



Application

BluePippinを用いたライブラリの長鎖 (>7kb) サイズ セレクションによるPacBioシーケンス結果の改善効果

製品名

自動DNA断片ゲル抽出システム BluePippin (BLU0001)

メーカー名

セージ サイエンス
Sage Science 社

このアプリケーションノートは、自然科学研究機構 基礎生物学研究所生物進化研究部門 柴田朋子様のご厚意により作成いたしました。

作製したSMRTBellライブラリについて、自動DNA断片ゲル抽出システムBluePippinを用いて7kb以上のサイズを回収(7kb以下のサイズを除去)し、PacBioにおけるシーケンス結果の改善効果を検証しました。

実験手順

● サンプルDNA

生物種 : InsectA
ゲノム抽出方法 : Qiagen Genomic-tip 100/G

● ゲノムDNA断片化条件

サンプルDNA量 : 10 μ g/150 μ l
g-Tube (Covaris) 5200rpm (1,814 \times g), 60sec
(遠心装置 : eppendorf MiniSpin plus使用)

1度遠心してサンプルが下に落ちなかったら 同じ遠心条件でサンプルが落ちるまで繰り返した。

その後、チューブをひっくり返して遠心し、これもサンプルが落ちるまで繰り返した。
ここまでの操作を1回として、これをもう1回繰り返した。

● BluePippinサイズセレクション条件

サンプル泳動量 : 30uL/lane (2.0 μ g/lane)
ゲルカセット : 0.75%ゲルカセット Marker S1 (BLF7510)
抽出条件 : high-passモード7kb-50kbで設定 (7kb以上を分画回収)

〈ワークフロー〉

InsectA からゲノム抽出



g-TubeによるゲノムDNA断片化



PacBio用ライブラリ調製



BluePippinを使って分画 >7kbの分子を回収



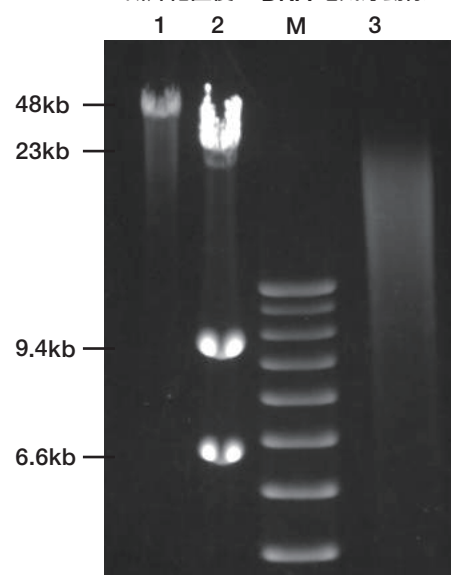
精製後、シーケンス



PacBio RSII

(Pacific Biosciences/トミーデジタルバイオロジー株式会社)

断片化直後の DNA 電気泳動像



泳動条件 : 0.5% agarose gel, 20V, 999min

- 1 : uncut lambda DNA
- 2 : lambda HindIII
- M : 1kb ladder
- 3 : 断片化後サンプル (138ng)

* 分画回収したサンプルは、以降のステップでのロスを考慮し、全量をシーケンス解析用に用いたため、サイズセレクション後のサイズ確認は実施しなかった。



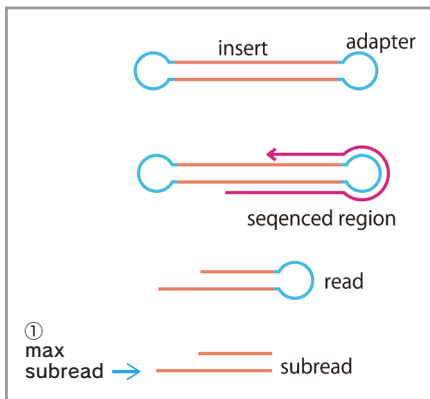
自動DNA断片ゲル抽出システム BluePippin

長鎖DNA断片のサイズセレクションに最適なパルスフィールド電気泳動が可能です。

結果

BluePippin による分画後ライブラリのシーケンス結果

JobID	BluePippin	Library_conc.	#bases	② Mean max subread length (bp)	#Reads	#Subreads	N50 subread length (bp)	③ Max max subread length (bp)	Mean Read Quality
1	使用	InsectA_bluepippin_100pM	208,015,129	5,857	35,513	38,398	8,445	24,012	0.833
2	使用	InsectA_bluepippin_150pM	213,735,908	5,414	39,478	42,146	7,808	24,135	0.826
3	使用	InsectA_bluepippin_200pM	239,420,335	5,454	43,895	47,485	7,488	22,071	0.813
4	不使用	InsectA_non-bluepippin_20pM	94,311,639	3,171	29,737	50,550	3,590	21,116	0.837



