UNIWERSYTET
WARMIŃSKO-MAZURSKI
W OLSZTYNIE

WYDZIAŁ: Biologii i Biotechnologii
Dziedzina: Nauki ścisłe i przyrodnicze
Dyscyplina: Nauki biologiczne

mgr inż. Sylwia Eryka Milarska

Rozprawa doktorska

Genetyczna charakterystyka wybranych alpejskich i arktyczno-alpejskich gatunków z rodzaju *Cerastium* L. (Caryophyllaceae)

Genetic characteristics of selected alpine and arctic-alpine species from genus *Cerastium* L. (Caryophyllaceae)

Praca wykonana w Katedrze Fizjologii, Genetyki i Biotechnologii Roślin

Promotor: dr hab. Piotr Androsiuk, Katedra Fizjologii, Genetyki i Biotechnologii Roślin, Wydział Biologii i Biotechnologii, Uniwersytet Warmińsko-Mazurski w Olsztynie

Olsztyn, 2024

SPIS TREŚCI

STRESZCZENIE ... 3

ABSTRACT ... 4

1. WYKAZ ORYGINALNYCH PRAC STANOWIĄCYCH PRZEDMIOT ROZPRAWY DOKTORSKIEJ ... 5

2. WSTĘP ... 6

3. HIPOTEZY I ZADANIA BADAWCZE ... 11

4. MATERIAŁY I METODY ... 13

4.1. Materiał roślinny ... 13

4.2. Pochodzenie materiału roślinnego ... 13

4.3. Analiza zmienności genetycznej *C. alpinum* z Babiej Góry ... 14

4.4. Uzyskanie kompletnych sekwencji genomów chloroplastowych *C. alpinum*, *C. arcticum* i *C. nigrescens* ... 16

5. OMÓWIENIE PUBLIKACJI STANOWIĄCYCH ROZPRAWĘ DOKTORSKĄ ... 19

5.1. Publikacja nr 1 ... 19

5.1.1. Ocena skuteczności zastosowanych markerów iPBS ... 19

5.1.2. Zróżnicowanie genetyczne analizowanych populacji *C. alpinum* ... 20

5.2. Publikacja nr 2 ... 21

5.2.1. Struktura i organizacja genomu chloroplastowego ... 22

5.2.2. Powtórzenia genomowe i sekwencje mikrosatelitarne ... 23

5.2.3. Substytucje synonimiczne (Ks) i niesynonimiczne (Ka) ... 24

5.2.4. Analizy porównawcze genomów chloroplastowych ... 25

5.2.5. Miejsca edycji RNA ... 26

5.2.6. Analizy filogenetyczne ... 27

6. PODSUMOWANIE I WNIOSKI ... 29

6.1. Zmienność genetyczna *C. alpinum* z Babiej Góry ... 29

6.2. Charakterystyka i analiza porównawcza genomów chloroplastowych *C. alpinum*, *C. arcticum* i *C. nigrescens* ... 29

7. LITERATURA ... 31

8. SPIS TABEL ... 36

9. KOPIE PRAC WCHODZĄCYCH W SKŁAD ROZPRAWY DOKTORSKIEJ ... 37

10. OŚWIADCZENIA WSPÓŁAUTORÓW ... 76

STRESZCZENIE

Rodzaj *Cerastium* (Caryophyllaceae) reprezentowany jest przez około 200 gatunków. *Cerastium alpinum* to gatunek arktyczno-alpejski występujący w północnej części Ameryki Północnej i Europy, notowany w piętrze subalpejskim, także w Polsce, na jedynym stanowisku na Babiej Górze (Beskid Żywiecki, Karpaty Zachodnie). Mimo, że *C. alpinum* należy w Polsce do gatunków krytycznie zagrożonych, dotychczas brakowało danych na temat zmienności genetycznej jego jedynego stanowiska w kraju. Badania wchodzące w skład niniejszej rozprawy doktorskiej wykazały, że pomimo izolacji geograficznej oraz doświadczenia w przeszłości silnej redukcji efektywnej wielkości populacji *C. alpinum* z Babiej Góry nadal stanowi ważny w skali gatunku rezerwuar zmienności genetycznej.

Relacje taksonomiczne w obrębie rodzaju *Cerastium* wciąż są przedmiotem dyskusji. Na szczególną uwagę zasługuje kompleks *C. alpinum – C. arcticum,* w którym wyróżnia się gatunki: *C. alpinum, C. arcticum* i *C. nigrescens*. W celu uzyskania danych umożliwiających jednoznaczną identyfikację tych gatunków oraz aby rzucić więcej światła na powiązania filogenetyczne pomiędzy nimi przeprowadzono sekwencjonowanie ich genomów chloroplastowych. Uzyskane dane wraz z już dostępnymi w bazie NCBI informacjami dla *C. glomeratum* i *C. arvense* wykorzystano do scharakteryzowania wielkości i struktury genomów, identyfikacji powtórzeń genomowych i mikrosatelitarnych oraz analizy ewolucji sekwencji kodujących białka. Dane genomiczne posłużyły także do rekonstrukcji relacji filogenetycznych w obrębie rodzaju *Cerastium* i rodziny Caryophyllaceae.

Słowa kluczowe: *Cerastium,* region arktyczno-alpejski, Babia Góra, zróżnicowanie genetyczne, genomy chloroplastowe

ABSTRACT

The genus *Cerastium* (Caryophyllaceae) consists of over 200 species. *Cerastium alpinum* is an arctic-alpine species that can be found in the northern part of North America and Europe, observed in the subalpine zone, also in Poland, where its only site of occurrence is on the Babia Góra (Żywiec Beskids, Western Carpathians). Although *C. alpinum* is a critically endangered species in Poland, so far there has been no data on the genetic variability of the species from its only Polish location. The results of the studies included in this doctoral dissertation have shown that despite geographical isolation and the past experience of a strong reduction of the effective population size, *C. alpinum* from Babia Góra is still an important reservoir of genetic variability on the species scale.

Taxonomic relationships within the genus *Cerastium* are still a subject of debate. The *C. alpinum* – *C. arcticum* complex, which consists of the *C. alpinum*, *C. arcticum* and *C. nigrescens* deserves special attention here. In order to obtain data enabling the unambiguous identification of the species and to shed more light on their phylogenetic relationships, the chloroplast genomes were sequenced. The obtained genomic data, together with information already available in the NCBI database for *C. glomeratum* and *C. arvense*, were used to characterize the size and structure of the chloroplast genomes, identify genomic repeats and microsatellites, and analyze the evolution of protein-coding sequences. The genomic data were also used to reconstruct phylogenetic relationships within the genus *Cerastium* and the family Caryophyllaceae.

Key words: *Cerastium*, arctic-alpine region, Babia Góra, genetic diversity, chloroplast genomes

1. WYKAZ ORYGINALNYCH PRAC STANOWIĄCYCH PRZEDMIOT ROZPRAWY DOKTORSKIEJ

Publikacja nr 1	**Milarska SE**, Androsiuk P, Bednarek PT, Larson K, Giełwanowska I. 2023. Genetic variation of *Cerastium alpinum* L. from Babia Góra, a critically endangered species in Poland. *Journal of Applied Genetics* 64: 37-53. DOI: 10.1007/s13353-022-00731-x **IF_{2022} – 2.400; MEiN – 140 pkt.**
Publikacja nr 2	**Milarska SE**, Androsiuk P, Paukszto Ł, Jastrzębski JP, Maździarz M, Larson K, Giełwanowska I. 2023. Complete chloroplast genomes of *Cerastium alpinum*, *C. arcticum* and *C. nigrescens*: genome structures, comparative and phylogenetic analysis. *Scientific Reports* 13: 18774. DOI: 10.1038/s41598-023-46017-y **IF_{2022} – 4.600; MEiN – 140 pkt.**

Sumaryczny IF – 7.000

Sumaryczna liczba punktów MEiN – 280 pkt.

2. WSTĘP

Przedmiotem niniejszej rozprawy doktorskiej jest genetyczna charakterystyka wybranych alpejskich i arktyczno-alpejskich gatunków z rodzaju *Cerastium*: *C. alpinum* L., *C. arcticum* Lange i *C. nigrescens* (H.C. Watson). Rodzaj *Cerastium* należący do rodziny goździkowatych (Caryophyllaceae) reprezentowany jest przez około 200 gatunków (The World Flora Online). Większość z nich występuje na półkuli północnej, w strefie klimatu umiarkowanego, a jedynie kilka z nich to gatunki typowo kosmopolityczne. W Polsce odnotowano 17 gatunków tego rodzaju, ale na szczególną uwagę zasługuje *C. alpinum*, uważany w Polsce za gatunek krytycznie zagrożony, wpisany do Polskiej Czerwonej Księgi Roślin (Parusel 2001, 2014), którego jedyne stanowiska znajdują się na Babiej Górze (Beskid Żywiecki, Karpaty Zachodnie) na obszarze Babiogórskiego Parku Narodowego (Parusel 2001). Za przyczynę zagrożenia tego gatunku w Polsce uważa się niewielki obszar występowania oraz przestrzenną izolację stanowiska (najbliższe stanowiska położone są w Alpach i Karpatach Wschodnich).

Aby zaplanować skuteczną strategię ochrony danej populacji, konieczna jest znajomość jej aktualnej kondycji genetycznej. Na zmienność genetyczną populacji ma wpływ wiele czynników, m.in. wielkość populacji, przepływ genów pomiędzy nimi, system kojarzenia, ale również takie zjawiska jak dobór naturalny i dryf genetyczny (Ellstrand i Elam 1993). W małych populacjach, w których obserwuje się wzrost częstości kojarzeń krewniaczych powstają warunki sprzyjające pojawieniu się depresji wsobnej. Zjawisko to może prowadzić do akumulacji szkodliwych mutacji, spadku dostosowania osobników i dalszej redukcji wielkości populacji (Ellstrand i Elam 1993; Lynch i in. 1995; Young i in. 1996). Kolejnym czynnikiem prowadzącym do degradacji genetycznej populacji jest jej izolacja przestrzenna, która może ograniczać lub nawet uniemożliwiać przepływ genów. Ze względu na ograniczoną liczebność oraz izolację geograficzną stanowiska *C. alpinum* z Babiej Góry mogą doświadczać wymienionych powyżej zagrożeń i być narażone na silną erozję genetyczną, a w konsekwencji również na niebezpieczeństwo wyginięcia. Dodatkowym zagrożeniem dla *C. alpinum* z Babiej Góry mogą być postępujące globalne zmiany klimatyczne. Obserwowane ocieplenie klimatu naraża rośliny występujące na większych wysokościach (piętro alpejskie)

na rosnącą konkurencję ze strony gatunków występujących dotychczas na niższych piętrach roślinności (Lenoir i in. 2008; Chen i in. 2011; Rumpf i in. 2018). Badania prowadzone w Tatrach na przestrzeni ostatnich 90 lat wykazały zmiany składu gatunkowego tego obszaru, właśnie na skutek zmian zasięgu występowania gatunków z poszczególnych pięter roślinności. Za najbardziej prawdopodobne przyczyny tego zjawiska uważa się podwyższoną temperaturę i zmiany w intensywności opadów (Czortek i in. 2018). Nie bez znaczenia dla ochrony *C. alpinum* jest również intensywnie rozwijający się ruch turystyczny, gdyż wraz ze wzrostem liczby turystów odwiedzających Babią Górę wzrasta niebezpieczeństwo zniszczenia jedynych w Polsce stanowisk tego gatunku.

Dotychczasowa wiedza na temat różnorodności genetycznej poszczególnych gatunków z rodzaju *Cerastium* jest bardzo ograniczona i opiera się na niewielkiej liczbie badań. Dotychczasowe prace obejmują głównie analizy z wykorzystaniem techniki AFLP dla *C. dinaricum* (Kutnjak i in. 2014), *C. decalvans* (Niketić i in. 2022), *C. sylvaticum, C. subtriflorum* (Skubic i in. 2018), *C. hekuravense* (Caković i in. 2018) oraz *C. grandiflorum, C. decalvans* i *C. dinaricum* (Đurović i in. 2021) oraz w mniejszym stopniu markerów RAPD i SCAR dla *C. arcticum* (Hagen i in. 2001). Z kolei badania z zakresu cytogenetyki wykazały spore zróżnicowanie liczby chromosomów pomiędzy wybranymi gatunkami *Cerastium*: *C. eriophorum* (2n = 36), *C. alpinum* (2n = 72), *C. transsylvanicum* (2n = 108), *C. arcticum* (2n = 108), *C. latifolium* (2n = 36), *C. carinthiacum* (2n = 36), *C. banaticum* (2n = 36), *C. arvense* ssp. *glandulosum* (2n = 36), *C. arvense* ssp. *arvense* (2n = 72) i *C. fontanum* (2n = 144) (Boşcaiu i in, 1999).

W przypadku *C. alpinum* większość badań molekularnych dotyczyła populacji występujących na Svalbardzie, Grenlandii, Islandii, Norwegii, Szwecji oraz Finlandii i opierała się na analizach polimorfizmu izoenzymatycznego. Badacze skupili się na identyfikacji i porównaniu fenotypów izoenzymatycznych z różnych populacji w celu opisania ich rozmieszczenia geograficznego i wyciągnięcia wniosków na temat historii gatunku (szlaków kolonizacji) w Fennoskandii, a także na problematyce hybrydyzacji i introgresji pomiędzy *C. alpinum* a innymi blisko spokrewnionymi gatunkami (*C. arcticum, C. nigrescens*) (Brysting i Borgen 2000; Hagen i in. 2002; Nyberg Berglund i in. 2006). Ponadto badania Scheen i in. (2004) pozwoliły na rekonstrukcję dróg rozprzestrzeniania się oraz analizę powiązań filogenetycznych

57 gatunków z rodzaju *Cerastium* w oparciu o sekwencje trzech niekodujących regionów plastydowego DNA. Wyniki sugerują, że gatunki z rodzaju *Cerastium* pochodzą ze Starego Świata i skolonizowały Amerykę Północną wskutek co najmniej dwóch zdarzeń migracyjnych. Ostatnia migracja do Ameryki Północnej, prawdopodobnie przez północny Atlantyk, miała miejsce w czwartorzędzie i doprowadziła do obecnego okołobiegunowego rozmieszczenia arktycznych gatunków *Cerastium*.

Pomimo swojego wyjątkowego charakteru jedyne polskie stanowisko *C. alpinum* na Babiej Górze nie było do tej pory przedmiotem badań mających na celu poznanie poziomu zmienności genetycznej. Analiza zmienności genetycznej tego stanowiska jest kluczowa do opracowania skutecznego programu ochrony tego gatunku w Polsce, a także stanowić będzie punkt wyjścia do monitorowania jego kondycji genetycznej w przyszłości.

Retrotranspozony występują powszechnie we wszystkich genomach eukariotycznych. W genomach roślinnych mogą stanowić od kilkunastu (*Arabidopsis thaliana*) do nawet 80% (*Hordeum vulgare*) materiału genetycznego (The Arabidopsis Genome Initiative 2000; Mascher i in. 2017). Ze względu na bogactwo retrotranspozonów oraz ich równomierne rozmieszczenie w genomie znalazły one zastosowanie jako źródło markerów molekularnych. W oparciu o nie opracowano m.in. system markerów iPBS (ang. *inter primer binding sites*), opisany przez Kalendara i in. (2010). Markery iPBS zostały z powodzeniem zastosowane do analizy zmienności genetycznej wielu gatunków roślin m.in. ciecierzycy (Andeden i in. 2013), grochu (Baloch i in. 2015), komosy ryżowej (Hossein-Pour i in. 2019), winorośli (Milovanov i in. 2019), cebuli (Khapilina i in. 2021a, 2021b), a także grzybów chorobotwórczych z rodzaju *Alternaria* (Turzhanova i in. 2020) i *Rhizoctonia* (Erper i in. 2021). Analogiczne badania przeprowadzone zostały z powodzeniem również przez nasz Zespół badawczy. Obiektem analiz tych prac były wybrane gatunki reprezentujące przedstawicieli rodziny Caryophyllaceae i Poaceae z obszarów polarnych – *Colobanthus quitensis* (Koc i in. 2018), *Poa annua* (Androsiuk i in. 2019) i *Deschampsia antarctica* (Androsiuk i in. 2020). Zastosowane markery pozwoliły skutecznie ocenić polimorfizm DNA, zarówno między osobnikami, jak i populacjami badanych gatunków. Nasze dotychczasowe doświadczenie wskazywało, że markery

iPBS mogą stanowić także odpowiednie narzędzie do poznania zmienności genetycznej *C. alpinum*.

Rodzaj *Cerastium* jest jedną z grup roślin, w przypadku której pozycja systematyczna poszczególnych gatunków nadal jest przedmiotem dyskusji. Jednym z przykładów budzących kontrowersje wśród systematyków jest kompleks *C. alpinum* – *C. arcticum*. Według części badaczy w obrębie tego kompleksu wyróżnia się wiele gatunków, podgatunków i odmian (m.in. Hultén 1956; Böcher 1977; Rønning 1996). Inni autorzy traktują kompleks jako jeden gatunek *C. alpinum* o dużym poziomie zmienności z lub bez wyróżniania taksonów niższego rzędu (Hooker 1884; Porsild i Cody 1980). Niemniej jednak, większość autorów wyróżnia w tej grupie dwa gatunki *C. alpinum* oraz *C. arcticum* (m.in. Clapham i in. 1962; Böcher 1977; Elven i Elvebakk 1996). Najbardziej problematycznym składnikiem tej grupy okazuje się *C. arcticum*. Najnowsze badania genetyczne i morfologiczne sugerują, że gatunek ten w rzeczywistości składa się z dwóch odrębnych taksonów: *C. arcticum* s. str. i *C. nigrescens* (Brysting i Hagen 1999; Brysting i Borgen 2000; Brysting i Elven 2000). Obszar występowania pierwszego z nich ogranicza się do obszarów kanadyjskiej Arktyki, Grenlandii, Svalbardu i północno-zachodnich wysp rosyjskiej Arktyki. Drugi natomiast jest charakterystyczny dla Wysp Brytyjskich, gór Fennoskandii, Wysp Owczych i Islandii. Niestety obserwowana duża zmienność w ramach tych gatunków powoduje trudności w ich identyfikacji. Istnieje zatem potrzeba badań, które ułatwią identyfikację i pozwolą wyjaśnić status poszczególnych taksonów. Aby wyjaśnić powiązania pomiędzy powyższymi gatunkami potrzebne są badania dające wgląd w ewolucję i historię rodzaju *Cerastium* oraz umożliwiające znalezienie uniwersalnych markerów do identyfikacji taksonów.

Rozwój i upowszechnienie technologii wysokoprzepustowego sekwencjonowania DNA pozwoliły na uzyskanie wysokiej jakości danych genomowych i bardzo dynamiczny wzrost liczby sekwencji, reprezentujących genomy jądrowe i organellowe, publikowanych w ogólnodostępnych bazach danych. W przypadku genetyki roślin, genomy chloroplastowe stały się źródłem danych powszechnie wykorzystywanych w badaniach porównawczych, biotechnologii, identyfikacji gatunków czy w analizach filogenetycznych (m.in. Jansen i in. 2007; Bortiri i in. 2008; Bock 2007; Kress i in. 2005; Chase i in. 2007; Raman i Park 2016; Zhai i in. 2023). Dotychczas w bazie NCBI dostępne były tylko dwie sekwencje

genomów chloroplastowych dla rodzaju *Cerastium* – kompletna sekwencja genomu chloroplastowego dla *C. glomeratum* (NC_066897) i częściowa sekwencja tego genomu dla *C. arvense* (MH627219). Z dostępnych danych wynika, że genom chloroplastowy w rodzaju *Cerastium* zachował klasyczną czterodzielną strukturę o wielkości i składzie genetycznym typowym dla okrytonasiennych. Oprócz wyżej wspomnianego *C. glomeratum*, kompletne sekwencje plastomów w bazie NCBI dostępne były dla 60 innych przedstawicieli rodziny Caryophyllaceae (stan na dzień 24 marca 2023 r., Publikacja nr 2) reprezentujących rodzaje: *Agrostemma*, *Arenaria*, *Colobanthus*, *Dianthus*, *Gymnocarpos*, *Gypsophila*, *Lychnis*, *Myosoton*, *Paronychia*, *Psammosilene*, *Pseudostellaria*, *Silene* i *Stellaria*. Obecnie liczba ta jest większa i obejmuje około 100 gatunków (13 marca 2024 r.). Biorąc pod uwagę fakt, że rodzina Caryophyllaceae liczy 40 rodzajów i obejmuje około 12 tys. gatunków, liczbę genomów chloroplastowych dostępnych obecnie dla tej grupy roślin należy uznać za nieznaczną.

3. HIPOTEZY I ZADANIA BADAWCZE

Dostępne dane literaturowe wskazują na ciągle bardzo ograniczoną wiedzę na temat zmienności genetycznej gatunków z rodzaju *Cerastium*, w tym zupełny brak danych na temat zmienności genetycznej *C. alpinum* z jedynego polskiego stanowiska tego gatunku na Babiej Górze.

Rodzaj *Cerastium* jest także jedną z tych grup roślin, których systematyka wciąż budzi kontrowersje. Przykładem skomplikowanej systematyki w obrębie rodzaju *Cerastium* jest kompleks *C. alpinum – C. arcticum*. Ogromne zróżnicowanie wewnątrz tego kompleksu doprowadziło do identyfikacji wielu gatunków, podgatunków i odmian w obrębie tej grupy roślin, wśród których możemy wyróżnić *C. alpinum, C. arcticum* i *C. nigrescens*. Dotychczas brakuje narzędzi umożliwiających jednoznaczną identyfikację powyższych gatunków. Niejednoznaczne wyniki dotychczasowych badań, niepewny status taksonomiczny gatunków oraz brak narzędzi umożliwiających ich jednoznaczną identyfikację wskazują potrzebę kontynuowania badań nad kompleksem *C. alpinum – C. arcticum*.

Biorąc pod uwagę powyższe, sformułowane zostały główne **cele** niniejszej rozprawy doktorskiej:

1. Analiza zmienności genetycznej *C. alpinum* z obszaru Babiej Góry.
2. Uzyskanie danych umożliwiających identyfikację oraz analizę powiązań filogenetycznych pomiędzy gatunkami kompleksu *C. alpinum – C. arcticum*.

Analiza danych literaturowych pozwoliła na sformułowanie następujących **hipotez badawczych**:

1. *C. alpinum* z Babiej Góry jako mała i izolowana geograficznie populacja na skutek doświadczenia w przeszłości silnej redukcji efektywnej wielkości populacji (efekt szyjki butelki) charakteryzuje się małą zmiennością genetyczną i wykazuje ślady erozji genetycznej.
2. Genomy chloroplastowe *C. alpinum*, *C. arcticum* i *C. nigrescens* wykazują cechy charakterystyczne dla większości plastomów roślin okrytonasiennych (tj. kolista, dwuniciowa cząsteczka o czterodzielnej strukturze), a ich charakterystyka

i analiza porównawcza umożliwia wskazanie cech unikalnych dla gatunku i przeprowadzenie rekonstrukcji powiązań filogenetycznych w ramach rodzaju *Cerastium* i rodziny Caryophyllaceae.

W celu weryfikacji postawionych hipotez zaplanowano następujące **zadania badawcze**:

1. Oszacowanie zmienności genetycznej *C. alpinum* z obszaru Babiej Góry. W celach porównawczych do analiz włączono przedstawicieli gatunku pochodzących z innych regionów geograficznych: z Arktyki (Kaffiøyra, Svalbard), Austrii (Piz Val Gronda) oraz ze Szwecji (masyw Nuolia, Park Narodowy Abisko). Uzyskane wyniki umożliwią określenie kondycji genetycznej *C. alpinum* z Babiej Góry i dostarczą informacji niezbędnych do opracowania skutecznej strategii ochrony zasobów genetycznych gatunku z tej lokalizacji (**Publikacja nr 1**).
2. Uzyskanie, charakterystyka i analiza porównawcza kompletnych sekwencji genomów chloroplastowych trzech gatunków: *C. alpinum*, *C. arcticum* i *C. nigrescens*; określenie wielkości i struktury tych genomów, identyfikacja powtórzeń genomowych i mikrosatelitarnych, przeanalizowanie kierunku ewolucji sekwencji kodujących białka chloroplastów, a także zrekonstruowanie powiązań filogenetycznych w obrębie rodzaju *Cerastium* i rodziny Caryophyllaceae (**Publikacja nr 2**).

4. MATERIAŁY I METODY

4.1. Materiał roślinny

Materiałem badawczym niniejszej rozprawy doktorskiej były trzy gatunki z rodzaju *Cerastium* (Caryophyllaceae):

- *Cerastium alpinum* L.
- *Cerastium arcticum* Lange
- *Cerastium nigrescens* H.C. Watson

4.2. Pochodzenie materiału roślinnego

Cerastium alpinum zebrano w czterech lokalizacjach: w Polsce, Szwajcarii, Arktyce i Szwecji (Tabela 1). Nasiona *C. alpinum* z Polski zostały zebrane przez nasz Zespół w 2020 roku w Babiogórskim Parku Narodowym po uzyskaniu zgody Ministra Środowiska. Nasiona *C. alpinum* i *C. arcticum* z Arktyki zostały zebrane przez prof. Michała Węgrzyna z Instytutu Botaniki Uniwersytetu Jagiellońskiego w Krakowie podczas wyprawy arktycznej do Stacji Polarnej UMK na Spitsbergenie w 2012 roku. Zebrane nasiona przechowywano w papierowych kopertach w temperaturze 4°C i wykorzystano do założenia uprawy roślin w szklarni Katedry Fizjologii, Genetyki i Biotechnologii Roślin, Wydziału Biologii i Biotechnologii Uniwersytetu Warmińsko-Mazurskiego w Olsztynie. W Szwecji w 2020 roku w masywie Nuolia na terenie Parku Narodowego Abisko zebrano przedstawicieli *C. alpinum* oraz *C. nigrescens*. Zbioru dokonał dr Keith W. Larson z Uniwersytetu w Umeå (Szwecja). Z kolei okazy *C. alpinum* ze Szwajcarii zebrano w rejonie Piz Val Gronda w 2020 roku. Materiał dostarczył prof. Peter Schönswetter z Uniwersytetu w Innsbrucku (Austria). Próby roślin *C. alpinum* zebrane w Szwajcarii i Szwecji oraz *C. nigrescens* zostały dostarczone do Polski w wysuszonej formie.

Tabela 1. Miejsce zbioru materiału roślinnego

Gatunek	Miejsce zbioru
Cerastium alpinum	**Polska** (49°34'N, 19°32'E; 49°34'N, 19°31'E) Karpaty Zachodnie, Babia Góra
	Szwajcaria (46°55'N, 10°17'E) Graubünden, Samnaun-Gruppe, Piz Val Gronda
	Arktyka (78°40′N, 11°49′E) Archipelag Svalbard, Spitsbergen, Kaffiøyra
	Szwecja (68°22’N, 18°44'E) Norrbotten, Park Narodowy Abisko, masyw Nuolja
Cerastium arcticum	**Norwegia** (78°40′N, 11°49′E) Archipelag Svalbard, Spitsbergen, Kaffiøyra
Cerastium nigrescens	**Szwecja** (68°22’N, 18°44'E) Norrbotten, Park Narodowy Abisko, masyw Nuolja

4.3. Analiza zmienności genetycznej *C. alpinum* z Babiej Góry

Do realizacji tego zadania wykorzystano populacje *C. alpinum* z Babiej Góry w Polsce, Szwajcarii Arktyce i Szwecji (Tabela 2). Każda populacja była reprezentowana przez 14 do 28 osobników (łącznie 133). Wielkość próby każdorazowo była uzależniona od wielkości populacji.

Tabela 2. Pochodzenie badanych populacji *Cerastium alpinum*

Populacja	Miejsce zbioru	Wysokość (m n.p.m.)	Liczba osobników
1	**Polska** (49°34'N, 19°32'E; 49°34'N, 19°31'E) Karpaty Zachodnie, Babia Góra	1592	23
2		1620	14
3		1618	28
4		1629	16
5	**Szwajcaria** (46°55'N, 10°17'E) Graubünden, Samnaun-Gruppe, Piz Val Gronda	2750	18
6	**Norwegia** (78°40′N, 11°49′E) Archipelag Svalbard, Spitsbergen, Kaffiøyra	5	17
7	**Szwecja** (68°22’N, 18°44'E) Norrbotten, Park Narodowy Abisko, masyw Nuolja	654-1164	17

W trakcie badań wstępnych przetestowano 34 startery iPBS. Do analiz zasadniczych wybrano osiem (Tabela 3). Informacje na temat zastosowanej metodyki przedstawiono w Tabeli 4.

Tabela 3. Startery iPBS wybrane do badań

Starter	Sekwencja	Temperatura przyłączania startera do matrycy (°C)
2217	5′-ACTTGGATGTCGATACCA-3′	56
2228	5′-CATTGGCTCTTGATACCA-3′	60
2229	5′-CGACCTGTTCTGATACCA-3′	60
2232	5′-AGAGAGGCTCGGATACCA-3′	58
2251	5′-GAACAGGCGATGATACCA-3′	60
2272	5′-GGCTCAGATGCCA-3′	56
2277	5′-GGCGATGATACCA-3′	56
2378	5′-GGTCCTCATCCA-3′	52

Tabela 4. Wykaz metod badawczych wykorzystanych w Publikacji 1

Lp.	Zadanie	Metoda
1.	Ekstrakcja genomowego DNA	Zestaw do izolacji DNA Genomic Mini AX Plant Spin (A&A Biotechnology)
2.	Określenie ilości i czystości DNA	Metoda spektrofotometryczna (NanoDrop)
3.	Amplifikacja DNA	PCR z użyciem starterów iPBS wg oryginalnej metodyki (Kalendar i in. 2010) z modyfikacjami (Androsiuk i in. 2015) Protokół reakcji PCR: 94°C – 3 min. 94°C – 15 s. 50-60°C* – 1 min. } 30 cykli 68°C – 1 min. 72°C – 5 min.
4.	Analiza produktów amplifikacji	Elektroforeza horyzontalna w 1,5% żelu agarozowym z 0,5 μg/ml bromku etydyny
5.	Analiza statystyczna	Analiza danych genetycznych przy użyciu wyspecjalizowanych programów używanych w badaniach z zakresu genetyki populacji (GenAlEx v.6.5, PAST, Arlequin v.3.5, Bottleneck v.1.2.02, Structure v.2.3.4, Structure Harvester v.0.6.94)

* w zależności od startera

4.4. Uzyskanie kompletnych sekwencji genomów chloroplastowych *C. alpinum*, *C. arcticum* i *C. nigrescens*

Ekstrakcję DNA genomowego na potrzeby sekwencjonowania nowej generacji przeprowadzono z:

- *Cerastium alpinum* zebranego w szklarni Wydziału Biologii i Biotechnologii Uniwersytetu Warmińsko-Mazurskiego w Olsztynie (Polska) wyhodowanego z nasion zebranych na Babiej Górze (Polska);

- *Cerastium arcticum* zebranego w szklarni Wydziału Biologii i Biotechnologii Uniwersytetu Warmińsko-Mazurskiego w Olsztynie (Polska) wyhodowanego z nasion zebranych na Svalbardzie (Norwegia);
- *Cerastium nigrescens* w postaci wysuszonej zebranego w masywie Nuolja (Szwecja).

W Tabeli 5 przedstawiono metody wykorzystane w celu realizacji powyższego zadania.

Tabela 5. Wykaz metod badawczych wykorzystanych w Publikacji 2

Lp.	Zadanie	Metoda
1.	Ekstrakcja genomowego DNA	Zestaw do izolacji DNA Maxwell 16 LEV Plant DNA Kit (Promega)
2.	Określenie ilości, czystości i jakości DNA	Metoda spektrofotometryczna oraz elektroforeza horyzontalna w 1,5% żelu agarozowym z 0,5 μg/ml bromku etydyny
3.	Przygotowanie bibliotek i sekwencjonowanie nowej generacji	Przygotowanie bibliotek – library kit: TruSeq DNA PCR Free (350) (Illumina) Sekwencjonowanie na platformie Illumina NovaSeq6000
4.	Określenie jakości surowych odczytów	Program FastQC
5.	Adnotacja genomu	Programy Geneious v.R7, PlasMapper
5.	Przygotowanie mapy genomu	Program OrganellarGenome DRAW
6.	Identyfikacja powtórzeń genomowych i sekwencji mikrosatelitarnych	Programy REPuter, Phobos v.3.3.12
7.	Analiza zróżnicowania genetycznego	Programy MAUVE v.1.1.1, MAFFT v.7.310, DnaSP v.6.10.04
8.	Identyfikacja substytucji synonimicznych i niesynonimicznych	Program DnaSP v.6.10.04
9.	Analiza częstości występowania poszczególnych kodonów	Program Geneious v.R7
10.	Identyfikacja miejsc edycji RNA	Program PREPACT 3.0
11.	Analizy filogenetyczne	Programy MAFFT v.7.310, Mega v.7, MrBayes v.3.2.6

5. OMÓWIENIE PUBLIKACJI STANOWIĄCYCH ROZPRAWĘ DOKTORSKĄ

5.1. Publikacja nr 1

Milarska SE, Androsiuk P, Bednarek PT, Larson K, Giełwanowska I. 2023. Genetic variation of *Cerastium alpinum* L. from Babia Góra, a critically endangered species in Poland. *Journal of Applied Genetics* 64: 37-53. DOI: 10.1007/s13353-022-00731-x

Jedynym miejscem występowania *C. alpinum* w Polsce jest piętro alpejskie Babiej Góry, gdzie rośnie w rozproszeniu na kopule szczytowej w zakresie wysokości 1629-1717 m n.p.m. (Parusel 2012). Gatunek ten występuje na murawach wysokogórskich, a także w szczelinach skalnych i kamienistych rumowiskach, tworząc jednogatunkowe skupiska. Ze względu na ograniczony zasięg, niewielką liczebność i izolację geograficzną od innych stanowisk tego gatunku, *C. alpinum* w Polsce uznawany jest za gatunek krytycznie zagrożony. Mimo to, dotychczas nie był przedmiotem szerszych badań, szczególnie analiz genetycznych.

W pracy doktorskiej podjęto się oceny kondycji genetycznej *C. alpinum* z Babiej Góry. W tym celu dokonano analizy zmienności genetycznej tego stanowiska w porównaniu z populacjami *C. alpinum* z innych regionów geograficznych: Szwecji, Szwajcarii oraz Norwegii. Ze względu na potrzebę przeprowadzenia jak najdokładniejszej charakterystyki genetycznej *C. alpinum* z Babiej Góry, stanowisko to podzielono na cztery części (populacje) wyróżnione na podstawie ich przestrzennego rozmieszczenia oraz ukształtowania powierzchni. Uzyskane dane stanowić będą istotne źródło informacji niezbędnych do opracowania skutecznych strategii ochrony *C. alpinum* w Polsce oraz jego monitoringu.

5.1.1. Ocena skuteczności zastosowanych markerów iPBS

Zastosowane w badaniach markery iPBS okazały się skuteczne zarówno w ujawnianiu polimorfizmu genetycznego pomiędzy osobnikami z poszczególnych populacji, jak i pomiędzy analizowanymi populacjami. Badania z wykorzystaniem ośmiu starterów iPBS ujawniły 262 produkty amplifikacji, przy czym najwięcej (38) uzyskano dla startera 2228, a najmniej (26) dla startera 2229. Średnio na starter przypadało 32,75 amplikonów. Spośród zidentyfikowanych locii 208 (79,39%) było

polimorficznych. Ponadto, wśród zaobserwowanych produktów amplifikacji 20 (7,63%) występowało tylko w jednej populacji i nie występowało w pozostałych. Najwięcej takich unikalnych produktów amplifikacji (4) ujawnił starter 2228. Najwięcej z nich (6) zidentyfikowano w populacji 6 (Svalbard, Arktyka). Brak tego typu produktów stwierdzono jedynie w jednej populacji (populacja 2 z Babiej Góry).

