**Aula – Montagem de genomas**

1. Abra um terminal em sua máquina Linux e crie uma pasta mkdir aula\_montagem e entre nela cd aula\_montagem. No Windows, crie uma pasta!

2. Agora abra uma conexão com a pinguim [ssh bioufmg@143.107.223.182](mailto:ssh@143.107.223.182)

3. Entre em sua pasta na pinguim cd eusoujacu crie um diretório mkdir aula\_montagem e entre nela cd aula\_montagem

4. Copie o material para essa pasta cp /home/treinamento/velvet\_aula/\* . (olha o ponto!)

5. Dá um ls e veja que você tem três arquivos, dois arquivos de sequenciamento (*reads\_forward.fastq* e *reads\_reverse.fastq*) e um arquivo com o genoma de referência (*genoma.fasta*). Esse sequenciamento foi do tipo “*paired end*” (dois lados de um fragmento).

6. Primeiro vamos olhar a qualidade dos reads com o programa **fastQC**, ele vai gerar várias saídas. Para ver as saídas temos uma novidade. Na biodados (<http://biodados.icb.ufmg.br> [you@pinguim](http://143.107.223.182/)) há um link para public\_html, que mostra na web o conteúdo de todas as pastas em bioufmg. Assim, toda figura gerada lá pode ser vista com o firefox!

7. Para observar a qualidade das bases nos arquivos *reads\_forward.fastq* e *reads\_reverse.fastq*, crie o diretório sem\_trim e execute o script fastqc (com –t especifique rodar com um único core)

mkdir sem\_trim (o programa exige que o diretório de output seja criado antes...) e depois:

fastqc -o sem\_trim -t 1 reads\_forward.fastq reads\_reverse.fastq

8. Veja a saída abrindo a public\_html, abrindo sua pasta e seguindo o caminho: aula\_montagem/sem\_trim/reads\_forward\_fastqc/fastqc\_report.html para informações do reads forward

e

aula\_montagem/sem\_trim/reads\_reverse\_fastqc/fastqc\_report.html para informações do reads reverse

9. Vamos filtrar esses reads com Trimmomatic assim:

/usr/local/bin/trimmomatic PE -phred33 reads\_forward.fastq reads\_reverse.fastq trimmed\_forward\_paired.fastq trimmed\_forward\_unpaired.fastq trimmed\_reverse\_paired.fastq trimmed\_reverse\_unpaired.fastq LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36

Os parâmetros para trimming são:

PE: indica *paired end*

LEADING: Remove bases com baixa qualidade no início da sequencia (qualidade abaixo de 3)

TRAILING: Remove bases com baixa qualidade no final da sequencia (qualidade abaixo de 3)

SLIDINGWINDOW: Utiliza uma janela deslizando com 4 bases e corta quando a qualidade média é menor que 15

MINLEN: Remove sequencias menores que 36 bases

10. Agora refaça o processo do fastQC com os reads filtrados

mkdir com\_trim

fastqc -o com\_trim -t 1 trimmed\_forward\_paired.fastq trimmed\_reverse\_paired.fastq

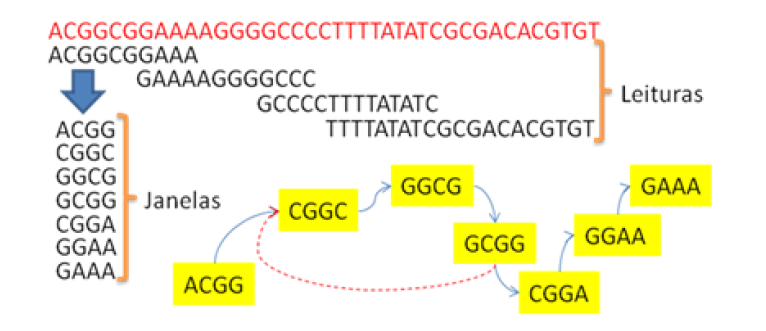
Veja o report do programa no firefox, em public\_html abra sua pasta e siga o caminho:

aula\_montagem/sem\_trim/trimmed\_forward\_paired\_fastqc/fastqc\_report.html

aula\_montagem/sem\_trim/trimmed\_reverse\_paired\_fastqc/fastqc\_report.html

Melhorou?

