

samtools faidx

fastp

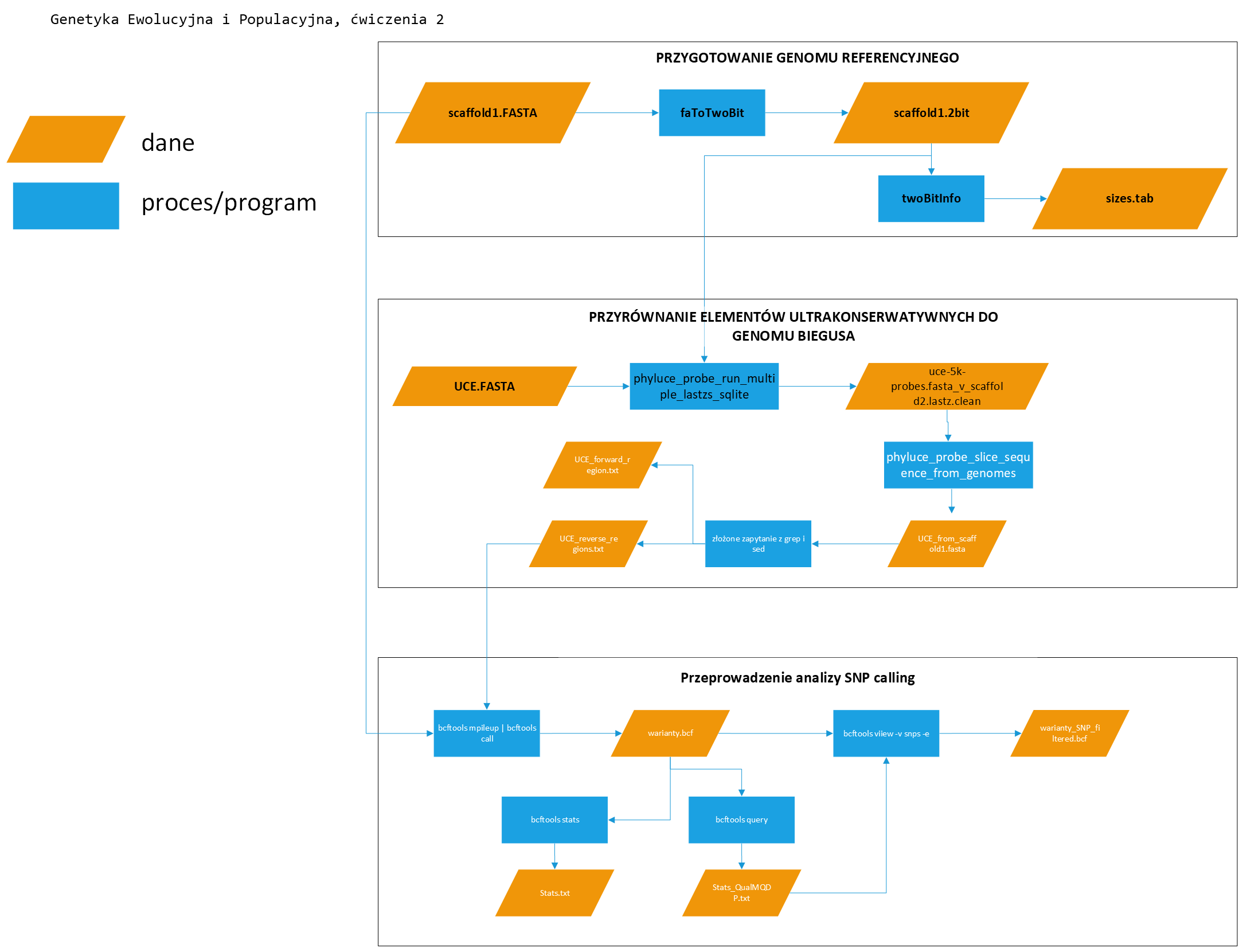
fastp

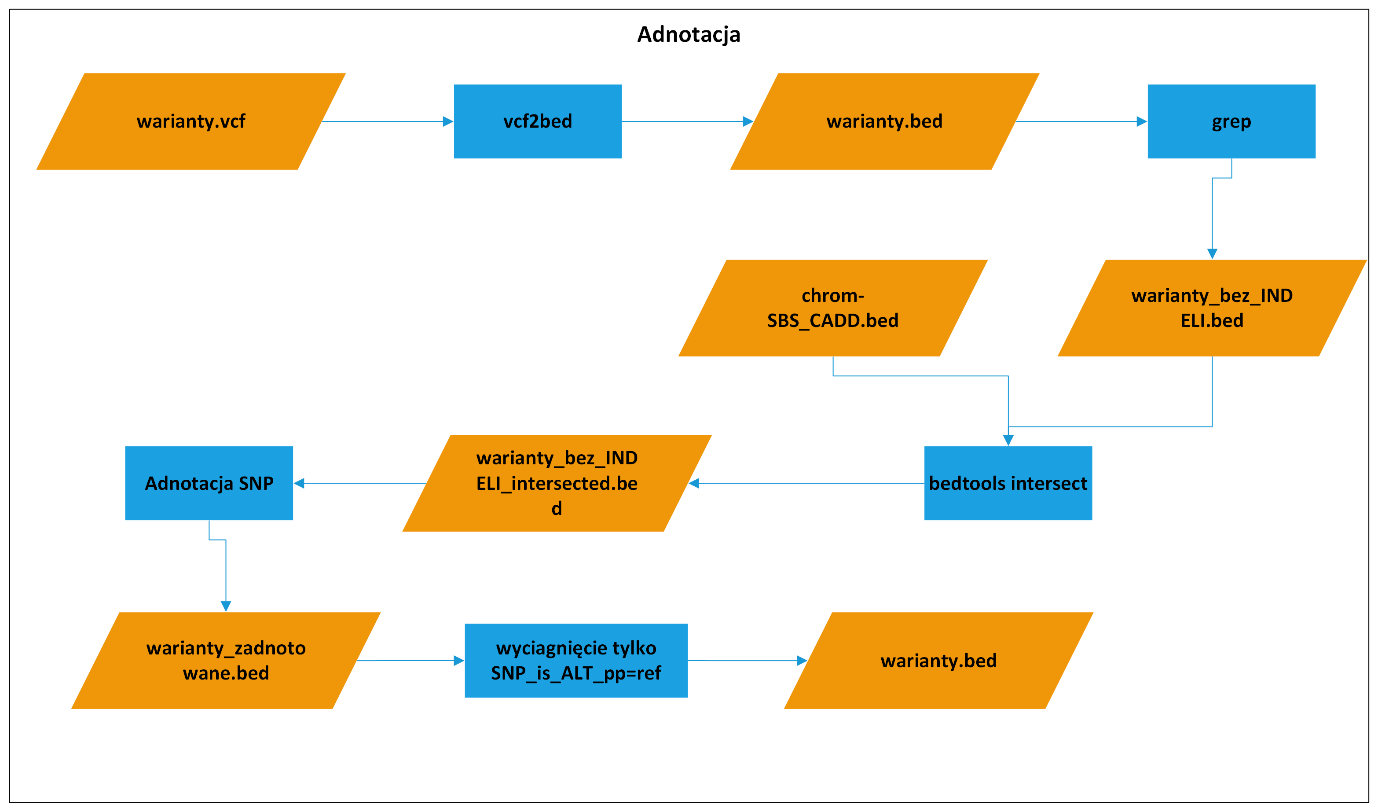
samtools sort

