随机抽取1million基因组中长度L=1kb的随机间隔，计算所有间隔之间相似度，间隔长度L1=（1+a）L=（1+0.15）\*1000=1150bp，假定插入删除比率为0.15，

在锚点15,20,25的的这些长度中，找到间隔中>=s-similar的数量，绘制直方图，

从这个图中，可以解释在使用锚点映射read时必须搜索的间隔数。

从四个图中，能得到间隔长度，所选的锚点长度和匹配上的锚点数量s-similar，了解到97.5%的样本需要至少20个锚点相似，才能将候选位置限定在100及以上很小范围。

我们比较人类基因组中锚点相似性值的分布和匹配锚点期望分布

We compared the distribution of values of anchor similarity from the human genome with values ofNumConfigurations(M,N,K,L)/(LM) to see how the mapability of sequences compares to the expected distributions of matching anchors.

有些基因组间隔中的reads 与基因组其他间隔具有很低锚相似度，因此假的匹配簇就少，得到特异匹配，然而有的间隔reads与其他间隔之间具有高相似度，那就会有很多匹配簇。

表四表明对于任何reads至少有8个长度为20或更大的锚

The green points in C show the number of matching intervals when using a similar set of parameters: at least 10 anchors of length 20. Importantly, 95% of the samples match uniquely in the genome.

然后模拟多个物种基因组，发现95%的长为1000碱基的reads能够正确匹配到基因组中

For example, 95% of 1000-base reads from the human genome simulated with a 15% error rate map to the correct location in the genome.

在图四中，K=15时，选择>=1-similar的相似度时，只有22%的reads能将候选间隔位置限定在100左右，但如果相似度选择为10时，即两个间隔之间锚点至少十个相似时，那么在K=15时，就能让超过90%的reads确定特定的匹配

As shown in Figure[4](https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-13-238#Fig4)B, a read with a 15% error rate has a 97% chance of having 10 anchors of length 15 or more. The anchor similarity corresponding to these reads uses parameters δ = 0. 15,L = 1000, and k = 15, and is shown by the red curve in Figure[5](https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-13-238#Fig5)A. Over 90% of the sampled intervals only have one location with at least 10 anchors of length 15, indicating they map uniquely under this repeat under this repeat metric. The other two genomes, E. coli, and A. thaliana, are shown for guidance