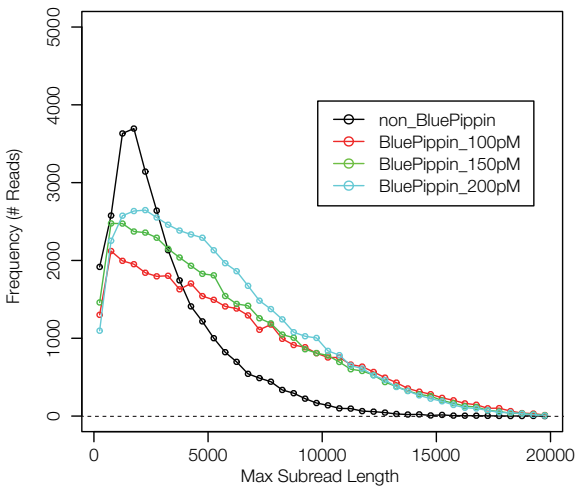
BluePippinによる分画を行ったサンプルについてローディング濃度100, 150, 200pMにそれぞれ調整し、1セルずつシーケンスした。BluePippin不使用サンプルはローディング濃度20pMで1セルシーケンスした。

BluePippin不使用の場合3.1kbだった平均最大サブリード長が、BluePippin使用時、5.4-5.8kbに増大した。また、N50サブリード長で大きく改善が見られたことから、小さい断片がBluePippinにより効果的に除去されたと推察された。更にサンプルのローディング濃度を高くするとリード数が増えたが、平均リードクオリティに低下が見られた。

用語説明：

- ① Max subread length : それぞれの分子のサブリードのうち、最長のものの長さ
- ② Mean max subread length : 1セル中の全てのリードのmax subread lengthの平均
- ③ Max max subread length : 1セル中の全てのリードのmax subread lengthの中で最長のもの

最大サブリード長の分布

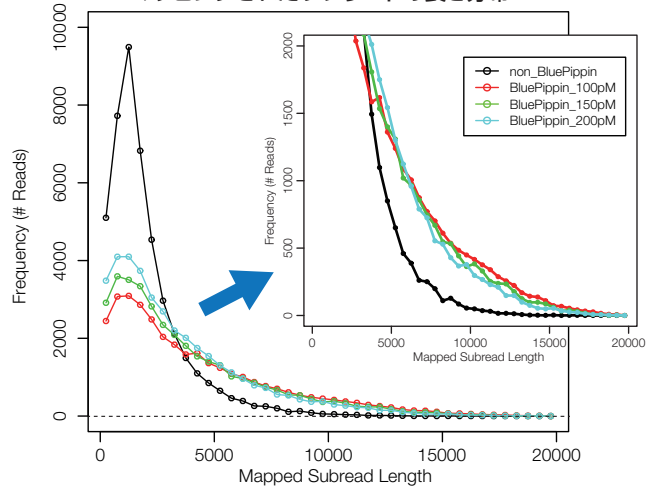


各リードの最大サブリードの長さの分布をグラフにした。BluePippin使用時、長さ>5kbのサブリード数は不使用時の3.1-3.6倍、>7kbのサブリード数4.7-5.0倍に増加した。

また、BluePippinを使用したライブラリの場合、ローディング濃度が高い方がリード数が多かった。

しかし、ローディング濃度が高いサンプルの場合、クオリティの低下が見られたため、有効なデータ量が最大になるかは不明である。そこで、参照ゲノムへのマッピング解析を行った。(右の図)

得られたサブリードのうち、参照配列にマッピングされたサブリードの長さ分布



参照配列にマッピングされたサブリードのアラインメント長分布をグラフにしたのが上図である。

全サブリードを、BLASR (PacBio用マッピングソフト) を用いて参照配列にマッピングした。参照配列は、サンガーシーケンスで作られたドラフトゲノムである。BluePippin使用時、アラインメント長>5kbのサブリード数は不使用時の3.2-3.6倍、>7kbのサブリード数4.7-6.3倍に増加した。

また、BluePippin使用サンプルと比較すると、参照配列にマップされたサブリード数はローディング濃度が低い場合の方が多く、リード数を比較した場合は逆の傾向が見られた。ローディング濃度が高い場合に見られたリードクオリティの低下が反映していると考えられる。

長い良好なデータを最大限得るために最適な濃度は、リード数が最大になる濃度より小さいことがわかった。最適な濃度を設定するには、サブリード長分布だけでなく、有効なデータの量にqualityなどが影響している点に注意が必要である。



お客様のコメント

PacBioは非常に長いシーケンスを得られますが、この長所を生かすにはインサートの長いライブラリを作成することが必要です。BluePippinは、短いライブラリを除去し、長いライブラリのみを回収するのに非常に有用であり、シーケンス結果も良好です。現在、私たちのプロトコルではBluePippinによるサイズセレクションは不可欠なステップになっています。