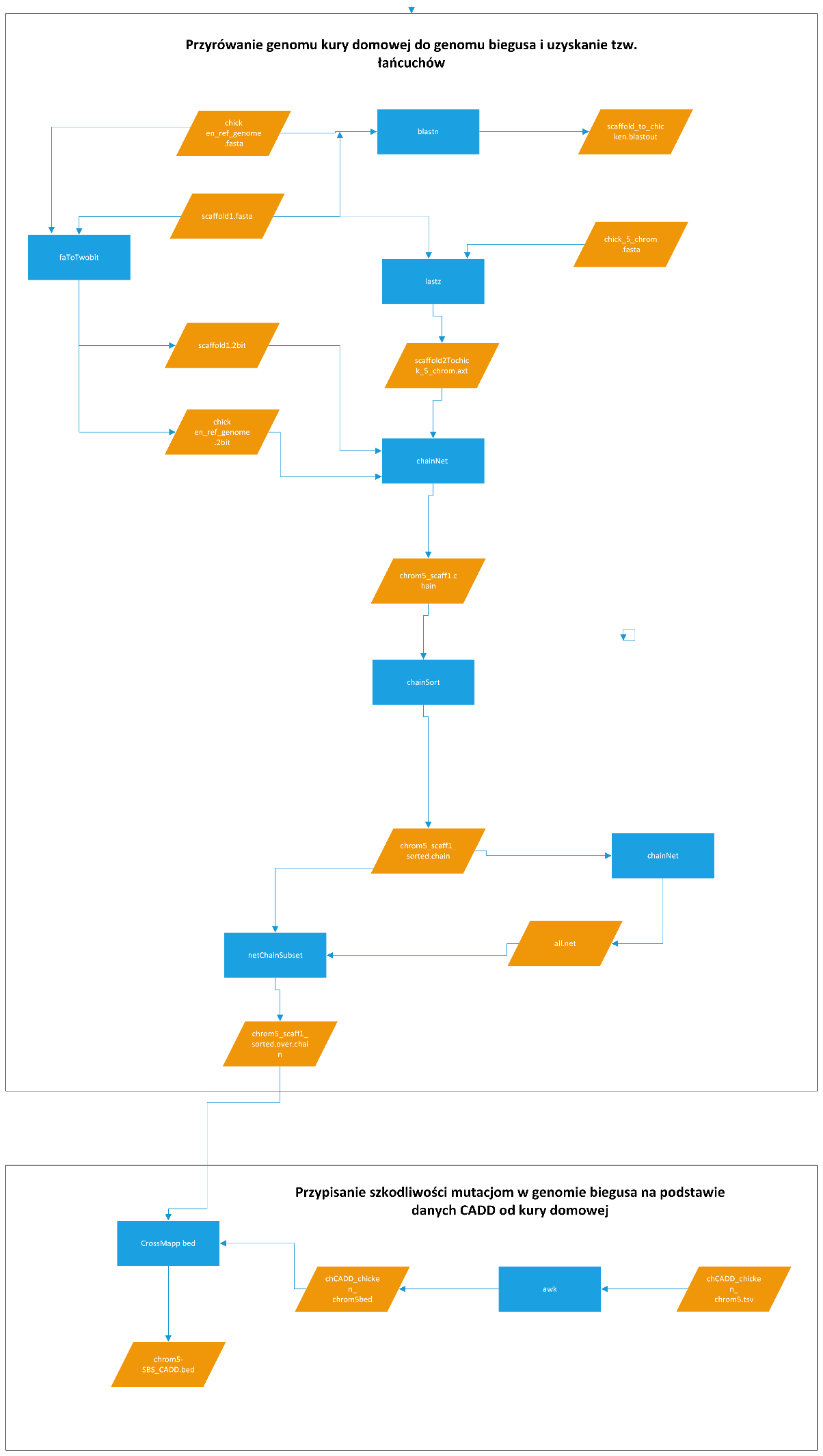
samtools sort

picard MarkDuplicates

samtools merge





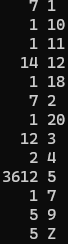


Analiza jakości odczytów sekwencjonowania:

Jakość odczytów była bardzo dobra, pomimo błędu przy sekwencjonowaniu - region na płytce z sekwencjami niskiej jakości.

