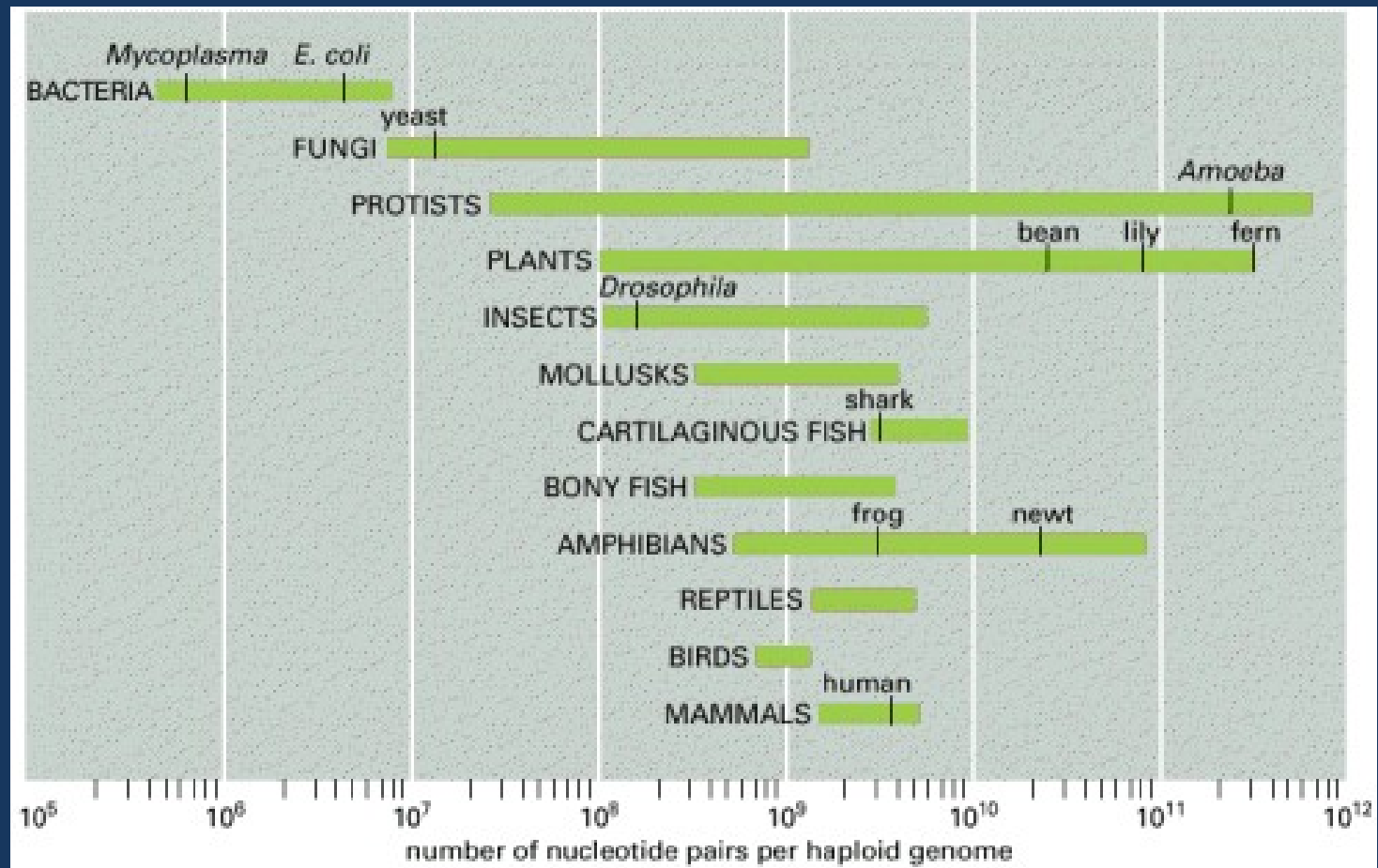


Sequenciamento de genoma e transcriptomas

Por que seqüenciar genomas?

- O seqüenciamento de genomas é o primeiro passo para obter uma descrição completa da composição molecular de cada organismo, pois todas informações necessária para construção estão presentes no DNA genômico (Entretanto a interpretação estas informações ainda é um problema)
- Comparação de genomas de diversos indivíduos permitira correlacionar características e síndromes com mutações de determinados *locus* do genoma (mesmo que não tenhamos idéia da função deste *locus*)
- Comparações entre genomas de espécies próximas permite o melhor entendimento dos mecanismos de evolução de genomas
- Um melhor conhecimento do genoma permite com que manipulemos este com maior facilidade

Tamanho de genomas



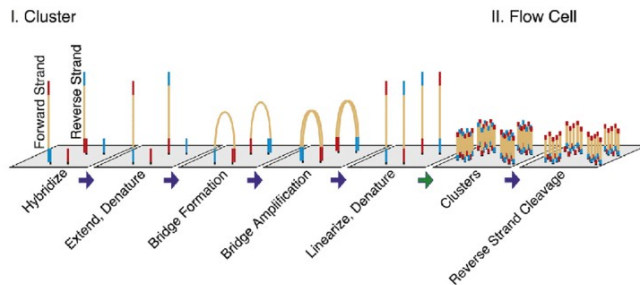
É possível notar que não há uma correlação direta entre tamanho do genoma e complexidade do organismo