5.1.2. Zróżnicowanie genetyczne analizowanych populacji *C. alpinum*

Liczba produktów amplifikacji ujawniona z wykorzystaniem markerów iPBS wynosiła od 172 w populacji 5 (Piz Val Gronda, Szwajcaria) do 236 w populacji 3 (Babia Góra). Najwyższą liczbę locii polimorficznych zaobserwowano dla populacji 3 (45,80%), a najniższą dla populacji 5 (13,36%). Zmienność genetyczną oszacowano na podstawie dwóch parametrów: współczynnika Shannona (I) i oczekiwanej heterozygotyczności (H_e). Najwyższe wartości obu parametrów stwierdzono dla populacji 3, podczas gdy najniższe dla populacji 5. Zauważono także, że średnie wartości parametrów I oraz H_e dla populacji babiogórskich wynosiły odpowiednio 0,159 oraz 0,106 i były wyższe od średniej wartości tych parametrów dla pozostałych populacji (odpowiednio 0,086 i 0,057).

Wyniki analizy głównych współrzędnych (PCoA) wykazały, że 51,56% zmienności wyjaśniają pierwsze trzy komponenty (odpowiednio 21,6, 18,83 i 11,13%). Przeprowadzone analizy wskazały na wysokie podobieństwo pomiędzy populacjami 5 (Szwajcaria) i 7 (Szwecja). Populacje 1-4 reprezentujące Babią Górę utworzyły jedną heterogeniczną grupę, podczas gdy populacja 6 (Arktyka) wykazała najmniejsze podobieństwo do pozostałych.

Badania prowadzone z wykorzystaniem programu Structure wykazały wyraźną strukturę genetyczną w badanym materiale. Z jednej strony ujawnił się indywidualny charakter populacji 5, 6 i 7, które również okazały się genetycznie jednorodne. Z kolei w przypadku populacji babiogórskich (1-4) zaobserwowano kombinację różnorodnych genotypów wspólnych dla większości tych populacji wskazującą na brak barier w przepływie genów pomiędzy nimi. Relacje pomiędzy analizowanymi populacjami *C. alpinum* określone zostały także za pomocą parametru F_{ST} oszacowanego pomiędzy wszystkimi możliwymi parami populacji (ang. *pairwise* F_{ST}). Największą wartość parametru F_{ST} (0,839) zaobserwowano pomiędzy populacjami 5 oraz 6, a najniższą (0,268) pomiędzy populacjami 1 oraz 3. Również w przypadku tej analizy populacje

babiogórskie wykazały się najmniejszym zróżnicowaniem (obserwowano dla nich najmniejsze wartości F_{ST}, które zawierały się w przedziale 0,268-0,538, dla porównania wartości F_{ST} pomiędzy populacjami 5-7 mieściły się w przedziale 0,537-0,839). Aby sprawdzić czy zróżnicowanie genetyczne pomiędzy populacjami wzrasta wraz z dzielącą je odległością przeprowadzono analizę IBD (ang. *isolation by distance*). Uzyskany wynik wskazał na istotną i pozytywną korelację (R^2 = 0,5073, p = 0,02) pomiędzy wartością dystansu genetycznego między populacjami, wyrażoną za pomocą wartości $F_{ST}/(1-F_{ST})$, a odległością geograficzną pomiędzy nimi (logarytm dziesiętny z odległości). Dane uzyskane za pomocą markerów iPBS wykorzystane zostały także, aby sprawdzić czy badane populacje doświadczyły w przeszłości istotnej redukcji efektywnej wielkości populacji. Wykorzystanie programu Bottleneck pozwoliło na wykrycie śladów efektu wąskiego gardła u wszystkich badanych populacji.

Analiza wariancji molekularnej (AMOVA) wykazała, że większość zidentyfikowanej zmienności (61,69%) zlokalizowana jest pomiędzy badanymi populacjami, podczas gdy pozostałe 38.31% opisuje zmienność między poszczególnymi osobnikami wewnątrz populacji.

5.2. Publikacja nr 2

Milarska SE, Androsiuk P, Paukszto Ł, Jastrzębski JP, Maździarz M, Larson K, Giełwanowska I. 2023. Complete chloroplast genomes of *Cerastium alpinum, C. arcticum* and *C. nigrescens*: genome structures, comparative and phylogenetic analysis. *Scientific Reports* 13: 18774. DOI: 10.1038/s41598-023-46017-y

Rodzaj *Cerastium* jest jednym z taksonów, których struktura wciąż pozostaje tematem licznych sporów i dyskusji. Obiektem badań niniejszej pracy doktorskiej stał się kompleks *C. alpinum – C. arcticum*. Większość badaczy wyróżnia w jego obrębie dwa gatunki: *C. alpinum* i *C. arcticum*. Najbardziej problematycznym składnikiem tej grupy zdaje się być *C. arcticum*, w ramach którego część badaczy wyróżnia dwa odrębne taksony: *C. arcticum* s. str. i *C. nigrescens* (m.in. Clapham i in. 1962; Böcher 1977; Elven i Elvebakk 1996). Wyniki dotychczasowych badań nad kompleksem *C. alpinum – C. arcticum* bazujące na analizach cech morfologicznych oraz danych molekularnych są niejednoznaczne, dlatego istnieje potrzeba dalszych badań z zastosowaniem bardziej zaawansowanych metod. W niniejszej pracy doktorskiej,

w celu dostarczenia danych umożliwiających identyfikację gatunków wchodzących w skład kompleksu *C. alpinum – C. arcticum* oraz rekonstrukcji powiązań filogenetycznych pomiędzy nimi zdecydowano się sięgnąć po dane genomiczne reprezentujące genom chloroplastowy. Obecnie dzięki technologii wysokoprzepustowego sekwencjonowania genomu, można uzyskać wysokiej jakości dane sekwencyjne, użyteczne w badaniach porównawczych, identyfikacji gatunków oraz analizach filogenetycznych. Niestety, dane dotyczące genomów chloroplastowych przedstawicieli rodzaju *Cerastium* są bardzo ograniczone. Dotychczas w bazie NCBI zdeponowano sekwencje genomów chloroplastowych dla tylko dwóch gatunków tego rodzaju: pełną sekwencję dla gatunku *C. glomeratum* Thuill. oraz niekompletną dla *C. arvense* L.

W trakcie badań służących realizacji niniejszej pracy doktorskiej zsekwencjonowano po raz pierwszy genomy chloroplastowe trzech gatunków z rodzaju *Cerastium*: *C. alpinum, C. arcticum* i *C. nigrescens,* a do analiz porównawczych włączono dostępne w bazie NCBI sekwencje genomów chloroplastowych gatunków *C. glomeratum* i *C. arvense*. Uzyskane sekwencje zdeponowano w bazie NCBI pod następującymi numerami: OQ696752 (*C. alpinum*), OQ696753 (*C. arcticum*) i OQ696754 (*C. nigrescens*). W badaniach filogenetycznych oprócz pięciu gatunków *Cerastium* uwzględniono także 54 reprezentantów rodziny Caryophyllaceae, dla których dostępna była kompletna sekwencja genomu chloroplastowego (stan na dzień 24 marca 2023 r.).

5.2.1. Struktura i organizacja genomu chloroplastowego

Genomy chloroplastowe *C. alpinum, C. arcticum* i *C. nigrescens* miały postać kolistej, dwuniciowej cząsteczki o typowej czterodzielnej strukturze, w której można wyróżnić regiony SSC (ang. *small single copy*) i LSC (ang. *large single copy*), oddzielone od siebie dwiema kopiami regionów IR (ang. *inverted repeats*) o identycznym składzie, lecz odwrotnej orientacji. Wielkość genomu wyniosła 147 945 pz u *C. alpinum,* 147 940 pz u *C. nigrescens* i 148 722 pz u *C. arcticum*. Zawartość GC była prawie identyczna u wszystkich analizowanych gatunków i wyniosła: 36,51% dla *C. alpinum,* 36,46% dla *C. arcticum* i 36,52% dla *C. nigrescens*. Genom chloroplastowy badanych gatunków *Cerastium* dzielił identyczny zestaw 113 genów, na który składa się 79 genów kodujących białka, 30 genów tRNA oraz cztery geny

rRNA. Większość genów kodujących białka miała standardowy kodon ATG jako kodon inicjacyjny, a całkowita liczba kodonów dla wszystkich genów kodujących białka wyniosła 26 116 dla *C. alpinum*, 26 115 dla *C. arcticum* i 26 220 dla *C. nigrescens*. Dominującym aminokwasem okazała się leucyna (10,7%), natomiast najrzadziej spotykanym cysteina (1,2%). Najliczniej występującym kodonem (4,46%) był ATT, a najmniej licznymi (0,004%) TTG (u wszystkich gatunków) oraz ATA i CTG u *C. nigrescens*. Większość genów w analizowanych genomach chloroplastowych nie zawierała intronów, jedynie 14 z nich zawierało jeden intron (*atpF*, *ndhA*, *ndhB*, *petB*, *petD*, *rpl16*, *rpoC1*, *rps16*, *trnA*-UGC, *trnG*-UCC, *trnI*-GAU, *trnK*-UUU, *trnL*-UAA, *trnV*-UAC), podczas gdy kolejne trzy geny składały się z trzech eksonów (*clpP*, *ycf3*, *rps12*). W regionie LSC znajdowało się 58 genów kodujących białka, 22 geny tRNA i dwie ORF (ang. *Open Reading Frames*) oznaczone jako *ycf3* i *ycf4*. Region SSC zawierał 11 genów kodujących białka, jeden gen tRNA i jedną chloroplastową ORF (*ycf1*). Natomiast region IR zawierał 6 genów kodujących białka, 7 genów tRNA, 4 geny rRNA i jedną chloroplastową ORF (*ycf2*).

5.2.2. Powtórzenia genomowe i sekwencje mikrosatelitarne

Identyfikację powtórzeń genomowych i sekwencji mikrosatelitarnych przeprowadzono dla gatunków *Cerastium*, dla których dysponowano kompletną sekwencją genomu chloroplastowego, tj. dla trzech nowo zsekwencjonowanych gatunków (*C. alpinum*, *C. arcticum* i *C. nigrescens*) oraz dla *C. glomeratum*. Analiza wykazała obecność 79 powtarzalnych sekwencji o długości od 30 do 170 pz. Najwyższą liczbę powtórzeń (23) ujawniono u *C. arcticum*, a najniższą (16) u *C. glomeratum*. Wśród zidentyfikowanych sekwencji dominowały powtórzenia palindromowe (od 47,8% u *C. arcticum* do 62,5% u *C. glomeratum*). Najwięcej sekwencji powtórzonych (80%) znaleziono w regionie LSC, pozostałe były równomiernie rozmieszczone (po 10%) w regionach SSC i IR. U każdego z analizowanych gatunków najczęściej występowały powtórzenia o długości od 30 do 40 pz (od 13 u *C. glomeratum* do 18 u pozostałych).

Badania genomów chloroplastowych ujawniły również obecność mono-, di-, tri-, tetra-, penta- i heksanukleotydowych sekwencji mikrosatelitarnych (SSR, ang. *simple sequence repeats*). Zidentyfikowano od 20 (*C. nigrescens*) do 23 (*C. arcticum*) sekwencji mikrosatelitarnych, przy czym dominowały sekwencje mononukleotydowe

składające się z jednostek A/T (od 36,4% u *C. glomeratum* do 43,5% u *C. arcticum*). Drugim najczęściej obserwowanym motywem był AAT/TAA z częstotliwością od 27,3% (*C. glomeratum*) do 35% (*C. nigrescens)*. Tetranukleotydowe motywy (AAAT/TAAA, AATT/TTAA, ACCT/TCCA, AGAT/TAGA i AAAG/GAAA) występowały z częstotliwością od 19% (*C. alpinum*) do 31,8% (*C. glomeratum*). Ponadto zidentyfikowano również pojedynczo występujący motyw dinukleotydowy (AT/TA, *C. glomeratum*), pentanukleotydowy (AATAT/TATAA, *C. alpinum*) oraz heksanukleotydowy (AAATCC/CCTAAA, *C. arcticum*). Większość SSRs zaobserwowano w regionie LSC (od 72,7% u *C. glomeratum* do 76,2% u *C. alpinum*). W regionie SSC zidentyfikowano od 14,3% (*C. alpinum*) do 18,2% (*C. glomeratum*) sekwencji SSRs, a najmniej w regionach IR (od 8,7% u *C. arcticum* do 10% u *C. nigrescens*). Sekwencje mikrosatelitarne zlokalizowane były głównie w przestrzeniach międzygenowych (od 55% u *C. nigrescens* do 61,9% u *C. alpinum*), natomiast pozostałe mikrosatelity były rozmieszczone w eksonach (od 31,8% u *C. glomeratum* do 35% u *C. nigrescens*) i intronach (od 4,8% u *C. alpinum* do 10% u *C. nigrescens*).

5.2.3. Substytucje synonimiczne (Ks) i niesynonimiczne (Ka)

W celu przeanalizowania presji selekcyjnej oddziałującej na sekwencje kodujące białka w genomach chloroplastowych badanych gatunków *Cerastium* przeanalizowano łącznie 77 genów, dla których oszacowano współczynniki substytucji synonimicznych (Ks) i niesynonimicznych (Ka), a także stosunek Ka/Ks. W analizach nie uwzględniono genów *infA* oraz *rpl23* ze względu na ich brak w genomie chloroplastowym *C. glomeratum*, który w tej analizie pełnił rolę genomu referencyjnego. Wartość współczynników substytucji różniła się w zależności od genu w każdej grupie funkcjonalnej i wahała się od 0 do 0,151 i od 0 do 0,0858 odpowiednio dla Ka i Ks. W sekwencjach nukleotydowych 11 genów nie zaobserwowano żadnych różnic (Ka=0 i Ks=0), natomiast w sekwencjach 18 genów zaobserwowano jedynie substytucje synonimiczne (Ka=0). Stosunek Ka/Ks był mniejszy niż 1 dla wszystkich genów, z wyjątkiem *ndhB* (2,725 u *C. arvense*). Stosunkowo wysokie wartości Ka/Ks zaobserwowano również w *rpl22* (0,8673) u wszystkich badanych gatunków oraz w *rps14* (0,8776) dla *C. arvense*. W pozostałych przypadkach wartość Ka/Ks nie przekraczała 0,75.

Współczynniki Ka, Ks oraz Ka/Ks są ważnymi parametrami w badaniach ewolucji genów. Z kolei stosunek Ka/Ks pozwala wnioskować o kierunku ewolucji sekwencji na poziomie białka: Ka/Ks > 1 oznacza selekcję pozytywną, Ka/Ks < 1 oznacza selekcję negatywną (czyszczącą), podczas gdy Ka/Ks = 1 oznacza ewolucję neutralną. Generalnie, w większości regionów kodujących wartość Ks dominuje nad Ka. Taką sytuację zaobserwowano także w niniejszej pracy, co skutkowało wysokim konserwatyzmem analizowanych sekwencji. Selekcja czyszcząca utrzymuje dotychczasową strukturę i funkcję białek, podczas gdy selekcja pozytywna faworyzuje nowe warianty genów, co może być korzystne dla adaptacji organizmu do zmieniających się warunków środowiskowych. W przypadku analizowanych gatunków można stwierdzić, że większość genów zostało poddane selekcji czyszczącej. Wyjątek stanowi gen *ndhB*. Dane literaturowe wskazują, że chloroplastowa dehydrogenaza NADH jest wrażliwa na silny stres świetlny i może chronić rośliny przed fotoinhibicją lub stresem oksydacyjnym poprzez stabilizację kompleksu NADH i zapobiegać związanemu z suszą spadkowi wydajności fotosyntezy i opóźnieniu wzrostu.

5.2.4. Analizy porównawcze genomów chloroplastowych

Przeprowadzone badania wykazały wysoce konserwatywną strukturę genomów chloroplastowych *C. alpinum, C. arcticum, C. nigrescens* i *C. arvense*, w przypadku których nie wykryto żadnych rearanżacji. Tylko w przypadku *C. glomeratum* zaobserwowano przeciwną orientację całego regionu SSC. Takie zjawisko opisane zostało już wcześniej dla roślin reprezentujących różne grupy systematyczne, u których obserwuje się typ heteroplazmii, w której genomy chloroplastowe występują w równych proporcjach w postaci izomerów inwersyjnych różniących się orientacją regionu SSC.

Analiza porównawcza sekwencji nukleotydowej genomów chloroplastowych gatunków *Cerastium*, dla których dysponowano ich kompletnymi sekwencjami wykazała niski poziom zróżnicowania (π = 0,00493). Jednocześnie badania wykazały obecność dziewięciu wysoce zmiennych (π > 0,015) regionów: *rpl32–trnL*-UAG, *ndhA* (intron), *rps16* (intron), *trnD*-GUC–*trnY*-GUA, *trnF*-GAA–*ndhJ*, *ndhC–trnV*-UAC, *petA–psbJ*, *psbE–petL* i *trnP*-UGG–*psaJ*. Wysoce zmienne sekwencje występujące w genomach chloroplastów są często źródłem markerów molekularnych odpowiednich do analiz filogenetycznych i identyfikacji gatunków. W przypadku analizowanych gatunków *Cerastium* największą wartość π (0,02708) zaobserwowano dla regionu *trnD-*

GUC–*trnY*-GUA. Wszystkie wyżej wymienione regiony zidentyfikowano w regionach niekodujących, tj. w przestrzeniach międzygenowych i intronach. Ponadto, większość wysoce zmiennych regionów stwierdzono w regionie LSC, mniej w SSC, z kolei w regionie IR nie zidentyfikowano żadnego.

5.2.5. Miejsca edycji RNA

Program PREPACT 3.0 umożliwił identyfikację w genomach chloroplastów *C. alpinum, C. arcticum* i *C. nigrescens* od 578 do 588 potencjalnych miejsc edycji w 63 genach kodujących białka. Najmniejszą liczbę miejsc edycji RNA stwierdzono u *C. alpinum* i *C. nigrescens*, a największą u *C. glomeratum*. W przypadku 14 genów nie stwierdzono takich zmian. Wśród znalezionych miejsc edycji zaobserwowano zarówno konwersje C do U, jak i U do C. Podstawienia C do U odpowiadały za 43,05% do 43,54% wszystkich miejsc edycji RNA, podczas gdy konwersje U do C były odpowiedzialne za 56,46% do 56,95% zdarzeń edycji RNA. Wszystkie zidentyfikowane miejsca edycji powodowały niesynonimiczne mutacje. 47% substytucji stwierdzono dla pierwszej pozycji kodonu, 53% podstawień dla drugiej pozycji kodonu, natomiast nie stwierdzono żadnych miejsc edycji w trzeciej pozycji kodonu. Największą liczbę potencjalnych miejsc edycji RNA przewidziano dla genów *ycf1* (85-88), *ycf2* (77) i *rpoC2* (64-65). Najczęściej występującą substytucją u wszystkich gatunków była zmiana fenyloalaniny na leucynę (16,48-16,75%), podczas gdy najrzadziej występującymi były zmiany proliny na fenyloalaninę i argininy na tryptofan (odpowiednio 0,353-0,358% i 0,881-0,896%).

Analogiczne badania przeprowadzono dla *C. arvense*, jednak ze względu na niekompletną sekwencję genomu chloroplastowego dostępną w bazie NCBI nie uwzględniono w analizach następujących genów: *rpl20, rpoB, rpoC1, rpoC2, ycf1* i *ycf2*. W 17 z 71 analizowanych genów nie zidentyfikowano miejsc edycji RNA. W pozostałych 54 genach znaleziono 286 potencjalnych miejsc edycji. U tego gatunku również dominowała konwersja U do C (56,46%). Najwięcej podstawień stwierdzono na pierwszej (53,06%) i drugiej (46,94%) pozycji kodonu, nie stwierdzono miejsc edycji dotyczących trzeciej pozycji kodonu. Największą liczbę miejsc edycji RNA zidentyfikowano w sekwencjach genów *matK* (40) i *ndhF* (37). Wszystkie zaobserwowane miejsca edycji powodowały mutacje niesynonimiczne. Najliczniejszą

substytucją (18,53%) była zmiana z fenyloalaniny na leucynę, natomiast najmniej liczną – leucyny na prolinę i argininy na cysteinę (0,7%).

5.2.6. Analizy filogenetyczne

Uzyskane sekwencje genomów chloroplastowych dla *C. alpinum, C. arcticum* i *C. nigrescens* umożliwiły rekonstrukcję powiązań filogenetycznych w ramach rodzaju *Cerastium* oraz rodziny Caryophyllaceae. Topologia uzyskanego drzewa filogenetycznego była zgodna z pozycją systematyczną analizowanych gatunków. Co więcej, uzyskane wyniki pozwoliły na jednoznaczne rozróżnienie wszystkich analizowanych gatunków, w tym pięciu przedstawicieli rodzaju *Cerastium* (*C. alpinum, C. arcticum, C. nigrescens, C. glomeratum* i *C. arvense*). Wszystkie gatunki *Cerastium* tworzyły jeden klad, w którym największe podobieństwo wykazały gatunki *C. nigrescens* oraz *C. arvense*, do których w dalszym przebiegu drzewa dołączyły pozostałe gatunki tego rodzaju w następującej kolejności *C. alpinum, C. arcticum* oraz *C. glomeratum*. W przypadku pozostałych gatunków zaobserwowano, że wszystkie gatunki *Silene* wraz z *Lychnis wilfordii* tworzyły jeden klad; kolejny utworzyli przedstawiciele Caryophylleae, czyli pięć gatunków *Dianthus*, trzech przedstawicieli rodzaju *Gysophila* oraz *Psammosilene tunicoides*; trzeci klad obejmował wszystkich przedstawicieli Alsineae (osiem gatunków *Pseudostellaria*, *Stellaria dichotoma* var. *lanceolata, Myosoton aquaticum* i wszystkie badane gatunki *Cerastium*, tworzące jedną podgrupę), czwarty klad składał się z ośmiu gatunków *Colobanthus* (Sagineae), natomiast *Spergula arvensis* (Sperguleae) oraz *Paronychia argentea* wraz *Gymnocarpos przewalskii* (Paronychieae) utworzyły dwie odrębne gałęzie. Najbardziej odległe stanowisko zajął *Arabidopsis thaliana*, który w analizach filogenetycznych został użyty jako grupa zewnętrzna.

Zaprezentowane w niniejszej pracy wyniki szacowania czasu dywergencji sugerują, że różnicowanie się taksonów w obrębie rodziny Caryophyllaceae rozpoczęło się ok. 74,46 mln lat temu (czyli jeszcze przed końcem kredy, najmłodszego okresu ery mezozoicznej) i było kontynuowane przez cały paleogen i neogen z największą intensywnością w ciągu ostatnich 10 mln lat. Zgodnie z naszymi obserwacjami plemię Alsinae, do którego zalicza się rodzaj *Cerastium*, zaczęło różnicować się już ok. 20,6 mln lat temu, podczas gdy początek tego procesu dla rodzaju *Cerastium* datowany jest na 3,66 mln lat temu. Uzyskane wyniki sugerują, że spośród przeanalizowanych przez

nas gatunków *Cerastium*, *C. glomeratum* oddzielił się najwcześniej, podczas gdy pozostali przedstawiciele znajdują się na wczesnych etapach dywergencji.

6. PODSUMOWANIE I WNIOSKI

Uzyskane wyniki umożliwiły weryfikację postawionych hipotez oraz sformułowanie następującego podsumowania.

6.1. Zmienność genetyczna *C. alpinum* z Babiej Góry

1. Markery DNA oparte na retrotranspozonach (iPBS) okazały się być wydajnym narzędziem do ujawnienia zmienności genetycznej populacji *Cerastium alpinum*.
2. Dane molekularne ujawniły niski poziom wewnątrzpopulacyjnej zmienności *C. alpinum* przy jednoczesnym wysokim zróżnicowaniu genetycznym populacji.
3. Istnieje pozytywna i istotna korelacja pomiędzy wartością dystansu genetycznego między badanymi populacjami, a odległością geograficzną pomiędzy nimi.
4. Populacje *C. alpinum* z Babiej Góry stanowią odrębną, aczkolwiek niejednorodną grupę o generalnie większej zmienności genetycznej niż populacje z innych regionów.
5. Większość zidentyfikowanej zmienności genetycznej zlokalizowana jest pomiędzy badanymi populacjami.
6. Identyfikacja kombinacji różnorodnych genotypów wspólnych dla większości populacji babiogórskich wskazuje na brak barier w przepływie genów pomiędzy nimi.
7. Mimo izolacji geograficznej oraz doświadczenia w przeszłości silnej redukcji efektywnej wielkości populacji, *C. alpinum* z Babiej Góry nadal stanowi rezerwuar cennej zmienności genetycznej w ramach gatunku.

6.2. Charakterystyka i analiza porównawcza genomów chloroplastowych *C. alpinum, C. arcticum* i *C. nigrescens*

1. Genomy chloroplastowe *C. alpinum, C. arcticum* i *C. nigrescens* okazały się wysoce konserwatywne pod względem struktury i składu genetycznego.

2. Poziom substytucji synonimicznych i niesynonimicznych wykazał wysoki konserwatyzm badanych genów.
3. Zaobserwowane wysokie podobieństwo badanych genomów chloroplastowych oraz bardzo niskie wartości szacowanego czasu dywergencji można traktować jako możliwe dowody nadal słabo wykształconych barier reprodukcyjnych pomiędzy tymi gatunkami, które umożliwiają spontaniczną hybrydyzację.
4. Uzyskane dane genomiczne dla genomów chloroplastowych *C. alpinum*, *C. arcticum* i *C. nigrescens* umożliwiły identyfikację cech unikalnych dla poszczególnych gatunków.
5. Ujawnione powiązania filogenetyczne w obrębie rodziny Caryophyllaceae były zgodne z pozycją taksonomiczną badanych gatunków, a w przypadku rodzaju *Cerastium* potwierdziły odrębny charakter każdego z badanych gatunków (w tym *C. alpinum*, *C. arcticum* i *C. nigrescens*).

7. LITERATURA

Andeden EE, Baloch FS, Derya M, Kilian B, Ozkan H (2013) iPBS-Retrotransposons-based genetic diversity and relationship among wild annual *Cicer* species. *J Plant Biochem Biotechnol* 22:453-466

Androsiuk P, Chwedorzewska KJ, Dulska J, Milarska S, Giełwanowska I (2020) Retrotransposon-based genetic diversity of *Deschampsia antarctica* Desv. from King George Island (Maritime Antarctic). *Ecol Evol* 11:648-663

Androsiuk P, Chwedorzewska KJ, Szandar K, Giełwanowska I (2015) Genetic variability of *Colobanthus quitensis* from King George Island (Antarctica). *Pol Polar Res* 36:281-295

Androsiuk P, Koc J, Chwedorzewska KJ, Górecki R, Giełwanowska I (2019) Retrotransposon-based genetic variation of *Poa annua* populations from contrasting climate conditions. *PeerJ* 7:e6888

Baloch FS, Alsaleh A, Sáenz de Miera LE, Hatipoğlu R, Çiftçi V, Karaköy T, Yıldız M, Özkan H (2015) DNA based iPBS-retrotransposon markers for investigating the population structure of pea (*Pisum sativum*) germplasm from Turkey. *Biochem Syst Ecol* 61:244-252

Böcher TW (1977) *Cerastium alpinum* and *C. arcticum*, a mature polyploid complex. *Bot Not* 130:303-309

Bock R (2007) Plastid biotechnology: prospects for herbicide and insect resistance, metabolic engineering and molecular farming. *Curr Opin Biotechnol* 18:100-106

Bortiri E, Coleman-Derr D, Lazo G, Anderson O, Gu Y (2008) The complete chloroplast genome sequence of *Brachypodium distachyon*: sequence comparison and phylogenetic analysis of eight grass plastomes. *BMC Research Notes* 1:61

Boşcaiu M, Vicente O, Ehrendorfer F (1999) Chromosome numbers, karyotypes and nuclear DNA contents from perennial polyploid groups of *Cerastium* (Caryophyllaceae). *Pl Syst Evol* 218: 13-21

Brysting AK, Borgen L (2000) Isozyme analysis of the *Cerastium alpinum-C. arcticum* complex (Caryophyllaceae) supports a splitting of *C. arcticum* Lange. *Plant Syst Evol* 220:199-221

Brysting AK, Elven R (2000) The *Cerastium alpinum-C. arcticum* complex (Caryophyllaceae): numerical analyses of morphological variation and taxonomical revision of *C. arcticum* Lange. *Taxon* 49:186-216

Brysting AK, Hagen A (1999) Species in polyploid complexes? The *Cerastium alpinum-C. arcticum* complex. *Det Norske Videnskaps Akademi. I. Mat. Nat. Kl. Avh. Ny Serie* 38:183-190

Brysting AK, Mathiesen C, Marcussen T (2011) Challenges in polyploid phylogenetic reconstruction: A case story from the arctic-alpine *Cerastium alpinum* complex. *Taxon* 60:333-347

Caković D, Stešević D, Schönswetter P, Frajman B (2018) Long neglected diversity in the Accursed Mountains of northern Albania: *Cerastium hekuravense* is genetically and morphologically divergent from *C. dinaricum*. *Plant Syst Evol* 304:57-69

Chase MW, Cowan RS, Hollingsworth PM, Van den Berg C, Madriñán S, Petersen G, Seberg O, Jørgsensen T, Cameron KM, Carine M, Pedersen N, Hedderson TAJ, Conrad F, Salazar GA, Richardson JE, Hollingsworth ML, Barraclough TG, Kelly L, Wilkinson M (2007) A proposal for standardized protocol to barcode all land plants. *Taxon* 56:295-299

Chen IC, Hill JK, Ohlemüller R, Roy DB, Thomas CD (2011) Rapid range shifts of species associated with high levels of climate warming. *Sci* 333:1024-1026

Clapham AR, Tutin TG, Warburg EF (1962) Flora of the British Isles, ed. 2. Cambridge University Press, Cambridge

Czortek P, Kapfer J, Delimat A, Eycott AE, Grytnes JA, Orczewska A, Ratyńska H, Zięba A, Jaroszewicz B (2018) Plant species composition shifts in the Tatra Mts as a response to environmental change: a resurvey study after 90 years. *Folia Geobot* 53:333-348

Đurović SZ, Temunović M, Niketić M, Tomović G, Schönswetter P, Frajman B (2021) Impact of Quaternary climatic oscillations on phylogeographic patterns of three habitat-segregated Cerastium taxa endemic to the Dinaric Alps. *J Biogeogr* 48:2022-2036

Ellstrand NC, Elam DR (1993) Population Genetic Consequences of Small Population Size: Implications for Plant Conservation. *Annu Rev Ecol Evol* Syst 24:217-242

Elven R, Elvebakk A (1996) Part 1. Vascular plants. In: Elvebakk A, Prestrud P (eds.) A catalogue of Svalbard plants, fungi, algae, and cyanobacteria. *Norsk Polarinst Skr.* 198:9-55

Erper I, Ozer G, Kalendar R, Avci S, Yildirim E, Alkan M, Turkkan M (2021) Genetic diversity and pathogenicity of *Rhizoctonia* spp isolates associated with red cabbage in Samsun (Turkey). *J Fungi* 7:234

Hagen AR, Giese H, Brochmann C (2001) Trans-Atlantic Dispersal and Phylogeography of *Cerastium arcticum* (Caryophyllaceae) Inferred from RAPD and SCAR Markers. *Am J Bot* 88:103-112

Hagen AR, Sæther T, Borgen L, Elven R, Stabbetorp OE, Brochmann C (2002) The arctic-alpine polyploids *Cerastium alpinum* and *C. nigrescens* (Caryophyllaceae) in a sympatric situation: breakdown of species integrity? *Plant Syst Evol* 230:203-219

Hooker JD (1884) The Student's flora of the British Islands, ed. 3 Macmillan &Co, London

Hossein-Pour A, Haliloglu K, Özkan G, Tan M (2019) Genetic diversity and population structure of quinoa (*Chenopodium quinoa* Willd.) using iPBS-retrotransposons markers. *Appl Ecol Environ Res* 17:1899-1911

Hultén E (1956) The *Cerastium alpinum* complex. A case of world-wide introgressive hybridization. *Svensk Bot Tidskr* 50:411-495

Jansen RK, Cai Z, Raubeson LA, Daniell H, dePamphilis CW, Leebens-Mack J, Müller KF, Guisinger-Bellian M, Haberle RC, Hansen AK, Chumley TW, Lee SB, Peery R, McNeal JR, Kuehl JV, Boore JL (2007) Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc. Natl. Acad. Sci. U.S.A.* 104:19369-19374

Kalendar R, Antonius K, Smýkal P, Schulman AH (2010) iPBS: a universal method for DNA fingerprinting and retrotransposon isolation. *Theor Appl Genet* 121:1419-1430

Khapilina O, Raiser O, Danilova A, Shevtsov V, Turzhanova A, Kalendar R (2021) DNA profiling and assessment of genetic diversity of relict species *Allium altaicum* Pall on the territory of Altai. *PeerJ* 9:e10674

Khapilina O, Turzhanova A, Danilova A, Tumenbayeva A, Shevtsov V, Kotukhov Y, Kalendar R (2021b) Primer Binding Site (PBS) profiling of genetic diversity of natural populations of endemic species *Allium ledebourianum* Schult. *Biotech* 10:23

Koc J, Androsiuk P, Chwedorzewska KJ, Cuba-Díaz M, Górecki R, Giełwanowska I (2018) Range-wide pattern of genetic variation in *Colobanthus quitensis*. *Polar Biol* 41:2467-2479

Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH (2005) Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci. U.S.A.* 102:8369-8374

Kutnjak D, Kuttner M, Niketić M, Dullinger S, Schönswetter P, Frajman B (2014) Escaping to the summits: Phylogeography and predicted range dynamics of *Cerastium dinaricum*, an endangered high mountain plant endemic to the western Balkan Peninsula. *Mol Phylogenet Evol* 78:365-374

Lenoir J, Gégout JC, Marquet PA, de Ruffray P, Brisse H (2008) A significant upward shift in plant species optimum elevation during the 20th century. Sci 320:1768-1771

Lynch M, Conery J, Bürger R (1995) Mutational meltdown in sexual populations. *Evolution* 49:1067-1080

Mascher M, Gundlach H, Himmelbach A, Beier S, Twardziok SO, Wicker T, Radchuk V, Dockter C, Hedley PE, Russell J, et al. (2017) A chromosome conformation capture ordered sequence of the barley genome. *Nature* 544:427-33

Milovanov A, Zvyagin A, Daniyarov A, Kalendar R, Troshin L (2019) Genetic analysis of the grapevine genotypes of the Russian *Vitis* ampelographic collection using iPBS markers. *Genetica* 147:91-101

Niketić M, Đurović SZ, Tomović G, Schönswetter P, Frajman B (2022) Diversification within ploidy-variable Balkan endemic *Cerastium decalvans* (Caryophyllaceae) reconstructed based on genetic, morphological and ecological evidence. *Bot J Linn Soc* 199:578-608

Nyberg Berglund AB, Saura A, Westerbergh A (2006) Electrophoretic Evidence for Disomic Inheritance and Allopolyploid Origin of the Octoploid *Cerastium alpinum* (Caryophyllaceae). *J Hered* 97:296-302

Parusel J (2001) *Cerastium alpinum* L. s.s. Rogownica alpejska. In: Kaźmierczakowa R, Zarzycki K (eds.) Polska Czerwona Księga Roślin. Instytut Botaniki im. W. Szafera i Instytut Ochrony Przyrody PAN, Kraków, 97-99

Parusel J (2014) *Cerastium alpinum* Rogownica alpejska. In: Kaźmierczakowa R, Zarzycki K, Mirek Z (eds.) Polska Czerwona Księga Roślin. Paprotniki i rośliny kwiatowe. Instytut Ochrony Przyrody PAN, Kraków, 121-123

Porsild AE, Cody WJ (1980) Vascular plants of Continental Northwest Territories, Canada. National Museum of Natural Sciences, National Museums of Canada, Ottawa

Raman G, Park S (2016) The complete chloroplast genome sequence of *Ampelopsis*: Gene organization, comparative analysis, and phylogenetic relationships to other angiosperms. *Front Plant Sci* 7:341

Rønning OI (1996) The flora of Svalbard. Norsk Polarinst, Polarhåndbook 10, Oslo

Rumpf SB, Hulber K, Klonner G, Moser D, Schutz M, Wessely J, Dullinger S (2018) Range dynamics of mountain plants decrease with elevation. *Proc Natl Acad Sci U S A* 115:1848-1853

Scheen AC, Brochmann C, Brysting AK, Elven R, Morris A, Soltis DE, Soltis PS, Albert VA (2004) Northern hemisphere biogeography of *Cerastium* (Caryophyllaceae): insight from phylogenetic analysis of noncoding plastid nucleotide sequences. *Am J Bot* 91:943-952

Skubic M, Schönswetter P, Frajman B (2018) Diversification of *Cerastium sylvaticum* and *C. subtriflorum* on the margin of the south-eastern Alps. *Plant Syst Evol* 304:1101-1115

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature* 408:796-815

The World Flora Online. Available at http://www.worldfloraonline.org (accessed on September 2022)

Young AG, Boyle T, Brown T (1996) The population genetic consequences of habitat fragmentation for plants. *Trends Ecol Evol* 11:413-418

Zhai Y, Zhang T, Guo Y, Gao C, Zhou L, Feng L, Zhou T, Xumei W (2023) Phylogenomics, phylogeography and germplasms authentication of the *Rheum palmatum* complex based on complete chloroplast genomes. *J Plant Res* 136:291-304

8. SPIS TABEL

Tabela 1. Miejsce zbioru materiału roślinnego

Tabela 2. Pochodzenie badanych populacji *Cerastium alpinum*

Tabela 3. Startery iPBS wybrane do badań

Tabela 4. Wykaz metod badawczych wykorzystanych w Publikacji 1

Tabela 5. Wykaz metod badawczych wykorzystanych w Publikacji 2

9. KOPIE PRAC WCHODZĄCYCH W SKŁAD ROZPRAWY DOKTORSKIEJ

Publikacja nr 1

Journal of Applied Genetics (2023) 64:37–53
https://doi.org/10.1007/s13353-022-00731-x

PLANT GENETICS • ORIGINAL PAPER

Genetic variation of *Cerastium alpinum* L. from Babia Góra, a critically endangered species in Poland

Sylwia Eryka Milarska[1] · Piotr Androsiuk[1] · Piotr Tomasz Bednarek[2] · Keith Larson[3] · Irena Giełwanowska[1]

Received: 15 August 2022 / Revised: 5 October 2022 / Accepted: 10 October 2022 / Published online: 2 November 2022
© The Author(s) 2022

Abstract

Babia Góra massif is the only site of occurrence of the *Cerastium alpinum* L. in Poland, an arctic-alpine perennial plant with a wide distribution in North America, northwestern Asia, and Europe. To determine whether the isolated Polish populations are genetically distinct, we have performed an evaluation of *C. alpinum* from Babia Góra with the use of iPBS markers. A total number of 133 individuals of *C. alpinum* from seven populations representing four localizations of the species were analyzed, i.e., from Babia Góra (Poland), Alps (Switzerland), Nuolja massif (Sweden), and Kaffiøyra (Svalbard, Norway). Genetic analysis of all *C. alpinum* samples using eight PBS primers identified 262 bands, 79.4% of which were polymorphic. iPBS markers revealed low genetic diversity (average $H_e = 0.085$) and high population differentiation ($F_{ST} = 0.617$). AMOVA results confirmed that the majority of the genetic variation (62%) was recorded among populations. The grouping revealed by PCoA showed that *C. alpinum* from Svalbard is the most diverged population, *C. alpinum* from Switzerland and Sweden form a pair of similar populations, whereas *C. alpinum* from the Babia Góra form a heterogeneous group of four populations. Results of isolation by distance analysis suggested that the spatial distance is the most probable cause of the observed differentiation among populations. Although significant traces of a bottleneck effect were noted for all populations of *C. alpinum* from Babia Góra, the populations still maintain a low but significant level of genetic polymorphism. These results are of great importance for developing conservation strategies for this species in Poland.