|  |  |  |  |
| --- | --- | --- | --- |
| Osobnik | Filtrowanie | Run 1 **SRR7054135** | Run 2 **SRR7054162** |
| C\_ruf\_09 | Przed | Długość odczytów - 125 pz. Wyniki w większości bardzo dobre. Jedna sekwencja jest nad reprezentowana. **Blastowana daje różne dziwne wyniki… ale najpewniej to primer Illuminy** Na płytce znajduje się miejsce z sekwencjami o niskiej jakości. | Stała długość odczytów - 125 pz. Wyniki w większości bardzo dobre. Jedna sekwencja jest nad reprezentowana. **To samo** |
| Po | Odczyty po filtracji mają długość od 42 do 125 pz, głównie > 123 pz. Filtrowanie usunęło tylko małą część nadreprezentowanych sekwencji | Odczyty po filtracji mają długość od 42 do 125 pz, głównie > 123 pz. Filtrowanie usunęło tylko małą część nadreprezentowanych sekwencji |
| Osobnik | Filtrowanie | Run 1 **SRR7054133** | Run 2 **SRR7054147** |
| C\_pyg\_26 | Przed | Generalnie dobre wyniki. Doszło jednak do problemu na płytce. **Te nadreprezentowane sekwencje są spodziewane?**. Więcej odczytów zawiera GC na poziomie 42% niż przewidziano w teoretycznym rozkładzie. | Wyniki dobre, wykryto nadmiar nadreprezentowanych sekwencji. Więcej odczytów zawiera GC na poziomie 42% niż przewidziano w teoretycznym rozkładzie. |
| Po | Usunięto nadreprezentatywne sekwencje. Liczba odczytów z 42% zawartością GC nadal wysoka. | Nadal dużo odczytów z 42% zawartością GC nadal wysoka. Usunięto nadreprezentowane sekwencje |

Wyniki

Analizowany scaffold referencyjny nr 1 miał długość 10631746. W jego obrębie znaleziono wyłącznie sekwencje UCE w orientacji reverse. Scaffold 1 biegusa łyżkodziobego jest homologiczny do chromosomu 5 kury domowej. Wynik przyrównania scaffoldu do genomu kury domowej:  


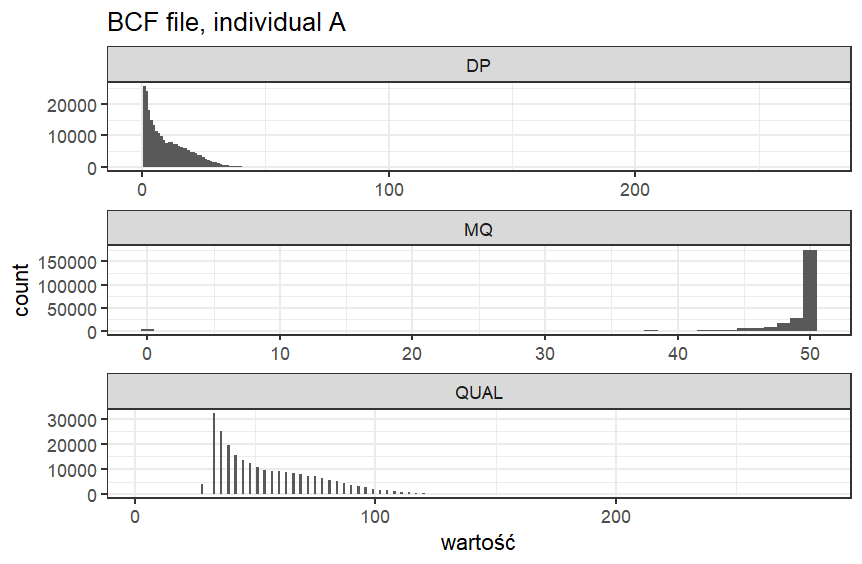
Porównując pliki BCF uzyskane po SNP callingu dla pojedynczych osobików i obu osobników pokazują, że oba osobniki współdzielą wiele SNP. Jednak osobnik biegusa rdzawoszyjego ma mniej SNP.

Obraz zawierający tekst, zrzut ekranu, Czcionka, linia

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.  
  