Sequenciamento de genomas

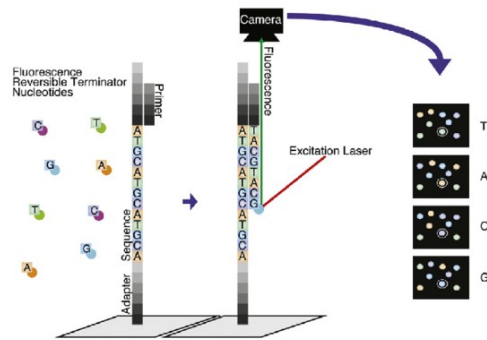
- Genomas possuem grande numero de bases em seu genoma (10^5 a 10^{12}) e ate o momento as técnicas existentes de seqüenciamento conseguem amostrar apenas algumas centenas de bases por reação.
- Deste modo, o genoma tem que ser seqüenciado de forma descontinua com milhares de reações sendo realizadas em paralelo para obtenção da informação necessária
- Isto gera uma grande quantidade de seqüências derivadas do genoma que se apresentam de forma desconexa, visto que não existe nenhuma propriedade intrínseca que permite a ordenação inequívoca destas
- Deste modo um dos grandes desafio ao seqüenciar genomas é a montagem destas sequencias de modo que elas possam reproduzir a ordenação encontrada nos cromossomos

Técnicas de sequenciamento em larga escala

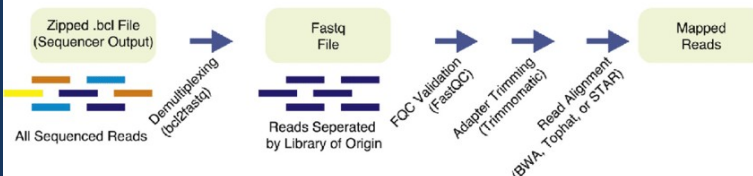
A. Clustering



B. High-throughput sequencing



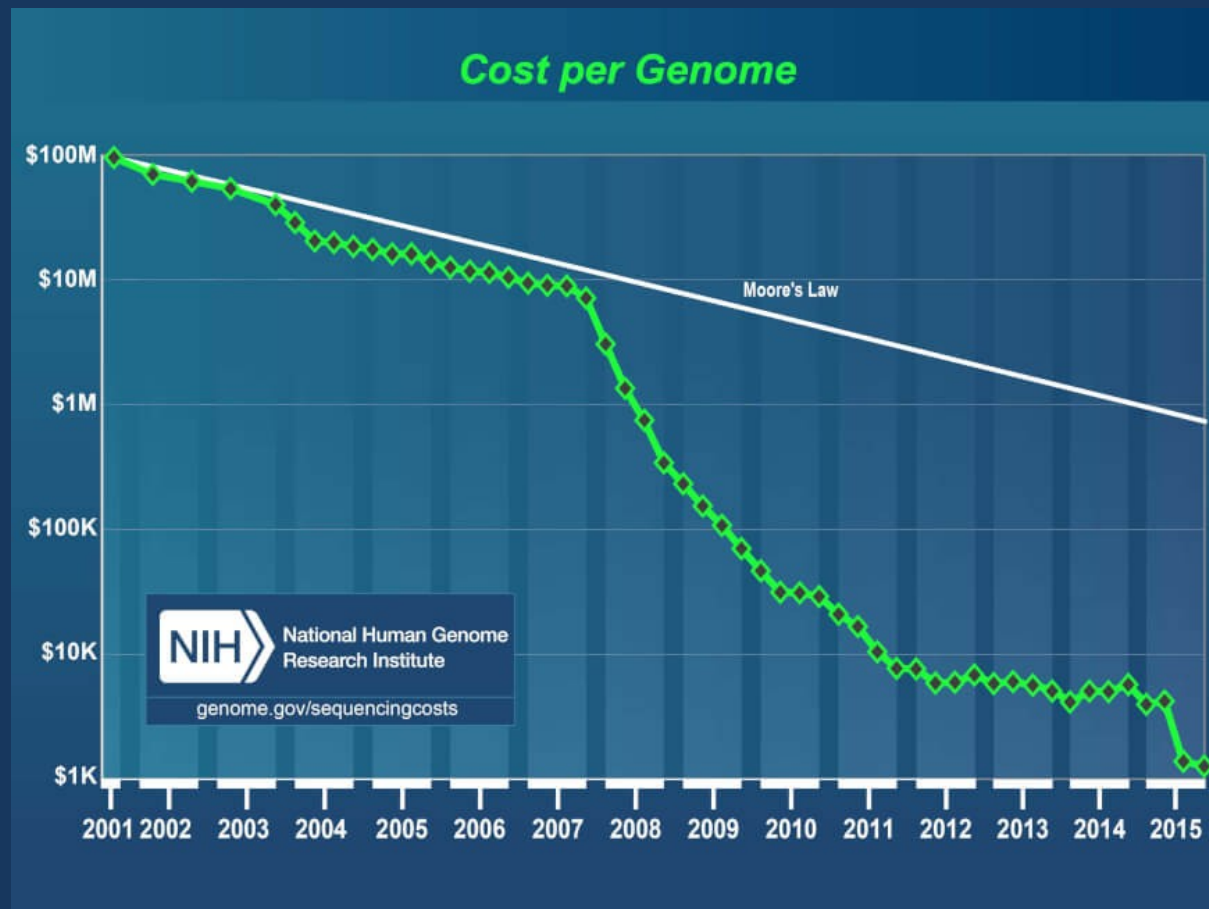
C. Demultiplexing samples and read mapping