Keywords Arctic-alpine perennial · Population differentiation · Genetic structure · Retrotransposon-based iPBS markers · Species conservation

Introduction

The genus *Cerastium* (Caryophyllaceae) consists of over 200 species (Plants of the World Online, http://www.plantsoftheworldonline.org), the majority of which are spread in the Northern Hemisphere in the temperate climate zone, but only a few of them are cosmopolitan species. *Cerastium alpinum* is a perennial that grows in rocky, sandy, or gravely substrates and is characterized by its matted growth form, hairy basal leaves, flowering stems, white, five-petalled flowers about 20 mm in diameter, and a branched root system (Porsild and Cody 1980; Clapham et al. 1987; Rønning 1999). Flowering is initiated relatively late, from June to August. The species is generally cross-pollinated by insects, although self-pollination occurs (Totland and Schulte-Herbrüggen 2003; Parusel 2008). Seeds are dispersed by gravity—they are sown mainly within a close vicinity of the source plant (Parusel 2008). In Poland, its only natural sites are located in the Babia Góra massif (Żywiec Beskids, Western Carpathians) within the area of the Babia Góra National Park (Parusel 2001). *C. alpinum* can be found on high mountain grasslands where it forms one-species aggregations on rock shelves, in rock crevices, and on stony debris (Parusel 2014). Most of these sites were recorded in the subalpine zone, from 1480 to 1680 m.a.s.l. (Parusel 2001). Due to

Communicated by Izabela Pawłowicz

✉ Piotr Androsiuk
piotr.androsiuk@uwm.edu.pl

[1] Department of Plant Physiology, Genetics and Biotechnology, Faculty of Biology and Biotechnology, University of Warmia and Mazury in Olsztyn, ul. M. Oczapowskiego 1A, 10-719 Olsztyn, Poland

[2] Plant Breeding and Acclimatization Institute – National Research Institute, Radzików, 05-870 Błonie, Poland

[3] Climate Impacts Research Centre, Department of Ecology and Environmental Sciences, Umeå University, 901 87 Umeå, Sweden

the limited area of occurrence and spatial isolation of its sites (the closest populations are in the Alps and the Eastern Carpathians; Parusel 2001), *C. alpinum* was included in the Polish Red Data Book of Plants and received the status of a critically endangered species in Poland (Parusel 1993, 2014; Mirek and Piękoś-Mirkowa 2008). This makes it a critical component of Polish flora that requires special attention, i.e., increased interdisciplinary activity to develop effective strategies for species conservation and restoration.

The subject of conservation genetic studies is to understand the dynamics of genes in populations threatened by extinction, to apply the techniques and concepts of genetics, and to use them to address problems in conservation biology (Hedrick and Miller 1992). To plan an effective conservation strategy, it is necessary to know the current genetic condition of the population. Genetic diversity is influenced by population size, gene flow, the mating system, and other facultative factors like genetic drift, founder effects or bottleneck, and subsequent inbreeding (Ellstrand and Elam 1993). Inbreeding depression, consequently, can lead to the accumulation of deleterious mutations, a reduction in plant fitness, and a further reduction in population size (Ellstrand and Elam 1993; Lynch et al. 1995; Young et al. 1996). Another factor leading to genetic degradation is spatial isolation, which may restrict or prevent gene flow. Due to being rare and in a remote area, *C. alpinum* from Babia Góra may be exposed to the above problems, which could make the risk of local extinction higher.

In recent decades, various PCR-based techniques such as randomly amplified polymorphic DNA (RAPD; Williams et al. 1990), inter simple sequence repeat (ISSR; Zietkiewicz et al. 1994), amplified fragment length polymorphism (AFLP; Vos et al. 1995), or simple sequence repeats (SSR; Tautz 1989) have been widely used for assessment of the genetic diversity of many plants (e.g., Arif et al. 2010; Amiteye 2021) and animal (e.g., Arif and Khan 2009; Yang et al. 2013) species. Knowledge of the level of genetic polymorphism maintained in a particular population and information about the relatedness between individuals is essential for establishing efficient conservation management programs.

Unfortunately, the knowledge of the genetic diversity of *C. alpinum* and the whole genus *Cerastium* is minimal. The genetic studies for the genus *Cerastium* include the RAPD and SCAR markers for *C. arcticum* (Hagen et al. 2001) or AFLP technique for *C. dinaricum* (Kutnjak et al. 2014), *C. decalvans* (Niketić et al. 2022), *C. sylvaticum*, *C. subtriflorum* (Skubic et al. 2018), *C. hekuravense* (Caković et al. 2018) and *C. grandiflorum*, *C. decalvans* and *C. dinaricum* (Đurović et al. 2021). Most molecular studies of *C. alpinum* involved populations found in Svalbard, Greenland, Iceland, Norway, Sweden, and Finland and relied on analyses of isoenzymatic polymorphism (Nyberg Berglund et al. 2006). Nonetheless, they concentrated on the issue of hybridization and introgression between *C. alpinum* and other closely related species (i.e., *C. arcticum*, *C. nigrescens*) (Brysting and Borgen 2000; Hagen et al. 2002), as well as the identification and comparison of various isoenzymatic phenotypes from different populations to describe their geographic distribution and infer about the species' history (colonization routes) in Fennoscandia. Scheen et al. (2004) also wrote about the phylogenetic relationships and biogeography of the genus *Cerastium*. Brysting (2000) and Brysting et al. (2011) researched how the number of chromosomes might vary in some *Cerastium* species.

The iPBS (inter primer binding sites) method described by Kalendar et al. (2010) is based on the virtually universal presence of a tRNA complement as a reverse transcriptase primer binding site (PBS) in long terminal repeat (LTR) retrotransposons. It does not require prior knowledge of the sequence, making it suitable for many objects, including "orphan crops" with underdeveloped marker systems. This marker system has been successfully used for the molecular characterization of different plant species, such as chickpea (Andeden et al. 2013), pea (Baloch et al. 2015), quinoa (Hossein-Pour et al. 2019), grapevine (Milovanov et al. 2019), lianas from genus *Gnetum* (Doungous et al. 2020), onion (Khapilina et al. 2021a, 2021b), or pathogenic fungi, e.g., representatives of genus *Alternaria* (Turzhanova et al. 2020) and *Rhizoctonia* (Erper et al. 2021). This method has also found application in phylogenetic research of saffron (Bayat et al. 2018) and emmer wheat (Arystanbekkyzy et al. 2019). Our previous studies demonstrated the usefulness of iPBS markers for genetic diversity studies in non-model plants, such as *Colobanthus quitensis* (Koc et al. 2018), *Poa annua* (Androsiuk et al. 2019), and *Deschampsia antarctica* (Androsiuk et al. 2020), for which knowledge on their genome sequence and organization is limited. The retrotransposon-based DNA marker system used in these studies effectively assessed DNA polymorphism between individuals and among populations.

The objectives of this study were as follows: (1) to test the suitability of iPBS to detect genetic polymorphism within and among *C. alpinum* populations; (2) to estimate genetic differentiation and relationships among studied *C. alpinum* populations from different geographic regions; and (3) to evaluate the genetic condition of *C. alpinum* from the Babia Góra massif to provide information essential for the development of effective conservation strategies for the genetic resources of *C. alpinum* in Babia Góra.

Materials and methods

Material

The research material consisted of 133 individuals of *Cerastium alpinum* representing seven sampling sites (from now

on referred to as populations) from Poland, Switzerland, Sweden, and Svalbard (Norway) (Table 1). Plants from Switzerland and Sweden were obtained in a dried form. In the case of *C. alpinum* from Svalbard, the original material consisted of seeds collected during the Arctic expedition to Nicolaus Copernicus University Polar Station in Spitsbergen in 2012. Polish *C. alpinum* populations were collected in Babia Góra National Park after obtaining the permission of the Polish Ministry of Environment. All Polish sampling sites are scattered on the Babia Góra massif. To avoid damaging the *C. alpinum* individuals from these unique and valuable sampling sites, we collected only bags with mature seeds from 14 to 28 individuals from each localization; the sampled plants were ca. 10–20 m apart from each other to avoid collecting material representing the clonal genets as it is reported that *C. alpinum* can reproduce vegetatively (Rune 1957; Parusel 2008). Plants grown from seeds were maintained in the greenhouse of the Department of Plant Physiology, Genetics and Biotechnology, Faculty of Biology and Biotechnology at the University of Warmia and Mazury in Olsztyn, Poland. In the case of *C. alpinum* from Babia Góra massif, one seedling obtained from three to five germinated seeds originating from one seed bag, representing each sampled specimen, was selected for DNA extraction.

The molecular analysis was performed using 133 *C. alpinum* individuals representing seven populations, ranging from 14 to 28 individuals per population. The size of a given population limited the number of sampled individuals.

DNA extraction and iPBS genotyping

Genomic DNA from each individual was extracted with Genomic Mini AX Plant Spin kit (A&A Biotechnology). The amount and purity of isolated DNAs were estimated spectrophotometrically (NanoDrop). Additionally, the quality of DNA was verified on 1.5% (w/v) agarose staining with 0.5 µg/ml ethidium bromide.

Initially, 34 PBS primers (2074, 2076, 2079, 2080, 2085, 2217, 2220, 2221, 2224, 2228, 2229, 2231, 2232, 2237, 2238, 2240, 2241, 2242, 2249, 2251, 2253, 2272, 2277, 2373, 2374, 2376, 2378, 2381, 2389, 2391, 2393, 2395, 2399, 2415) were screened for *C. alpinum* (Kalendar et al. 2010), with the annealing temperature set at 56 °C for all primers. During this step, the quality of the band pattern was the factor that decided whether the particular PBS primer could be selected for further studies or not. To evaluate the quality of the band pattern for each tested PBS primer, we applied the scale of PCR efficiency described in the original paper by Kalendar et al. (2010). According to that scale, we could distinguish five types of iPBS band profiles: 0, no bands; 1, few and weak bands; 2, a few strong bands; 3, ≈10 strong bands; 4, many bands (good primer); 5, many strong and equally amplifying bands. Based on our observations for further steps of our study, we selected eight PBS primers with the highest score (Table S1). In the next step, we experimentally determined the optimal annealing temperature for selected primers, by performing PCR with the gradient of annealing temperature (50–60 °C). When the optimal annealing temperature for each selected PBS primer was determined (Table 2), the reproducibility of band profiles for each of selected PBS primer was verified. This verification was based on a comparison of the electrophoretic profiles for eleven individuals representing population 1 from Babia Góra, Poland. Data were generated and compared in two replicates. Gels were then checked to identify iPBS amplicons (bands) in one or both replicates.

Eight PBS primers that gave clearly identifiable and repeatable amplification products (bands) were selected for subsequent genetic diversity studies (Table 2). The iPBS amplification was performed according to the procedure described by Kalendar et al. (2010) with modifications (Androsiuk et al. 2015). Amplification products were analyzed by gel electrophoresis in 1.5% (w/v) agarose with

Table 1 The origin of *Cerastium alpinum* populations used in the study

Population	Sampling site	Coordinates (GPS)	Altitude (m.a.s.l.)	No. individuals
1	Poland Western Carpathians, Babia Góra National Park	49°34′28.54″ N; 19°32′13.95″ E	1592	23
2		49°34′27.04″ N; 19°32′4.64″ E	1620	14
3		49°34′27.21″ N; 19°31′57.80″ E	1618	28
4		49°34′24.14″ N; 19°31′47.27″ E	1629	16
5	Switzerland Graubünden, Samnaun-Gruppe, Piz Val Gronda	46°55′18″ N; 10°17′16″ E	2750	18
6	Svalbard (Norway) Svalbard archipelago, Spitsbergen, Kaffiøyra	78°40′34″ N; 11°49′38″ E	5	17
7	Sweden Norrbotten, Abisko National Park, Nuolja massif	68°22′04″ N; 18°44′2.60″ E	654–1164	17

Table 2 The sequence and specification of primers applied in the study

Primer	Sequence	Annealing temp. (°C)	Number of scored bands	Number of polymorphic bands
2217	5′-ACTTGGATGTCGATACCA-3′	56	35	24
2228	5′-CATTGGCTCTTGATACCA-3′	60	38	32
2229	5′-CGACCTGTTCTGATACCA-3′	60	26	17
2232	5′-AGAGAGGCTCGGATACCA-3′	58	34	30
2251	5′-GAACAGGCGATGATACCA-3′	60	34	31
2272	5′-GGCTCAGATGCCA-3′	56	30	17
2277	5′-GGCGATGATACCA-3′	56	34	34
2378	5′-GGTCCTCATCCA-3′	52	31	23
Total			262	208

1 × TBE electrophoresis buffer at 100 V for 2 h and visualized with 0.5 µg/ml ethidium bromide.

Data analysis

The amplified bands were scored as either 1 (present) or 0 (absent) across genotypes. Based on the obtained binary matrix of amplification products, the following genetic parameters were estimated with the use of GenAlEx 6.5 software (Peakall and Smouse 2006; 2012): the total number of bands per population (N_B), percentage of polymorphic bands (P), Shannon's information index (I), and expected heterozygosity (H_e).

The genetic structure of the studied populations was inferred based on the Bayesian model-based clustering method implemented in Structure ver. 2.3.4 (Pritchard et al. 2000). The model probabilistically assigns individual multilocus genotypes to a user-defined number of clusters (K), achieving linkage equilibrium within clusters (Pritchard et al. 2000). We ran ten replicates with a burn-in of 100,000 iterations and with 500,000 iterations for Markov chain Monte Carlo for K ranging from 1 to 10 possible clusters. The analysis, with the implemented admixture model, was conducted without prior information on the original population of each sampled individual. The optimal number of clusters and ad hoc statistics ΔK (Evanno et al. 2005) was evaluated in Structure Harvester ver.0.6.94 (Earl and vonHoldt 2012). The genetic subdivision patterns of the analyzed *C. alpinum* populations were also investigated by principal coordinate analysis (PCoA) based on the matrix of Euclidean distances between individuals from all analyzed populations. This analysis was performed in PAST software (Hammer et al. 2001).

Analysis of molecular variance (AMOVA) was performed on Arlequin 3.5 software (Excoffier et al. 2005). For that analysis, the iPBS dataset was treated as haplotypic, consisting of a combination of alleles at one or several loci (Excoffier et al. 2005). The significance of the fixation indices was tested using a nonparametric permutation approach (Excoffier et al. 1992). The significance test was based on 1023 permutations.

The possible effect of increased spatial distance on the genetic structure of the studied populations of *C. alpinum* was also investigated. The spatial genetic structure was estimated by testing the significance of isolation by distance (IBD) using a Mantel test with 9999 permutations of the relationship between the matrix of pairwise $F_{ST}/(1-F_{ST})$ and that of the logarithm of geographical distance between populations (Rousset 1997), using GenAlEx 6.5. The pairwise F_{ST} values were estimated in Arlequin 3.5 software, and their significance was also tested based on 110 permutations.

Tajima's D, Fu's F_S neutrality test, mismatch distribution, and demographic processes affecting populations were also performed using Arlequin 3.5 software (Excoffier et al. 2005).

Bottleneck ver. 1.2.02 (Cornuet and Luikart 1996) software was used to investigate the recent reduction of effective population size based on allele data frequencies for each *C. alpinum* population (Cornuet and Luikart 1996; Piry et al. 1999). This effect was studied using dominant markers in the infinite allele model (IAM) to test the mutation-drift hypothesis against the bottleneck hypothesis (Tero et al. 2003). The significance of potential bottleneck was estimated in the SIGN test, the standardized differences test, and the one-tailed Wilcoxon sign-rank test for heterozygosity excess with 10,000 simulations.

Results

Efficiency of applied PCR primers

Our analysis of *Cerastium alpinum* populations, using eight PBS primers, yielded 262 distinguished amplification products (bands) (Table 2, Table S2, File S1). The highest number of bands (38) was revealed by primer 2228, and the

lowest number of bands (26) was obtained for primer 2229. On average, 32.75 bands were obtained per primer. Of the identified loci, 208 (79.39%) were polymorphic (Table 2).

Of a total of 262 scored bands, 20 (7.63%) were represented as private bands that were found only within one population and absent in the others (Table 3). The highest number of private bands (4) was revealed by primer 2228. The Svalbard population (6) appeared as the most abundant in private bands—six of them were scored in individuals representing that population, while for population 2, no private bands were found.

Genetic diversity and differentiation

The iPBS markers revealed the presence of genetic polymorphism between individuals within a population and genetic variation among populations (Table 4). The number of iPBS amplification products ranged from 172 for population 5 to 236 for population 3. The highest number of polymorphic bands was scored for population 3 (45.80%) and the lowest for population 5 (13.36%). On average (over loci and populations), 25.35% of iPBS loci were polymorphic. The genetic variation was assessed with two parameters: Shannon's information index (I) and expected heterozygosity (H_e). Population 3 had the highest values for both parameters, while population 5 had the lowest values.

Table 3 Private alleles (bands) per locus and their respective frequencies revealed in studied *Cerastium alpinum* populations

Locus	Frequency	Population
2272_10	0.008	1
2228_11	0.105	3
2229_20	0.105	3
2232_23	0.030	3
2232_31	0.053	3
2272_21	0.015	3
2229_12	0.008	4
2251_21	0.120	4
2251_22	0.120	4
2217_10	0.053	5
2378_23	0.135	5
2228_12	0.120	6
2228_24	0.105	6
2229_16	0.120	6
2232_10	0.128	6
2272_18	0.015	6
2378_12	0.120	6
2217_4	0.060	7
2228_1	0.128	7
2378_26	0.008	7

Table 4 Population genetic characteristics for analyzed populations of *Cerastium alpinum*: number of bands (N_B), percentage of polymorphic bands (P), Shannon's information index (I), and expected heterozygosity (H_e)

Population	N_B	P (%)	I	H_e
1	215	25.95	0.140	0.095
2	219	28.24	0.148	0.099
3	236	45.80	0.241	0.162
4	218	22.14	0.107	0.070
5	172	13.36	0.076	0.051
6	206	25.57	0.095	0.060
7	191	16.41	0.088	0.060
Mean over loci and populations	208.14	25.35	0.128	0.085

PCoA indicates that 51.56% of the variation was explained by the first three components (21.6, 18.83, and 11.13%, respectively). The projection of the analyzed populations on the first two axes is shown in Fig. 1. The grouping revealed by PCoA pointed to the high degree of similarity between populations 5 and 7. The Babia Góra populations (1–4) can be put together into one group, but population 6 is very different from the others along coordinates 1 and 2.

The Bayesian clustering analysis performed in Structure and Structure Harvester software revealed that ΔK has a maximum at $K=7$, but a marked secondary peak at $K=5$ was also found (Fig. 2a). Both populations' structure bar plots generated by Structure software for $K=5$ (Fig. 2b) and $K=7$ (Fig. 2c) reveal a clear separation between populations 5, 6, and 7, which also appeared to be genetically homogenous. In the case of populations 1–4, a combination of diverse genotypes shared by all these populations can be observed, indicating admixture between these localizations. However, even then (for $K=7$), in the case of populations 1 and 4, the predicted values of the proportion of membership of these populations to separate, individual clusters were >0.7. The highest admixture was observed for populations 2 and 3, for which the predicted proportion of membership in the certain cluster was ≤ 0.499. For $K=5$, an analogous pattern of population subdivision is observed. Pairwise F_{ST} analysis was also performed. The highest F_{ST} value (0.839) was observed between populations 5 and 6, whereas the lowest (0.268) was between population 1 and population 3 (Table 5). All the pairwise F_{ST} values appeared to be significant ($p<0.05$).

The results of AMOVA showed that most of the identified genetic variation occurred among populations (61.69%), whereas the remaining 38.31% was attributed to variation within populations (Table 6).

The IBD analysis revealed a significant correlation between genetic divergence and the logarithm of the geographic distance between populations ($R^2=0.5073$,

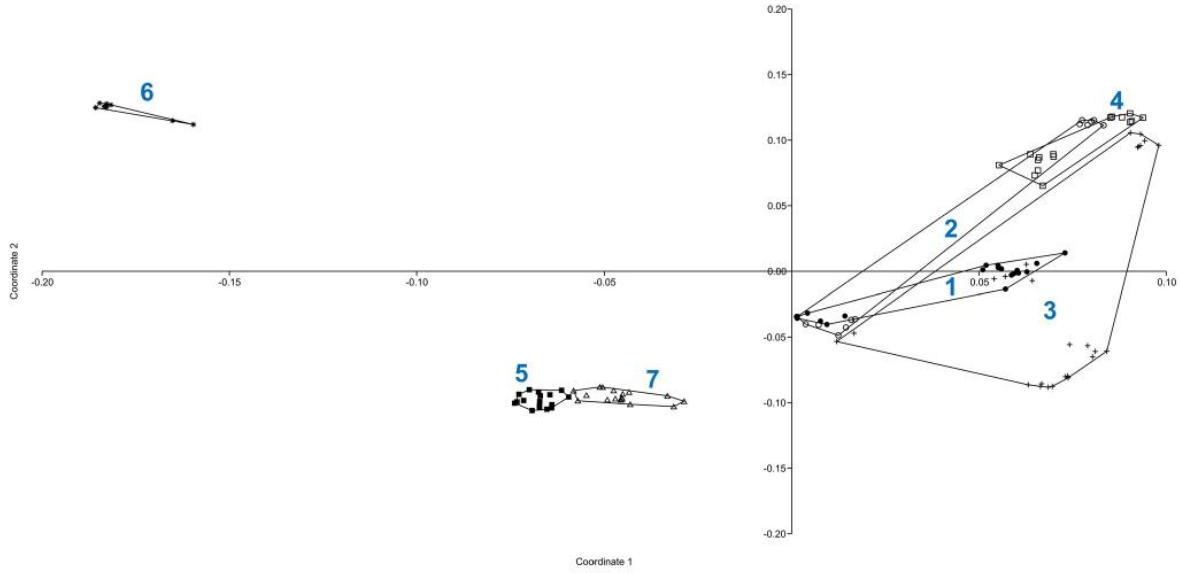


Fig. 1 Plot of coordinate 1 versus coordinate 2 obtained by principal coordinates analysis (PCoA) based on Euclidean distances between all individuals from seven *Cerastium alpinum* populations. Populations are numbered according to Table 1: 1 (filled circle); 2 (open circle); 3 (plus sign); 4 (open square); 5 (filled square); 6 (asterisk); 7 (open triangle)

$p = 0.02$), which means that the genetic distance between populations, expressed as pairwise $F_{ST}/(1 - F_{ST})$ values, increases with the spatial distance between them (Fig. 3, Table 5).

Neutrality tests and demography

Tajima's D did not show any deviation from zero except for population 6, which was negative and significant. Fu's F_S was negative for all populations (except population 6), but only for populations 5 and 7 did it significantly deviate from zero (Table 7). The F_S statistic interpretation was performed according to Fu (1997), who indicated that it should be considered significant at the 5% level if its p value is below 0.02 but not below 0.05. In the mismatch analysis for population growth and geographic expansion, SSD values were not statistically significant, and all samples had a low raggedness index (Table 8).

Three tests (SIGN, standardized differences, and Wilcoxon sign-rank) were used to analyze the bottleneck effect in studied populations (Table 9). In each case, the values were statistically significant, except for population 6 (SIGN test), population 4 (standardized differences test), and populations 4 and 6 (one-tailed Wilcoxon test for heterozygosity excess). Also, the results of all three tests showed that there were significant signs of bottleneck effects in populations 1, 2, 3, 5, and 7.

Discussion

Genetic diversity and differentiation of *Cerastium alpinum*

Cerastium alpinum occurs mainly in the Northern Hemisphere, North America (Canada and Greenland), Europe, and Western Siberia (Hultén 1956). Our research focused on the only natural Polish stand of *C. alpinum* (in the Babia Góra National Park) and populations of the species from Switzerland (Piz Val Gronda, Alps), Sweden (Nuolja massif, Abisko National Park), and Norway (Svalbard, Kaffiøyra).

Cerastium alpinum is a common perennial in Fennoscandia. In alpine regions, it has a continuous distribution on alpine heat and serpentine soils, while in boreal forests, it has a scattered distribution on ultramafic soils and steep slopes (Nyberg Berglund et al. 2001; Nyberg Berglund and Westerbergh 2001). Ultramafic soils (often called serpentine) are characterized by low concentrations of plant nutrients and high, potentially toxic, concentrations of Mg and Ni (Proctor 1999). Although Mg is one of the vital nutrients associated with chlorophyll health, the increased concentration of Mg creates conflict with other ions, like Ca, and may decrease their uptake (Nyberg Berglund et al. 2003). In the case of Ni, increased concentrations of these ions were reported to be responsible for the inhibition of

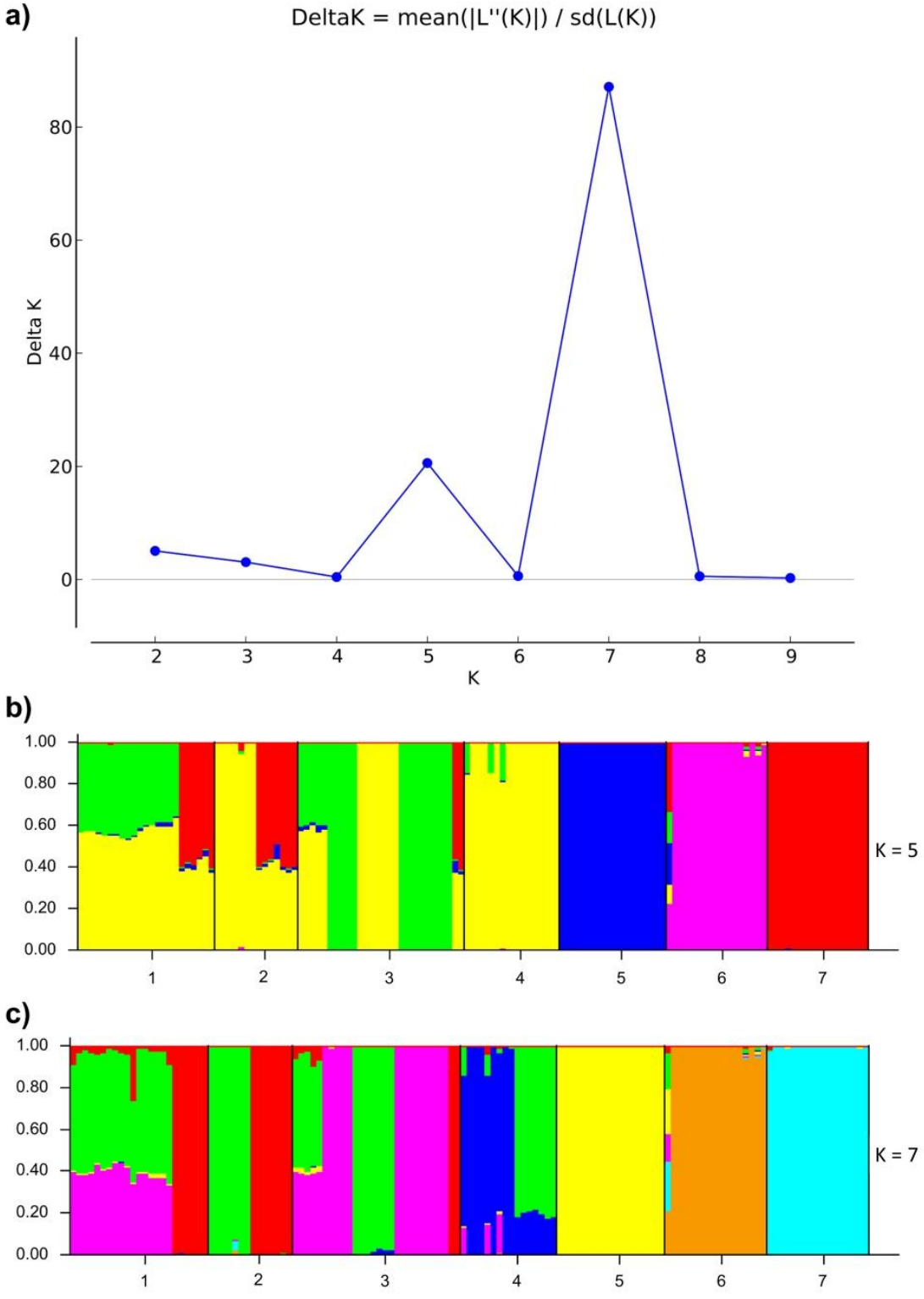
cell division at root meristems (Robertson 1985). Moreover, nickel's negative effect on photosynthesis and transpiration regulation was observed (Carlson et al. 1975). Research on *C. alpinum* showed that individuals growing in serpentine soils produce a lower number of seeds, display a more dwarfed structure, and show a higher tolerance to Ni and Mg stress (Grundt et al. 1999; Nyberg Berglund et al. 2003). In the Alps, the species has a scattered distribution and can be found on stony grasslands and ridges of the alpine and subalpine zones (The National Data and Information Center on the Swiss Flora, http://www.infoflora.ch). In Poland, *C. alpinum* occurs only in one place, on the Babia Góra massif in the Babia Góra National Park. This species is critically endangered in Poland due to its small range and geographical isolation (Parusel 2001).

Environmental and geographic isolation can strongly influence a population's genetic structure and diversity (Wang and Bradburd 2014). This isolation can lead to decreased genetic variation and increased inter-population genetic differences due to reduced gene flow. That could be especially observed in alpine landscapes, which, due to their heterogeneous topography, may create numerous geographically and ecologically isolated habitats for alpine plants (Körner 2003). The terrain relief with steep valleys and high mountain ridges may limit gene flow and lead to more genetic differentiation between isolated plant populations compared to plants from less isolated habitats or lower altitudes (Till-Bottraud and Gaudeul 2002). Many factors, such as population size and the mating system, can affect genetic variation. Genetic diversity in small populations may also be limited by genetic drift and/or bottleneck during colonization by a small number of colonists and subsequent inbreeding (Ellstrand and Elam 1993; Lowe et al. 2004; Stöcklin et al. 2009). As a result, the lower ability of a population to respond to environmental changes might be observed, associated with the accumulation of deleterious recessive alleles and increased mortality of seeds and seedlings, which increases the probability of population extinction (Young et al. 1996; Lowe et al. 2005). High-mountain areas are also characterized by many stress factors, such as extreme temperatures, shorter vegetative periods, strong winds, heavy rainfall, and increased UV radiation, which affect plant growth and development (Körner 2003). Adaptive evolution driven by these factors favors individuals (genotypes) with traits that enhance survival and reproduction and consequently shapes the genetic diversity of populations (Bertel et al. 2018). The reproductive system also influences the genetic diversity of a plant population. Outcrossing species usually have higher within-population diversity and lower differentiation among populations than selfing species (Loveless and Hamrick 1984).

