illumina

MiSeq[®]系统上的de novo细菌测序

高质量、末端配对的读取确保优异的de novo基因组装配

简介

MiSeq系统利用Illumina久经考验的TruSeq[™]边合成边测序 (SBS)可逆终止碱基试剂,与HiSeq[™]2000运行的数据质量相媲 美。本应用指南描述了在HiSeq2000和MiSeq平台上对从大肠杆 菌参考菌株制备而来的同一个文库进行测序,以及*de novo*装配。 将MiSeq仪器上的细菌基因组测序结果与近期Ion Torrent个人化 操作基因组测序仪(PGM)的大肠杆菌数据组进行了比较,显示 出高质量、末端配对读取对*de novo*基因组装配的重要性。

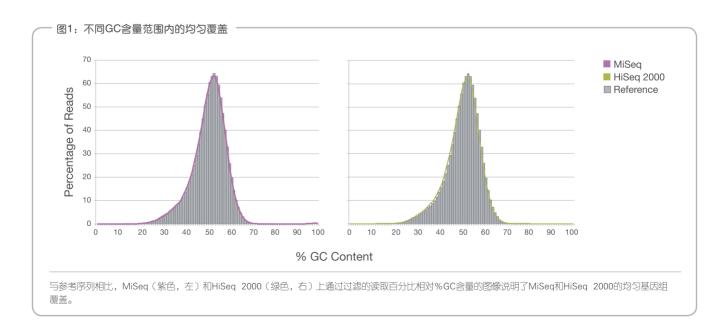
测序和de novo装配流程

从研究透彻的大肠杆菌MG1655菌株中分离出基因组DNA,并利用IIIumina的TruSeq样品制备试剂制备测序文库。对于MiSeq仪器上的测序,样品置于试剂舱中,并与流动槽一起上样至仪器内。后续的所有步骤在仪器上开展,包括簇生成和2×150末端配对测序,时间不到27小时。MG1655文库还用于簇生成,在HiSeq流动槽中利用cBot™和TruSeq v3 clustering试剂,接着在HiSeq 2000上开展2×100 bp的末端配对测序,10.5天完成。

对于重测序,直接在MiSeq内置的计算机上开展数据分析,无需特殊的服务器或计算设备。利用CASAVA 1.8a5分析HiSeq和MiSeq的两组碱基检出文件,并利用Velvet完成*de novo*装配。对于MiSeq和Ion Torrent之间的*de novo*装配比较,对MiSeq 50×

- 表1.大肠杆菌MG1655文库的MiSeq和HiSeq 2000 运行和装配指标

| 指标 | MiSeq | HiSeq 2000 |
|------------------|----------------------|----------------------|
| 原始簇 | 840/mm² | 794/mm ² |
| 通过过滤的簇 | 799/mm² | 726/mm ² |
| 高于Q30的碱基百分比 | 读取1:91.9 读取2:87.5 | 读取1:89.3 读取2:86.1 |
| 高于030的总碱基百分比 | 89.7 | 87.7 |
| 装配大小 | 4,573,128 | 4,571,539 |
| Contigs数量 | 125 | 125 |
| N50(bp(Contigs)) | 148,627 (11) | 148,770 (12) |
| 平均Contig大小(bp) | 36,585 | 36,571 |
| Scaffolds中缺口百分比 | 0.07 | 0.07 |
| 最大Contig(bp) | 311,761 | 305,482 |
| GC百分比 | 50.71 | 50.71 |





抽样数据使用开放获取的装配软件MIRA¹和Ray²,并与Ion Torrent读取的整个数据组进行比较³。据报告,这些开源的装配工 具能很好地用于Illumina和Ion Torrent的数据³,并产生了与处理 MiSeq数据的Velvet相当的结果。

结果与数据分析

MiSeq和HiSeq系统产生的数据表现出相似的簇密度以及通过过 滤的簇数量。HiSeq和MiSeq读取的*de novo*装配指标非常相似 (表1)。HiSeq和MiSeq数据与参考序列的比较说明了GC含量 范围内的相同覆盖(图1)。MiSeq装配与大肠杆菌参考序列重 叠的数据是以Circos曲线显示的(图2),说明了整个基因组范 围的极佳覆盖。

将2×150 bp MiSeq运行的*de novo*装配数据与Ion Torrent的数据³进行了比较。为了平等的比较,MiSeq的50×覆盖度抽样数据,包含231 Mb,约为数据的1/7。与整个Ion Torrent的数据组相比,抽样的MiSeq数据在最大contig长度和N50值上都明显更佳(图3A和3B)。

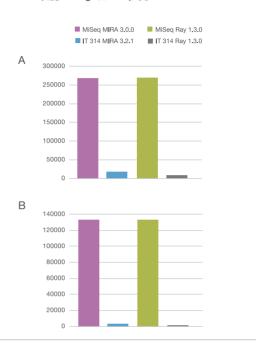
Illumina China

上海办公室

北京办公室

地址:北京市朝阳区曙光西里甲5号凤凰置地广场H座写字楼1103A室,100028 电话:+86-10-84554866 传真:+86-10-84554855 2012年5月22日印刷

83. de novo装配Contigs和N50长度



A. 利用MIRA v3.0.0和Ray v1.3.0 (绿色)的MiSeq装配(紫色)与利用 MIRA v3.2和Ray v1.3.0 (灰色)的Ion Torrent装配(蓝色)的最大contig 长度比较(单位为bp)。B. N50是数据组中最小contig的长度,它包含了 最少(最大)的contig,它们的累加长度达到至少50%的装配。

结论

利用从细菌DNA制备而来的同一文库,MiSeq上的测序与HiSeq 非常相近;两个平台都产生了高质量的数据,其中85%以上的 碱基高于Q30,且有着均匀的GC覆盖。这些数据的de novo装配 也产生了相似的结果,极好地覆盖了参考序列。MiSeq系统上产 生的测序结果很好地预测了高通量HiSeq 2000测序平台所带来 的结果,让MiSeq成为试行大规模研究或开展需要速度和准确性 的独立实验的理想选择。与Ion Torrent相比,高质量、配对末端 MiSeq读取对de novo装配的重要性是显而易见的。MiSeq末端 配对读取所产生的高质量装配表明,更好的数据带来基因组的更 准确图像。

参考资料

- 1. http://www.chevreux.org/projects_mira.html
- 2. http://sourceforge.net/projects/denovoassembler/files
- 3. http://pathogenomics.bham.ac.uk/blog/2011/05/first-look-at-ion-torrentdata-de-novo-assembly

Illumina, Inc. • 9885 Towne Centre Drive, San Diego, CA 92121 USA • 1.800.809.4566 toll-free • 1.858.202.4566 tel • techsupport@illumina.com • illumina.com

FOR RESEARCH USE ONLY

© 2011 Illumina, Inc. All rights reserved. Illumina, illuminaDx, BeadXrray, BeadXpress, cBot, CSPro, DASL, Eco, GAllx, Genetic Energy, Genome Analyzer, GenomeStudio, GoldenGate, HiScan, HiSeq, Infinium, iSelect, MiSeq, Nextera, Sentrix, Solexa, TruSeq, VeraCode, the pumpkin crange color, and the Genetic Energy streaming bases design are trademarks or registered trademarks of Illumina, Inc. All other brands and names contained herein are the property of their respective owners. Pub. No. 770-2011-009 Current as of 8 August 2011