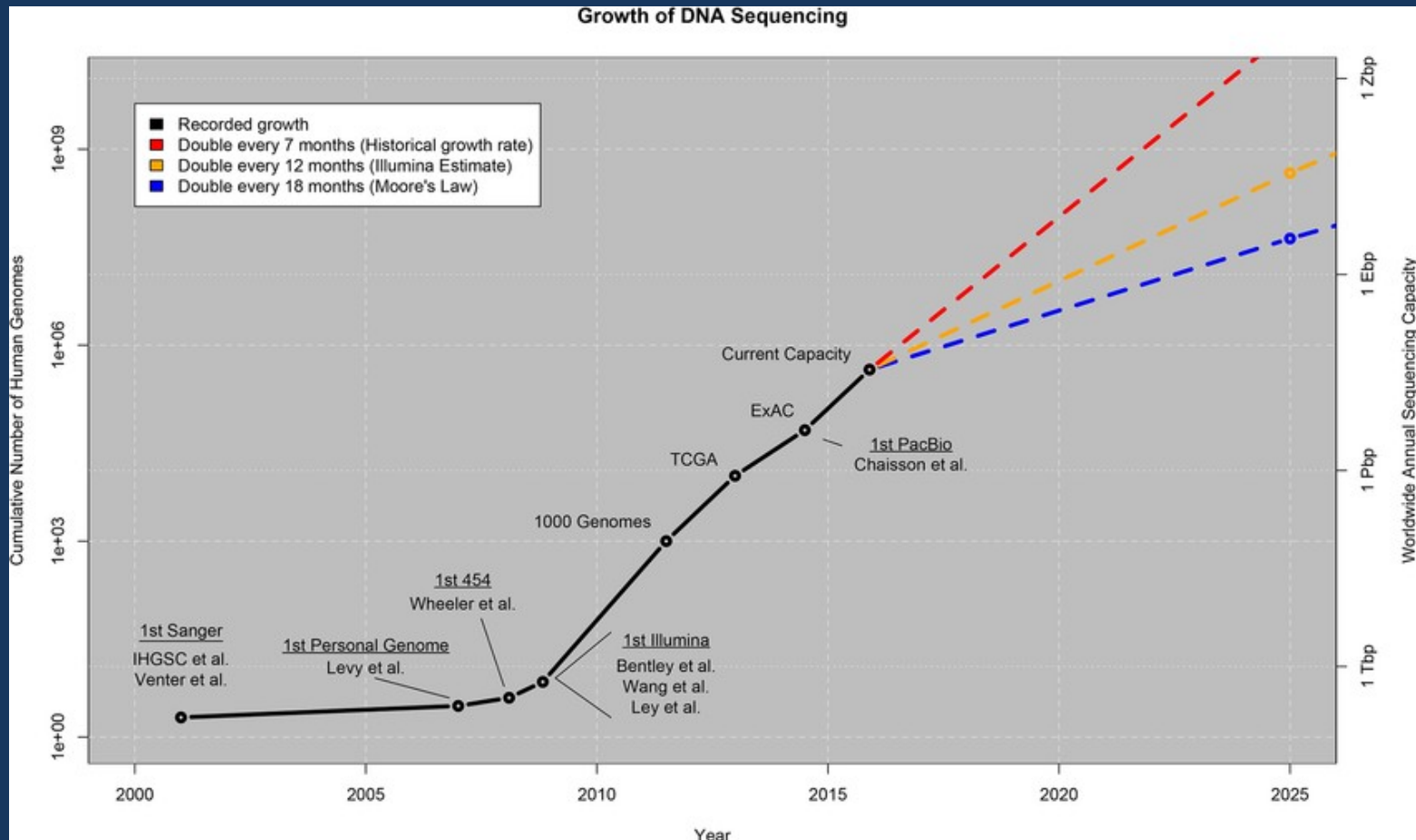
Em meados da década de 2000 foram desenvolvidos métodos alternativos ao sequenciamento de Sanger que foram denominados “next generation sequencing” que permitiram um grande barateamento e simplificação no sequenciamento em larga escala.

Isso permitiu com que o sequenciamento de genomas e transcriptomas fosse ampliado.

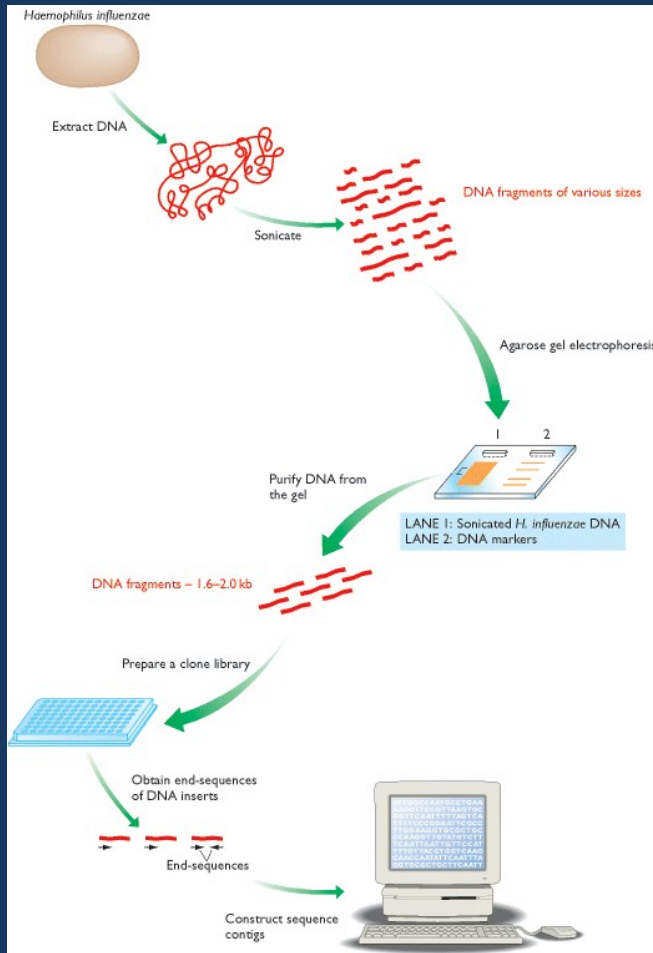
Aumento da quantidade de informações biológicas