All the factors mentioned above also exist for *C. alpinum* on the Babia Góra massif. However, the unique character of *C. alpinum* from that location has not received adequate attention, which is reflected in a deficient number of published studies. The earliest scientific report describing the occurrence of *C. alpinum* on Babia Góra was released in 1906 by Zapałowicz (1906). Later reports included more detailed descriptions of the species' morphology, occurrence, and abundance (Parusel 1993, 2001, 2008, 2014; Borysiak and Stachnowicz 2004, 2018). So far, no genetic studies have been performed on *C. alpinum* from Babia Góra, so we lack knowledge of its historical and current genetic composition. Genetic evaluation of these valuable genetic resources is essential not only to complement our state of knowledge but also for developing and managing effective conservation strategies for the species in Poland. Therefore, genetic characteristics of the *C. alpinum* from the Babia Góra massif seem to be a very urgent task, especially in light of the present and possible future threats, for example, intensive and increasing tourism in this area and the consequences of global climate change. The mountain environment is an ideal model for studying physiological and adaptive responses to global climate change (Körner 2003). In the next 50–80 years, the amount of rain and snow in Europe's alpine areas is expected to go down, while the temperature is expected to go up (Engler et al. 2011). In high-elevation areas, the growing season could even get 60 days longer (Raible et al. 2006).

Due to progressive climate warming, plants found at high altitudes are particularly vulnerable to increasing competition due to the upward migration of species previously located at lower altitudes (Lenoir et al. 2008; Chen et al. 2011; Rumpf et al. 2018). Unfortunately, no such research has been carried out so far for the Babia Góra massif and its vicinity. However, the increased temperature and changes in precipitation appeared as the most likely drivers of changes in plant species composition observed within the last 90 years in the Tatra Mountains (Czortek et al. 2018), the mountain range located about 50 km southeast of the Babia Góra massif. Czortek et al. (2018) found that species composition changed the most in snowbeds, mylonite grasslands, and hygrophilous tall herb communities.

To the best of our knowledge, no investigations using iPBS markers for *C. alpinum* or any other member of the genus *Cerastium* have been carried out. Results of our study proved the suitability of this technique for evaluation of genetic diversity of the species. Two hundred and eight (79.39%) of the detected amplification products in our study on the genetic diversity of *C. alpinum* using the iPBS method were polymorphic, and the average polymorphism was found to be at 25.35% (across loci and populations). From 13.36% in population 5 to 45.80% in population 3, this parameter's value was found. There were even higher levels of polymorphism in earlier studies using iPBS markers, such as 82.35% for *Nicotiana tabacum* (Yaldiz et al. 2018), 85.7%


a)
DeltaK = mean(|L''(K)|) / sd(L(K))
Delta K
80
60
40
20
0
2
3
4
5
6
7
8
9
K
b)
1.00
0.80
0.60
0.40
0.20
0.00
1
2
3
4
5
6
7
K = 5
c)
1.00
0.80
0.60
0.40
0.20
0.00
1
2
3
4
5
6
7
K = 7

◂**Fig. 2** **a** The uppermost hierarchical level of genetic structure of *Cerastium alpinum* using STRUCTURE software (Pritchard et al. 2000); the optimal *K* value generated by Structure Harvester: **b** $K=5$ and **c** $K=7$

for *Psidium guajava* (Mehmood et al. 2013), 86.3% for *Vitis* varieties (Guo et al. 2014), 97.4% for *Myrica rubra* (Fang-Yong and Ji-Hong 2014), and 98.7% for Peruvian rosewood (Baloch et al. 2022). For Antarctic plants, however, a substantially lower level of polymorphism was noted, such as an average of 12.5% for *Deschampsia antarctica* (Androsiuk et al. 2020) and 9.57% for *Colobanthus quitensis* (Koc et al. 2018). Even lower variability between clones of the apricot cultivar was found by Baránek et al. (2012) at 4.88%. Twenty of the total 208 polymorphic bands were found to be private bands, meaning they only appeared in one population and were not present in the others. Populations 3 and 6 had the most private alleles (5 and 6, respectively), but population 1 only had one such amplification result. The only population for which none of these factors was scored was population 2. Unfortunately, these unique amplification products cannot be treated as diagnostic markers as they were not observed in all individuals representing particular populations.

Our study revealed low genetic diversity within populations of *C. alpinum* (mean $H_e=0.085$). The highest value of that parameter was observed for population 3 ($H_e=0.162$) and the lowest for population 5 ($H_e=0.051$). Several authors found similarly low levels of heterozygosity in other isolated plant populations. The arctic-alpine species *Dryas octopetala* forms isolated populations in mountainous regions on northern Greece's southernmost edge of the species' distribution range. Studies by Varsamis et al. (2021) aimed at evaluating genetic variation within three selected populations of the species and analyzing its overall genetic structure. The results indicated relatively low intrapopulation genetic diversity (mean value of $H_e=0.156$). In *Pilosella alpicola* subsp. *ullepitschii*, an endemic plant in the Carpathians, in the Western Carpathians (Slovakia and Poland), this species occurs at many sites, but only four sites are known in the Eastern and Southern Carpathians (Romania). Studies by Šingliarová et al. (2008) with the use of allozyme markers indicated a significant loss of genetic diversity ($H_e=0.134$) in isolated populations of the Eastern and Southern Carpathians in comparison to populations from the Western part ($H_e=0.235$). The results suggested that this population experienced a robust genetic bottleneck, probably due to a founder effect. On the other hand, studies of the influence of alpine habitat isolation on the genetic diversity of the rare and isolated population of *Campanula thyrsoides* showed a large genetic variation ($H_e=0.762$; Ægisdóttir et al. 2009; $H_e=0.714$; Frei et al. 2012) within the populations and a relatively low inbreeding coefficient. This rare monocarpic perennial's high genetic diversity within its own population is best explained by the fact that its long-lived individuals do not breed with each other and that its generations overlap.

Significant traces of a recent reduction in effective population size were noted for all *C. alpinum* populations in this study. However, negative and significant Fu's F_S statistics were observed only for populations 5 and 7, whereas negative and significant Tajima's *D* was only for population 6, confirming a demographic expansion of the populations mentioned above. It might suggest that the studied *C. alpinum* populations might experience different demographic histories or extinction-recolonization events.

The results of our study revealed that despite a low level of polymorphism revealed by iPBS markers (on average, 25.35% of iPBS bands revealed differences between individuals from the particular population) and low genetic diversity (average $H_e=0.085$), the analyzed populations of *C. alpinum* were characterized by very high genetic differentiation ($F_{ST}=0.6169$, results of Structure software which assigned individuals into seven probable clusters, and high population dispersal revealed by PCoA). The significant pairwise F_{ST} values between all pairs of studied populations and the significant results of IBD analysis suggest that this population differentiation could be attributed to limited gene flow between them. This is, however, a surprising observation in the case of *C. alpinum* populations from the Babia Góra massif, which are not separated by high geographical distances (the pairwise distance between populations from this site ranges from 0.14 to 0.55 km). Therefore, the most likely explanation is the seed dispersal method (gravity) and the fact that the *C. alpinum* seeds are not morphologically adapted to wind dispersal.

Isoenzymatic analysis of *C. alpinum* from 31 populations from Fennoscandia also revealed their high genetic differentiation ($F_{ST}=0.526$) (Nyberg Berglund et al. 2006). The authors also observed that some populations were fixed for some alleles and/or were characterized by inbreeding. The isolation of these populations was pointed out as the most likely explanation for the observed pattern of genetic variation (Nyberg Berglund et al. 2006).

Implication for conservation

Genetic diversity is a fundamental source of biodiversity and can be defined as any measure that quantifies the magnitude of genetic variability within a population (Hughes et al. 2008). Therefore, evaluating genetic diversity and population differentiation is vital for species conservation (Carvalho et al. 2019). There are two complementary conservation strategies for plant species: in situ (on-site) conservation that protects species in their native habitat, while ex situ (off-site) protects endangered species outside their natural habitat in artificial environments, e.g., botanical gardens, arboreta,

Table 5 Pairwise F_{ST} values (below diagonal) and geographic distance in kilometers (above diagonal) between studied *Cerastium alpinum* populations

	1	2	3	4	5	6	7
1	–	0.19	0.33	0.55	745.89	3254.63	2092.4
2	0.304	–	0.14	0.36	745.7	3254.67	2092.44
3	0.268	0.300	–	0.23	745.57	3254.65	2092.44
4	0.538	0.381	0.375	–	745.34	3254.74	2092.53
5	0.680	0.658	0.559	0.777	–	3535.48	2434.46
6	0.756	0.718	0.661	0.786	0.839	–	1166.2
7	0.657	0.625	0.537	0.753	0.773	0.831	–

Table 6 Partitioning of diversity found in *Cerastium alpinum* from all analyzed populations using AMOVA (F_{ST}=0.6169)

Source of variation	Degrees of freedom	Sum of squares	Variance components	Percentage of variation
Among population	6	2265.891	19.421	61.69
Within population	126	1519.839	12.062	38.31
Total	132	3785.729	31.483	

Significance tests (1023 permutations); p=0.001

or as a germplasm collection in gene banks (Maunder et al. 2001; Schoen and Brown 2001; Heywood and Iriondo 2003; Guerrant et al. 2004; Havens et al. 2006; Hardwick et al. 2011; Miller et al. 2016). The general concept of in situ conservation is the conservation of living resources in the surroundings where they have developed and to maintain certain species in a dynamic relationship with the natural habitat (Edwards and Kelbessa 1999). This strategy enables the conservation of a large amount of natural genetic diversity. However, it requires adequate human and financial resources to ensure the effectiveness of protected areas in maintaining biodiversity (Zegeye 2017). However, because of climate change, possible pressure from invasive species, or habitat degradation, we cannot always completely prevent the extinction of a species in its natural environment. In that case, ex situ conservation seems to be the only alternative.

The popularity of seed banking is due to the ease of storage and immediate access for scientists, as well as resistance to habitat destruction, diseases, and predators (Roberts 1991; Schoen and Brown 2001). In addition, the deposited material can be used for species reintroduction to their natural habitat or as a genetic rescue, i.e., to increase the diversity of genetically depleted populations (Schoen and Brown 2001). However, despite the many advantages of ex situ conservation, it is not indifferent to the protected resources—it limits genetic variability and eliminates natural processes (e.g., evolutionary and ecological processes) that act on the particular species and are responsible for the adaptation of the species to changing environmental conditions. One

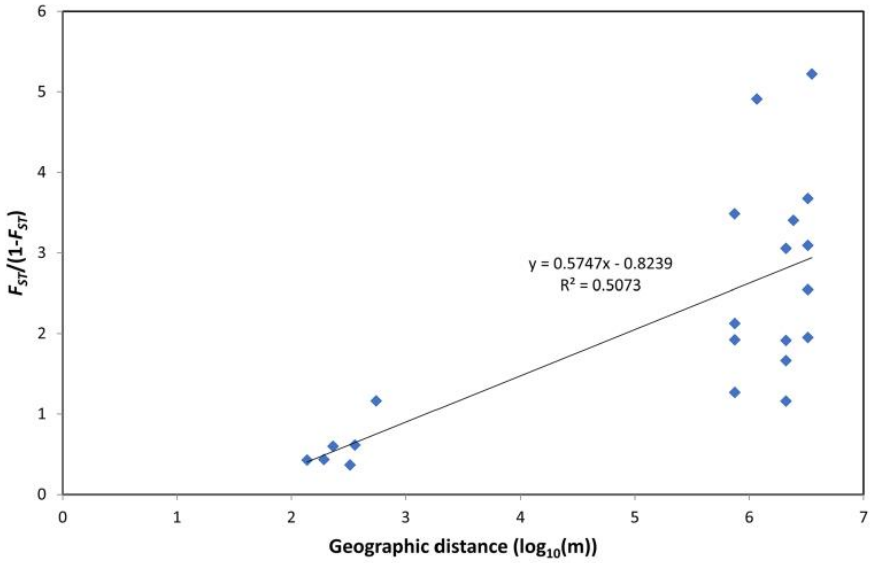


Fig. 3 Plot of the isolation by distance (IBD) analysis. The Mantel test scatterplot shows the pairwise $F_{ST}/(1-F_{ST})$ as a function of the logarithm of the geographical distances between studied populations of *Cerastium alpinum*

Table 7 Tajima's D test and Fu's F_S neutrality tests of seven populations of *Cerastium alpinum*

Test	Description	Populations							Statistics	
		1	2	3	4	5	6	7	Mean	SD
Tajima's D test	S	68	74	120	58	35	67	43	66.429	27.501
	P_i	22.506	33.341	42.942	21.967	12.046	11.735	14.353	22.698	11.740
	Tajima's D	0.877	1.919	1.523	1.093	0.745	−1.719	0.533	0.710	1.171
	Tajima's D p value	0.854	0.989	0.952	0.900	0.809	0.031	0.757	0.756	0.330
Fu's F_S test	Theta p_i	22.506	33.341	42.942	21.967	12.046	11.735	14.353	22.698	11.740
	Exp. no. of alleles	16.101	11.838	21.756	12.233	11.315	10.811	11.490	13.649	3.983
	F_S	−3.047	−0.703	−5.073	−4.488	−6.382	0.290	−7.103	−3.787	2.786
	F_S p value	0.125	0.309	0.033	0.023	0.007	0.564	0.003	0.152	0.212

S, number of segregating sites; P_i, mean number of pairwise differences; Theta p_i, Watterson's theta based on S; F_S, Fu's F_S; *SD*, standard deviation

Table 8 Estimates of mismatch analysis of seven populations of *Cerastium alpinum*

Model	Statistics	Populations							Mean	SD
		1	2	3	4	5	6	7		
Demographic expansion	SSD	0.014	0.046	0.009	0.022	0.008	0.035	0.010	0.021	0.015
	Model (SSD) p value	0.380	0.130	0.740	0.460	0.370	0.200	0.400	0.383	0.196
	Raggedness index	0.009	0.025	0.005	0.014	0.013	0.048	0.023	0.020	0.014
	Raggedness p value	0.790	0.620	0.890	0.890	0.590	0.370	0.270	0.631	0.245
Spatial expansion	SSD	0.014	0.038	0.011	0.023	0.008	0.034	0.010	0.020	0.012
	Model (SSD) p value	0.560	0.230	0.510	0.450	0.280	0.160	0.300	0.356	0.151
	Raggedness index	0.009	0.025	0.005	0.014	0.013	0.048	0.023	0.020	0.014
	Raggedness p value	0.910	0.910	1.000	0.990	0.600	0.470	0.230	0.730	0.300

SD, standard deviation; *SSD*, sum of squared deviations; *Raggedness index*, Harpending's raggedness index

example of ex situ conservation efforts that did not meet expectations is associated with *Cochlearia polonica* (Brassicaceae). It is an endemic plant species that is extinct in the wild in Poland and is known only from one transplanted population (by the Centuria River, right bank tributary of the Biała Przemsza river, in Southern Poland) with similar habitat conditions to its native site. Some individuals were transplanted to the Botanical Garden of the Polish Academy of Sciences in Warsaw to protect the species from complete extinction. After 18 years of ex situ conservation, Rucińska and Puchalski (2011) studied genetic diversity within the garden and source population. They observed a decrease in genetic diversity within the botanical garden population (Nei's gene diversity = 0.1007) compared with the source population (0.1558). Similar results were observed for *Silene otites* (Lauterbach et al. 2012) and *Metasequoia glyptostroboides* (Li et al. 2005). So, in situ and ex situ conservation strategies should not be seen as alternatives, but rather as ways to protect endangered species that work well together. Both have their pros and cons, but both have been shown to be important and effective ways to do so.

Populations with more genetic diversity are more likely to be able to adapt to changing environmental conditions because they have more genetic variation (Willi et al., 2006). This makes them better candidates for conservation efforts. But there are also examples of both plant (e.g., *Arabidopsis lyrata*; Huber et al. 2014, Takou et al. 2021) and animal (e.g., *Oncorhynchus mykiss*; Willoughby et al. 2018, *Salvelinus fontinalis*; Yates et al. 2019) species that show that even populations with very little genetic diversity (e.g., because of a bottleneck) can have distinct signatures of local adaptation. Therefore, the low genetic diversity of *C. alpinum* populations from Babia Góra should not be treated as a direct threat to the population's survival. What is more, *C. alpinum* from Babia Góra is characterized by higher genetic diversity than other studied populations of the species from Sweden, Switzerland, or Norway (Svalbard). Even lower values of H_e were observed with the application of the same DNA markers (iPBS) for populations of *C. quitensis* ($H_e = 0.036$; Koc et al. 2018), *D. antarctica* ($H_e = 0.021$; Androsiuk et al. 2020), and *Poa annua* ($H_e = 0.063$; Androsiuk et al. 2019). In all the above-mentioned cases, the populations' adaptation to

Table 9 Testing the bottleneck versus mutation-drift equilibrium hypotheses for all analyzed populations (IAM mutation model)

Population	SIGN test	Standardized test	Wilcoxon test		
			One tail for heterozygosity deficiency	One tail for heterozygosity excess	Two tails for heterozygosity excess and deficiency
1	Hee = 29.73 Hd = 22 He = 45 p = 0.00014	T2 = 5.423 p = 0.00000	1.00000	0.00000	0.00000
2	Hee = 33.37 Hd = 18 He = 55 p = 0.00000	T2 = 2.500 p = 0.00621	0.99739	0.00266	0.00531
3	Hee = 57.01 Hd = 28 He = 91 p = 0.00000	T2 = 7.053 p = 0.00000	1.00000	0.00000	0.00000
4	Hee = 27.72 Hd = 19 He = 38 p = 0.00457	T2 = 1.355 p = 0.08769	0.70143	0.30131	0.60262
5	Hee = 17.28 Hd = 6 He = 28 p = 0.00014	T2 = 3.897 p = 0.00005	0.99994	0.00007	0.00014
6	Hee = 32.89 Hd = 33 He = 33 p = 0.53768	T2 = − 2.838 p = 0.00227	0.00127	0.99876	0.00253
7	Hee = 20.94 Hd = 12 He = 30 p = 0.00376	T2 = 3.098 p = 0.00097	0.99087	0.00946	0.01892

Hee, expected heterozygosity excess; *Hd*, heterozygosity deficiency; *He*, heterozygosity excess

harsh climatic conditions of the Maritime Antarctic was indicated as the most probable explanation of observed patterns of genetic variation. This may also be true for *C. alpinum* populations whose genetic pools have been subjected to strong selection from climatic and edaphic factors found in mountainous habitats.

Considering the low genetic diversity and high population differentiation revealed in this study, we suggest that all *C. alpinum* populations from the Babia Góra massif should be treated the same during the development of its conservation strategy, as only then most of the genetic variability can be preserved. Furthermore, the results of botanical inventory studies of the Babia Góra massif clearly show that the area of occurrence of *C. alpinum* is shrinking (Perzanowska 2012), so there is an urgent need to develop an efficient strategy for species protection. Fortunately, all Polish *C. alpinum* sites are located within the protected area of Babia Góra National Park, which will help to maximize the effects of activities associated with species conservation. Nevertheless, due to the constant threat from tourism (trampling, unauthorized plant collection), regular monitoring of the species abundance and population condition should be continued.

Conclusions

Currently, we observe the increasing recognition of the value of genetic data to support the decisions associated with developing and managing species conservation strategies. In this paper, the retrotransposon-based DNA markers (iPBS) proved helpful in revealing the genetic variation of *Cerastium alpinum* populations. The results showed a relatively low level of genetic diversity and high population differentiation. Furthermore, the genetic structure analyses reveal that the *C. alpinum population* from Babia Góra forms a distinct and heterogeneous group with greater genetic diversity than populations from other regions. Although *C. alpinum* populations from Babia Góra revealed significant traces of a recent bottleneck, like all other studied populations of the species, they still represent valuable genetic resources deserving protection. This has become especially important

in light of the recent observations pointing to the shrinking area of its occurrence. Therefore, immediate action is required to develop an efficient conservation strategy to protect this species's only Polish location.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s13353-022-00731-x.

Acknowledgements We are very grateful to Prof. Peter Schönswetter from University of Innsbruck (Austria), Prof. Xiao-Ru Wang from Umeå University (Sweden), Prof. Michał Węgrzyn from Jagiellonian University in Kraków (Poland) for their help in obtaining the research material, and employees of the Babia Góra National Park for their great help in our field work. Preliminary report on this study was presented in the form of poster at the 59th Congress of the Centenary of the Polish Botanical Society Warsaw, June 26–July 3, 2022 (Milarska S., Androsiuk P., Giełwanowska I. "Genetic variation of critically endangered *Cerastium alpinum* L. (Caryophyllaceae) population from Babia Góra National Park, Poland"). Abstracts of lectures and posters presented during the Congress are available at https://pbsociety.org.pl/repository/bitstream/handle/20.500.12333/320/2022.0003.pdf?isAllowed=y&sequence=3.

Author contribution Conceptualization: P.A. and I.G.; supervision: P.A.; methodology: P.A.; resources: P.A., K.L., and I.G.; investigation: S.E.M.; data analysis and interpretation: S.E.M., P.A., and P.T.B.; data curation: P.A.; writing—original draft preparation: S.E.M. and P.A.; visualization: S.E.M. and P.A.; writing—review and editing: P.T.B., K.L., and I.G. All authors have read and agreed to the published version of the manuscript.

Declarations

Ethics approval The experiments in this study comply with the current laws of Poland, Switzerland, Sweden, and Norway.

Consent for publication The corresponding author declares, on behalf of all co-authors, that the work described is original, previously unpublished research, and not under consideration for publication elsewhere.

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Ægisdóttir HH, Kuss P, Stöcklin J (2009) Isolated populations of a rare alpine plant show high genetic diversity and considerable population differentiation. Ann Bot 104:1313–1322. https://doi.org/10.1093/aob/mcp242

Amiteye S (2021) Basic concepts and methodologies of DNA marker systems in plant molecular breeding. Heliyon 7:e080932. https://doi.org/10.1016/j.heliyon.2021.e08093

Andeden EE, Baloch FS, Derya M, Kilian B, Ozkan H (2013) iPBS-retrotransposons-based genetic diversity and relationship among wild annual *Cicer* species. J Plant Biochem Biotechnol 22:453–466. https://doi.org/10.1007/s13562-012-0175-5

Androsiuk P, Chwedorzewska K, Szandar K, Giełwanowska I (2015) Genetic variability of *Colobanthus quitensis* from King George Island (Antarctica). Pol Polar Res 36:281–295. https://doi.org/10.1515/popore-2015-0017

Androsiuk P, Koc J, Chwedorzewska KJ, Górecki R, Giełwanowska I (2019) Retrotransposon-based genetic variation of *Poa annua* populations from contrasting climate conditions. PeerJ 7:e6888. https://doi.org/10.7717/peerj.6888

Androsiuk P, Chwedorzewska KJ, Dulska J, Milarska S, Giełwanowska I (2020) Retrotransposon-based genetic diversity of *Deschampsia antarctica* Desv. from King George Island (Maritime Antarctic). Ecol Evol 11:648–663. https://doi.org/10.1002/ece3.7095

Arif IA, Khan HA (2009) Molecular markers for biodiversity analysis of wildlife animals: a brief review. Anim Biodivers Conserv 32(1):9–17. https://doi.org/10.32800/abc.2009.32.0009

Arif IA, Bakir MA, Khan HA, Al Farhan AH, Al Homaidan AA, Bahkali AH, Sadoon MA, Shobrak M (2010) A brief review of molecular techniques to assess plant diversity. Int J Mol Sci 11:2079–2096. https://doi.org/10.3390/ijms11052079

Arystanbekkyzy M, Nadeem MA, Aktaş H, Yeken MZ, Zencirci N, Nawaz MA, Ali F, Haider MS, Tunc K, Chung G, Baloch FS (2019) Phylogenetic and taxonomic relationship of Turkish wild and cultivated emmer (*Triticum turgidum* ssp. *dicoccoides*) revealed by iPBS-retrotransposons markers. Int J Agri Biol 21:155–163. https://doi.org/10.17957/IJAB/15.0876

Baloch FS, Alsaleh A, Sáenz de Miera LE, Hatipoğlu R, Çiftçi V, Karaköy T, Yıldız M, Özkan H (2015) DNA based iPBS-retrotransposon markers for investigating the population structure of pea (*Pisum sativum*) germplasm from Turkey. Biochem Syst Ecol 61:244–252. https://doi.org/10.1016/j.bse.2015.06.017

Baloch FS, Guizado SJV, Altaf MT, Yüce I, Çilesiz Y, Bedir M, Nadeem MA, Hatipoglu R, Gómez JCC (2022) Applicability of inter-primer binding site iPBS-retrotransposon marker system for the assessment of genetic diversity and population structure of Peruvian rosewood (*Aniba rosaeodora* Ducke) germplasm. Mol Biol Rep 49:2553–2564. https://doi.org/10.1007/s11033-021-07056-8

Baránek M, Meszáros M, Sochorová J, Čechová J, Raddová J (2012) Utility of retrotransposon-derived marker systems for differentiation of presumed clones of the apricot cultivar Velkopavlovická. Sci Hortic 143:1–6. https://doi.org/10.1016/j.scienta.2012.05.022

Bayat M, Amirnia R, Özkan H, Gedik A, Ates D, Tanyulac B, Rahimi M (2018) Diversity and phylogeny of saffron (*Crocus sativus* L.) accessions based on iPBS markers. Genetika 50:33–44. https://doi.org/10.2298/GENSR1801033B

Bertel C, Rešetnik I, Frajman B, Erschbamer B, Hülber K, Schönswetter P (2018) Natural selection drives parallel divergence in the mountain plant *Heliosperma pusillum* s.l. Oikos 127:1355–1367. https://doi.org/10.1111/oik.05364

Borysiak J, Stachnowicz W (2018) Flora roślin naczyniowych Babiogórskiego Parku Narodowego (Karpaty Zachodnie, Polska). The vascular flora of the Babia Góra National Park (The Western Carpathian, Poland) In: Holeksa J, Szwagrzyk J (ed) Rośliny Babiej Góry. Grafpol Agnieszka Blicharz-Krupińska, Wrocław-Zawoja, pp 63–102

Borysiak J, Stachnowicz W (2004) Zarys flory roślin naczyniowych Babiogórskiego Parku Narodowego. Babiogórski Park

Narodowy. Monografia Przyrodnicza, Outline of the vascular flora of the Babiogórski National Park (Poland)

Brysting AK (2000) Chromosome number variation in the polyploid *Cerastium alpinum-C. arcticum* complex (Caryophyllaceae). Nord J Bot 20:149–156. https://doi.org/10.1111/j.1756-1051.2000.tb01558.x

Brysting AK, Borgen L (2000) Isozyme analysis of the *Cerastium alpinum-C. arcticum* complex (Caryophyllaceae) supports a splitting of *C. arcticum* Lange. Plant Syst Evol 220:199–221. https://doi.org/10.1007/BF00985046

Brysting AK, Mathiesen C, Marcussen T (2011) Challenges in polyploid phylogenetic reconstruction: a case story from the arctic-alpine *Cerastium alpinum* complex. Taxon 60:333–347. https://doi.org/10.1002/tax.602004

Caković D, Stešević D, Schönswetter P, Frajman B (2018) Long neglected diversity in the Accursed Mountains of northern Albania: *Cerastium hekuravense* is genetically and morphologically divergent from *C. dinaricum*. Plant Syst Evol 304:57–69. https://doi.org/10.1007/s00606-017-1448-1

Carlson RW, Bazzazet FA, Rolfe GL (1975) The effect of heavy metals on plants. II. Net photosynthesis and transpiration of whole corn and sunflower plants treated with Pb, Cd. Ni and Tl Environ Res 10:113–120. https://doi.org/10.1016/0013-9351(75)90077-8

Carvalho YGS, Vitorino LC, de Souza UJB, Bessa LA (2019) Recent trends in research on the genetic diversity of plants: implications for conservation. Divers 11:62. https://doi.org/10.3390/d11040062

Chen IC, Hill JK, Ohlemüller R, Roy DB, Thomas CD (2011) Rapid range shifts of species associated with high levels of climate warming. Sci 333:1024–1026. https://doi.org/10.1126/science.1206432

Clapham AR, Tutin TG, Moore DM (1987) Flora of the British Isles, 3rd edn. Cambridge University Press, Cambridge

Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics 144:2001–2014. https://doi.org/10.1093/genetics/144.4.2001

Czortek P, Kapfer J, Delimat A, Eycott AE, Grytnes JA, Orczewska A, Ratyńska H, Zięba A, Jaroszewicz B (2018) Plant species composition shifts in the Tatra Mts as a response to environmental change: a resurvey study after 90 years. Folia Geobot 53:333–348. https://doi.org/10.1007/s12224-018-9312-9

Doungous O, Kalendar R, Filippova N, Ngane BK (2020) Utility of iPBS retrotransposons markers for molecular characterization of African Gnetum species. Plant Biosyst 154:587–592. https://doi.org/10.1080/11263504.2019.1651782

Đurović SZ, Temunović M, Niketić M, Tomović G, Schönswetter P, Frajman B (2021) Impact of Quaternary climatic oscillations on phylogeographic patterns of three habitat-segregated *Cerastium* taxa endemic to the Dinaric Alps. J Biogeogr 48:2022–2036. https://doi.org/10.1111/jbi.14133

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359–361. https://doi.org/10.1007/s12686-011-9548-7

Edwards S, Kelbessa E (1999) Indicators to determine the level of threat to tree species. In: Edwards S, Demissie A, Bekele T, Haase G (ed) Forest genetic resources conservation: principles, strategies and actions. Proceedings of the national forest genetic resources conservation strategy workshop, 21-22 June 1999, Addis Abeba, Ethiopia. Institute of Biodiversity Concervation and Research, Addis Abeba, pp 101–133

Ellstrand NC, Elam DR (1993) Population genetic consequences of small population size: implications for plant conservation. Annu Rev Ecol Evol Syst 24:217–242. https://doi.org/10.1146/annurev.es.24.110193.001245

Engler R, Randin CF, Thuiller W et al (2011) 21st century climate change threatens mountain flora unequally across Europe. Glob Chang Biol 17:2330–2341. https://doi.org/10.1111/j.1365-2486.2010.02393.x

Erper I, Ozer G, Kalendar R, Avci S, Yildirim E, Alkan M, Turkkan M (2021) Genetic diversity and pathogenicity of Rhizoctonia spp isolates associated with red cabbage in Samsun (Turkey). J Fungi 7:234. https://doi.org/10.3390/jof7030234

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Excoffier L, Smouse P, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction sites. Genetics 131:479–491. https://doi.org/10.1093/genetics/131.2.479

Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol Bioinform 1:47–50. https://doi.org/10.1177/117693430500100003

Fang-Yong C, Ji-Hong L (2014) Germplasm genetic diversity of *Myrica rubra* in Zhejiang Province studied using inter-primer binding site and start codon-targeted polymorphism markers. Sci Hortic 170:169–175. https://doi.org/10.1016/j.scienta.2014.03.010

Frei ES, Scheepens JF, Stöcklin J (2012) High genetic differentiation in populations of the rare alpine plant species *Campanula thyrsoides* on a small mountain. Alp Bot 122:23–34. https://doi.org/10.1007/s00035-012-0103-2

Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147:915–925. https://doi.org/10.1093/genetics/147.2.915

Grundt HH, Borgen L, Elven R (1999) Aspects of reproduction in *Cerastium alpinum* on calcic and ultramafic soils in Central Norway. Nord J Bot 19:447–453. https://doi.org/10.1111/j.1756-1051.1999.tb01228.x

Guerrant EO Jr, Havens K, Maunder M (2004) *Ex situ* plant conservation: supporting species survival in the wild. Island, Washington, DC

Guo DL, Guo MX, Hou XG, Zhang GH (2014) Molecular diversity analysis of grape varieties based on iPBS markers. Biochem Syst Ecol 52:27–32. https://doi.org/10.1016/j.bse.2013.10.008

Hagen AR, Giese H, Brochmann C (2001) Trans-Atlantic dispersal and phylogeography of *Cerastium arcticum* (Caryophyllaceae) inferred from RAPD and SCAR markers. Am J Bot 88:103–112. https://doi.org/10.2307/2657131

Hagen AR, Sæther T, Borgen L, Elven R, Stabbetorp OE, Brochmann C (2002) The arctic-alpine polyploids *Cerastium alpinum* and *C. nigrescens* (Caryophyllaceae) in a sympatric situation: breakdown of species integrity? Plant Syst Evol 230:203–219. https://doi.org/10.1007/s006060200005

Hammer Ø, Harper DAT, Ryan PD (2001) PAST: paleontological statistics software package for education and data analysis. Palaeontol Electron 4:1–9

Hardwick KA, Fiedler P, Lee LC et al (2011) The role of botanic gardens in the science and practice of ecological restoration. Conserv Biol 25:265–275. https://doi.org/10.1111/j.1523-1739.2010.01632.x

Havens K, Vitt P, Maunder M, Guerrant EOJ, Dixon K (2006) *Ex situ* plant conservation and beyond. Biosci 56:525–531. https://doi.org/10.1641/0006-3568(2006)56[525:ESPCAB]2.0.CO;2

Hedrick PW, Miller PS (1992) Conservation genetics: techniques and fundamentals. Ecol Appl 2:30–46. https://doi.org/10.2307/1941887

Heywood VH, Iriondo JM (2003) Plant conservation: old problems, new perspectives. Biol Conserv 113:321–335. https://doi.org/10.1016/S0006-3207(03)00121-6

Hossein-Pour A, Haliloglu K, Özkan G, Tan M (2019) Genetic diversity and population structure of quinoa (*Chenopodium Quinoa* Willd.) using iPBS-retrotransposons markers. Appl Ecol Environ Res 17:1899–1911. https://doi.org/10.15666/aeer/1702_18991911

Huber CD, Nordborg M, Hermisson J, Hellmann I (2014) Keeping it local: evidence for positive selection in Swedish *Arabidopsis thaliana*. Mol Biol Evol 31:3026–3039. https://doi.org/10.1093/molbev/msu247

Hughes AR, Inouye BD, Johnson MTJ, Underwood N, Vellend M (2008) Ecological consequences of genetic diversity. Ecol Lett 11:609–623. https://doi.org/10.1111/j.1461-0248.2008.01179.x

Hultén E (1956) The *Cerastium alpinum* complex. A case of introgressive hybridization. Sven Bot Tidskr 50:411–495

Kalendar R, Antonius K, Smýkal P, Schulman AH (2010) iPBS: a universal method for DNA fingerprinting and retrotransposon isolation. Theor Appl Genet 121:1419–1430. https://doi.org/10.1007/s00122-010-1398-2

Khapilina O, Raiser O, Danilova A, Shevtsov V, Turzhanova A, Kalendar R (2021) DNA profiling and assessment of genetic diversity of relict species Allium altaicum Pall on the territory of Altai. PeerJ 9:e10674. https://doi.org/10.7717/peerj.10674

Khapilina O, Turzhanova A, Danilova A, Tumenbayeva A, Shevtsov V, Kotukhov Y, Kalendar R (2021b) Primer Binding Site (PBS) profiling of genetic diversity of natural populations of endemic species *Allium ledebourianum* Schult. Biotech 10:23. https://doi.org/10.3390/biotech10040023

Koc J, Androsiuk P, Chwedorzewska KJ, Cuba-Díaz M, Górecki R, Giełwanowska I (2018) Range-wide pattern of genetic variation in *Colobanthus quitensis*. Polar Biol 41:2467–2479. https://doi.org/10.1007/s00300-018-2383-5

