

Rendiconto OLEA – I Anno

U.O. FEM-IASMA

Responsabile: ...Dr. Riccardo VELASCO.

WorkPackage 2 – Genomica strutturale

Obiettivi della ricerca

(Max 2.000 caratteri spazi inclusi)

Il progetto di sequenziamento del genoma dell'olivo prevede il raggiungimento di 15 genoma equivalenti con la tecnologia 454 FLX Titanium, con un sequenziamento misto di singole reads e di pair ends da librerie di 3, 8 e 20 kb. La tecnologia in oggetto prevede una copertura di massima di 2/3 con pair ends, nelle proporzioni più appropriate dalle librerie di frammenti di dimensioni crescenti, e 1/3 di single reads, inizialmente con la tecnologia disponibile al febbraio 2011 (400 bps di media), in un secondo momento con la tecnologia che si avvale di una nuova chimica che raggiunge oltre i 700 bps nelle librerie a single reads. Mentre le single reads garantiscono una maggiore copertura totale con una attesa facilitazione nell'assemblare o confermare regioni di relativamente piccole dimensioni, le pair reads dovrebbero garantire l'assemblaggio di contigs di discrete dimensioni costituendo i ponti tra pool di sequenze assemblate con la dovuta copertura in termine di genomi equivalenti.

Attività svolte

(Max 15.000 caratteri spazi inclusi)

L'attività di sequenziamento ha visto in primis la realizzazione di librerie sia per single reads che per sequenziamento in pair ends. L'estrazione del DNA genomico non ha costituito particolare difficoltà se realizzato con kit Qiagen su colonnine da 96 e relativa eluizione. La purezza ottenuta ha immediatamente consentito di procedere alle prime reazioni di sequenza. Sono stati anche ottenuti alcuni microgrammi di DNA isolato con gradiente di cloruro di cesio hanno dato origine a DNA notevolmente purificato tale da, ipoteticamente, poter superare eventuali difficoltà nell'ottenimento di long reads da librerie pair ends, ed aumentare la resa nel sequenziamento di frammenti di librerie di grandi dimensioni. Inizialmente sono state testate 3 librerie per la realizzazione delle single reads, prima dell'upgrade della macchina con la tecnologia che ha permesso una lettura fino a 700 basi per read. La realizzazione delle 3 librerie è stata inizialmente causata dalla anomala lunghezza e relativa bassa quantità di sequenze lette dal software, a seguito della lettura e pulizia dei dati grezzi di sequenza. Inizialmente, è stato un risultato relativamente insoddisfacente, in quanto le dimensioni dei frammenti sono stati mediamente del 30% più corti delle librerie dalle altre specie sequenziate come pero e lampone, (290 contro 450 basi), il che avrebbe comportato una quantità di corse necessarie per il completamento del progetto oltre il doppio di quanto preventivato. Si è proceduto quindi alla realizzazione di ulteriori due librerie che hanno purtroppo confermato le maggiori difficoltà ad ottenere lunghezze medie ormai standardizzate su altre specie (vite, melo, fragola e lampone) con una lunghezza media tuttavia crescente che ha raggiunto quasi i livelli delle abituali medie (390 contro 450). Si è proceduto quindi all'utilizzo per i sequenziamenti in single reads con la terza delle 3 librerie fino ad una copertura iniziale di 2.5 genomi equivalenti (2.5x). Alla fine di luglio si è ottenuto l'aggiornamento della macchina con la tecnologia FLX plus, che ha portato alla lettura media di 590 basi nel mese di settembre. In totale sono stati sequenziati oltre 9 Gb per una copertura totale complessivo di 10 genomi equivalenti. Si è proceduto poi alla realizzazione di due librerie per la lettura in pair-ends, con frammenti rispettivamente di 3, 8 e 20 kb. Per la prima dimensione sono state ottenute circa 1.5

Gb con 1 libreria, sono in corso di sequenziamento altre 3 librerie per un totale di altre 3-4 Gb (2-2.5x genoma equivalenti). La libreria di frammenti da 20 kb ha mostrato i maggiori problemi nella produzione di dati utilizzabili, si è proceduto tuttavia nella produzione delle librerie necessarie. Attualmente sono in corso di sequenziamento librerie da 3 kb e 8 kb, con l'attuale risultato di 1.5 Gb da frammenti di 3 kb e 1.2 Gb da frammenti di 8 kb. Sono previste ulteriori 5 librerie da 8 kb e 7-8 librerie da 20 kb per il completamento entro fine aprile del sequenziamento anche pair-ends.