Aumento da quantidade de informações biológicas



Seqüenciamento de DNA por fragmentação

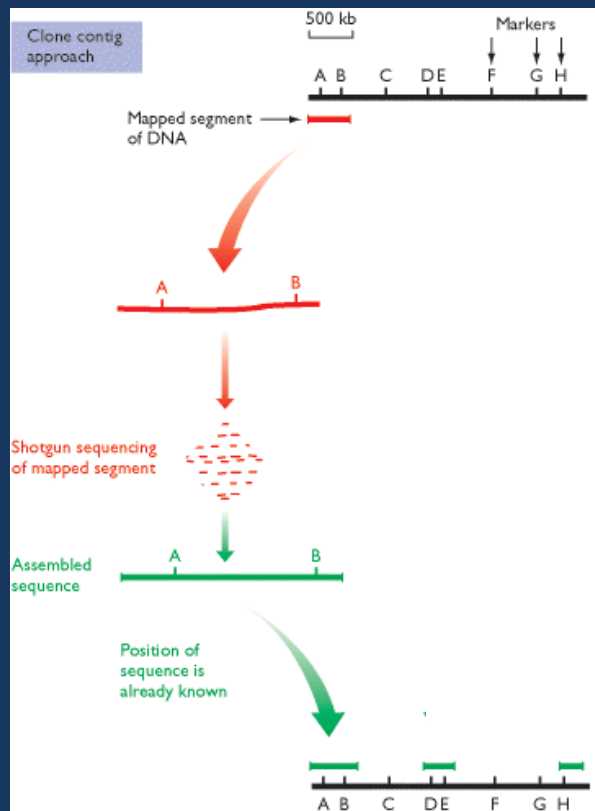


Esquematização da estratégia básica de seqüenciamento de genomas

Quebra do DNA pode ser feita com enzimas de restrição ou via nebulização

Clonagem de fragmentos com tamanho aproximado conhecido é importante para a posterior montagem das seqüências

Estratégias de seqüenciamento de genomas

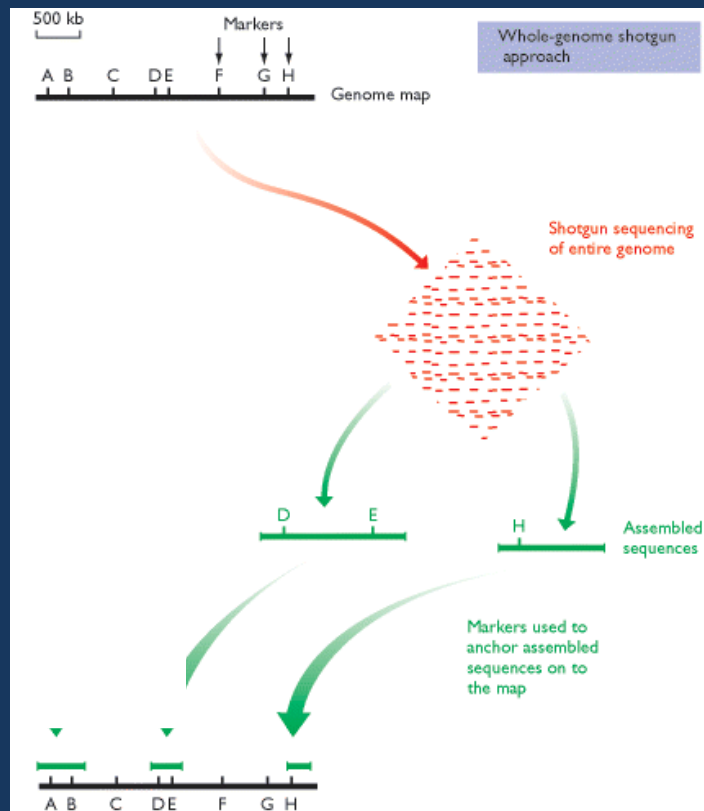


Existem duas estratégias básicas de seqüenciamento de genomas :

Seqüenciamento de clones- DNA é cortado em fragmentos grandes (utilizando enzimas de restrição) e clonado em BACs (Cromossomo Artificial de Bactéria), que aceitam fragmentos de até ~200 mil bases

Após isso são selecionados clones que são separadamente cortados (por nebulização) e sub-clonados em plamideos. Seqüenciamento destes sub-clones permitirá a reconstituição do clone do BAC