Körner C (2003) Alpine plant life: functional plant ecology of high mountain ecosystems. Springer, Berlin

Kutnjak D, Kuttner M, Niketić M, Dullinger S, Schönswetter P, Frajman B (2014) Escaping to the summits: phylogeography and predicted range dynamics of *Cerastium dinaricum*, an endangered high mountain plant endemic to the western Balkan Peninsula. Mol Phylogenet Evol 78:365–374. https://doi.org/10.1016/j.ympev.2014.05.015

Lauterbach D, Burkart M, Gemeinholzer B (2012) Rapid genetic differentiation between *ex situ* and their *in situ* source populations: an example of the endangered *Silene otites* (Caryophyllaceae). Bot J Linn Soc 168:64–75. https://doi.org/10.1111/j.1095-8339.2011.01185.x

Lenoir J, Gégout JC, Marquet PA, de Ruffray P, Brisse H (2008) A significant upward shift in plant species optimum elevation during the 20th century. Sci 320:1768–1771. https://doi.org/10.1126/science.1156831

Li YY, Chen XY, Zhang X, Wu TY, Lu HP, Cai YW (2005) Genetic differences between wild and artificial populations of *Metasequoia glyptostroboides*: implications for species recovery. Conserv Biol 19:224–231. https://doi.org/10.1111/j.1523-1739.2005.00025.x

Loveless MD, Hamrick JL (1984) Ecological determinants of genetic structure in plant populations. Annu Rev Ecol Evol Syst 15:65–95. https://doi.org/10.1146/annurev.es.15.110184.000433

Lowe A, Harris S, Ashton P (2004) Ecological genetics: design, analysis, and application. Blackwell publishing, Oxford

Lowe AJ, Boshier D, Ward M, Bacles CFE, Navarro C (2005) Genetic resource impacts of habitat loss and degradation; reconciling empirical evidence and predicted theory for neotropical trees. Heredity 95:255–273. https://doi.org/10.1038/sj.hdy.6800725

Lynch M, Conery J, Bürger R (1995) Mutational meltdown in sexual populations. Evol 49:1067–1080. https://doi.org/10.1111/j.1558-5646.1995.tb04434.x

Maunder M, Higgens S, Culham A (2001) The effectiveness of botanic garden collections in supporting plant conservation: a European case study. Biodivers Conserv 10:383–401. https://doi.org/10.1023/A:1016666526878

Mehmood A, Jaskani MJ, Ahmad S, Ahmad R (2013) Evaluation of genetic diversity in open pollinated guava by iPBS primers. Pak J Agric Sci 50:591–597

Miller JS, Lowry PP, Aronson J, Blackmore S, Havens K, Maschinski J (2016) Conserving biodiversity through ecological restoration: the potential contributions of botanical gardens and arboreta. Candollea 71:91–98. https://doi.org/10.15553/c2016v711a11

Milovanov A, Zvyagin A, Daniyarov A, Kalendar R, Troshin L (2019) Genetic analysis of the grapevine genotypes of the Russian *Vitis* ampelographic collection using iPBS markers. Genetica 147:91–101. https://doi.org/10.1007/s10709-019-00055-5

Mirek Z, Piękoś-Mirkowa H (2008). Czerwona Księga Karpat Polskich - Rośliny Naczyniowe. Instytut Botaniki im. W. Szafera PAN, Kraków, pp 114–119

Niketić M, Đurović SZ, Tomović G, Schönswetter P, Frajman B (2022) Diversification within ploidy-variable Balkan endemic *Cerastium decalvans* (Caryophyllaceae) reconstructed based on genetic, morphological and ecological evidence. Bot J Linn Soc 199:578–608. https://doi.org/10.1093/botlinnean/boab037

Nyberg Berglund AB, Saura A, Westerbergh A (2001) Genetic differentiation of a polyploid plant on ultramafic soils in Fennoscandia. S Afr J Sci 97:533–535

Nyberg Berglund AB, Westerbergh A (2001) Two postglacial immigration lineages of the polyploid *Cerastium alpinum* (Caryophyllaceae). Hereditas 134:171–183. https://doi.org/10.1111/j.1601-5223.2001.00171.x

Nyberg Berglund AB, Dahlgren S, Westerbergh A (2003) Evidence of parallel evolution and site-specific selection of serpentine tolerance in *Cerastium alpinum* during the colonization of Scandinavia. New Phytol 161:199–209. https://doi.org/10.1046/j.1469-8137.2003.00934.x

Nyberg Berglund AB, Saura A, Westerbergh A (2006) Electrophoretic evidence for disomic inheritance and allopolyploid origin of the octoploid *Cerastium alpinum* (Caryophyllaceae). J Hered 97:296–302. https://doi.org/10.1093/jhered/esj029

Parusel J (1993) Czerwona lista roślin Babiogórskiego Parku Narodowego (2) Rogownica alpejska *Cerastium alpinum*. Parki Narodowe i Rezerwaty Przyrody 12:53

Parusel J (2014) *Cerastium alpinum* Rogownica alpejska. In: Kaźmierczakowa R, Zarzycki K, Mirek Z (eds) Polska Czerwona Księga Roślin. Paprotniki i rośliny kwiatowe, Instytut Ochrony Przyrody PAN, Kraków, pp 121–123

Parusel J (2001) *Cerastium alpinum* L. s.s. Rogownica alpejska. In: Kaźmierczakowa R, Zarzycki K (ed) Polska Czerwona Księga Roślin. Instytut Botaniki im. W. Szafera PAN, Instytut Ochrony Przyrody PAN, Kraków, pp 97–99

Parusel J (2008) Rogownica alpejska *Cerastium alpinum* L. In: Mirek Z, Piękoś–Mirkowa H (ed) Czerwona księga Karpat polskich. Instytut Botaniki im. W. Szafera PAN, Kraków, pp 118–120

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6:288–295. https://doi.org/10.1111/j.1471-8286.2005.01155.x

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. Bioinforma 28:2537–2539. https://doi.org/10.1093/bioinformatics/bts460

Perzanowska J (2012) Rogownica alpejska *Cerastium alpinum* L. In: Perzanowska J (ed) Monitoring gatunków roślin. Przewodnik metodyczny. Część III, GIOŚ, Warszawa, pp 192–202

Piry S, Luikart G, Cornuet JM (1999) BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. J Hered 90:502–503. https://doi.org/10.1093/jhered/90.4.502

Porsild AE, Cody WJ (1980) Vascular of the continental North-west Territories. Canada, National Museum of Natural Sciences, Ottawa

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959. https://doi.org/10.1093/genetics/155.2.945

Proctor J (1999) Toxins, nutrient shortage and droughts: the serpentine challenge. Trends Ecol Evol 14:334–335

Raible CC, Casty C, Luterbacher J, Pauling A, Esper J, Frank DC, Büntgen U, Roesch AC, Tschuck P, Wild M, Vidale PL, Schär C, Wanner H (2006) Climate variability - observations reconstructions and model simulations for the Atlantic-European and Alpine region from 1500–2100 AD In: Wanner H Grosjean M Röthlisberger R Xoplaki E (eds.) Climate variability, predictability and climate risks. Springer, Dordrecht. 9–29 https://doi.org/10.1007/978-1-4020-5714-4_2

Roberts EH (1991) 3. Genetic conservation in seed banks. Biol J Linn Soc 43:23–29. https://doi.org/10.1111/j.1095-8312.1991.tb00580.x

Robertson AI (1985) The poisoning of roots of *Zea mays* by nickel ions, and the protection afforded by magnesium and calcium. New Phytol 100:173–189. https://doi.org/10.1111/j.1469-8137.1985.tb02769.x

Rønning OI (1999) The flora of Svalbard. Norsk Polarinstitutt, Oslo, pp 33–34

Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. Genetics 145:1219–1228. https://doi.org/10.1093/genetics/145.4.1219

Rucińska A, Puchalski J (2011) Comparative molecular studies on the genetic diversity of an *ex situ* garden collection and its source population of the critically endangered polish endemic plant *Cochlearia polonica* E. Fröhlich Biodivers Conserv 20:401–413. https://doi.org/10.1007/s10531-010-9965-z

Rumpf SB, Hulber K, Klonner G, Moser D, Schutz M, Wessely J, Dullinger S (2018) Range dynamics of mountain plants decrease with elevation. Proc Natl Acad Sci U S A 115:1848–1853. https://doi.org/10.1073/pnas.1713936115

Rune O (1957) De serpentinicola elementen i Fennoscandiens flora. Sven Bot Tidskr 51:43–105

Scheen AC, Brochmann C, Brysting AK, Elven R, Morris A, Soltis DE, Soltis PS, Albert VA (2004) Northern hemisphere biogeography of *Cerastium* (Caryophyllaceae): insight from phylogenetic analysis of noncoding plastid nucleotide sequences. Am J Bot 91:943–952. https://doi.org/10.3732/ajb.91.6.943

Schoen DJ, Brown ADH (2001) The conservation of wild plant species in seed banks: attention to both taxonomic coverage and population biology will improve the role of seed banks as conservation tools. Biosci 51:960–966. https://doi.org/10.1641/0006-3568(2001)051[0960:TCOWPS]2.0.CO;2

Šingliarová B, Chrtek J, Mráz P (2008) Loss of genetic diversity in isolated populations of an alpine endemic *Pilosella alpicola* subsp. *ullepitschii*: effect of long-term vicariance or long-distance dispersal? Plant Syst Evol 275:181–191. https://doi.org/10.1007/s00606-008-0058-3

Skubic M, Schönswetter P, Frajman B (2018) Diversification of *Cerastium sylvaticum* and *C. subtriflorum* on the margin of the south-eastern Alps. Plant Syst Evol 304:1101–1115. https://doi.org/10.1007/s00606-018-1535-y

Stöcklin J, Kuss P, Pluess AR (2009) Genetic diversity, phenotypic variation and local adaptation in the alpine landscape: case studies with alpine plant species. Bot Helv 119:125–133. https://doi.org/10.1007/s00035-009-0065-1

Takou M, Hämälä T, Koch EM (2021) Maintenance of adaptive dynamics and no detectable load in a range-edge outcrossing plant population. Mol Biol Evol 38:1820–1836. https://doi.org/10.1093/molbev/msaa322

Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Res 17:6463–6471. https://doi.org/10.1093/nar/17.16.6463

Tero N, Aspi J, Siikamäki P, Jäkäläniemi A, Tuomi J (2003) Genetic structure and gene flow in a metapopulation of an endangered plant species, *Silene tatarica*. Mol Ecol 12:2073–2085. https://doi.org/10.1046/j.1365-294X.2003.01898.x

Till-Bottraud I, Gaudeul M (2002) Intraspecific genetic diversity in alpine plants. In: Körner C, Spehn EM (eds) Mountain biodiversity: a global assessment. Parthenon Publishing, New York, NY, pp 23–34

Totland Ø, Schulte-Herbrüggen B (2003) Breeding system, insect flower visitation, and floral traits of two Alpine *Cerastium* species in Norway. Arct Antarct Alp Res 35:242–247. https://doi.org/10.1657/1523-0430(2003)035[0242:BSIFVA]2.0.CO;2

Turzhanova A, Khapilina O, Tumenbayeva A, Shevtsov V, Raiser O, Kalendar R (2020) Genetic diversity of Alternaria species associated with black point in wheat grains. PeerJ 8:e9097. https://doi.org/10.7717/peerj.9097

Varsamis G, Merou T, Karapatzak E, Papageorgiou AC, Fotiadis G, Tsiftsis S (2021) Genetic diversity of alpine *Dryas octopetala* populations at their southern distribution limit in Europe. Nord J Bot 39:1–8. https://doi.org/10.1111/njb.03150

Vos P, Hogers R, Bleeker M, Reijans M, Lee T, Hornes M, Friters A, Pot J, Paleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414. https://doi.org/10.1093/nar/23.21.4407

Wang IJ, Bradburd GS (2014) Isolation by environment. Mol Ecol 23:5649–5662. https://doi.org/10.1111/mec.12938

Willi Y, van Buskirk J, Hoffmann AA (2006) Limits to the adaptive potential of small populations. Annu Rev Ecol Evol Sys 37:433–458. https://doi.org/10.1146/annurev.ecolsys.37.091305.110145

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 18:6531–6535. https://doi.org/10.1093/nar/18.22.6531

Willoughby JR, Harder AM, Tennessen JA, Scribner KT, Christie MR (2018) Rapid genetic adaptation to a novel environment despite a genome-wide reduction in genetic diversity. Mol Ecol 27:4041–4051. https://doi.org/10.1111/mec.14726

Yaldiz G, Çamlica M, Nadeem MA, Nawaz MA, Baloch FS (2018) Genetic diversity assessment in *Nicotiana tabacum* L. with iPBS-retrotransposons. Turk J Agric For 42. https://doi.org/10.3906/tar-1708-32

Yang W, Kang X, Yang Q, Lin Y, Fang M (2013) Review on the development of genotyping methods for assessing farm animal diversity. J Anim Sci Biotechnol 4:2. https://doi.org/10.1186/2049-1891-4-2

Yates MC, Bowles E, Fraser DJ (2019) Small population size and low genomic diversity have no effect on fitness in experimental translocations of a wild fish. Proc Royal Soc B 286:20191989. https://doi.org/10.1098/rspb.2019.1989

Young AG, Boyle T, Brown T (1996) The population genetic consequences of habitat fragmentation for plants. Trends Ecol Evol 11:413–418. https://doi.org/10.1016/0169-5347(96)10045-8

Zapałowicz H (1906) Krytyczny przegląd roślinności Galicyi. PAU, Kraków

Zegeye H (2017) *In situ* and *ex situ* conservation: complementary approaches for maintaining biodiversity. Int J Res Env Std 4:1–12. https://doi.org/10.33500/ijres.2017.4.001

Zietkiewicz E, Rafalski A, Labuda D (1994) Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. Genomics 20:176–183. https://doi.org/10.1006/geno.1994.1151

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Publikacja nr 2

scientific reports

Check for updates

OPEN

Complete chloroplast genomes of *Cerastium alpinum*, *C. arcticum* and *C. nigrescens*: genome structures, comparative and phylogenetic analysis

Sylwia E. Milarska[1], Piotr Androsiuk[1]✉, Łukasz Paukszto[2], Jan P. Jastrzębski[1], Mateusz Maździarz[2], Keith W. Larson[3] & Irena Giełwanowska[1]

The genus *Cerastium* includes about 200 species that are mostly found in the temperate climates of the Northern Hemisphere. Here we report the complete chloroplast genomes of *Cerastium alpinum*, *C. arcticum* and *C. nigrescens*. The length of cp genomes ranged from 147,940 to 148,722 bp. Their quadripartite circular structure had the same gene organization and content, containing 79 protein-coding genes, 30 tRNA genes, and four rRNA genes. Repeat sequences varied from 16 to 23 per species, with palindromic repeats being the most frequent. The number of identified SSRs ranged from 20 to 23 per species and they were mainly composed of mononucleotide repeats containing A/T units. Based on Ka/Ks ratio values, most genes were subjected to purifying selection. The newly sequenced chloroplast genomes were characterized by a high frequency of RNA editing, including both C to U and U to C conversion. The phylogenetic relationships within the genus *Cerastium* and family Caryophyllaceae were reconstructed based on the sequences of 71 protein-coding genes. The topology of the phylogenetic tree was consistent with the systematic position of the studied species. All representatives of the genus *Cerastium* were gathered in a single clade with *C. glomeratum* sharing the least similarity with the others.

The genus *Cerastium* belonging to the Caryophyllaceae family contains over 200 species[1] and consists of herbaceous plants, annuals, and perennials[2,3] that occur mainly in the Northern Hemisphere. The genus is most common in temperate and cold regions, especially at high elevations, with Eurasia serving as its center of diversity. The majority of the representatives of the genus *Cerastium* have a limited range with only a few species characterized by a cosmopolitan distribution[3]. The current state of knowledge about the genetic diversity of the genus *Cerastium* is based on a rather limited number of studies. Genetic investigations employed analysis of isoenzymatic polymorphism[4,5] and different molecular markers like RAPD and SCAR[6], AFLP[7,8] and iPBS[9]. Apart from traditional genetic diversity studies, there are also papers that focused on the role of hybridization and introgression events in the evolution of the genus *Cerastium*[10–12]. One example of the intricated systematics within the genus *Cerastium* is the *C. alpinum*–*C. arcticum* complex. Vast physical variation within that complex has resulted in the identification of numerous species, subspecies, and varieties within that group of plants[13–16], among which *C. alpinum* L., *C. arcticum* Lange and *C. nigrescens* (H.C. Watson) can be found[17–19]. *C. alpinum* is an arctic-alpine species that occurs in the northern part of North America and Europe. Moreover, it is recorded in Europe on high mountain grasslands, mostly in the subalpine zone (from 1480 to 1680 m.a.s.l.), where it forms one-species aggregations[20,21]. *C. arcticum* appeared here as the most problematic component of the species group. Latest genetic and morphological analyses suggest that the species conventionally known as *C. arcticum* actually consists of two separate taxa: *C. arcticum* s. str. and *C. nigrescens*[10,18,19]. The first of them is restricted to arctic areas (the Canadian Arctic, Greenland, Svalbard, north-western arctic Russian islands), while the other

[1]Department of Plant Physiology, Genetics and Biotechnology, Faculty of Biology and Biotechnology, University of Warmia and Mazury in Olsztyn, ul. M. Oczapowskiego 1A, 10-719 Olsztyn, Poland. [2]Department of Botany and Nature Protection, Faculty of Biology and Biotechnology, University of Warmia and Mazury in Olsztyn, Pl. Łódzki 1, 10-721 Olsztyn, Poland. [3]Climate Impacts Research Centre, Umeå University, 90187 Umeå, Sweden. ✉email: piotr.androsiuk@uwm.edu.pl

is characteristic to fell regions (the British Isles, Fennoscandian mountains, Faeroe Islands, Iceland). Despite the intensive studies delimitation of these taxa is still problematic on a large geographic scale[11]. Consequently, a novel approach is needed to find a universal marker for taxon identification.

Due to the recent progress observed in molecular sciences, high-throughput genome sequencing technologies have become widely available and provide a relatively fast and inexpensive way of obtaining high-quality genomic data. In case of the plant genetics, chloroplast (cp) genomes became a source of data commonly used in comparative studies[22,23], biotechnology[24], species identification[25,26] or in analyses addressing phylogenetic questions[27,28]. It was shown that the complete cp genome contains roughly equivalent amount of information as the *cox1* gene used in animals, so it has the potential to provide enough distinguishing differences that enable molecular identification of even closely related species[29]. Using the entire chloroplast genome as a super-barcode is a novel approach that could potentially address the limitations of conventional two-locus barcoding[30]. Traditional barcoding primarily relies on sequence variation within two regions of the chloroplast genome, *matK* and *rbcL*, which is not always sufficient for precise species delimitation. To date, there are only two publicly available chloroplast genome sequences for the genus *Cerastium*, i.e. complete cp genome sequence for *C. glomeratum* and partial genome sequence for *C. arvense* (NC_066897 and MH627219, respectively; NCBI). The available data revealed that the *Cerastium* chloroplast genome has conserved quadripartite structure with size and gene content typical for angiosperms.. Except for the above-mentioned *C. glomeratum*, there are 60 other species (representing the following genera: *Agrostemma*, *Arenaria*, *Colobanthus*, *Dianthus*, *Gymnocarpos*, *Gypsophila*, *Lychnis*, *Myosoton*, *Paronychia*, *Psammosilene*, *Pseudostellaria*, *Silene,* and *Stellaria*) for which complete plastome sequences are available in the NCBI database (accessed on March 24, 2023). Considering the fact, that the Caryophyllaceae family consists of 40 genera and includes about 12 500 species, the number of chloroplast genomes currently available for this group of plants should be treated as very low.

The complete chloroplast genomes of three *Cerastium* species (*C. alpinum*, *C. arcticum* and *C. nigrescens*) have been sequenced and annotated for the first time in this paper. The specific objectives of this study included: (1) determination of the size and structure of cp genomes for *C. alpinum*, *C. arcticum* and *C. nigrescens*, (2) identification of genomic repeats, including forward, reverse, palindromic and complementary sequences among *Cerastium* chloroplast genomes, (3) identification and characterization of simple sequence repeats (SSRs) in newly sequenced *Cerastium* plastomes, (4) analysis of the evolution and dynamics of chloroplast protein-coding sequences, (5) comparative study of all available *Cerastium* chloroplast genomes, and (6) reconstruction of the phylogenetic relationships within genus *Cerastium* and family Caryophyllaceae based on plastome sequences.

Results

Organization of chloroplast genomes

NovaSeq Illumina platform was applied for chloroplast genome sequencing of three *Cerastium* species. The highest number of raw reads (15,370,470) was obtained for *C. arcticum*, whereas in the case of *C. nigrescens* and *C. alpinum* sequencing yielded 13,970,724 and 13,586,446 reads, respectively. The raw reads were then mapped separately to the reference chloroplast genome of *C. glomeratum.* As a result, 297,869 mapped reads with a mean coverage of 304 were observed for *C. nigrescens*, while in the case of two other species these value were more than twice as high and amounted to 650,075 reads and 664 coverage for *C. alpinum* and 664,454 reads and 675 coverage for *C. arcticum* (Supplementary FigureS1). The size of reported cp genomes was 147,945 for *C. alpinum*, 147,940 bp for *C. nigrescens* and 148,722 bp for *C. arcticum*. Each chloroplast genome appeared as a circular, double-stranded DNA molecule with a traditional quadripartite structure composed of Large Single Copy (LSC) and Small Single Copy (SSC) separated by a pair of Inverted Repeats (IR) regions which have identical sequences but opposite orientation (Fig. 1). The overall GC content was nearly identical in all *Cerastium* species: 36.51% for *C. alpinum*, 36.46% for *C. arcticum* and 36.52% for *C. nigrescens* (Table 1). Additionally, variant calling analysis revealed no heteroplasmy in reported chloroplast genomes.

All three reported *Cerastium* chloroplast genomes contained an identical set of 113 genes composed of 75 protein-coding genes, 30 transfer RNA genes, four ribosomal RNA genes, and four conserved chloroplast ORFs (*ycf1*, *ycf2*, *ycf3*, *ycf4*) (Table 2). We have also identified in each IR region a sequence for the *rpl23* gene which due to the internal, premature termination codon was retained rather as a nonfunctional pseudogene. Most protein-coding genes have the standard AUG as the initiation codon. The total number of codons for all protein-coding genes in the reported cp genomes was 26,115 for *C. arcticum*, 26,116 for *C. alpinum* and 26,220 for. *C. nigrescens.* All studied species shared similar pattern of codon usage and amino acid frequency. Leucine appeared as the dominant amino acid (10.7%), whereas cysteine was less frequently encountered (1.2%). The most abundant codon (4.46%) was ATT and the last (0.004%) were TTG (all species), ATA and CTG (*C. nigrescens*). CTG codon appeared only in *C. nigrescens* (Supplementary Table S1). Most of the genes in analyzed chloroplast genomes did not contain introns, 14 others contained one intron (*atpF*, *ndhA*, *ndhB*, *petB*, *petD*, *rpl16*, *rpoC1*, *rps16*, *trnA-UGC*, *trnG-UCC*, *trnI-GAU*, *trnK-UUU*, *trnL-UAA*, *trnV-UAC*), whereas only three genes consisted of three exons (*clpP*, *ycf3,* and *rps12*). Our data confirmed that the *rps12* gene, coding plastid ribosomal protein S12, is a trans-splicing gene. This gene was split into three exons: the first exon (5'end of the sequence) was located in the LSC, while the second and third exons in the IRs. The smallest intron was found in the *trnL-UAA* (518 bp for *C. arcticum* and 520 bp for *C. alpinum* and *C. nigrescens*), whereas the biggest was in the *trnK-UUU* (2479 bp for *C. arcticum* and 2480 bp for *C. alpinum* and *C. nigrescens*) gene. The *matK* gene was positioned inside the intron of *trnK*-UUU. Fifty-eight protein-coding genes, 22 tRNA genes, and two conserved chloroplast ORFs (*ycf3* and *ycf4*) were located in the LSC region, SSC region contained eleven protein-coding genes, one tRNA gene, and one chloroplast ORF (*ycf1*, located on the boundary between SSC and IR_B), whereas repeated IR region contained six protein-coding genes (including *rps19* gene located on the boundary between IR_A and LSC), seven tRNA genes, four rRNA genes, and one chloroplast ORF (*ycf2*).

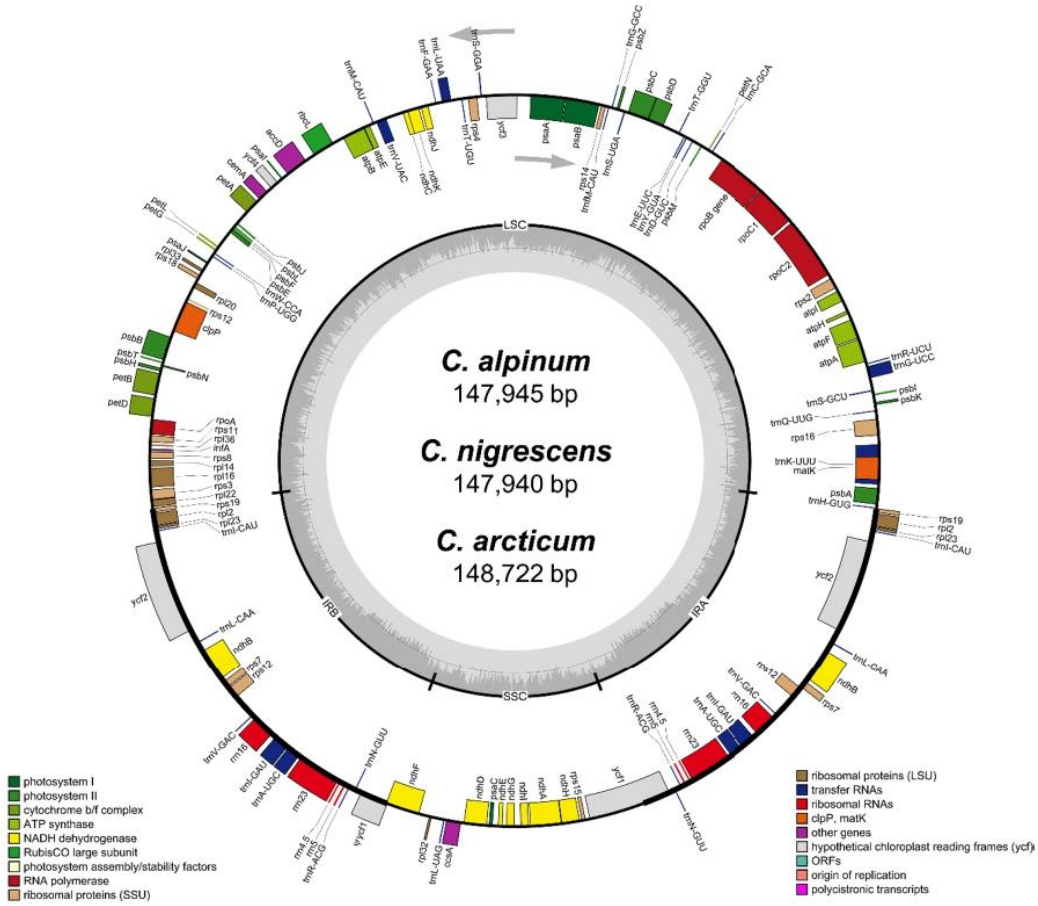


Figure 1. Gene map of the three *Cerastium* chloroplast genomes. Genes drawn inside the circle are transcribed clockwise, and those outside are transcribed counterclockwise (indicated by arrows). Differential functional gene groups are color-coded. GC content variations is shown in the middle circle.

Genome features	*C. alpinum*	*C. arcticum*	*C. nigrescens*
Raw data reads no	13,586,446	15,370,470	13,970,724
Mapped reads no	650,075	664,454	297,869
Percent of chloroplast genome reads (%)	4.78	4.32	2.13
Mean coverage (x)	664	675	304
Size (bp)	147,945	148,722	147,940
LSC length (bp)	80,080	80,835	80,076
SSC length (bp)	16,851	16,861	16,850
IR length (bp)	25,507	25,513	25,507
Protein-coding genes	79	79	79
tRNA genes	30	30	30
rRNA genes	4	4	4
Number of genes duplicated in IR	18	18	18
Overall GC content (%)	36.51	36.46	36.52

Table 1. Summary of chloroplast genome characteristics of studied *Cerastium* species.

Category	Group of gene	Name of genes
Photosynthesis	Photosystem I	*psaA, psaB, psaC, psaI, psaJ*
	Photosystem II	*psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ*
	Cytochrome complex	*petA, petB, petD, petG, petL, petN*
	ATP synthase	*atpA, atpB, atpE, atpF, atpH, atpI*
	NADH dehydrogenase	*ndhA, ndhB* (×2), *ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK*
	Large subunit of RUBISCO	*rbcL*
DNA replication and protein synthesis	Ribosomal RNA	*rrn4.5* (×2), *rrn5* (×2), *rrn16* (×2), *rrn23* (×2)
	Small subunit ribosomal proteins	*rps2, rps3, rps4, rps7* (×2), *rps8, rps11, rps12* (×2), *rps14, rps15, rps16, rps18, rps19* (×2)
	Large subunit ribosomal proteins	*rpl2* (×2), *rpl14, rpl16, rpl20, rpl22, rpl32, rpl33, rpl36*
	RNA polymerase subunits	*rpoA, rpoB, rpoC1, rpoC2*
	Transfer RNA	*trnA-UGC* (×2), *trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnfM-CAU, trnG-GCC, trnG-UCC, trnH-GUG, trnI-CAU* (×2), *trnI-GAU* (×2), *trnK-UUU, trnL-CAA* (×2), *trnL-UAA, trnL-UAG, trnM-CAU, trnN-GUU* (×2), *trnP-UGG, trnQ-UUG, trnR-ACG* (×2), *trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC* (×2), *trnV-UAC, trnW-CCA, trnY-GUA*
Other genes	Conserved hypothetical chloroplast ORF	*ycf1* (×2), *ycf2* (×2), *ycf3*[a], *ycf4*[a]
	Other proteins	*accD, ccsA, cemA, clpP, infA, matK*
Pseudogenes		*rpl23* (×2)

Table 2. List of genes present in chloroplast genome of *Cerastium*. [a]Genes associated with Photosystem I.

The boundaries between IR and LSC/SSC regions were identified (Fig. 2). In the case of plastomes of *C. alpinum, C. arcticum,* and *C. nigrescens* the complete sequence of *ycf1* gene was located on the boundary between IR_A and SSC, and its incomplete copy on IR_B/SSC boundary where it functions as a pseudogene (*Ψycf1*). *Ψycf1* was overlapped (89 bp) with the *ndhF* gene. The IR_A/SSC boundary was located within *ycf1* sequence 1822 bp from its 5' end. The IR_B/LSC boundary was found within the *rps19* gene (52–54 bp from its 3' end, depending on the species). Its shorter copy was located at the IR_A/LSC boundary, where it acts as a pseudogene (*Ψrps19*). *Ψrps19* was overlapped (19 bp) with *trnH* gene. The *trnH* gene was near the IR_A/LSC border (11 bp apart in case of *C. alpinum* and *C. nigrescens* and 9 bp for *C. arcticum*). The localization of IR and LSC/SSC boundaries was also analyzed for *C. glomeratum*. In the case of this species the analyzed boundaries were identified within the same genetic elements. The IR_B/LSC boundary was located within the *rps19* gene (65 bp from its 3' end) and pseudogene *Ψrps19* was found at the IR_A/LSC boundary. However, *Ψrps19* did not overlap with *trnH*. Analysis of the IR_A/SSC and IR_B/SSC boundaries revealed the inversion of the entire SSC region in *C. glomeratum* cp genome. The IR_B/SSC border was located within *ycf1* gene (1867 bp from its 5' end) whereas the IR_A/SSC was within the *ndhF* gene (57 bp from its 3' end). The sequence for *Ψycf1* was not annotated in the analyzed plastome. Finally, the *trnH* gene was located 25 bp apart from the IR_A/LSC border.

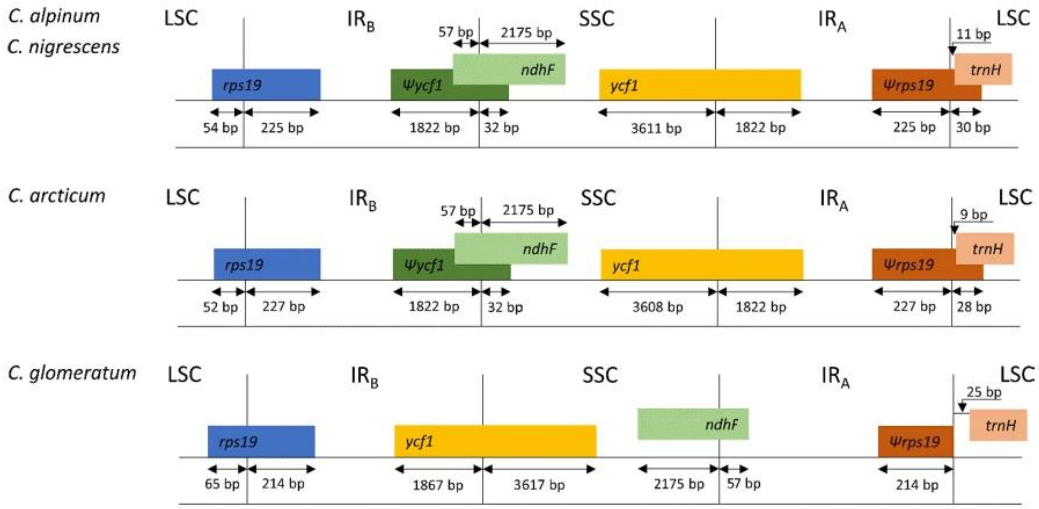


Figure 2. Comparison of LSC, SSC, and IR boundaries of four *Cerastium* chloroplast genomes.

Repetitive sequences and SSRs

The analysis of genomic repeats in the cp genomes of studied *Cerastium* species (*C. alpinum*, *C. arcticum*, *C. nigrescens*, and *C. glomeratum* revealed 79 repetitive sequences with lengths ranging from 30 to 170 bp (Supplementary Table S2A–D). The number of repeats was the highest (23) in *C. arcticum* and the lowest (16) in *C. glomeratum*. Palindromic repeats dominated among identified sequences (from 47.8% in *C. arcticum* to 62.5% in *C. glomeratum*), followed by forward repeats (from 31.3% in *C. glomeratum* to 47.8% in *C. arcticum*) and reverse repeats (from 4.3% in *C. arcticum* to 6.3% in *C. glomeratum*) (Fig. 3b). No complementary repeats were found in analyzed chloroplast genomes. Most repeat sequences (80%) were found in the LSC region, and the remaining repeats were equally distributed (10%) in IR and SSC regions (Fig. 3c). Repeats with a length of 30–40 bp were the most frequent in each species (from 13 in *C. glomeratum* to 18 in *C. alpinum*, *C. arcticum*, and *C. nigrescens*) (Fig. 3a).