11. Com os reads já trimados, fazer a montagem sem referencia (do zero) com o programa **Velvet**



12. Rode o velvet assim:

Crie os índices (o parâmetro numérico refere-se ao tamanho do *k-mer* utilizado, sempre números ímpares)

velveth montagem 31 -fastq -shortPaired -separate trimmed\_forward\_paired.fastq trimmed\_reverse\_paired.fastq

Execute a montagem

velvetg montagem/ -exp\_cov auto -scaffolding yes

Após a execução, o Velvet apresenta algumas informações da montagem. Os nodes são o número de contigs gerados. O n50 diz que metade do genoma está representada em contigs maiores que esse valor.

**Screenshot - 4_9_2014 , 5_00_56 AM.png**

13. Verifique o arquivo de montagem

cd montagem

more contigs.fa

14. Agora vamos fazer o outro tipo de montagem que é "com genoma de referencia" usando o programa Bowtie para ver os reads alinhando. Primeiro volte ao diretório aula\_montagem cd ..

Rode o bowtie assim:

Crie um bowtie index usando o genoma de ancoragem (cts.fasta) fornecido:

bowtie2-build genoma.fasta genoma.build

Alinhe os reads no index do genoma de referência (genoma.build), criando um arquivo de alinhamentos reads\_mapeados.sam no formato chamado SAM (-S manda criar o SAM)

bowtie2 -x genoma.build -1 trimmed\_forward\_paired.fastq -2 trimmed\_reverse\_paired.fastq -S reads\_mapeados.sam

15. Para ver o resultado teremos que abrir um programa Java na sua máquina Linux (ou na windows) chamado Tablet. Você pode rodar ele da web usando este [link](http://bioinf.hutton.ac.uk/tablet/webstart/tablet.jnlp). Ele vai abrir um ambiente gráfico. Se não funcionar na sua máquina pessoal, você pode fazer o download do programa [aqui](http://ics.hutton.ac.uk/tablet/download-tablet/).

16. Temos que pegar o arquivo de saída do Bowtie (*reads\_mapeados.sam*) e o genoma de referência (*genoma.fasta*) e trazer para **a sua máquina**. No Linux da aula, o melhor a fazer é achar o resultado usando o [you@pinguim](mailto:you@pinguim" \t "_blank) (link na biodados) e "copiar link", e depois, no terminal da sua máquina, dar um wget, assim:

wget http://143.107.223.182/public\_html/eusoujacu/aula\_montagem/genoma.fasta

wget http://143.107.223.182/public\_html/eusoujacu/aula\_montagem/reads\_mapeados.sam

17. Dá um ls pra ver o arquivo. E quem está no Windows? Nesse caso você já sabe, é só baixar tudo como de costume e abrir pelo modo gráfico no Tablet.

18. Aberto o genoma de referencia e o arquivo de saída do Bowtie, você vai ver um monte de reads ancorados na referência. Explore ao longo do genoma a cobertura que vc tem dessa forma:

Clique em: "Open Assembly"

Na primeira janela selecione o arquivo: “reads\_mapeados.sam”

Na segunda janela selecione o genoma de referência: “genoma.fasta”

Clique em open.

19. Não é aula de transcriptômica... então vamos aprender a exportar os resultados do Bowtie usando o SAMTools. De volta à pinguim:

Converta o arquivo SAM para a versão binária BAM

samtools view -S reads\_mapeados.sam -b -o reads\_mapeados.bam

Ordene os reads

samtools sort reads\_mapeados.bam reads\_mapeados\_ordenados

Crie um índice para o BAM

samtools index reads\_mapeados\_ordenados.bam

Crie os *contigs* que terão o consenso do mapeamento que vc viu com Bowtie

samtools mpileup -uf genoma.fasta reads\_mapeados\_ordenados.bam | bcftools view -cg - | vcfutils.pl vcf2fq > montagem\_por\_referencia.fastq

Visualize a montagem

more montagem\_por\_referencia.fastq

A aula de montagem traz muitas sintaxes típicas dos programas, todavia permanece a mesma rotina de trabalho de controle de programas em servidores. Vc executou dois protocolos de montagem com reads de sequenciamento Illumina *paired-end*. E pode agora anotar os contigs com Artemis.