Estratégias de seqüenciamento de genomas

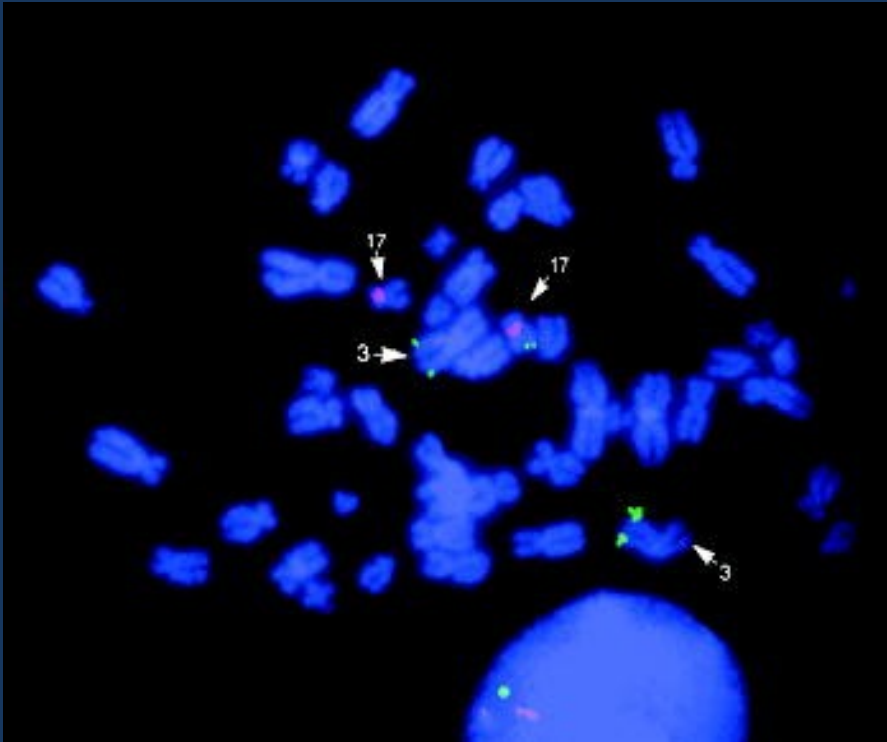


Whole Genome Shotgun (WGS)- O genoma inteiro é picotado em pedaços de alguns poucos milhares de pares de bases (por nebulização) e clonado em plasmídeos. São realizadas clonagens de fragmentos maiores em vetores apropriados (BACs, Fosmídeos, etc..) mas somente as pontas são seqüenciadas.

Seqüenciamento das duas extremidades de cada clone é realizada e utilizando a informação de seqüência e a estimativa de distância entre as duas pontas do clone busca-se montar o genoma inteiro

Após a montagem das seqüências obteremos pedaços que são denominados *contigs*, no melhor cenário cada *contig* deveria representar um cromossomo, mas normalmente o que ocorre é a formação de mais de 1 contig por cromossomo devido a regiões que apresentam problemas e que não são montadas de modo apropriado

Mapeamento físico de seqüências por FISH

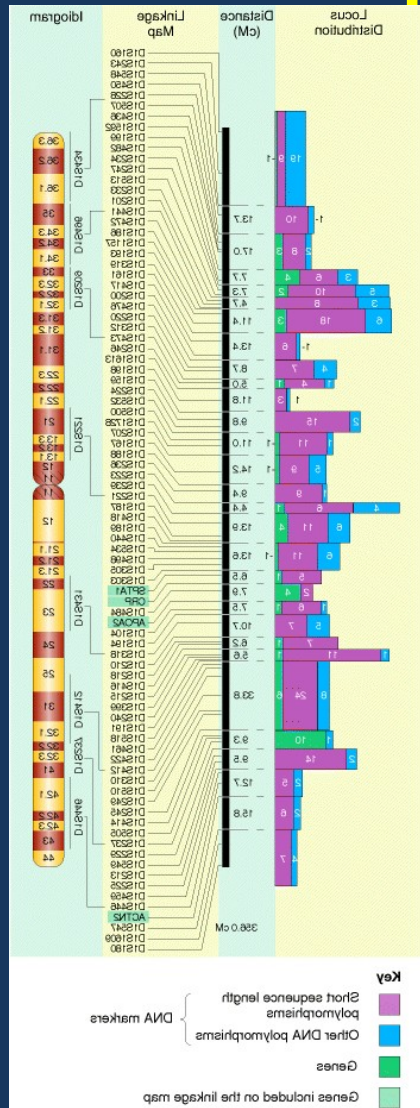


Um vez que obtemos os *contigs* é importante a determinação da região de que cromossomo esta seqüência é localizada

Para isso utiliza-se uma técnica de hibridização em cromossomos (FISH), que utiliza uma sonda que represente uma região única no genoma que esteja representada no contig de interesse, o resultado mostrará a região de qual cromossomo esta seqüência se localiza

