroplastowy	174 587	9160	5,25%	9157	5,23%	0,01386