Risultati ottenuti

(Max 6.000 caratteri spazi inclusi)

Complessivamente, sono stati ottenuti 15 Gb di reads a single reads per una copertura di 10 genoma equivalenti (10x) e sono previsti complessivamente altri 7-8 Gb di reads in pair-ends per un totale parziale di 5 genoma equivalenti (5x) ed un totale complessivo di 15x, come preventivato. Tali risultati sono stati ottenuti inizialmente con la strumentazione presente al momento dell'inizio del progetto, strumentazione e chimica 454/Roche che è stata aggiornata dalla ditta nel luglio del 2011, ben oltre la metà del primo anno del progetto, creando non poche difficoltà nella realizzazione delle coperture genoma equivalenti attese, con rese intorno alla metà del previsto. Dopo l'agosto 2011, l'implementazione della chimica FLX plus ha garantito, almeno nella single reads, letture di circa il doppio della lunghezza iniziale garantendo per ogni singola corsa dello strumento quasi 0.5 genoma equivalenti (circa 5-700 Mb per run) portando così la resa ad una quantità tale da consentire una buona copertura almeno vicina alle rese previste. Sulla base di questi risultati si è deciso di procedere alla produzione di single reads per circa 10 genoma equivalenti, mentre si sono ridotti a 5 genoma equivalenti le runs di pair-ends poiché la nuova chimica non è applicabile e le letture delle librerie pair-ends sono rimaste ad un massimo di 400 bp per singola read con una resa circa la metà dell'atteso.

Ricapitolando, sono stati raggiunti 15 Gbp ottenute da librerie shot gun, di cui 4 Gbp con reads lunghe 400 bp, 9 Gbp con reads lunghe 700 bp da DNA genomico estratto con tecnologia Qiagen e 2 Gbp con DNA estratto in gradiente di cloruro di cesio. Queste ultime non hanno dato una resa superiore pertanto si è rimasti alla metodologia Qiagen anche per le librerie da sequenziare in pair-ends.

Inoltre sono stati ottenuti 1,5 Gbp di reads lunghe 400 bp ottenute da 1 libreria a largo inserto da 3 kbp e 1,2 Gbp da una libreria con inserto lungo 8 kbp, sempre con reads da 400 bp. Sono in corso di sequenziamento il cui completamento è previsto per fine aprile 4 librerie da 3 kbp con 2 runs cadauna per un totale di 8 runs per complessivi 4 Gbp, 8 librerie da 8 kbp con 1 run cadauna per un totale di 8 runs e circa 4 Gbp. Sono invece da realizzare 7-8 librerie da 20 kbp che si sono dimostrate le più difficili da ottenere per la resa eccessivamente bassa in DNA totale.

Ostacoli incontrati e azioni correttive

(Max 6.000 caratteri spazi inclusi)

Il maggiore problema è stato il ritardo nella commercializzazione della nuova chimica e l'upgrade dello strumento da parte della ditta 454/Roche avvenuto in luglio e non all'inizio dell'anno 2011 come previsto. Ulteriori ritardi si sono avuti durante la messa a punto della nuova chimica che ha causato il mancato uso dello strumento per circa 3 mesi complessivi tra agosto 2011 e febbraio 2012, tuttavia la quasi totalità della copertura in genoma equivalenti è stata ottenuta nei tempi previsti.

Un secondo problema non risolto è relativo alla lunghezza delle reads nelle librerie da sequenziare in pari-ends, per le quali la nuova chimica non è stata aggiornata dalla ditta e la lettura è rimasta intorno alle 400 bp. Questo ha portato all'inversione della quantità totale delle reads che è passata da 10x in pair-ends e 5x in single reads a 10x in single reads e 5x in pair-ends, riducendo ad un terzo la lettura in pair-ends, che risulterà in una complicazione nell'assemblaggio finale che non possiamo risolvere. Una complicazione nella realizzazione delle librerie da 20 kbp è risultata per adesso insormontabile, pertanto si sono aumentate le quantità delle librerie da 3 e 8 kbp.

Pubblicazioni / Eventi divulgativi / Eventi formativi

-

Timbro Istituzione	Firma Responsabile Scientifico U.O.