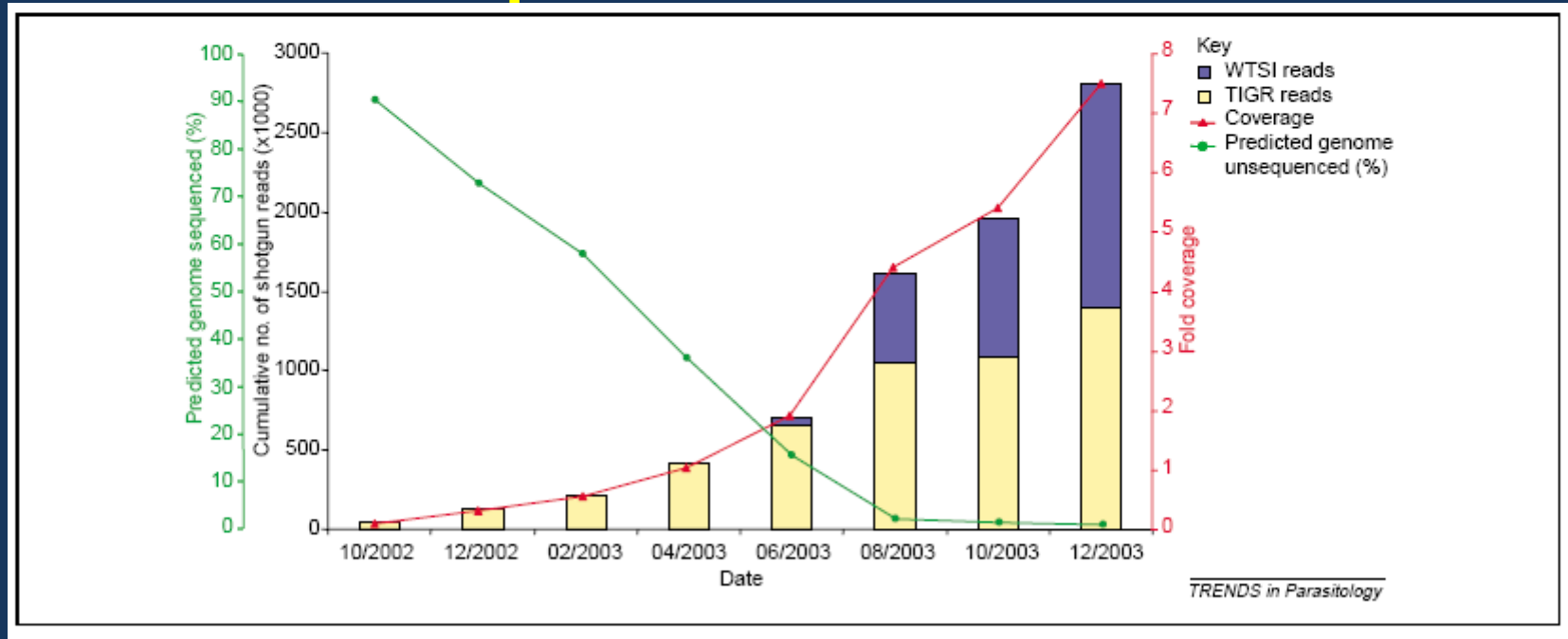
Com esta informação é possível utilizar os contig para montar um esqueleto de cada cromossomo.

Mapeamento de sequencias por segregação



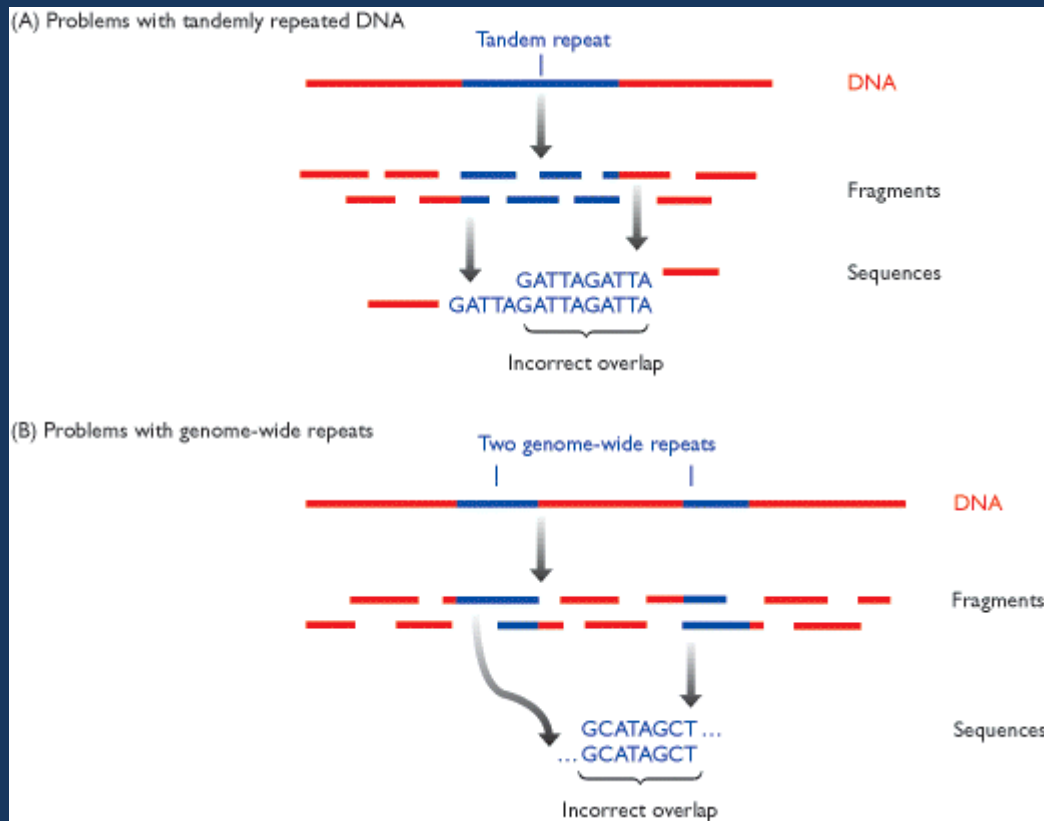
- Construído a partir das taxas de recombinação meiótica (crossing-over) de marcadores no genoma
- Centimorgan- probabilidade de 1% que dois genes se separem em evento de mitose.