Application of the Phobos software revealed from 20 (*C. nigrescens*) to 23 (*C. arcticum*) chloroplast microsatellites (Fig. 4a), including mono-, di-, tri-, tetra-, penta- and hexanucleotide SSRs (Fig. 4b, Supplementary Table S3A–D). The mononucleotide SSRs, all composed of A/T repeat units, were the most common in each species with a frequency ranging from 36.4% (*C. glomeratum*) to 43.5% (*C. arcticum*). The second most common motif among identified SSRs was AAT/TAA with a frequency ranging from 27.3% (*C. glomeratum*) to 35% (*C. nigrescens*). Tetranucleotide SSRs, with frequency ranging from 19% (*C. alpinum*) to 31.8% (*C. glomeratum*) were composed of AAAT/TAAA, AATT/TTAA, ACCT/TCCA, AGAT/TAGA, and AAAG/GAAA motifs. Among identified chloroplast microsatellites there was only one SSR that contained a dinucleotide motif (AT/TA, *C. glomeratum*), one SSR with pentanucleotide motif (AATAT/TATAA, *C. alpinum*), and one SSR built of hexanucleotide motif (AAATCC/CCTAAA, *C. arcticum*). A substantial number of SSRs were identified in the LSC region (from 72.7% in *C. glomeratum* to 76.2% in *C. alpinum*), followed by SSC (from 14.3% in *C. alpinum* to 18.2% in *C. glomeratum*) and IR regions (from 8.7% in *C. arcticum* to 10% in *C. nigrescens*) (Fig. 4c). SSRs were mainly located within intergenic spacers (from 55% in *C. nigrescens* to 61.9% in *C. alpinum*), whereas the remaining microsatellites were distributed within exons (from 31.8% in *C. glomeratum* to 35% in *C. nigrescens*) and introns (from 4.8% in *C. alpinum* to 10% in *C. nigrescens*) (Fig. 4d).

Synonymous (Ks) and non-synonymous (Ka) substitution rate analysis

The substitution rate varied across genes in each functional group and ranged from 0 to 0.151 and from 0 to 0.0858 for Ka and Ks, respectively (Supplementary Table S4). The highest average value of Ka (0.0062) was noted in the group of "other genes" and the lowest (0.0012 and 0.0014) in genes related to the cytochrome b/f complex

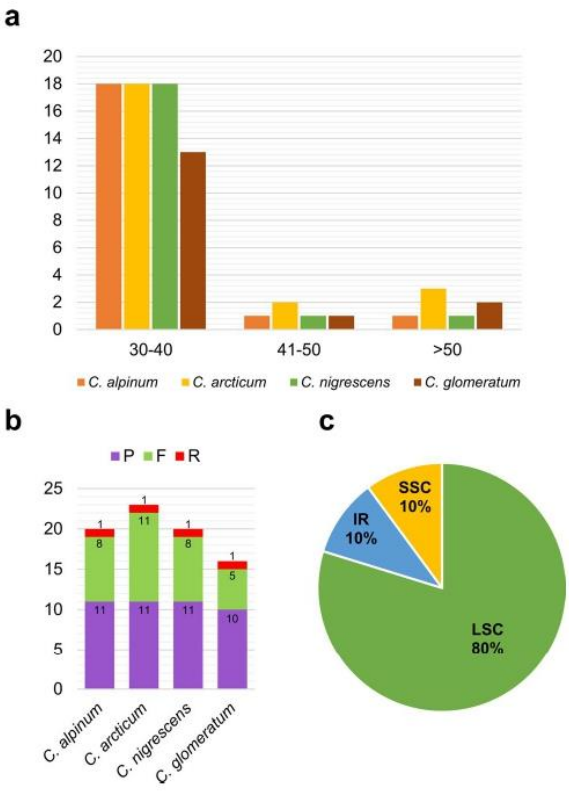


Figure 3. Number of repeat types and their distribution in four *Cerastium* species. (**a**) Length of the repeats; (**b**) types of repeats; (**c**) location of repeat sequences. F, P, R represent forward, palindromic and reverse repeats.

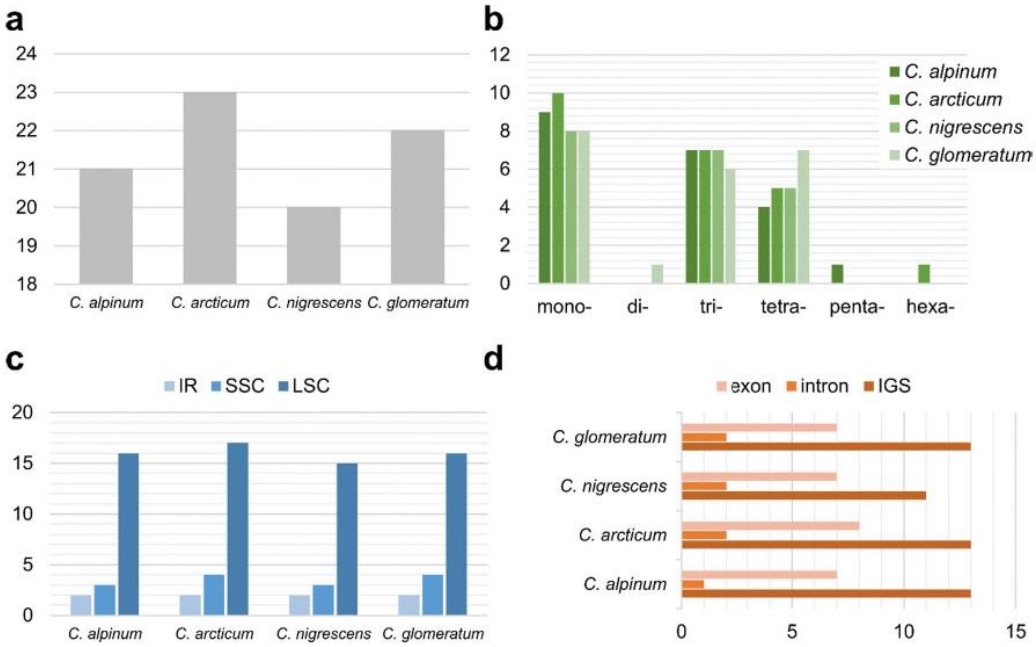


Figure 4. The distribution and type of simple sequence repeats (SSRs) in cp genomes of four *Cerastium* species. (**a**) Number of different SSRs types; (**b**) distribution of SSR motifs in different repeat class types; (**c**) location of different SSRs in IR, SSC and LSC regions; (**d**) partition of SSRs among IGS, introns and exons.

and photosystem II, respectively. The highest average value of Ks (0.0247) was noted in gene for RubisCO large subunit, and the lowest in genes associated with the small subunit of ribosome (0.0159) and subunits of ATP synthase (0.0160). In summary, no differences (Ka = 0 and Ks = 0) were observed in the sequences of 11 genes, whereas only synonymous substitutions (Ka = 0) were observed in 18 genes. The Ka/Ks ratio was less than 1 in all genes, excluding *ndhB* (2.7250 for *C. arvense*). Relatively high values of Ka/Ks were observed in *rpl22* (0.8673) for all studied species and in *rps14* (0.8776) for *C. arvense*. In the remaining cases, the values did not exceed 0.75 (Fig. 5).

Genomic comparative and nucleotide diversity analyses

The MAUVE results revealed a highly conservative structure of chloroplast genomes of *C. alpinum*, *C. arcticum*, *C. nigrescens,* and *C. arvense* for which no rearrangements (inversions or translocations) were detected. Only in the case of *C. glomeratum* the opposite orientation of the whole SSC region was observed (Supplementary Fig. S2).

Nucleotide diversity (π) in the analyzed cp genomes of *Cerastium* species was determined at 0.00493. The results of sliding window analysis showed that the π value for studied *Cerastium* cp genomes varied from 0 to 0.02708 (Fig. 5). Nine highly variable ($\pi > 0.015$) regions were identified in analyzed cp genomes: *rpl32–trnL*-UAG, *ndhA* (intron), *rps16* (intron), *trnD*-GUC–*trnY*-GUA, *trnF*-GAA–*ndhJ*, *ndhC–trnV*-UAC, *petA–psbJ*, *psbE–petL,* and *trnP*-UGG–*psaJ*. The highest π value (0.02708) was observed for *trnD*-GUC–*trnY*-GUA region. All of these divergent hotspots were identified in non-coding regions i.e. intergenic spacers and introns. Furthermore, the majority of highly variable regions (7) were identified in LSC, followed by two such regions in SSC, and none in the IR region (Fig. 5).

Prediction of RNA editing sites

Prediction of RNA editing sites with the use of PREPACT 3.0 tool revealed from 578 to 588 editing sites in 63 protein-coding genes (Fig. 5, Supplementary Table S5A–D). The lowest number of predicted RNA editing sites (578) was found for *C. alpinum* and *C. nigrescens*, whereas the highest was for *C. glomeratum*. In the case of the *C. arcticum* the number of RNA editing sites was 583. Among identified editing events both C to U and U to C conversions were found. In the case of 14 genes no such changes were identified. The C to U conversion accounted for 43.05% to 43.54% of total RNA editing sites, while U to C substitutions were responsible for 56.46% to 56.95% of the identified editing events. All predicted RNA editing sites resulted in non-synonymous mutations. Forty-seven (47.17–47.28%) percent of the substitutions were found at the first position of the codon, 53% (52.72–52.83%) were found at the second position, and none were found at the third position. Among predicted RNA editing events there were also conversions that involved two sites of RNA editing within one codon. Eighteen such editing events were identified in the case of *C. alpinum* and *C. nigrescens,* and 20 for *C. arcticum* and *C. glomeratum*. Most of these events involved conversions of UCU and UCC codons for serine (S) into CUU and CUC triplets for leucine (L) and back from leucine to serine, and also conversion of UUU and UUC for phenylalanine (F)

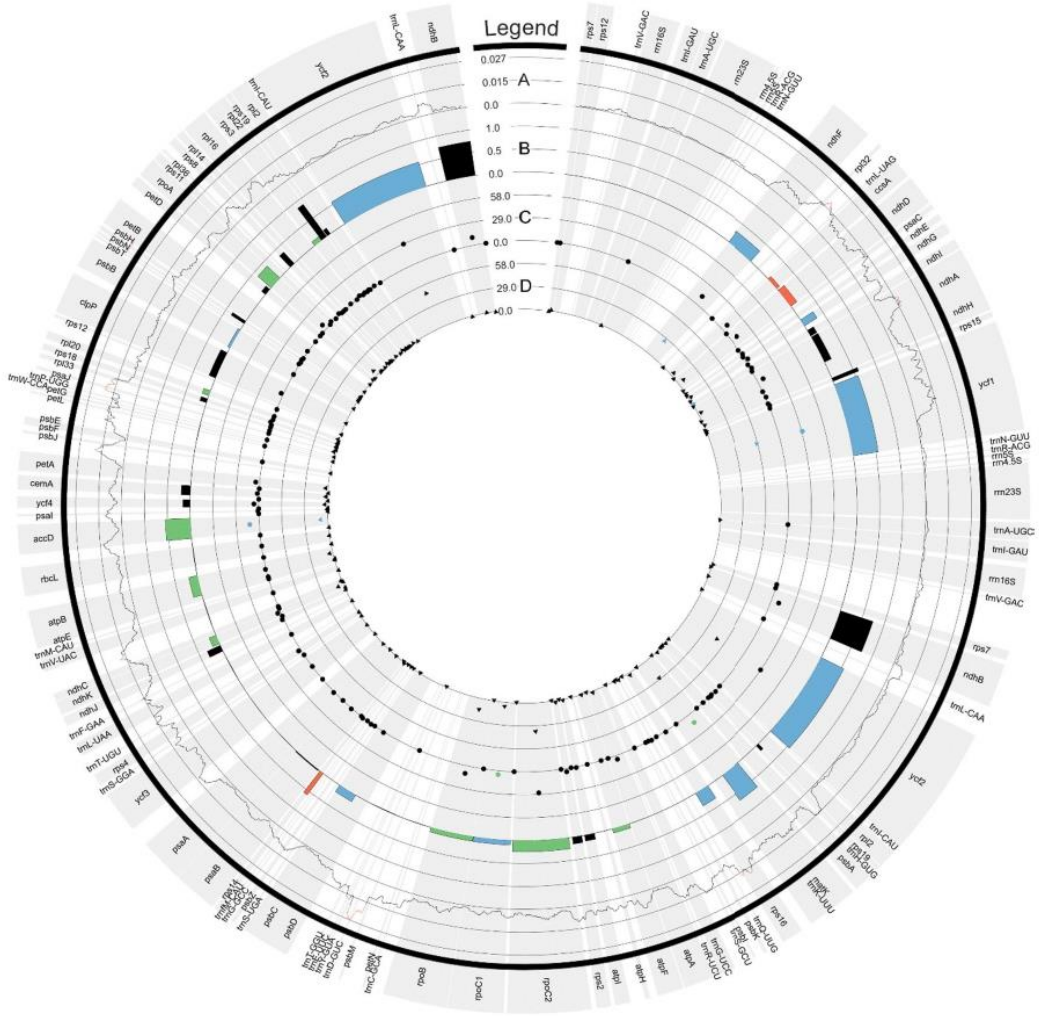


Figure 5. Circular visualization of the plastome comprehensive analyses of three *Cerastium* species (*C. nigrescens, C. arcticum,* and *C. alpinum*). The first outer track represents the chloroplast gene symbols. The second line track (**A**) shows haplotype diversity (π) values calculated for sliding window equal to 800 bp. The red part of the line plot depicts regions with the highest diversity (π > 0.015). Histograms (**B**) show comparative Ka/Ks ratio values for *Cerastium* species, where blue, red, green and black colors depict the dominant Ka/Ks values in *C. arcticum, C. nigrescens*, equal for *C. alpinum* and *C. nigrescens*, and equal for all three species, respectively. Both scatter plots show the number of potential C > U and U > C editing sites within each plastid gene (**C**,**D**, respectively). The colors describe higher numbers of RNA editing sites in *C. arcticum* (blue points) and *C. nigrescens* and *C. alpinum* (green points) in comparison to other compared species.

to CCU and CCC for proline (P), and in the opposite direction i.e., from proline to phenylalanine. The highest number of predicted RNA editing sites were reported for *ycf1* (85–88), *ycf2* (77), and *rpoC2* (64–65) genes. The most often substitution in each species was phenylalanine (F) to leucine (L) change (16.48–16.75%), whereas P (proline) to F (phenylalanine) and R (arginine) to W (tryptophan) changes were observed with the lowest frequencies (0.353–0.358% and 0.881–0.896%, respectively). Additionally, the conversion of the termination codon UAA to CAA triplet encoding glutamine was found to be created by RNA editing in *ndhI* gene for *C. arcticum*.

Additionally, we conducted the same investigation for chloroplast genes of *C. arvense*. Unfortunately, due to incomplete sequences available for *rpl20*, *rpoB*, *rpoC1*, *rpoC2*, *ycf1,* and *ycf2,* these genes were not included in the analysis. In 17 out of 71 analyzed genes, we did not identify potential RNA editing sites. In the remaining 54 genes we found 286 editing sites (Fig. 5, Supplementary Table S5E). Further, for this species, both C to U and U to C conversions were found, but U to C edition dominated (56.46%). The highest number of substitutions were observed for the first (53.06%) and the second (46.94%) position of the codon, whereas they were absent in the third position. Analogous to the situation described above for *C. alpinum, C. arcticum, C. nigrescens,* and *C. glomeratum* also here, for C. *arvense,* among predicted RNA editing events we found conversions that involved two sites of RNA editing within one codon. There were seven situations in which CUU and CUA codons for

leucine (L) were changed into UCU and UCA for serine (S), and backward from serine to leucine. The highest number of predicted RNA editing sites were identified within sequences for *matK* (40) and *ndhF* (37) genes. All the identified RNA edition events caused non-synonymous mutations. The change from phenylalanine (F) to leucine (L) was the most abundant substitution (18.53%), whereas leucine (L) to proline (P) and arginine (R) to cysteine (C) were observed with the lowest frequency (0.7%).

Phylogenetic analysis

In the BI tree, a very high Bayesian posterior probability value (≥ 0.92) was reached in 96.4% of the nodes (53 out of 55). The reconstructed tree supported the taxonomic position of the studied species and revealed the following relationships: all *Silene* species together with *Lychnis wilfordii* formed one clad which gathered only representatives of Sileneae tribe; a second clad was formed by the representatives of the Caryophylleae tribe i.e., five *Dianthus* species, three representatives of genus *Gysophila* and *Psammosilene tunicoides*; a third clad consisted of all representatives of Alsineae tribe (eight *Pseudostellaria* species, *Stellaria dichotoma var. lanceolata*, *Myosoton aquaticum* and all studied *Cerastium* species which formed one subgroup; a fourth clad consisted of eight *Colobanthus* species (Sagineae tribe) whereas *Spergula arvensis* (Sperguleae) and *Paronychia argentea* with *Gymnocarpos przewalski* (Paronychieae) form two separate branches. The most diverged position on the tree was occupied by *A. thaliana* which was used here as an outgroup (Fig. 6).

Results of divergence time estimation suggested that the family Caryophyllaceae started to diversify ca. 74.46 millions-years ago (Mya). Later, subsequent radiation within the family Caryophyllaceae occurred: ca. 51.47 Mya Sperguleae tribe splits from the other sister clades and at ca. 48.32 Mya diversification of Sagineae tribe was observed (represented here only by one genus *Colobanthus*). At ca. 46.26 Mya the evolutionary paths of Alsineae and Arenariae tribes diverged from Caryophylleae and Sileneae tribes. At ca 42.26 Mya Alsineae and Arenariae split apart and c.a. 41.93 Mya Caryophylleae split from Sileneae. Diversification events at the lower taxonomic level e.g. within tribe Caryophylleae, Sileneae and Alsineae started at 30.87, 26.0 and 20.6 Mya, respectively. The genus *Cerastium* began to diversify at 3.66 Mya (Fig. 7).

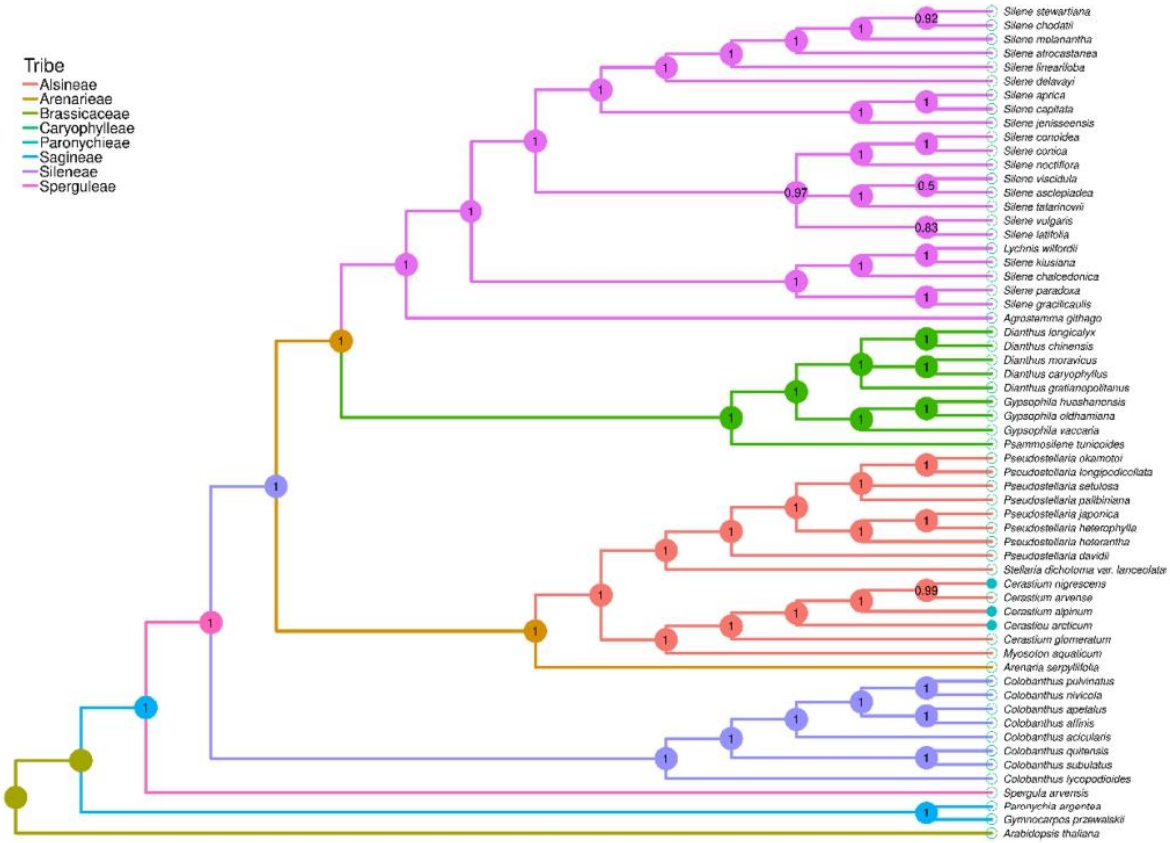


Figure 6. Phylogenetic tree (cladogram) based on sequences of sheared 71 protein-coding genes from five *Cerastium* species and 54 other Caryophyllaceae representatives using Bayesian posterior probabilities (PP). Bayesian PP are given at each node.

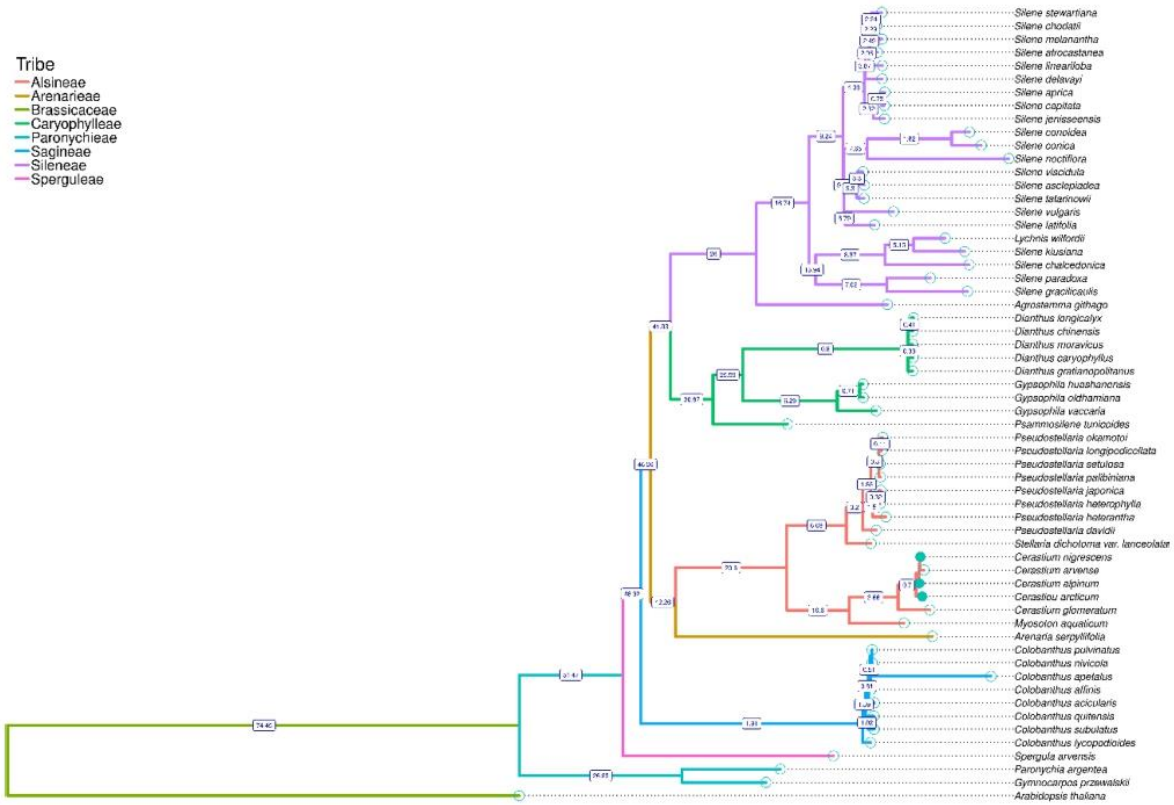


Figure 7. Divergence time estimation of selected Caryophyllaceae taxa. The numbers next to the nodes represent the divergence time (*Mya* millions years ago).

Discussion

Chloroplast genomes are a relevant resource for many genomic and biotechnological applications[31]. Its unique features, like lack of recombination and slower mutation rate in comparison to nuclear genomes, make the chloroplast genome a frequently used source of data in evolutionary biology[32]. Moreover, common use of chloroplast genome in phylogeographic studies is observed due to its uniparental inheritance that exhibits geographical structure[33].

Although the genus *Cerastium* consists of more than 200 species[1], the availability of the genomic data for this group of plants is very limited and, to date there is only one complete chloroplast genome sequence in the NCBI database for *C. glomeratum*. There is also a chloroplast genome sequence for another *Cerastium* species (*C. arvense*), but due to the several gaps in the intergenic spacers and lack of complete sequences for six protein-coding genes (*rpl20*, *rpoB*, *rpoC1*, *rpoC2*, *ycf1,* and *ycf2*), create constraints for the utilization of this sequence. To fill the gap in the knowledge concerning the genomics of the genus *Cerastium* we sequenced and annotated the plastid genomes of three species: *C. alpinum*, *C. arcticum,* and *C. nigrescens*. The size of reported cp genomes ranged from 147,940 (*C. nigrescens*) to 148,722 bp (*C. arcticum*) and was similar to the plastome of *C. glomeratum* (148,643 bp) and other angiosperms[34]. All three studied cp genomes share the same gene content and order and typical quadripartite structure, with a pair of inverted repeats (IR) separated by a SSC and a LSC region). Length variation in cp genomes in different groups of plants is often caused by expansions and contractions of IR regions[35]. In extreme cases, IR regions were completely lost by chloroplast genomes of some algae[36] or one of its copies is not observed in some representatives of leguminous plants[37]. Consequently, the analysis of the distribution of IR/LSC and IR/SSC borders became a standard element of plastome characteristics. Obtained results revealed that their locations may differ among various species, even between closely related genera[38]. Analysis of reported here chloroplast genomes of *C. alpinum, C. arcticum,* and *C. nigrescens* revealed that IR/LSC and IR/SSC boundaries were located within sequences of *ycf1* and *rps19* genes (Fig. 2), which is analogous to the situation observed for most angiosperms[39]. The location of IR boundaries was identical for *C. alpinum* and *C. nigrescens*, whereas a minor shift (two bases shift within *rps19* and three bases within the *ycf1* gene) was observed for *C. arcticum*. The length of IR and SSC regions in reported plastomes was very similar and ranged from 25,507 to 25,513 bp and from 16,850 to 16,861 bp, respectively. Higher variation was found for the LSC region where the difference between the longest and the shortest LSC is 782 bp (*C. arcticum* vs. *C. nigrescens*). Nevertheless, the sizes of all three plastome regions values are consistent with previous reports for other dicotyledons[40,41]. For comparative purposes, the IR borders within the chloroplast genome of *C. glomeratum* were also examined. In this case, more differences were observed. Although the IR borders were also located within the *rps19* and *ycf1*

genes, the eleven base shift for *rps19* and 45 base shift for *ycf1* was found. Additionally, only one copy of *ycf1* can be found within *C. glomeratum* plastome at the IR_B/SSC border as its incomplete copy (*ψycf1*) between IR_A/SSC was not annotated. However, the main difference is associated with the opposite orientation of the whole SSC region. This interesting phenomenon was originally reported for *Phaseolus* vulgaris[42]. The author with the use of restriction enzyme analysis revealed, that the individual plants' chloroplast DNA demonstrates a type of heteroplasmy in which the plastomes occurs in two equimolar states (i.e., inversion isomers) that differ in the orientation of the SSC region. Later this phenomenon was confirmed in other species, e.g., *Heterosigma akashiwo*[43], *Lasthenia burkei*[44], and *Artemisia frigida*[45].

Chloroplast genomes of *C. alpinum*, *C. arcticum,* and *C. nigrescens* contained an identical set of 113 genes which appeared to be identical with *C. arvense*. In the case of the cp genome of *C. glomeratum* lack of the *psbL* gene was noticed during the analyses, but reannotation of the plastome allowed us to identify the *psbL* sequence between *psbJ* and *psbF* genes. Furthermore, two additional genes, i.e., *infA* (coding translation initiation factor I) and *rpl23* (encoding ribosomal protein L23) were not annotated in *C. glomeratum* plastome. Detailed analysis of the chloroplast genome for the species enable identification of these sequences, but their pseudogenization (i.e., the presence of internal, premature termination codons) was the most probable reason why their annotations were not considered by the original authors of the sequence. In the case of *C. alpinum*, *C. arcticum,* and *C. nigrescens rpl23* was also identified as a pseudogene, whereas a complete sequence of *infA* gene was found and annotated. Loss of the *infA* gene was also observed in other species within the Caryophyllales[46]. In some cases, the *infA* gene was found to be a pseudogene, i.a. in *Nicotiana tabacum*[47], *Arabidopsis thaliana*[48], *Oenothera elata*[41] and several *Allium* species[49]. In the chloroplast genomes of another Caryophyllaceae representative, i.e. *Dianthus superbus var. longicalyncinus*, both *infA* and *rpl23* were retained as pseudogenes[50]. Pseudogenization of the *rpl23* gene was also previously reported in various species, i.a. within the genus *Triticum*[51], *Hordeum*[52] and *Secale*[53]. The studied *Cerastium* cp genomes had a GC-content of 36.46–36.52%, which is comparable with other Caryophyllaceae – 36.32% in *Dianthus caryophyllus*[54], 36.4% in *Silene jenisseensis*[55], 36.5% in *Pseudostellaria palibiniana*[56], *P. okamotoi*[57], *P. heterophylla*[58], *P. longipedicellata*[59] and *Gymnocarpos przewalskii*[60] and 36.7% in *Colobanthus quitensis*[61].

The repeat regions of the genomes are of particular importance in sequence rearrangement and recombination[62]. The genomic repeats identified within chloroplast genomes of *C. alpinum*, *C. arcticum*, *C. nigrescens,* and *C. glomeratum* ranged from 30 to 170 bp in length and they were identified predominantly (56.3–69.6%) within non-coding regions. Similar values were reported in other Caryophyllaceae, such as *C. quitensis* (53.3%[63];), *Silene capitata* (56.0%) and *Lychnis wilfordii* (69.2%)[64]. The majority of the repeats (78–90%) in all four *Cerastium* genomes are between 30 and 40 bp in length. Similar values were reported in other angiosperms—legumes (*Glycine*, *Lotus*, *Medicago*[65]) and cotton (*Gossypium hirsutum*[66]).

Chloroplast simple sequence repeats, or microsatellites, are repetitive genomic elements that typically consist of tandemly repeated multiple copies of mono- to hexanucleotide motifs which are usually found in the non-coding regions[67]. Due to their high abundance, random distribution within the genome and high polymorphism information content, they are also widely used for high-throughput genotyping[68]. These markers proved their usefulness in population genetics and evolutionary studies[69,70]. In the analyzed plastomes of four *Cerastium* species, the mononucleotide (A/T) repeats were the most abundant SSR motif (36.4–43.5%). The dominance of mononucleotide chloroplast SSRs has been also observed in other Caryophyllaceae, where it ranged from 44.8% in *Colobanthus apetalus*[63] or 55.3% in *C. lycopodioides*[71] up to 76.8% in *Silene capitata* or 77.6% in *Lychnis wilfordii*[64]. In turn, di- (AT/TA), penta- (AATAT/TATAA) and hexanucleotide (AAATCC/CCTAAA) microsatellites were least abundant, and only one such element was identified in *C. glomeratum*, *C. alpinum,* and *C. arcticum*, respectively.

The synonymous (Ks) and non-synonymous (Ka) substitution rate and their ratio (Ka/Ks) are important parameters in gene evolution studies[72]. Generally, in most of the coding regions synonymous nucleotide substitutions dominate over non-synonymous changes[73]. This was also observed in our study, where Ks values dominated over Ka which resulted in high sequence conservation. Nevertheless, there were also sequences for which considerable variation was found due to the high Ka values. The highest Ka values were observed for *rpl32* (average Ka = 0.0151) and *matK* (average Ka = 0.0134). High variation of the *matK* sequence has been widely documented and it is recognized as one of the most promising barcoding sites for systematic and evolutionary studies in plants[74,75]. There are also studies reporting high genetic diversity in the immediate vicinity of the *rpl32* gene (*ndhF–rpl32* or *rpl32–trnL*)[76,77] and the role of *rpl32* gene in the evolution of chloroplast genomes which involve its complete loss, substitution or transfer to the nucleus (for review see[78]). Assessment of the ratio of nonsynonymous (Ka) to synonymous (Ks) substitution is widely accepted approach used to infer about the direction of the sequence evolution at the protein level (Ka/Ks > 1 indicates a positive selection, Ka/Ks < 1 indicates a negative or purifying selection, whereas Ka/Ks = 1 indicates a neutral evolution)[79,80]. Protein functions are maintained through purifying selection, whereas positive selection favors new gene variants which may be beneficial for organism adapting to changing environmental conditions. In the case of our study, Ka/Ks ratio of all genes was less than 1, except for *ndhB* (2.7250 for *C. arvense*), implying that this gene evolved at a faster rate and underwent positive selection. The same pattern of selection (Ka/Ks > 1.0) for *ndhB* gene was also reported in various species representing the family Gentianaceae (*Gentiana lawrencei*[81]), Orchidaceae (*Calanthe delavayi*[82]) and Cupressaceae (*Cupressus* and *Juniperus* species[83]). The group of *ndh* genes, encoding subunits of NADH dehydrogenase, play a key role in the use of light energy and electron transfer chain to produce ATP, an essential component for photosynthesis[84]. Chloroplast NADH dehydrogenase is sensitive to strong light stress and can protect plants from photoinhibition or photooxidation stress by stabilizing the NADH complex and preventing drought-related declines in photosynthetic rate and growth delay[85]. These observations may suggest that NADH dehydrogenase genes are involved in adaptation to environmental stresses by optimization of photosynthesis. An excess of functionally adaptive amino acid substitutions within NADH dehydrogenase genes was described

previously for Poaceae[86]. Authors observed there the signals of positive selection acting on one-third of all chloroplast protein-coding genes (25 out of 76), including nine of the eleven genes encoding subunits of NADH dehydrogenase. In the case of our study, the signal of positive selection detected for the *ndhB* gene in *C. arvense* which might be interpreted as one of the mechanisms of physical adaptation which enabled this cosmopolitan species to colonize vast areas of Europe and North America.

Highly variable sequences found within chloroplast genomes appeared as a common source of molecular markers suitable for phylogenetic analyses and species identification[87]. Although traditional barcoding chloroplast regions, like *matK, rbcL* or intergenic spacer *trnH-psbA* revealed lower than expected genetic diversity, our genome-wide comparative analysis of plastomes of four *Cerastium* species (*C. alpinum, C. arcticum, C. nigrescens,* and *C. glomeratum*) allowed us to identify nine fast evolving regions. Among these divergent hotspots ($\pi > 0.015$) there were seven regions (*trnD*-GUC–*trnY*-GUA, *trnF*-GAA–*ndhJ*, *ndhC*–*trnV*-UAC, *petA*–*psbJ*, *psbE*–*petL*, *trnP*-UGG–*psaJ* and intron within *rps16* sequence) located within LSC region and two others (*rpl32*–*trnL*-UAG and intron within *ndhA* sequence) identified within SSC region. To the best of our knowledge, none of these chloroplast genome regions have been used to date for phylogeny reconstruction within the genus *Cerastium*. Nevertheless, there are several phylogenetic studies performed within various groups of plant species, including the family Caryophyllaceae, in which at least some of the regions listed above were used, e.g. intron of *rps16*[88], *petA*–*psbJ*[89] or *rpl32*–*trnL*-UAG[90].