Obraz zawierający tekst, Czcionka, zrzut ekranu, linia

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

**Statystki plików bcf dla osobnika C\_pyg\_26:**  
Przed filtrowanie w pliku bcf było 244504 wariantów, w tym 324 SNP.  
Parametry filtrowania: QUAL < 20 || MQ < 40 || FORMAT/DP < 3 || FORMAT/ DP > 100  
Po filtrowaniu pozostały tylko 242 SNP.



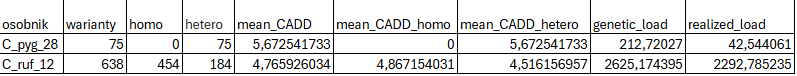
Statystki plików bcf dla osobnika C\_ruf\_09.   
Przed filtrowanie w pliku bcf było 240385 wariantów, w tym 1669 SNP   
Parametry filtrowania: QUAL < 25 || MQ < 30 || FORMAT/DP < 2 || FORMAT/ DP > 100  
Po filtrowaniu pozostało 1314 wariantów, w tym 1314 SNP.

Obraz zawierający tekst, zrzut ekranu, diagram, linia

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Po połączeniu sekwencji uzyskanych z przyrównania scaffoldu 1 i chromosomu 5 kury uzyskano 1666 łańcuchów. W wyniku filtrowania łańcuchów usunięto 5 łańcuchów.

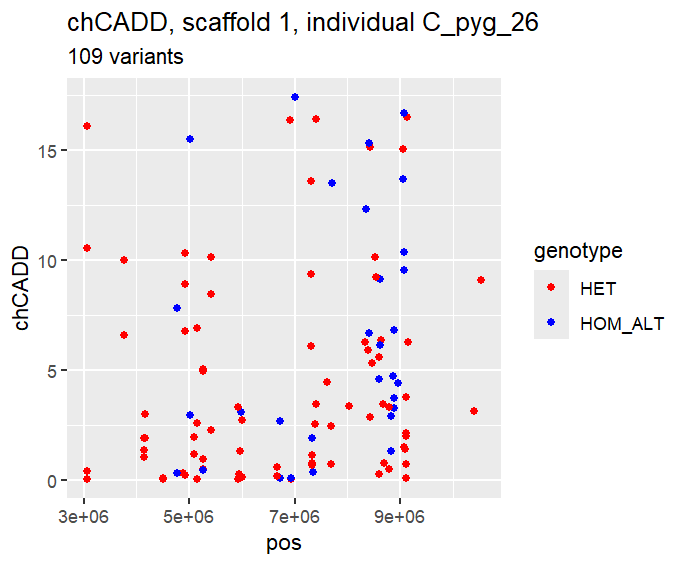
Analiza SNP u biegusa łyżkodziobego i biegusa rdzawoszyjego