Cobertura de sequenciamento



- Devido ao caráter randômico de sequenciamento utilizando a técnica de WGS é necessário sequenciar uma quantidade de bases muito maior do que o numero de bases do genoma (pelo menos 8X mais) devido a redundância do sequenciamento

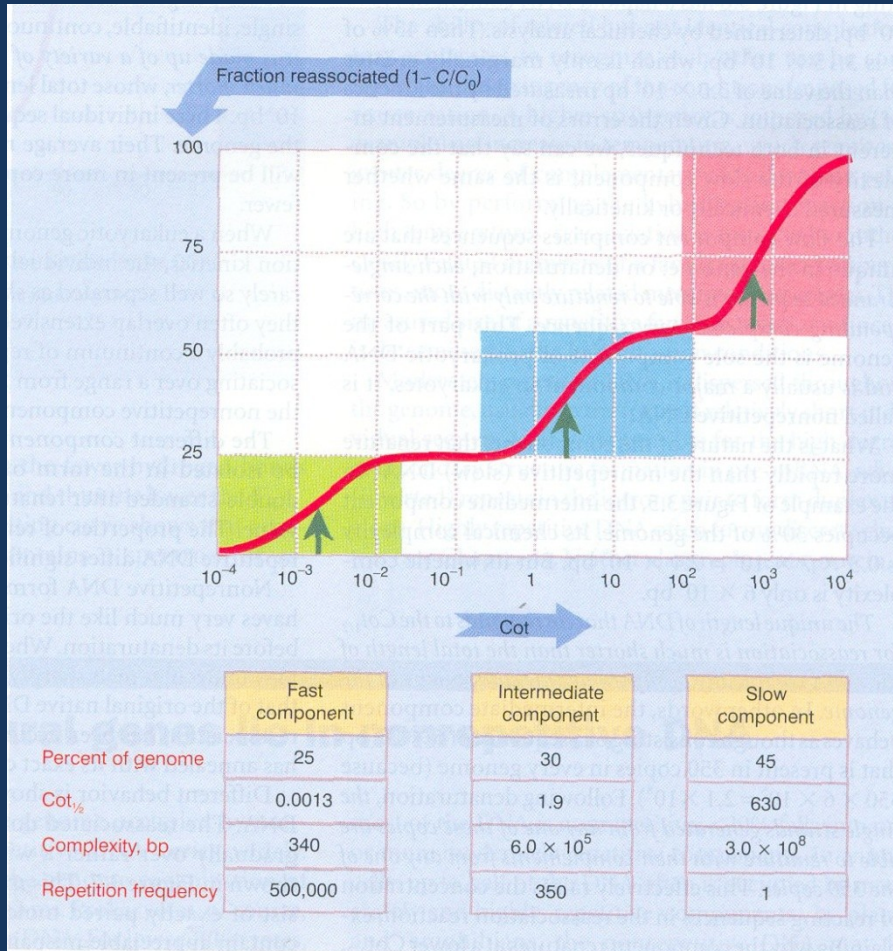
Problemas na montagem do DNA



Devido ao alto numero de seqüências repetitivas existe uma alta probabilidade que duas seqüências não contiguas mas que possuam a mesma repetição em sua extremidade possam ser sobrepostas na montagem

Par evitar isso em adição ao algoritmo de alinhamento de DNA utiliza-se a informação de distancia entre as seqüências das pontas de cada clone na montagem do genoma

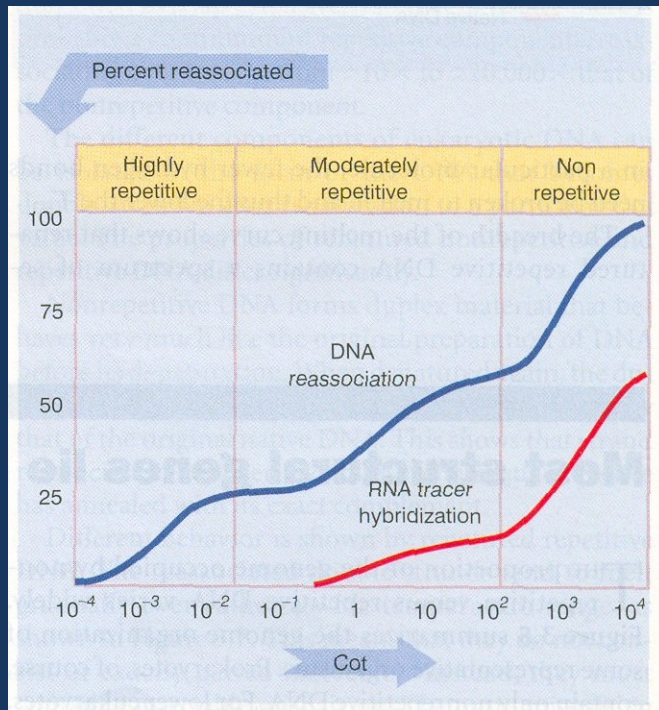
Complexidade de genomas



Experimentos de cinética de reassociação permitem obter uma estimativa da complexidade de cada genoma