RNA editing is one of the most important post-transcriptional modifications which mainly occurs in mitochondrial and chloroplast transcripts[91,92]. RNA editing is described as a process involved in the correction of a missense mutation of genes at the RNA level. This mechanism could alter the nucleotide sequence through insertion, deletion, or substitution of nucleotides[93,94] to preserve the function of encoded proteins[95]. The first report of RNA editing was documented for the *cox2* gene in the protozoan parasite *Trypanosoma brucei*[96], whereas in plants RNA editing was first discovered in the sequence of *cox2* of *Triticum aestivum*[97] and then in *rpl2* in maize[98]. Several editing sites have been reported in many other species, i.a. *A. thaliana*[93], *N. tabacum*[99], *Oryza sativa*[100], *Pisum sativum*[101] and *Manihot esculenta*[102]. RNA editing that converts cytidine into uridine (C into U) is widespread in plant organelles and occurs mostly at the first or second positions of codons[103]. Whereas the reverse U to C conversions is more restricted in occurrence. In studied *Cerastium* species the presence of both C to U and U to C editing has been revealed. RNA editing by U to C is rather rare in terrestrial plants, but it has been found in some species i.a. *A. thaliana*[104], hornworts[105], lycophytes[106] and ferns[107].

One of the plant groups that has been intensively studied in terms of its phylogeny is the family Caryophyllaceae. Traditionally, Caryophyllaceae was divided into three subfamilies: Alsinoideae, Caryophylloideae, and Paronychioideae[108]. However, the traditional taxonomy of the family encountered many difficulties, i.e., most of the genera appeared to be polyphyletic probably because many of the morphological characters evolved in parallel[109]. More recently, a new classification of Caryophyllaceae family based on three chloroplast regions (*matK, trnL-trnF,* and *rps16*) was proposed which divided this group into 11 tribes[110]. Unfortunately, only two *Cerastium* species (*C. arvense* and *C. fontanum*) were represented in this study and based on their molecular characteristics they were nested within the Alsineae tribe, together with representatives of the following genera: *Stellaria, Pseudostellaria, Myosoton, Plettkea, Holosteum, Moenchia,* and *Lepyrodictis*. *Cerastium* is one of the Caryophyllaceae genera whose structure is still intensively debated. Even determining the number of species distinguished within this group of plants is problematic and vary from 60[111] or 100[3,112] up to 200[113] species. Phylogenetic analyses employing multiple nuclear and plastid DNA sequences have established *Cerastium's* monophyly[13,114]. However, there are still some issues associated with *Cerastium* systematics that need clarification, for example, the status of the *C. alpinum*–*C. arcticum* complex which includes *C. alpinum, C. arcticum,* and *C. nigrescens*. Several evolutionary lineages were identified within that complex in earlier research based on morphology, isozymes, and DNA markers[6,10,19]. It was reported that the origin and evolution of these taxa are most likely related to the fluctuations of ice sheet range during the Quaternary glaciations which caused the extensive migrations of the species and enabled multiple hybridization and introgression events[11,19,115]. This hypothesis is consistent with the results of studies reporting no variation in chloroplast *trnL-trnF* and *psbA-trnH* sequences among representatives of the arctic-alpine *C. alpinum*–*C. arcticum* complex and members of the boreal-temperate *C. tomentosum* and *C. arvense* groups[13].

In our study, phylogenetic analysis was based on 71 concatenated protein-coding gene sequences. Revealed phylogenetic relationships between analyzed representatives of the Caryophyllaceae family were in concordance with the taxonomic position of studied species and previous phylogenies of this group[109,116]. Moreover, obtained results allowed us to undoubtedly discriminate all analyzed species, including five representatives of the genus *Cerastium* (*C. alpinum, C. arcticum, C. nigrescens, C. glomeratum,* and *C. arvense*). This is in agreement with the previous observation that a phylogenetic network that combines several genes is preferable to a single-gene tree, as the latter is typically insufficient to reveal reliable phylogenetic relationships[117]. All *Cerastium* species were gathered in one clade, but *C. glomeratum* appeared to be the most divergent from the other species.

Our divergence time analysis confirmed the results of the previous studies on molecular and temporal diversification of the Caryophyllace family. Analogous to the results of the latest research based on nuclear ITS region and four plastid sequences (*matK, rbcL, rps16* and *trnL*-F)[118] our studies suggested that the family Caryophyllaceae began to diversify before the end of Crecateous (ca. 74.46 Mya) and this process continued through the Paleogene and Neogene with the highest intensity of the diversification in the last 10 Mya[119]. According to our observations Alsineae tribe, which includes the genus *Cerastium*, started to diversify at 20.6 Mya, whereas the beginning of that process for the genus *Cerastium* was dated on ca. 3.66 Mya. Our results suggested that *C. glomeratum* split earliest from the other representatives of this genus, whereas the other species appeared to be on the early stages of diversification. The high similarity of studied *Cerastium* plastome sequences may be treated as possible evidence for weak barriers to breeding between these species which enabled spontaneous hybridization between them. Previously, interspecific hybridization events were reported for many *Cerastium* species[8,120].

Although a close relationship between *C. nigrescens* and *C. arcticum* was previously reported[10,18,19], our study suggested a more divergent character of more geographically distant *C. arcticum* and closer genetic relationships between *C. nigrescens* and *C. alpinum*. These observations and results of previous studies pointing to possible hybridization between these two sympatric species (*C. nigrescens* and *C. alpinum*)[11] showed the complexity of evolution which can take place across a broad range of scenarios and spatial circumstances[121].

C. arvense was unexpectedly grouped with species from the *C. alpinum–C. arcticum* complex. This is probably because the publicly available partial sequence of *C. arvense* chloroplast genome that we used in phylogenetic studies lacked complete sequences of six genes (*rpl20*, *rpoB*, *rpoC1*, *rpoC2*, *ycf1,* and *ycf2*) thus the phylogeny reconstruction was performed on the limited number (71) of plastid genes. The absence of these genes might be then responsible for the underestimation of genetic divergence between *C. arvense* and other *Cerastium* species. The application of a complete sequence of chloroplast genome appeared here as the alternative method for distinguishing the true phylogenetic relationships between these closely related taxa. This approach has already proved its usefulness for taxa with a relatively short time since the divergence event or a low rate of evolution resulted in low sequence variation[31,122]. Nevertheless, in both cases resequencing of *C. arvense* is required.

Although complete chloroplast genomes of three *Cerastium* species (*C. alpinum, C. arcticum* and *C. nigrescens*) were reported and characterized here for the first time further research is required to investigate and finally resolve the taxonomic issues associated with the genus *Cerastium* and the *C. alpinum–C. arcticum* complex. Subsequent studies should include not only analyses of chloroplast genomes but also nucleic regions because when hybridization and polyploidy are common the resolution that chloroplast genome sequence can provide for phylogenomics research may be limited[123]. Nevertheless, our results proved the suitability of chloroplast genome sequences as reliable and effective DNA barcodes for *Cerastium* species.

Conclusion

The chloroplast genomes of *Cerastium alpinum*, *C. arcticum,* and *C. nigrescens* were sequenced and characterized for the first time. The reported chloroplast genomes appeared to be highly conserved in terms of the gene content and order as well as their quadripartite structure. Highly divergent regions (*rpl32–trnL*-UAG, *ndhA* intron, *rps16* intron, *trnD*-GUC–*trnY*-GUA, *trnF*-GAA–*ndhJ*, *ndhC–trnV*-UAC, *petA–psbJ*, *psbE–petL* and *trnP*-UGG–*psaJ*) and microsatellite sequences that could be potentially used as markers in genetic diversity or phylogenetic studies were identified. Reconstruction of phylogenetic relationships within the family Caryophyllaceae confirmed the previously reported systematic relations within that group of plants and supported the position of *Cerastium* species as a separate clad within the tribe Alsineae. Although obtained data provide insight into the evolution and biogeographic history of the genus *Cerastium* further studies are needed to finally elucidate the relationships between species from the *C. alpinum–C. arcticum* complex.

Methods

Plant material, DNA extraction and chloroplast genome sequencing

The research material consisted of three *Cerastium* species–*C. alpinum*, *C. arcticum,* and *C. nigrescens*. Fresh leaves of *C. alpinum* and *C. arcticum* were harvested from plants grown from seeds in a greenhouse (Department of Plant Physiology, Genetics and Biotechnology, University of Warmia and Mazury in Olsztyn, Poland). The seeds of *C. alpinum* were collected in 2020 in Babia Góra National Park (Poland) after obtaining permission from the Polish Ministry of Environment. In the case of *C. arcticum*, seeds were collected by Michał Węgrzyn from the Institute of Botany of Jagiellonian University in Kraków, Poland, during the Arctic expedition to Nicolaus Copernicus University Polar Station in Spitsbergen in 2012. In turn, *C. nigrescens* individuals were collected by Keith W. Larson from Climate Impacts Research Centre, Umeå University, Sweden, in Nuolja massif (Sweden) and delivered to Olsztyn in dried form. The species identification included analysis of both vegetative and generative organs. In the case of *C. alpinum* and *C. arcticum* identification was performed by Irena Giełwanowska, whereas *C. nigrescens* status was verified by Keith W. Larson. Voucher specimens of each studied species have been deposited in the Vascular Plants Herbarium of the Department of Botany and Nature Protection at the University of Warmia and Mazury in Olsztyn, Poland (OLS), under the following numbers: *C. alpinum* (No. OLS 33837), *C. arcticum* (No. OLS 33840) and *C. nigrescens* (No. OLS 33841). The photographs of the representatives of each studied species were provided as the supplementary material: *C. alpinum* (Supplementary Fig. S3), *C. arcticum* (Supplementary Fig. S4) and *C. nigrescens* (Supplementary Fig. S5).

Total genomic DNA was extracted from the fresh or dried material of a single plant using Maxwell 16 LEV Plant DNA Kit (Promega, Madison, WI). The amount and purity of isolated DNA was estimated spectrophotometrically (NanoDrop ND-1000 UV/Vis; NanoDrop Technology). Additionally, the quality of DNA was verified on 1.5% (w/v) agarose in the presence of 0.5 μg/ml ethidium bromide (wavelength 300 nm; Ultra-Lum EB-20 Electronic UV Transilluminator).

The appropriate genome libraries (library kit: TruSeq DNA PCR Free (350), prepared from high-quality genomic DNA, were sequenced on Illumina NovaSeq6000 platform (Illumina Inc., San Diego, CA, USA) with a 150 bp paired-end read.

Annotation and genome analysis

The quality of raw reads was checked with the FastQC tool. Raw reads were trimmed (5 bp of each read end, regions with more than 5% probability of error per base) and mapped to the reference chloroplast genome of *C. glomeratum* (NC_066897) using Geneious v.R7 software[124] with medium–low sensitivity settings. The details on subsequent procedures for chloroplast genome assembly and annotation were described in our previous study[78]. The chloroplast genomes were annotated using PlasMapper[125] with manual adjustment and circular maps of

chloroplast genomes were drawn using the OrganellarGenome DRAW tool[126]. Each chloroplast genome assembly was validated using GetOrganelle v.1.7.7.0[127].

Additionally, to check for the possible presence of heteroplasmy variant calling analysis was performed in Geneious software using "Find Variations/SNPs (Single Nucleotide Polymorphism)" feature with the following parameters: minimum variant frequency = 0.1; minimum coverage = 10, p-value cut off = 0.0001 and default values for the remaining parameters.

Genomic repeats and SSR analysis

The genomic repeats, including forward, reverse, palindromic and complementary sequences were identified using REPuter software[128] with the following settings: minimal repeat size of 30 bp, Hamming distance of 3, and 90% sequence identity. Chloroplast simple sequence repeats (SSR), also called microsatellites, were identified in Phobos v.3.3.12[129]. Only perfect SSRs with a motif size of one to six nucleotide units were considered. Additionally, we applied the standard thresholds for chloroplast SSRs' identification[130]: minimum number of repeat units were set to 12, 6, 4, 3, 3, and 3 for mono-, di-, tri-, tetra-, penta- and hexanucleotides, respectively. A single IR region was used to eliminate the influence of doubled IR regions, and redundant results were deleted manually.

Comparative analysis of chloroplast genomes

Chloroplast genome sequences of three *Cerastium* species (*C. alpinum, C. arcticum, C. nigrescens*) reported in this paper and plastome sequence of *C. glomeratum* (NC_066897) and *C. arvense* (MH627219) acquired from NCBI database were used for the genome synteny analysis which was performed with the use of MAUVE v.1.1.1[131]. Furthermore, the sequences were aligned in MAFFT v.7.310[132] to perform sliding window analysis and evaluate nucleotide diversity (π) in chloroplast genomes using DnaSP v.6.10.04[133]. The step size was set to 50 base pairs, and the window length was set to 800 base pairs. Here, only complete chloroplast genome sequences were used—*C. arvense* plastome which has several gaps in its sequence was excluded from this analysis. The results were visualized with the CIRCOS software package v.0.69–9[134].

The selective pressure for genes identified in chloroplast genomes of *C. alpinum, C. arcticum, C. nigrescens, C. glomeratum,* and *C. arvense* was also analyzed. A total number of 77 protein-coding genes were selected for which synonymous (Ks) and non-synonymous (Ka) substitution rates, as well as Ka/Ks ratio, were estimated using DnaSP v.6.10.04. *Cerastium glomeratum chloroplast* genome was used as a reference. During the analyses, lack of *psbL* gene was noticed in *C. glomeratum*. Reannotation of the *C. glomeratum* plastome allowed us to identify the sequence for this lacking gene in its traditional position i.e., between *psbJ* and *psbF* (detailed location: 61391..61507). In the *C. arvense* cp genome all genes which were annotated in plastomes of *C. alpinum, C. arcticum, C. nigrescens* were also present, but unknown nucleotides (n) were recorded in six (*rpl20*, *rpoB*, *rpoC1*, *rpoC2*, *ycf1,* and *ycf2*), therefore these sequences were excluded from calculations for this species. The results were visualized with the CIRCOS software package v.0.69-9[134].

The junction sites between LSC, SSC, and IRs regions were also identified and compared. Additionally, data on the codon usage distribution was acquired from the Geneious v.R7 statistic panel.

Prediction of RNA editing sites

Potential RNA editing sites in the protein-coding genes from chloroplast genomes of *C. alpinum, C. arcticum, C. nigrescens, C. glomeratum,* and *C. arvense* were predicted using PREPACT 3.0 tool[135]. *Arabidopsis thaliana* (NC_000932) was used as a reference for BLASTx prediction, both forward (C to U) and reverse (U to C) editing options were selected, while the remaining settings were kept at default (0.001 e-value cutoff and 30% filter threshold). In the case of *C. arvense rpl20*, *rpoB*, *rpoC1*, *rpoC2*, *ycf1,* and *ycf2* genes were excluded from the analysis as unknown nucleotides (n) were recorded in their sequences. The results were visualized with the CIRCOS software package v.0.69–9[134].

Phylogenetic analysis

Chloroplast genome sequences of three *Cerastium* species (*C. alpinum, C. arcticum,* and *C. nigrescens*) reported in this paper, as well as 56 plastomes of other representatives of family Caryophyllaceae (including *C. glomeratum* and *C. arvense*) and *A. thaliana* (outgroup), were used for phylogenetic analysis (Table 3). Initially, the sequences of 71 protein-coding genes shared by all these species were extracted using a custom R script. Then, the concatenated sequences of 71 genes were aligned in MAFFT v7.310 and used for phylogeny reconstruction by Bayesian Inference (BI). The Mega v.7 software[136] was used to determine the best-fitting substitution model, and the GTR + G + I model was selected. The BI analysis was conducted using MrBayes v.3.2.6[137,138], according to the parameter's settings described in our previous paper[63]. The obtained phylogenetic tree was used as a starting tree for divergence time analysis performed using RelTimeML feature in MEGA 7 with GTR model. The divergence time between *Cerastium arvense* and *Myosoton aquaticum* (6.2–38.1 Mya), *Arenaria serpyllifolia* and *Pseudostellaria japonica* (20.3–83.4 Mya) and *Dianthus chinensis* and *Silene latifolia* (20.3–46.7 Mya) obtained in TimeTree[139] were used as calibration constraints in calculations.

Ethics declarations

Authors confirm that the use of plants in the present study complies with international, national and/or institutional guidelines.

Species	Accession number	Species	Accession number
Arabidopsis thaliana	NC_000932	*Pseudostellaria heterantha*	NC_058231
Agrostemma githago	NC_023357	*Pseudostellaria heterophylla*	NC_044183
Arenaria serpyllifolia	NC_065316	*Pseudostellaria japonica*	OP526393
Cerastium alpinum	QC696752	*Pseudostellaria longipedicellata*	NC_039454
Cerastium arcticum	QC696753	*Pseudostellaria okamotoi*	NC_039974
Cerastium arvense	MH627219	*Pseudostellaria palibiniana*	NC_041166
Cerastium glomeratum	NC_066897	*Pseudostellaria setulosa*	NC_041462
Cerastium nigrescens	QC696754	*Silene aprica*	MN097700
Colobanthus acicularis	NC_053724	*Silene asclepiadea*	NC_061183
Colobanthus affinis	NC_053722	*Silene atrocastanea*	NC_061184
Colobanthus apetalus	NC_036424	*Silene capitata*	NC_035226
Colobanthus lycopodioides	NC_053721	*Silene chalcedonica*	NC_023359
Colobanthus nivicola	NC_053720	*Silene chodatii*	NC_061185
Colobanthus pulvinatus	NC_053719	*Silene conica*	NC_016729
Colobanthus quitensis	NC_028080	*Silene conoidea*	NC_023358
Colobanthus subulatus	NC_053723	*Silene delavayi*	NC_061186
Dianthus caryophyllus	NC_039650	*Silene gracilicaulis*	NC_061187
Dianthus chinensis	NC_053731	*Silene jenisseensis*	MN723869
Dianthus gratianopolitanus	LN877392	*Silene kiusiana*	NC_048886
Dianthus longicalyx	NC_050834	*Silene latifolia*	NC_016730
Dianthus moravicus	LN877396	*Silene linearilobа*	NC_061188
Gymnocarpos przewalskii	NC_036812	*Silene melanantha*	NC_061189
Gypsophila huashanensis	OP094658	*Silene noctiflora*	NC_016728
Gypsophila oldhamiana	NC_058757	*Silene paradoxa*	NC_023360
Gypsophila vaccaria	NC_040936	*Silene stewartiana*	NC_061190
Lychnis wilfordii	NC_035225	*Silene tatarinowii*	NC_061191
Myosoton aquaticum	MZ570968	*Silene viscidula*	NC_061192
Paronychia argentea	NC_066008	*Silene vulgaris*	NC_016727
Psammosilene tunicoides	MK684403	*Spergula arvensis*	NC_041240
Pseudostellaria davidii	OP526392	*Stellaria dichotoma var. lanceolata*	MN718731

Table 3. GenBank accession numbers and references for chloroplast genomes used in this study. Species list arranged alphabetically.

Data availability

The complete chloroplast genomes reported in this paper have been submitted to the NCBI database under the following accession numbers: QC696752 for *C. alpinum*, QC696753 for *C. arcticum,* and QC696754 for *C. nigrescens*.

Received: 7 June 2023; Accepted: 26 October 2023
Published online: 31 October 2023

References

1. The World Flora Online. http://www.worldfloraonline.org. Accessed Sep 2022 (2022).
2. Jalas, J., Wyse Jackson, M. B., Sell, P. D. & Whitehead, F. H. *Cerastium* L. *Flora Europaea.* 2nd Ed. Vol. 1 (ed. Tutin, T. G. *et al.*). 164–175. (Cambridge University Press, 1993).
3. Dequan, L. & Morton, J. *Cerastium* L. *Flora of China.* Vol. 6 (*Caryophyllaceae Through Lardizabalaceae*) (ed. Wu, Z. & Raven, P. H). (Science Press/Missouri Botanical Garden Press, 2001).
4. Quiroga, M. P., Premoli, A. C. & Ezcurra, C. Morphological and isozyme variation in *Cerastium arvense* (Caryophyllaceae) in the southern Andes. *Can. J. Bot.* **80**, 786–795 (2002).
5. Nyberg Berglund, A. B., Saura, A. & Westerbergh, A. Electrophoretic evidence for disomic inheritance and allopolyploid origin of the octoploid *Cerastium alpinum* (Caryophyllaceae). *J. Hered.* **97**, 296–302 (2006).
6. Hagen, A. R., Giese, H. & Brochmann, C. Trans-Atlantic dispersal and phylogeography of *Cerastium arcticum* (Caryophyllaceae) inferred from RAPD and SCAR markers. *Am. J. Bot.* **88**, 103–112 (2001).
7. Caković, D., Stešević, D., Schönswetter, P. & Frajman, B. Long neglected diversity in the accursed mountains of northern Albania: *Cerastium hekuravense* is genetically and morphologically divergent from *C. dinaricum*. *Plant Syst. Evol.* **304**, 57–69 (2018).
8. Niketić, M., Đurović, S. Z., Tomović, G., Schönswetter, P. & Frajman, B. Diversification within ploidy-variable Balkan endemic *Cerastium decalvans* (Caryophyllaceae) reconstructed based on genetic, morphological and ecological evidence. *Bot. J. Linn. Soc.* **199**, 578–608 (2022).
9. Milarska, S. E., Androsiuk, P., Bednarek, P. T., Larson, K. & Giełwanowska, I. Genetic variation of *Cerastium alpinum* L. from Babia Góra, a critically endangered species in Poland. *J. Appl. Genet.* **64**, 37–53 (2023).
10. Brysting, A. K. & Borgen, L. Isozyme analysis of the *Cerastium alpinum-C. arcticum* complex (Caryophyllaceae) supports a splitting of *C. arcticum* Lange. *Plant Syst. Evol.* **220**, 199–221 (2000).

11. Hagen, A. R. *et al.* The arctic-alpine polyploids *Cerastium alpinum* and *C. nigrescens* (Caryophyllaceae) in a sympatric situation: Breakdown of species integrity?. *Plant Syst. Evol.* **230**, 203–219 (2002).
12. Tolmachev, A. Die Gattung *Cerastium* in der Flora von Spitzbergen. *Skr. Svalb. Ishavet* **34**, 1–8 (1930).
13. Scheen, A. C. *et al.* Northern hemisphere biogeography of *Cerastium* (Caryophyllaceae): Insights from phylogenetic analysis of noncoding plastid nucleotide sequences. *Am. J. Bot.* **91**, 943–952 (2004).
14. Hultén, E. The *Cerastium alpinum* complex. A case of world-wide introgressive hybridization. *Svensk Bot. Tidskr.* **50**, 411–495 (1956).
15. Böcher, T. W. *Cerastium alpinum* and *C. arcticum*, a mature polyploid complex. *Bot. Not.* **130**, 303–309 (1977).
16. Rønning, O. I. *The Flora of Svalbard* (Norsk Polarinstitutt, 1996).
17. Elven, R. & Elvebakk, A. Part 1. Vascular plants. In *A Catalogue of Svalbard Plants, Fungi, Algae, and Cyanobacteria* (eds. Elvebakk, A. & Prestrud, P.). Vol. 198. 9–55 (Norsk Polarinst. Skr., 1996).
18. Brysting, A. K. & Hagen, A. Species in polyploid complexes? The *Cerastium alpinum-C. arcticum* complex. *Det Norske Videnskaps Akademi. I. Mat. Nat. Kl. Avh. Ny Serie* **38**, 183–190 (1999).
19. Brysting, A. K. & Elven, R. The *Cerastium alpinum*–C. *arcticum* complex (Caryophyllaceae): Numerical analyses of morphological variation and taxonomical revision of *C. arcticum* Lange. *Taxon* **49**, 186–216 (2000).
20. Parusel, J. *Cerastium alpinum* L. s.s. Rogownica alpejska. In *Polska Czerwona Księga Roślin* (ed. Kaźmierczakowa, R. & Zarzycki, K.). 97–99 (Instytut Botaniki im. W. Szafera PAN, Instytut Ochrony Przyrody PAN, 2001).
21. Parusel, J. *Cerastium alpinum* Rogownica alpejska. In *Polska Czerwona Księga Roślin. Paprotniki i Rośliny Kwiatowe* (eds. Kaźmierczakowa, R., Zarzycki, K. & Mirek, Z.). 121–123 (Instytut Ochrony Przyrody PAN, 2014).
22. Jansen, R. K. *et al.* Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc. Natl. Acad. Sci. USA* **104**, 19369–19374 (2007).
23. Bortiri, E., Coleman-Derr, D., Lazo, G., Anderson, O. & Gu, Y. The complete chloroplast genome sequence of *Brachypodium distachyon*: Sequence comparison and phylogenetic analysis of eight grass plastomes. *BMC Res. Notes* **1**, 61 (2008).
24. Bock, R. Plastid biotechnology: prospects for herbicide and insect resistance, metabolic engineering and molecular farming. *Curr. Opin. Biotechnol.* **18**, 100–106 (2007).
25. Kress, W. J., Wurdack, K. J., Zimmer, E. A., Weigt, L. A. & Janzen, D. H. Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci. USA* **102**, 8369–8374 (2005).
26. Chase, M. W. *et al.* A proposal for standardized protocol to barcode all land plants. *Taxon* **56**, 295–299 (2007).
27. Raman, G. & Park, S. The complete chloroplast genome sequence of *Ampelopsis*: Gene organization, comparative analysis, and phylogenetic relationships to other angiosperms. *Front. Plant Sci.* **7**, 341 (2016).
28. Zhai, Y. *et al.* Phylogenomics, phylogeography and germplasms authentication of the *Rheum palmatum* complex based on complete chloroplast genomes. *J. Plant. Res.* **136**, 291–304 (2023).
29. Parks, M., Cronn, R. & Liston, A. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biol.* **7**, 84 (2009).
30. Ohyama, K. *et al.* Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. *Nature* **322**, 572–574 (1986).
31. Daniell, H., Lin, C. S., Yu, M. & Chang, W. J. Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. *Genome Biol.* **17**, 134 (2016).
32. Gitzendanner, M. A., Soltis, P. S., Wong, G. K., Ruhfel, B. R. & Soltis, D. E. Plastid phylogenomic analysis of green plants: A billion years of evolutionary history. *Am. J. Bot.* **105**, 291–301 (2018).
33. Zhai, W. *et al.* Chloroplast genomic data provide new and robust insights into the phylogeny and evolution of the Ranunculaceae. *Mol. Phylogenet. Evol.* **135**, 12–21 (2019).
34. Dong, W., Xu, C., Cheng, T., Lin, K. & Zhou, S. Sequencing angiosperm plastid genomes made easy: A complete set of universal primers and a case study on the phylogeny of Saxifragales. *Genome Biol. Evol.* **5**, 989–997 (2013).
35. Zhu, A., Guo, W., Gupta, S., Fan, W. & Mower, J. P. Evolutionary dynamics of the plastid inverted repeat: The effects of expansion, contraction, and loss on substitution rates. *New Phytol.* **209**, 1747–1756 (2016).
36. Reith, M. & Munholland, J. Complete nucleotide sequence of the *Porphyra purpurea* chloroplast genome. *Plant Mol. Biol. Rep.* **13**, 333–335 (1995).
37. Wolfe, K. H. The site of deletion of the inverted repeat in pea chloroplast DNA contains duplicated gene fragments. *Curr. Genet.* **13**, 97–99 (1988).
38. Kim, K. J. & Lee, H. L. Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. *DNA Res.* **11**, 247–261 (2004).
39. Goulding, S. E., Olmstead, R. G., Morden, C. W. & Wolfe, K. H. Ebb and flow of the chloroplast inverted repeat. *Mol. Gen. Genet.* **252**, 195–206 (1996).
40. Goremykin, V. V., Hirsch-Ernst, K. I., Wölfl, S. & Hellwig, F. H. Analysis of the *Amborella trichopoda* chloroplast genome sequence suggests that *Amborella* is not a basal angiosperm. *Mol. Biol. Evol.* **20**, 1499–1505 (2003).
41. Hupfer, H. *et al.* Complete nucleotide sequence of the *Oenothera elata* plastid chromosome, representing plastome I of the five distinguishable *Euoenothera* plastomes. *Mol. Gen. Genet.* **263**, 581–585 (2000).
42. Palmer, J. D. Chloroplast DNA exists in two orientations. *Nature* **301**, 92–93 (1983).
43. Cattolico, R. A. *et al.* Chloroplast genome sequencing analysis of *Heterosigma akashiwo* CCMP452 (West Atlantic) and NIES293 (West Pacific) strains. *BMC Genom.* **9**, 211 (2008).
44. Walker, J. F., Zanis, M. J. & Emery, N. C. Comparative analysis of complete chloroplast genome sequence and inversion variation in *Lasthenia burkei* (Madieae, Asteraceae). *Am. J. Bot.* **101**, 722–729 (2014).
45. Wang, M., Cui, L., Feng, K. & Deng, P. Comparative analysis of Asteraceae chloroplast genomes: Structural organization, RNA editing and evolution. *Plant Mol. Biol. Rep.* **33**, 1526–1538 (2015).
46. Millen, R. S. *et al.* Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. *Plant Cell.* **13**, 645–658 (2001).
47. Shinozaki, K. *et al.* The complete nucleotide sequence of the tobacco chloroplast genome: Its gene organization and expression. *EMBO J.* **5**, 2043–2049 (1986).
48. Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E. & Tabata, S. Complete structure of the chloroplast genome of *Arabidopsis thaliana*. *DNA Res.* **6**, 283–290 (1999).
49. Scobeyeva, V. A. *et al.* Gene loss, pseudogenization in plastomes of genus *Allium* (Amaryllidaceae), and putative selection for adaptation to environmental conditions. *Front. Genet.* **12**, 674783 (2021).
50. Raman, G. & Park, S. Analysis of the complete chloroplast genome of a medicinal plant, *Dianthus superbus* var. *longicalyncinus*, from a comparative genomics perspective. *PLOS ONE* **10**, e0141329 (2015).
51. Ogihara, Y., Terachi, T. & Sasakuma, T. Molecular analysis of the hot spot region related to length mutations in wheat chloroplast DNAs. I. Nucleotide divergence of genes and intergenic spacer regions located in the hot spot region. *Genetics* **129**, 873–884 (1991).
52. Lencina, F. *et al.* The *rpl23* gene and pseudogene are hotspots of illegitimate recombination in barley chloroplast mutator seedlings. *Sci. Rep.* **9**, 9960 (2019).

53. Skuza, L. *et al.* Molecular structure, comparative and phylogenetic analysis of the complete chloroplast genome sequences of weedy rye *Secale cereale* ssp. *segetale*. *Sci. Rep.* **13**, 5412 (2023).
54. Li, G., Tembrock, L. R., Wu, Z. & Liu, F. Complete chloroplast genome of carnation (Caryophyllaceae: *Dianthus caryophyllus* L.). *Mitochondrial DNA Part B* **4**, 1463–1464 (2019).
55. Ling, L. Z. Characterization of the complete chloroplast genome and phylogenetic analysis of *Silene jenisseensis* (Caryophyllaceae). *Mitochondrial DNA Part B* **5**, 476–477 (2020).
56. Kim, Y., Heo, K. I., Lee, S. & Park, J. The complete chloroplast genome sequence of *Pseudostellaria palibiniana* (Takeda) Ohwi (Caryophyllaceae). *Mitochondrial DNA Part B* **4**, 973–974 (2019).
57. Kim, J. & Park, J. The complete chloroplast genome sequence of the *Pseudostellaria okamotoi* Ohwi (Caryophyllaceae). *Mitochondrial DNA Part B* **4**, 174–175 (2019).
58. Kim, Y., Xi, H. & Park, J. The complete chloroplast genome of Prince Ginseng, *Pseudostellaria heterophylla* (Miq.) Pax (Caryophyllaceae). *Mitochondrial DNA Part B* **4**, 2251–2253 (2019).
59. Kim, Y., Heo, K. I., Lee, S. & Park, J. Complete chloroplast genome sequence of the *Pseudostellaria longipedicellata* S. Lee, K. Heo & S. C. Kim (Caryophyllaceae). *Mitochondrial DNA Part B* **3**, 1296–1297 (2018).
60. Yang, Z., Zhang, Y., Pan, L. & Fu, C. Characterization of the complete chloroplast genome of *Gymnocarpos przewalskii*, an endangered species in China and Mongolia. *Conserv. Genet. Resour.* **10**, 717–721 (2018).
61. Kang, Y. *et al.* The complete chloroplast genome of Antarctic pearlwort, *Colobanthus quitensis* (Kunth) Bartl. (Caryophyllaceae). *Mitochondrial DNA Part A* **27**, 4677–4678 (2015).
62. Timme, R. E., Kuehl, J. V., Boore, J. L. & Jansen, R. K. A comparative analysis of the *Lactuca* and *Helianthus* (Asteraceae) plastid genomes: Identification of diverged regions and categorization of shared repeats. *Am. J. Bot.* **94**, 302–312 (2007).
63. Androsiuk, P. *et al.* The complete chloroplast genome of *Colobathus apetalus* (Labill.) Druce: Genome organization and comparison with related species. *PeerJ* **6**, e4723 (2018).
64. Kang, J. S., Lee, B. Y. & Kwak, M. The complete chloroplast genome sequences of *Lychnis wilfordii* and *Silene capitata* and comparative analyses with other Caryophyllaceae genomes. *PLOS ONE* **12**, e0172924 (2017).
65. Saski, C. *et al.* Complete chloroplast genome sequence of *Glycine max* and comparative analyses with other legume genomes. *Plant Mol. Biol.* **59**, 309–322 (2005).
66. Lee, S. B. *et al.* The complete chloroplast genome sequence of *Gossypium hirsutum*: Organization and phylogenetic relationships to other angiosperms. *BMC Genomics* **7**, 61 (2006).
67. Jakobsson, M., Säll, T., Lind-Halldén, C. & Halldén, C. Evolution of chloroplast mononucleotide microsatellites in *Arabidopsis thaliana*. *Theor. Appl. Genet.* **114**, 223–235 (2007).
68. Hirano, R. *et al.* Propagation management methods have altered the genetic variability of two traditional mango varieties in Myanmar, as revealed by SSR. *Plant Genet. Resour. C.* **9**, 404–410 (2011).
69. Takahashi, D., Sakaguchi, S., Isagi, Y. & Setoguchi, H. Comparative chloroplast genomics of series *Sakawanum* in genus *Asarum* (Aristolochiaceae) to develop single nucleotide polymorphisms (SNPs) and simple sequence repeat (SSR) markers. *J. For. Res.* **23**, 387–392 (2018).
70. Ping, J. *et al.* Molecular evolution and SSRs analysis based on the chloroplast genome of *Callitropsis funebris*. *Ecol. Evol.* **11**, 4786–4802 (2021).
71. Androsiuk, P. *et al.* Evolutionary dynamics of the chloroplast genome sequences of six *Colobanthus* species. *Sci. Rep.* **10**, 11522 (2020).
72. Liu, F. *et al.* The complete chloroplast genome and characteristics analysis of *Callistemon rigidus* R.Br. *Mol. Biol. Rep.* **47**, 5013–5024 (2020).
73. Makalowski, W. & Boguski, M. S. Evolutionary parameters of the transcribed mammalian genome: An analysis of 2,820 orthologous rodent and human sequences. *Proc. Natl. Acad. Sci. USA* **95**, 9407–9412 (1998).
74. Newmaster, S. G., Fazekas, A. J. & Ragupathy, S. DNA barcoding in land plants: evaluation of rbcL in a multigene tiered approach. *Can. J. Bot.* **84**, 335–341 (2006).
75. Saarela, J. M., Sokoloff, P. C., Gillespie, L. J., Consaul, L. L. & Bull, R. D. DNA barcoding the Canadian Arctic Flora: Core plastid barcodes (*rbcL* + *matK*) for 490 vascular plant species. *PLOS ONE* **8**, e77982 (2013).
76. Wang, Y. *et al.* Chloroplast genome variation and phylogenetic relationships of *Atractylodes* species. *BMC Genomics* **22**, 103 (2021).
77. Choi, K., Hwang, Y. & Hong, J. K. Comparative chloroplast genomics and phylogenetic analysis of *Persicaria amphibia* (Polygonaceae). *Diversity* **14**, 641 (2022).
78. Alqahtani, A. A. & Jansen, R. K. The evolutionary fate of *rpl32* and *rps16* losses in the *Euphorbia schimperi* (Euphorbiaceae) plastome. *Sci. Rep.* **11**, 7466 (2021).
79. Nei, M. & Kumar, S. *Molecular Evolution and Phylogenetics* (Oxford University Press, 2000).
80. Yang, Z. & Bielawski, J. P. Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* **15**, 496–503 (2000).
81. Fu, P. C., Zhang, Y. Z., Geng, H. M. & Chen, S. L. The complete chloroplast genome sequence of *Gentiana lawrencei* var. *farreri* (Gentianaceae) and comparative analysis with its congeneric species. *PeerJ* **4**, e2540 (2016).
82. Chen, Q., Wu, X. & Zhang, D. Comparison of the abilities of universal, super, and specific DNA barcodes to discriminate among the original species of *Fritillariae cirrhosae* bulbus and its adulterants. *PLOS ONE* **15**, e0229181 (2020).
83. Chen, C., Xia, X., Peng, J. & Wang, D. Comparative analyses of six complete chloroplast genomes from the genus *Cupressus* and *Juniperus* (Cupressaceae). *Gene* **837**, 146696 (2022).
84. Yamori, W. & Shikanai, T. Physiological functions of cyclic electron transport around photosystem I in sustaining photosynthesis and plant growth. *Annu. Rev. Plant Biol.* **67**, 81–106 (2016).
85. Rumeau, D., Peltier, G. & Cournac, L. Chlororespiration and cyclic electron flow around PSI during photosynthesis and plant stress response. *Plant Cell Environ.* **30**, 1041–1051 (2007).
86. Piot, A., Hackel, J., Christin, P. A. & Besnard, G. One-third of the plastid genes evolved under positive selection in PACMAD grasses. *Planta* **247**, 255–266 (2018).
87. Yang, L. *et al.* Analysis of complete chloroplast genome sequences and insight into the phylogenetic relationships of *Ferula* L.. *BMC Genomics* **23**, 643 (2022).
88. Yao, G. *et al.* Phylogenetic estimation and morphological evolution of Alsineae (Caryophyllaceae) shed new insight into the taxonomic status of the genus *Pseudocerastium*. *Plant Divers.* **43**, 299–307 (2020).
89. Sánchez-del Pino, I. *et al.* High phylogeographic and genetic diversity of *Tidestromia lanuginosa* supports full-glacial refugia for arid-adapted plants in southern and central Coahuila, Mexico. *Am. J. Bot.* **107**, 1296–1308 (2020).
90. Moonlight, P. W. *et al.* Dividing and conquering the fastest-growing genus: Towards a natural sectional classification of the mega-diverse genus *Begonia* (Begoniaceae). *Taxon* **67**, 267–323 (2018).
91. Steinhauser, S., Beckert, S., Capesius, I., Malek, O. & Knoop, V. Plant mitochondrial RNA editing. *J. Mol. Evol.* **48**, 303–312 (1999).
92. Knoop, V. When you can't trust the DNA: RNA editing changes transcript sequences. *Cell. Mol. Life Sci.* **68**, 567–586 (2011).
93. Tillich, M. *et al.* Editing of plastid RNA in *Arabidopsis thaliana* ecotypes. *Plant J.* **43**, 708–715 (2005).
94. Chen, H., Deng, L., Jiang, Y., Lu, P. & Yu, J. RNA editing sites exist in protein-coding genes in the chloroplast genome of *Cycas taitungensis*. *J. Integr. Plant Biol.* **53**, 961–970 (2011).