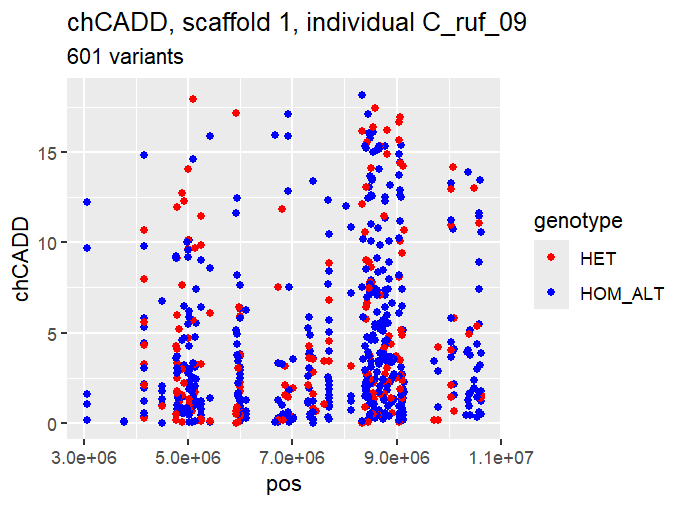


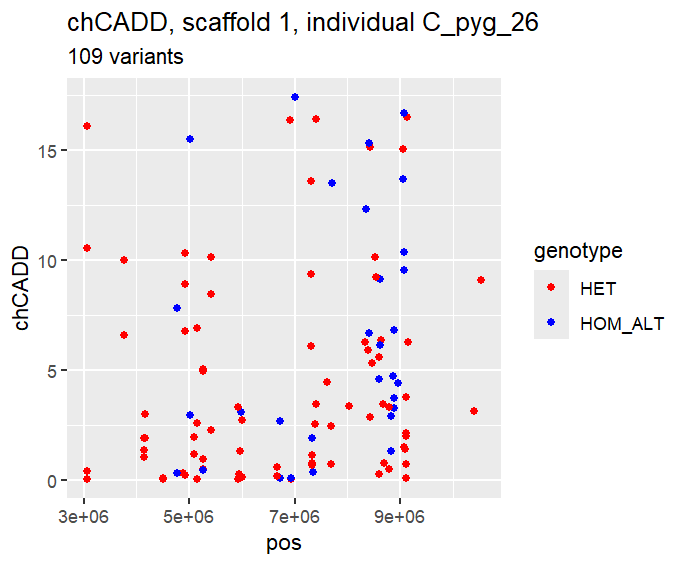
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Osobnik | VAR | HOM | HET | Mean\_CADD | Mean\_CADD\_HOM | Mean\_CADD\_HET | Genetic\_load | Realized load |
| C\_pyg\_26 | 109 | 30 | 79 |  |  |  |  |  |
| C\_ruf\_09 | 501 | 415 | 186 |  |  |  |  |  |

Analizowany biegus rdzawoszyi ma znacznie więcej homozygot w obrębie scafoldu 1, przez co może się ujawnić ich szkodliwy wpływ i tym samym obciążenie genetyczne jest bardzo duże. Natomiast u biegusa łyżkodziobego wykryto więcej heterozygot, dzięki czemu negatywny wpływ mutacji może pozostać ukryty, co skutkuje niską wartością obciążenia genetycznego.

Wartości CADD wzdłuż scaffoldu 1







Można zauważyć, że rozkład wartości CADD koreluje z pozycją w scafoldzie. Widać to u obu gatunków.

Dla obu osobników rozkład wartości CADD nie przyjął postaci rozkładu normalnego. Widać to też przy sprawdzaniu rozkładu wartości CADD dla heterozygot i homozygot biegusa rdzawoszyjego. Zostało to potwierdzone testem Shapiro. Dlatego do porównań poszczególnych grup wykorzystano test permutacyjny Manna-Whitneya.

|  |  |  |
| --- | --- | --- |
| Porównanie | Biegus łyżkodzioby vs biegus rdzawoszyjny | Biegus rdzawoszyjy homozygota vs heteroterozygota |
| p-value | 0.3286 | 0.2004 |

Uzyskane wyniki testu razem z wcześniejszymi danymi (o ilości wykrytych wariantów dla poszczególnych osobników - porównanie wielkości plików) pokazują, że nie ma statystycznie ważnej różnicy w średnich wartościach CADD dla heterozygot i homozygot biegusa rdzawoszyjego. Również nie widać statystycznie ważnej różnicy między średniki wartościami CADD między dwoma gatunkami biegusa. W tym przypadku może to jednak wynikać z małej liczby wykrytych SNP dla biegusa łyżkodziobego.

Wnioski:

Celem badania było oszacowanie obciążenia genetycznego u biegusa łyżkodziobego (mała populacja) i u biegusa rdzawoszyjego (duża populacja). Analiza pozwoliła stwierdzić, że biegus rdzawoszyi ma znacznie wyższe obciążenie genetyczne z powodu dużej ilości wykrytych homozygot. Należy jednak pamiętać, że analiza ograniczyła się tylko do scaffoldu 1. Nie wiadomo, jak obciążenie wygląda na całym genomie biegusa. Innym powodem takich wyników może być mała liczba wykrytych SNP, które przetrwały filtrowanie, dla biegusa łyżkodziobego.