Uvod

Gen je osnovna molekularna jedinica nasljeđivanja svih živih organizama. U genima čuvaju se kodirane upute potrebne za proizvodnju proteina ili RNA lanaca koji imaju aktivnu funkciju u organizmu. Kod većine živih bića geni se nalaze pohranjeni u dvostrukoj DNA uzvojnici koja se sastoji od niza nukleotida.

Skup svih gena nekog organizma naziva se genom. Genom sadrži sveukupnu nasljednu informaciju nekog organizma, stoga je određivanje slijeda nukleinskih baza unutar njega od velike koristi. Određivanje tog slijeda pomaže nam u boljem razumijevanju svih bioloških interakcija organizma, nasljednih bolesti, tumora i drugih zloćudnih pojava i stanja u organizmu te njihovom liječenju i dijagnosticiranju. Proces određivanja slijeda pojedinih nukleinskih baza unutar DNA ili RNA lanca nazivamo sekvenciranje. Rezultati sekvenciranja koriste se za utvrđivanje samoga genoma, utvrđivanje srodnosti i evolucije vrsta, utvrđivanje novih gena i njihovo pridruživanje bolestima, utvrđivanje potencijalnih ciljeva za lijekove te metagenomiku (istraživanje zajednica mikroba direktno iz okoliša). Kako broj nukleotida koji se može sekvencirat varira ovisno o metodi sekvenciranja i njenoj izvedbi, prije sekvenciranja potrebno je podijeliti dulje nukleotidne lance u kraće. Najdominantnija strategija sekvenciranja je *shotgun* sekvenciranje kod kojeg se višestruka očitanja DNA na slučajan način lome na kraće dijelove koji se kasnije pomoću računalnih programa sastavljaju u kontinuiranu sekvencu.

Za sastavljanje fragmenata nazad u genom koriste se razne metode od kojih su najznačajnije: algoritmi poravnanja, mapiranje na referentni genom i *de novo* sastavljanje. Algoritmi poravnanja koriste dinamičko programiranje kako bi opisali razliku između dva niza, odnosno način na koji se jedan niz može dobiti iz drugog. Najjednostavniji način implementiranja algoritma poravnanja je korištenje matrice dimenzija *(n + 1) \* (m + 1),* pri čemu su *n* i *m* duljine nizova koje poravnavamo. Vrijednost svake ćelije dobivamo iz rezultata dobivenih zbrajanjem gornje ćelije s cijenom zamjene baza, zbrajanjem lijeve ćelije s cijenom umetanja baze te zbrajanjem gornje lijeve ćelije s cijenom usklađenosti, odnosno neusklađenosti baza nakon čega uzimamo najveću vrijednost od svih dobivenih rezultata kao vrijednost ćelije. Vrijednosti ćelija prvog stupca i prvog retka ovise o vrsti poravnanja (globalno, poluglobalno i lokalno). Praćenjem puta kojim smo dobili ciljnu ćeliju (koja također ovisi o vrsti poravnanja) dobivamo optimalno poravnanje dvaju niza. Mapiranjem na referentni genom dobivamo područja na referentnom genomu koje najbolje odgovaraju pojedinom fragmentu. Najpopularniji način određivanja regije s kojim se fragment najbolje podudara je usporedbom *minimizera,* leksikografski najmanjih podnizova (tzv. *k-merova*) od svih *w* podnizova duljine *k* unutar niza. U slučaju kad ne postoji referentni genom onda govorimo o *de novo* sastavljanju koje traži najbolja preklapanja između fragmenata te ih tako slaže nazad u originalan lanac.

Postoje brojni alati koji koriste navedene metode za sastavljanje genoma. Neki od tih alata su: *Minimap2*, *Ram* i *Raven.* Cilj ovog rada je napraviti aplikaciju s grafičkim sučeljem koja pokreće procese sastavljanja genoma pomoću navedenih alata uz mogućnost dodavanja dodatnih argumenata ovisno o alatu te jednostavnu analizu i statistiku rezultata tih procesa ukoliko su uspješno završeni.