95. Takenaka, M., Zehrmann, A., Verbitskiy, D., Härtel, B. & Brennicke, A. RNA editing in plants and its evolution. *Annu. Rev. Genet.* **47**, 335–352 (2013).
96. Benne, R. *et al.* Major transcript of the frame shifted *coxII* gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. *Cell* **46**, 819–826 (1986).
97. Covello, P. S. & Gray, M. W. RNA editing in plant mitochondria. *Nature* **341**, 662–666 (1989).
98. Hoch, B., Maier, R. M., Appel, K., Igloi, G. L. & Kossel, H. Editing of a chloroplast mRNA by creation of an initiation codon. *Nature* **353**, 178–180 (1991).
99. Sasaki, T., Yukawa, Y., Miyamoto, T., Obokata, J. & Sugiura, M. Identification of RNA editing sites in chloroplast transcripts from the maternal and paternal progenitors of tobacco (*Nicotiana tabacum*): Comparative analysis shows the involvement of distinct trans-factors for *ndhB* editing. *Mol. Biol. Evol.* **20**, 1028–1035 (2003).
100. Corneille, S., Lutz, K. & Maliga, P. Conservation of RNA editing between rice and maize plastids: Are most editing events dispensable?. *Mol. Gen. Genet.* **264**, 419–424 (2000).
101. Miyamoto, T., Obokata, J. & Sugiura, M. Recognition of RNA editing sites is directed by unique proteins in chloroplasts: Biochemical identification of *cis*-acting elements and *trans*-acting factors involved in RNA editing in tobacco and pea chloroplasts. *Mol. Cell. Biol.* **22**, 6726–6734 (2002).
102. Daniell, H. *et al.* The complete nucleotide sequence of the cassava (*Manihot esculenta*) chloroplast genome and the evolution of *atpF* in Malpighiales: RNA editing and multiple losses of a group II intron. *Theor. Appl. Genet.* **116**, 723–737 (2008).
103. Germain, A., Hotto, A. M., Barkan, A. & Stern, D. B. RNA processing and decay in plastids. *Wiley Interdiscip. Rev. RNA* **4**, 295–316 (2013).
104. Ruchika, O. C., Sakari, M. & Tsukahara, T. Genome-wide identification of U-To-C RNA editing events for nuclear genes in *Arabidopsis thaliana*. *Cells* **10**, 635 (2021).
105. Xue, J., Liu, Y., Li, L., Wang, B. & Qiu, Y. The complete mitochondrial genome sequence of the hornwort *Phaeoceros laevis*: Retention of many ancient pseudogenes and conservative evolution of mitochondrial genomes in hornworts. *Curr. Genet.* **56**, 53–61 (2010).
106. Grewe, F. *et al.* A unique transcriptome: 1782 positions of RNA editing alter 1406 codon identities in mitochondrial mRNAs of the lycophyte *Isoetes engelmannii*. *Nucleic Acids Res.* **39**, 2890–2902 (2011).
107. Knie, N., Grewe, F. & Knoop, V. Monilophyte mitochondrial rps1 genes carry a unique group II intron that likely originated from an ancient paralog in rpl2. *RNA* **22**, 1338–1348 (2016).
108. Rabeler, R. K. & Bittrich, V. Suprageneric nomenclature in the Caryophyllaceae. *Taxon* **42**, 857–863 (1993).
109. Fior, S., Karis, P. O., Casazza, G., Minuto, L. & Sala, F. Molecular phylogeny of the Caryophyllaceae (Caryophyllales) inferred from chloroplast *matK* and nuclear rDNA ITS sequences. *Am. J. Bot.* **93**, 399–411 (2006).
110. Harbaugh, D. T. *et al.* A new lineage-based tribal classification of the family Caryophyllaceae. *Int. J. Pl. Sci.* **171**, 185–198 (2010).
111. Willis, J. C. *A Dictionary of the Flowering Plants and Ferns*. 8th Ed. 1246 + 66. (Cambridge University Press, 1973).
112. Pax, F. & Hoffmann, K. Caryophyllaceae. In *Die Natürlichen Pflanzenfamilien Nebst Ihren Gattungen und Wichtigeren Arten, Insbesondere den Nutzpflanzen 16c*. 2nd Ed. (eds. Engler, A. & Harms, H.). 275–364 (Wilhelm Engelmann, 1934).
113. Niketić M. *Cerastium* L. In *Flora Srbije*. Vol. 2. 2nd Ed. (ed Stevanović, V.). 270–334 (Srpska Akademija Nauka i Umetnosti, 2012).
114. Brysting, A. K., Mathiesen, C. & Marcussen, T. Challenges in polyploid phylogenetic reconstruction: A case story from the Arctic-alpine *Cerastium alpinum* complex. *Taxon* **60**, 333–347 (2011).
115. Brysting, A. K. Chromosome number variation in the polyploid *Cerastium alpinum-C. arcticum* complex (Caryophyllaceae). *Nord. J. Bot.* **20**, 149–156 (2008).
116. Greenberg, A. K. & Donoghue, M. J. Molecular systematics and character evolution in Caryophyllaceae. *Taxon* **60**, 1637–1652 (2011).
117. Pfeil, B. E., Schlueter, J. A., Shoemaker, R. C. & Doyle, J. J. Placing paleopolyploidy in relation to taxon divergence: A phylogenetic analysis in legumes using 39 gene families. *Syst. Biol.* **54**, 441–454 (2005).
118. Xue, B. *et al.* Phylogenetic analysis and temporal diversification of the tribe Alsineae (Caryophyllaceae) with the description of three new genera, *Hesperostellaria*, *Reniostellaria* and *Torreyostellaria*. *Front. Plant Sci.* **14**, 1127443 (2023).
119. Frajman, B., Eggens, F. & Oxelman, B. Hybrid origins and homoploid reticulate evolution within *Heliosperma* (Sileneae, Caryophyllaceae)—A multigene phylogenetic approach with relative dating. *Syst. Biol.* **58**, 328–345 (2009).
120. Vít, P., Wolfová, K., Urfus, T., Tájek, P. & Suda, J. Interspecific hybridization between rare and common plant congeners inferred from genome size data: assessing the threat to the Czech serpentine endemic *Cerastium alsinifolium*. *Preslia* **86**, 95–117 (2014).
121. Nosil, P. *Ecological Speciation* (Oxford Series in Ecology and Evolution, 2012).
122. Tonti-Filippini, J., Nevill, P. G., Dixon, K. & Small, I. What can we do with 1000 plastid genomes?. *Plant J.* **90**, 808–818 (2017).
123. McKain, M. R., Johnson, M. G., Uribe-Convers, S., Eaton, D. & Yang, Y. Practical considerations for plant phylogenomics. *Appl. Plant Sci.* **6**, e1038 (2018).
124. Kearse, M. *et al.* Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).
125. Dong, X., Stothard, P., Forsythe, I. J. & Wishart, D. S. PlasMapper: A web server for drawing and auto-annotating plasmid maps. *Nucleic Acids Res.* **32**, W660–W664 (2004).
126. Lohse, M., Drechsel, O. & Bock, R. OrganellarGenomeDRAW (OGDRAW)—A tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr. Genet.* **52**, 267–274 (2007).
127. Jin, J. J. *et al.* GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* **21**, 241 (2020).
128. Kurtz, S. *et al.* REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* **29**, 4633–4642 (2001).
129. Mayer, C. *2006–2010. Phobos 3.3.11. A Tandem Repeat Search Program. 2006–2010.*
130. Sablok, G. *et al.* ChloroMitoSSRDB 2.00: More genomes, more repeats, unifying SSRs search patterns and on-the-fly repeat detection. *Database* **2015**, bav084 (2015).
131. Darling, A. C. E., Mau, B., Blattner, F. R. & Perna, N. T. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* **14**, 1394–1403 (2004).
132. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
133. Rozas, J. *et al.* DnaSP 6: DNA sequence polymorphism analysis of large datasets. *Mol. Biol. Evol.* **34**, 3299–3302 (2017).
134. Krzywinski, M. *et al.* Circos: An information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).
135. Lenz, H., Hein, A. & Knoop, V. Plant organelle RNA editing and its specificity factors: Enhancements of analyses and new database features in PREPACT 3.0. *BMC Bioinform.* **19**, 255 (2018).
136. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).
137. Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755 (2001).
138. Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574 (2003).

95. Takenaka, M., Zehrmann, A., Verbitskiy, D., Härtel, B. & Brennicke, A. RNA editing in plants and its evolution. *Annu. Rev. Genet.* **47**, 335–352 (2013).
96. Benne, R. *et al.* Major transcript of the frame shifted *coxII* gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. *Cell* **46**, 819–826 (1986).
97. Covello, P. S. & Gray, M. W. RNA editing in plant mitochondria. *Nature* **341**, 662–666 (1989).
98. Hoch, B., Maier, R. M., Appel, K., Igloi, G. L. & Kossel, H. Editing of a chloroplast mRNA by creation of an initiation codon. *Nature* **353**, 178–180 (1991).
99. Sasaki, T., Yukawa, Y., Miyamoto, T., Obokata, J. & Sugiura, M. Identification of RNA editing sites in chloroplast transcripts from the maternal and paternal progenitors of tobacco (*Nicotiana tabacum*): Comparative analysis shows the involvement of distinct trans-factors for *ndhB* editing. *Mol. Biol. Evol.* **20**, 1028–1035 (2003).
100. Corneille, S., Lutz, K. & Maliga, P. Conservation of RNA editing between rice and maize plastids: Are most editing events dispensable?. *Mol. Gen. Genet.* **264**, 419–424 (2000).
101. Miyamoto, T., Obokata, J. & Sugiura, M. Recognition of RNA editing sites is directed by unique proteins in chloroplasts: Biochemical identification of *cis*-acting elements and *trans*-acting factors involved in RNA editing in tobacco and pea chloroplasts. *Mol. Cell. Biol.* **22**, 6726–6734 (2002).
102. Daniell, H. *et al.* The complete nucleotide sequence of the cassava (*Manihot esculenta*) chloroplast genome and the evolution of *atpF* in Malpighiales: RNA editing and multiple losses of a group II intron. *Theor. Appl. Genet.* **116**, 723–737 (2008).
103. Germain, A., Hotto, A. M., Barkan, A. & Stern, D. B. RNA processing and decay in plastids. *Wiley Interdiscip. Rev. RNA* **4**, 295–316 (2013).
104. Ruchika, O. C., Sakari, M. & Tsukahara, T. Genome-wide identification of U-To-C RNA editing events for nuclear genes in *Arabidopsis thaliana*. *Cells* **10**, 635 (2021).
105. Xue, J., Liu, Y., Li, L., Wang, B. & Qiu, Y. The complete mitochondrial genome sequence of the hornwort *Phaeoceros laevis*: Retention of many ancient pseudogenes and conservative evolution of mitochondrial genomes in hornworts. *Curr. Genet.* **56**, 53–61 (2010).
106. Grewe, F. *et al.* A unique transcriptome: 1782 positions of RNA editing alter 1406 codon identities in mitochondrial mRNAs of the lycophyte *Isoetes engelmannii*. *Nucleic Acids Res.* **39**, 2890–2902 (2011).
107. Knie, N., Grewe, F. & Knoop, V. Monilophyte mitochondrial rps1 genes carry a unique group II intron that likely originated from an ancient paralog in rpl2. *RNA* **22**, 1338–1348 (2016).
108. Rabeler, R. K. & Bittrich, V. Suprageneric nomenclature in the Caryophyllaceae. *Taxon* **42**, 857–863 (1993).
109. Fior, S., Karis, P. O., Casazza, G., Minuto, L. & Sala, F. Molecular phylogeny of the Caryophyllaceae (Caryophyllales) inferred from chloroplast *matK* and nuclear rDNA ITS sequences. *Am. J. Bot.* **93**, 399–411 (2006).
110. Harbaugh, D. T. *et al.* A new lineage-based tribal classification of the family Caryophyllaceae. *Int. J. Pl. Sci.* **171**, 185–198 (2010).
111. Willis, J. C. *A Dictionary of the Flowering Plants and Ferns*. 8th Ed. 1246 + 66. (Cambridge University Press, 1973).
112. Pax, F. & Hoffmann, K. Caryophyllaceae. In *Die Natürlichen Pflanzenfamilien Nebst Ihren Gattungen und Wichtigeren Arten, Insbesondere den Nutzpflanzen 16c*. 2nd Ed. (eds. Engler, A. & Harms, H.). 275–364 (Wilhelm Engelmann, 1934).
113. Niketić M. *Cerastium* L. In *Flora Srbije*. Vol. 2. 2nd Ed. (ed Stevanović, V.). 270–334 (Srpska Akademija Nauka i Umetnosti, 2012).
114. Brysting, A. K., Mathiesen, C. & Marcussen, T. Challenges in polyploid phylogenetic reconstruction: A case story from the Arctic-alpine *Cerastium alpinum* complex. *Taxon* **60**, 333–347 (2011).
115. Brysting, A. K. Chromosome number variation in the polyploid *Cerastium alpinum-C. arcticum* complex (Caryophyllaceae). *Nord. J. Bot.* **20**, 149–156 (2008).
116. Greenberg, A. K. & Donoghue, M. J. Molecular systematics and character evolution in Caryophyllaceae. *Taxon* **60**, 1637–1652 (2011).
117. Pfeil, B. E., Schlueter, J. A., Shoemaker, R. C. & Doyle, J. J. Placing paleopolyploidy in relation to taxon divergence: A phylogenetic analysis in legumes using 39 gene families. *Syst. Biol.* **54**, 441–454 (2005).
118. Xue, B. *et al.* Phylogenetic analysis and temporal diversification of the tribe Alsineae (Caryophyllaceae) with the description of three new genera, *Hesperostellaria*, *Reniostellaria* and *Torreyostellaria*. *Front. Plant Sci.* **14**, 1127443 (2023).
119. Frajman, B., Eggens, F. & Oxelman, B. Hybrid origins and homoploid reticulate evolution within *Heliosperma* (Sileneae, Caryophyllaceae)—A multigene phylogenetic approach with relative dating. *Syst. Biol.* **58**, 328–345 (2009).
120. Vít, P., Wolfová, K., Urfus, T., Tájek, P. & Suda, J. Interspecific hybridization between rare and common plant congeners inferred from genome size data: assessing the threat to the Czech serpentine endemic *Cerastium alsinifolium*. *Preslia* **86**, 95–117 (2014).
121. Nosil, P. *Ecological Speciation* (Oxford Series in Ecology and Evolution, 2012).
122. Tonti-Filippini, J., Nevill, P. G., Dixon, K. & Small, I. What can we do with 1000 plastid genomes?. *Plant J.* **90**, 808–818 (2017).
123. McKain, M. R., Johnson, M. G., Uribe-Convers, S., Eaton, D. & Yang, Y. Practical considerations for plant phylogenomics. *Appl. Plant Sci.* **6**, e1038 (2018).
124. Kearse, M. *et al.* Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).
125. Dong, X., Stothard, P., Forsythe, I. J. & Wishart, D. S. PlasMapper: A web server for drawing and auto-annotating plasmid maps. *Nucleic Acids Res.* **32**, W660–W664 (2004).
126. Lohse, M., Drechsel, O. & Bock, R. OrganellarGenomeDRAW (OGDRAW)—A tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr. Genet.* **52**, 267–274 (2007).
127. Jin, J. J. *et al.* GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* **21**, 241 (2020).
128. Kurtz, S. *et al.* REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* **29**, 4633–4642 (2001).
129. Mayer, C. *2006–2010. Phobos 3.3.11. A Tandem Repeat Search Program. 2006–2010.*
130. Sablok, G. *et al.* ChloroMitoSSRDB 2.00: More genomes, more repeats, unifying SSRs search patterns and on-the-fly repeat detection. *Database* **2015**, bav084 (2015).
131. Darling, A. C. E., Mau, B., Blattner, F. R. & Perna, N. T. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* **14**, 1394–1403 (2004).
132. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
133. Rozas, J. *et al.* DnaSP 6: DNA sequence polymorphism analysis of large datasets. *Mol. Biol. Evol.* **34**, 3299–3302 (2017).
134. Krzywinski, M. *et al.* Circos: An information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).
135. Lenz, H., Hein, A. & Knoop, V. Plant organelle RNA editing and its specificity factors: Enhancements of analyses and new database features in PREPACT 3.0. *BMC Bioinform.* **19**, 255 (2018).
136. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).
137. Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755 (2001).
138. Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574 (2003).

139. Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. TimeTree: A resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* **34**, 1812–1819 (2017).

Acknowledgements
We thank Michał Węgrzyn from the Institute of Botany of Jagiellonian University in Kraków, Poland, for supporting us with the samples of *Cerastium arcticum*.

Author contributions
The study was conceptualized and designed by P.A. Chloroplast genome characteristics and comparative analyses were performed by S.E.M. and P.A. Ł.P. and J.P.J. performed NGS data analysis. M.M. assisted in data analysis and generating manuscript figures. S.E.M. and P.A. wrote the manuscript with input from K.L. and I.G. All authors read and edited the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-46017-y.

Correspondence and requests for materials should be addressed to P.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2023

10. OŚWIADCZENIA WSPÓŁAUTORÓW

Olsztyn, 20.03.2024
(miejscowość, data)

mgr inż. Sylwia Eryka Milarska
(imię i nazwisko kandydata)

Nauki biologiczne
(dyscyplina naukowa)

Przewodniczący Rady Naukowej Dyscypliny
Nauki biologiczne
Uniwersytetu Warmińsko-Mazurskiego w Olsztynie

OŚWIADCZENIE
autora rozprawy doktorskiej

Oświadczam, że w pracy pod tytułem: **Milarska SE, Androsiuk P, Bednarek PT, Larson K, Giełwanowska I. 2023. Genetic variation of *Cerastium alpinum* L. from Babia Góra, a critically endangered species in Poland. Journal of Applied Genetics 64: 37-53** mój wkład merytoryczny w jej przygotowanie polegał na: genotypowaniu całej kolekcji *C. alpinum*, analizie i interpretacji danych, wizualizacji danych, przygotowaniu tekstu manuskryptu i odpowiedzi na uwagi Recenzentów.

Milarska Sylwia
podpis kandydata

Olsztyn, 20.03.2024
(miejscowość, data)

mgr inż. Sylwia Eryka Milarska
(imię i nazwisko kandydata)

Nauki biologiczne
(dyscyplina naukowa)

Przewodniczący Rady Naukowej Dyscypliny
Nauki biologiczne
Uniwersytetu Warmińsko-Mazurskiego w Olsztynie

OŚWIADCZENIE
autora rozprawy doktorskiej

Oświadczam, że w pracy pod tytułem: **Milarska SE, Androsiuk P, Paukszto Ł, Jastrzębski JP, Maździarz M, Larson K, Giełwanowska I. 2023. Complete chloroplast genomes of *Cerastium alpinum*, *C. arcticum* and *C. nigrescens*: genome structures, comparative and phylogenetic analysis. Scientific Reports 13: 18774** mój wkład merytoryczny w jej przygotowanie polegał na: charakterystyce i analizach porównawczych genomów chloroplastowych, przygotowaniu tekstu manuskryptu i odpowiedzi na uwagi Recenzentów.

Milarska Sylwia
.....
podpis kandydata

Olsztyn, 20.03.2024
(miejscowość, data)

dr hab. Piotr Androsiuk
(tytuł zawodowy/stopień/tytuł naukowy, imię i nazwisko)

Przewodniczący Rady Naukowej Dyscypliny
Nauki biologiczne
Uniwersytetu Warmińsko-Mazurskiego w Olsztynie

OŚWIADCZENIE
współautora

Oświadczam, że w pracach pod tytułem: **(1) Milarska SE, Androsiuk P, Bednarek PT, Larson K, Giełwanowska I. 2023. Genetic variation of *Cerastium alpinum* L. from Babia Góra, a critically endangered species in Poland. Journal of Applied Genetics 64: 37-53** oraz **(2) Milarska SE, Androsiuk P, Paukszto Ł, Jastrzębski JP, Maździarz M, Larson K, Giełwanowska I. 2023. Complete chloroplast genomes of *Cerastium alpinum*, *C. arcticum* and *C. nigrescens*: genome structures, comparative and phylogenetic analysis. Scientific Reports 13: 18774** mój wkład merytoryczny w ich przygotowanie polegał na: (praca 1) opracowaniu koncepcji i metodyki badań, uczestniczeniu w analizie danych oraz ich wizualizacji i interpretacji, udziale w przygotowaniu tekstu manuskryptu, jego korekcie oraz odpowiedzi na uwagi Recenzentów; (praca 2) opracowaniu koncepcji i metodyki badań, uczestniczeniu w analizach mających na celu charakterystykę i analizę porównawczą genomów chloroplastowych, udziale w przygotowaniu tekstu manuskryptu, jego korekcie oraz odpowiedzi na uwagi Recenzentów.

Jednocześnie wyrażam zgodę na przedłożenie w/w prac przez **mgr inż. Sylwię E. Milarską** jako rozprawy doktorskiej w formie spójnego tematycznie zbioru artykułów naukowych opublikowanych w czasopismach naukowych.

Oświadczam, iż samodzielna i możliwa do wyodrębnienia część ww. pracy wykazuje indywidualny wkład kandydata **mgr inż. Sylwii E. Milarskiej** polegający na: (praca 1) genotypowanie całej kolekcji *C. alpinum*, analiza i interpretacja danych, wizualizacja danych, przygotowanie tekstu manuskryptu i odpowiedź na uwagi Recenzentów; (praca 2) charakterystyka i analizy porównawcze genomów chloroplastowych, przygotowanie tekstu manuskryptu oraz odpowiedź na uwagi Recenzentów.

.....
podpis

Radzików, 18.03.2024r.

prof. dr hab. Piotr T. Bednarek

Przewodniczący Rady Naukowej Dyscypliny
Nauki Biologiczne
Uniwersytetu Warmińsko-Mazurskiego w Olsztynie

OŚWIADCZENIE
współautora

Oświadczam, że w pracy pod tytułem: **Milarska SE, Androsiuk P, Bednarek PT, Larson K, Giełwanowska I. 2023. Genetic variation of *Cerastium alpinum* L. from Babia Góra, a critically endangered species in Poland. Journal of Applied Genetics 64: 37-53** mój wkład merytoryczny w jej przygotowanie polegał na: uczestniczeniu w analizie i interpretacji danych oraz konsultacjach w trakcie przygotowania i ostatecznej redakcji tekstu manuskryptu.

Jednocześnie wyrażam zgodę na przedłożenie w/w pracy przez **mgr inż. Sylwię E. Milarską** jako część rozprawy doktorskiej w formie spójnego tematycznie zbioru artykułów naukowych opublikowanych w czasopismach naukowych.

Oświadczam, iż samodzielna i możliwa do wyodrębnienia część ww. pracy wykazuje indywidualny wkład kandydata **mgr inż. Sylwii E. Milarskiej** polegający na genotypowaniu całej kolekcji *C. alpinum*, analizie i interpretacji danych, wizualizacji danych, przygotowaniu tekstu manuskryptu i odpowiedzi na uwagi Recenzentów.

Bednarek P.T.

podpis

Umeå, 14th March 2024

dr Keith W. Larson

Chairman of the Scientific Council
of the Discipline of Biological Sciences
University of Warmia and Mazury in Olsztyn

Co-author Statement

I hereby declare that I am the co-author of the following scientific papers: **(1) Milarska SE, Androsiuk P, Bednarek PT, Larson K, Giełwanowska I. 2023. Genetic variation of *Cerastium alpinum* L. from Babia Góra, a critically endangered species in Poland. Journal of Applied Genetics 64: 37-53** and **(2) Milarska SE, Androsiuk P, Paukszto Ł, Jastrzębski JP, Maździarz M, Larson K, Giełwanowska I. 2023. Complete chloroplast genomes of *Cerastium alpinum*, *C. arcticum* and *C. nigrescens*: genome structures, comparative and phylogenetic analysis. Scientific Reports 13: 18774** and my contribution to the works above (1 and 2) included providing the research material and consultations during the preparation and final editing of the manuscripts' text.

Simultaneously, I grant MSc Sylwia E. Milarska permission to submit the aforementioned works as components of her PhD dissertation in the form of a thematically coherent set of scientific articles published in scientific journals.

I declare that an independent and separable part of the above-mentioned works demonstrates the candidate's individual contribution, including (1) genotyping of the whole *C. alpinum* collection, data analysis and interpretation, data visualisation, writing the manuscript and response to the Reviewers' comments; (2) chloroplast genome characteristics and comparative analyses, writing the manuscript and response to the Reviewers' comments.

Signature page

This document has been electronically signed using eduSign.

eduSign

Electronically signed by
eduSign
Keith Larson

Date and time of signature
2024-03-14 05:52 UTC

Authenticated by
Umea University

Validate signed documents at: https://validator.edusign.sunet.se

SUNET

Węgorzewo, 18.03.2024r.

prof. dr hab. Irena Giełwanowska

Przewodniczący Rady Naukowej Dyscypliny
Nauki Biologiczne
Uniwersytetu Warmińsko-Mazurskiego w Olsztynie

OŚWIADCZENIE
współautora

Oświadczam, że w pracach pod tytułem: **(1) Milarska SE, Androsiuk P, Bednarek PT, Larson K, Giełwanowska I. 2023. Genetic variation of *Cerastium alpinum* L. from Babia Góra, a critically endangered species in Poland. Journal of Applied Genetics 64: 37-53**oraz **(2) Milarska SE, Androsiuk P, Paukszto Ł, Jastrzębski JP, Maździarz M, Larson K, Giełwanowska I. 2023. Complete chloroplast genomes of *Cerastium alpinum*, *C. arcticum* and *C. nigrescens*: genome structures, comparative and phylogenetic analysis. ScientificReports 13: 18774** mój wkład merytoryczny w ich przygotowanie polegał na: (praca 1) udziale opracowaniu koncepcji i metodyki badań, konsultacjach w trakcie przygotowania i ostatecznej redakcji tekstu manuskryptu; (praca 2) konsultacjach w trakcie przygotowania i ostatecznej redakcji tekstu manuskryptu.

Jednocześnie wyrażam zgodę na przedłożenie w/w prac przez **mgr inż. Sylwię E. Milarską** jako rozprawy doktorskiej w formie spójnego tematycznie zbioru artykułów naukowych opublikowanych w czasopismach naukowych.

Oświadczam, iż samodzielna i możliwa do wyodrębnienia część ww. prac wykazuje indywidualny wkład kandydata **mgr inż. Sylwii E. Milarskiej** polegający na: (praca 1) genotypowanie całej kolekcji *C. alpinum*, analiza i interpretacja danych, wizualizacja danych, przygotowanie tekstu manuskryptu i odpowiedź na uwagi Recenzentów; (praca 2) charakterystyka i analizy porównawcze genomów chloroplastowych, przygotowanie tekstu manuskryptu oraz odpowiedź na uwagi Recenzentów.

Irena Giełwanowska
podpis

Olsztyn, 18.03.2024
(miejscowość, data)

dr Łukasz Paukszto
(tytuł zawodowy/stopień/tytuł naukowy, imię i nazwisko)

Przewodniczący Rady Naukowej Dyscypliny
Nauki biologiczne
Uniwersytetu Warmińsko-Mazurskiego w Olsztynie

OŚWIADCZENIE
współautora

Oświadczam, że w pracy pod tytułem: **Milarska SE, Androsiuk P, Paukszto Ł, Jastrzębski JP, Maździarz M, Larson K, Giełwanowska I. 2023. Complete chloroplast genomes of *Cerastium alpinum*, *C. arcticum* and *C. nigrescens*: genome structures, comparative and phylogenetic analysis. Scientific Reports 13: 18774** mój wkład merytoryczny w jej przygotowanie polegał na udziale w wykonaniu analizy bioinformatycznej danych NGS oraz wizualizacji wyników.

Jednocześnie wyrażam zgodę na przedłożenie w/w pracy przez **mgr inż. Sylwię E. Milarską** jako część rozprawy doktorskiej w formie spójnego tematycznie zbioru artykułów naukowych opublikowanych w czasopismach naukowych.

Oświadczam, iż samodzielna i możliwa do wyodrębnienia część ww. pracy wykazuje indywidualny wkład kandydata **mgr inż. Sylwii E. Milarskiej** polegający na charakterystyce i analizie porównawczej genomów chloroplastowych, przygotowaniu tekstu manuskryptu oraz odpowiedzi na uwagi Recenzentów.

Łukasz Paukszto
.....
podpis

Olsztyn, 20.03.2024
(miejscowość, data)

dr Jan P. Jastrzębski
(tytuł zawodowy/stopień/tytuł naukowy, imię i nazwisko)

Przewodniczący Rady Naukowej Dyscypliny
Nauki biologiczne
Uniwersytetu Warmińsko-Mazurskiego w Olsztynie

OŚWIADCZENIE
współautora

Oświadczam, że w pracy pod tytułem: **Milarska SE, Androsiuk P, Paukszto Ł, Jastrzębski JP, Maździarz M, Larson K, Giełwanowska I. 2023. Complete chloroplast genomes of *Cerastium alpinum*, *C. arcticum* and *C. nigrescens*: genome structures, comparative and phylogenetic analysis. Scientific Reports 13: 18774** mój wkład merytoryczny w jej przygotowanie polegał na udziale w wykonaniu analizy bioinformatycznej danych NGS.

Jednocześnie wyrażam zgodę na przedłożenie w/w pracy przez **mgr inż. Sylwię E. Milarską** jako część rozprawy doktorskiej w formie spójnego tematycznie zbioru artykułów naukowych opublikowanych w czasopismach naukowych.

Oświadczam, iż samodzielna i możliwa do wyodrębnienia część ww. pracy wykazuje indywidualny wkład kandydata **mgr inż. Sylwii E. Milarskiej** polegający na charakterystyce i analizie porównawczej genomów chloroplastowych, przygotowaniu tekstu manuskryptu oraz odpowiedzi na uwagi Recenzentów.

.....
podpis

Olsztyn, 18.03.2024
(miejscowość, data)

mgr Mateusz Maździarz
(tytuł zawodowy/stopień/tytuł naukowy, imię i nazwisko)

Przewodniczący Rady Naukowej Dyscypliny
Nauki biologiczne
Uniwersytetu Warmińsko-Mazurskiego w Olsztynie

OŚWIADCZENIE
współautora

Oświadczam, że w pracy pod tytułem: **Milarska SE, Androsiuk P, Paukszto Ł, Jastrzębski JP, Maździarz M, Larson K, Giełwanowska I. 2023. Complete chloroplast genomes of *Cerastium alpinum*, *C. arcticum* and *C. nigrescens*: genome structures, comparative and phylogenetic analysis. Scientific Reports 13: 18774** mój wkład merytoryczny w jej przygotowanie polegał na udziale w wykonaniu analizy bioinformatycznej danych NGS oraz wizualizacji wyników.

Jednocześnie wyrażam zgodę na przedłożenie w/w pracy przez **mgr inż. Sylwię E. Milarską** jako część rozprawy doktorskiej w formie spójnego tematycznie zbioru artykułów naukowych opublikowanych w czasopismach naukowych.

Oświadczam, iż samodzielna i możliwa do wyodrębnienia część ww. pracy wykazuje indywidualny wkład kandydata **mgr inż. Sylwii E. Milarskiej** polegający na charakterystyce i analizie porównawczej genomów chloroplastowych, przygotowaniu tekstu manuskryptu oraz odpowiedzi na uwagi Recenzentów.

.....
podpis