Osnovna ideja svodi na sekvenciranje dijelova podataka u cjelinu, koja predstavlja genom. Cilj algoritma za sastavljanje genoma jest pretvoriti očitanja, dobivenih iz sekvenci, u izlazni niz tj. genom, koristeći međusobna preklapanja tih očitanja. Preklapanje se svodi na preklapanje sufiksa jednog niza sa prefiksom drugog niza, odnosno preklapanje očitanja. Moderni algoritmi za sastavljanje genoma temelje se na algoritmima nad grafovima. Jedna od danas poznatih metoda jest metoda Preklapanje-Razmještaj-Konsenzus, a to je upravo tema ovog rada. Zamislimo graf, njega čine vrhovi i bridovi koji povezuju vrhove. Bridovima se može proći samo u jednom smjeru, u takvom slučaju graf zovemo usmjereni graf. Svaki usmjereni brid predstavlja vezu između izvorišnog i odredišnog vrha. Skupina bridova tvori šetnje prilikom kojih prolazimo preko vrhova u nekom redoslijedu, a sve na način da je odredišni vrh jednog brida zapravo izvorišni vrh za slijedeći brid. Graf konstruiramo na način da vrhovima pridjelimo očitanja, a ako postoji preklapanje između tih očitanja, onda se to očitanje manifestira u bridu između ta dva vrha.

OLC pristupa se svodi na traženje puta u grafu koji prolazi kroz sve vrhove u grafu jedanput, te pritom prelazeći preko očitanja i preklapanja tvori konačni niz, tj genom. Osnovni problem traženja takvoga puta, tzv. Hamiltonovog puta je vrijeme rješavanja problema. Ovaj problem kategorizira se kao NP potpun problem, nije ga moguće obaviti u dostupnom vremenu. Štoviše, za NP potpune probleme dokazano je da se ne mogu rješiti pomoću računala danas. Iz tog se razloga koriste heuristike kojima će se graf što je moguće više pojednostaviti i na taj način doći do rješenja.

U ovom radu programski je implementirana jedna od tri faze OLC-a, faza razmještaja. Faza razmještaja ostvarena je nekolicinom metoda, objašnjavat će se u istom redoslijedu kao što se one izvode u programu.

1Pretvaranje preklapanja u oblik sa „lastinim repom“

Ovu radnju potrebno je obaviti jer graf zahtjeva zapise sačinjene samo od tzv. lastinih repova. Ovaj korak je preduvjet za nastavak obrade podataka i korištenje nadolazećih algoritama.

2Filtriranje očitanja

Moguća je situacija u kojoj se jedno očitanje u potpunosti preklapa sa drugim očitanjem, tj jedno očitanje sadrži u sebi i neko drugo očitanje. Pritom je moguće u potpunosti maknuti očitanje koje je sadržano u većem očitanju. Ovaj korak jako pridonosi uklanjanju velike količine podataka, a da se pritom ne izgube nikakve bitne informaicije.

3Prilagodavanje lastinih repova

Ovaj algoritam izravno se nastavlja na prethodni, i to na na način da nastoji skratiti ili produljiti dijelove preklapanja. Ovaj korak ispravlja forsirano jednake duljine lastinih repova koje je stvorio prethodni korak.

4Filtriranje tranzitivnih preklapanja

Zamislimo tri očitanja, i sva imaju preklapanje jedno s drugim. Kako su nam za definiciju preklapanja dovoljna dva očitanja, treće očitanje na istom području ne donosi nikakvu novu informaciju, stoga ga je sigurno maknuti. Informacija o preklapanju će ostati sačuvana.

5Filtriranje kratkih preklapanja

Ovaj jednostavni algoritam uklanja preklapanja koja sačinjavaju jako mali postotak ukupnog očitanja, odnosno za nekoliko redova veličine su kraća od samog očitanja.

6Filtriranje pogrešnih preklapanja

Ovaj proces svodi se na uklanjanje pogrešno stvorenih preklapanja. Prilikom stvaranja preklapanja zapisuje se vjerojatnost pogreške, i ovaj algoritam uklanja sva preklapanja sa vjerojatnosti pogreške većom od dozvoljenje.

7Kreiranje grafa

Nakon što su podaci pripremljeni prethodnim algoritmima, vrijeme je za implementaciju grafa. Svako očitanje čini jedan čvor, a svako preklapanje je prikazano dvama bridovima.

8Pojednostavnjenje grafa

Izgrađeni graf moguće sadrži neke oblike pogrešaka ili strukture koje nam ne odgovoraju. Tada je potrebno primjeniti niz metoda koje će graf dovesti u odgovarajuć oblik.

8.1Uklanjanje čvorova bez bridova

8.2Uklanjanje ogranaka

8.3Uklanjanje mjehurića

9Ekstrahiranje unitig-a

Graf je prethodnim koracima maksimalno pojednostavljen, a da sadrži kvalitetnu informaciju. Graf je moguće običi u jednoj šetnji. Slijedeći korak je, kao što podnaslov ovog procesa glasi, izlučivanje podnizova. Podniz je ograničen dio niza u kojem nema grananja, to su slijedno povezani čvorovi koji imaju samo jednog susjeda prije i jednog susjeda poslije sebe.

10Ekstrahiranje kontiga

Kontizi se sastoje od niza uzastopnih očitanja koja se preklapaju. To su zapravo rješenja cjelokupnog algoritma i pretpostavka za konačnu sliku genoma.