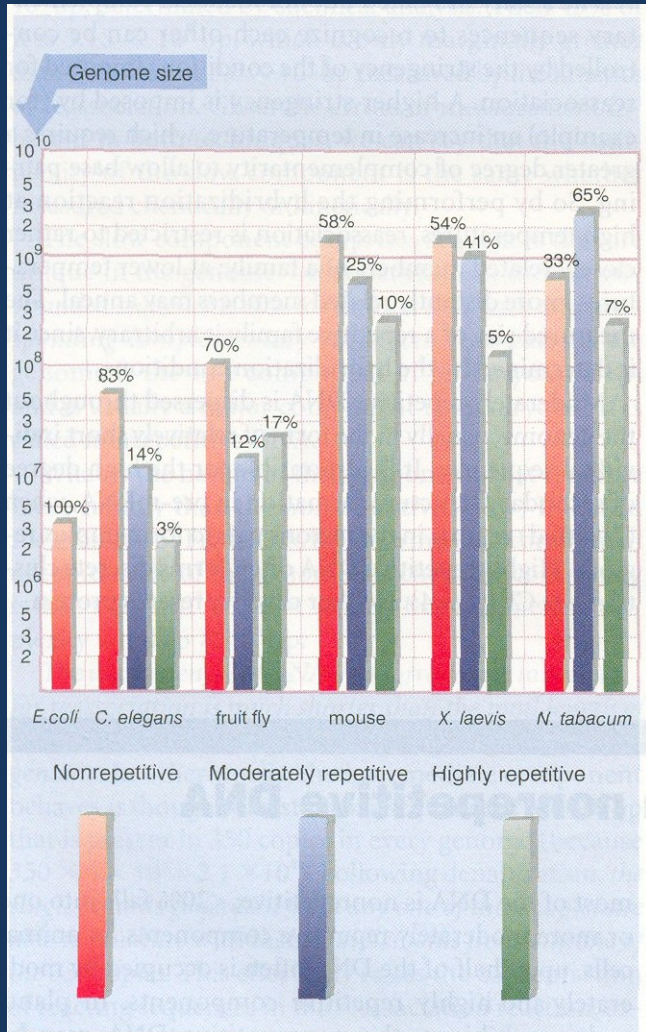
De modo geral genomas de eucariotos possuem uma grande fração de regiões repetitivas

Complexidade de genomas



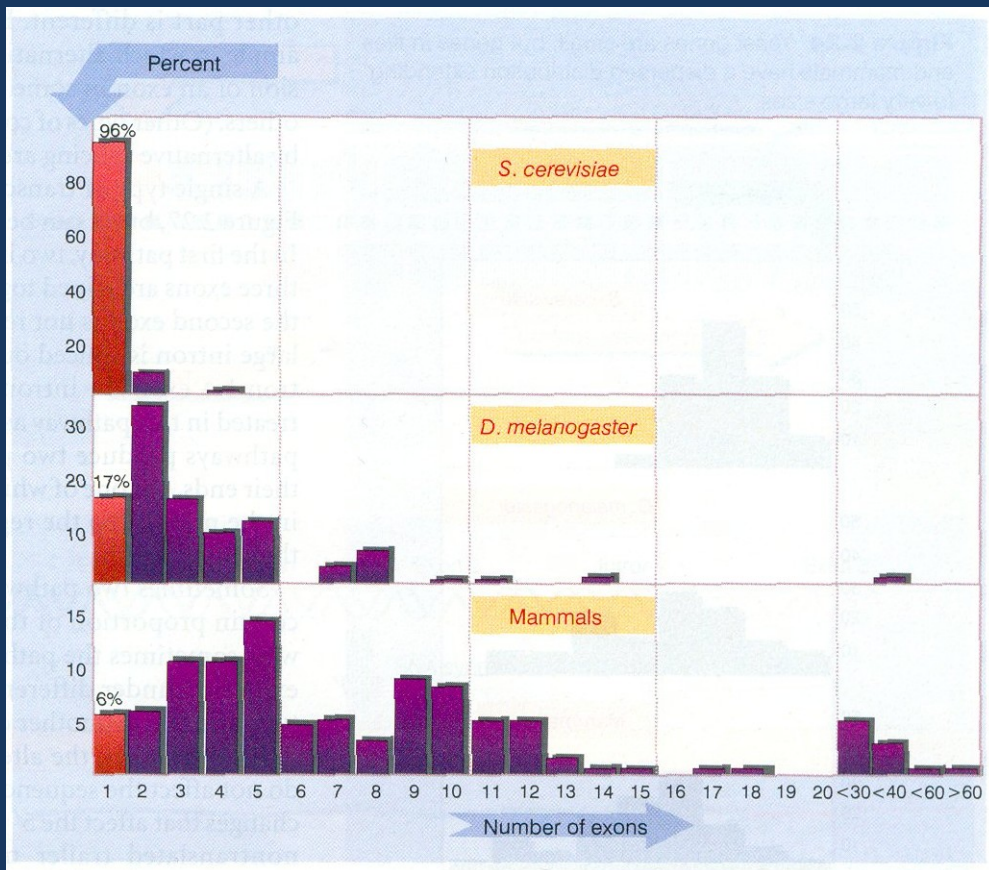
- Experimentos utilizados com pequena quantidade de moléculas de mRNA permitem ver que a maior fração do que é expresso pela célula provém de regiões não repetitivas

Complexidade de genomas



- A presença destas regiões repetitivas explica em parte porque o tamanho dos genomas não é proporcional a complexidade do organismo visto que uma serie de organismos menos completos possuem uma alta quantidade de DNA repetitivo.
- Este DNA repetitivo não possui uma função clara e por isso muitas vezes é referido como “junk DNA”

Correlação entre o genoma e a complexidade de organismos

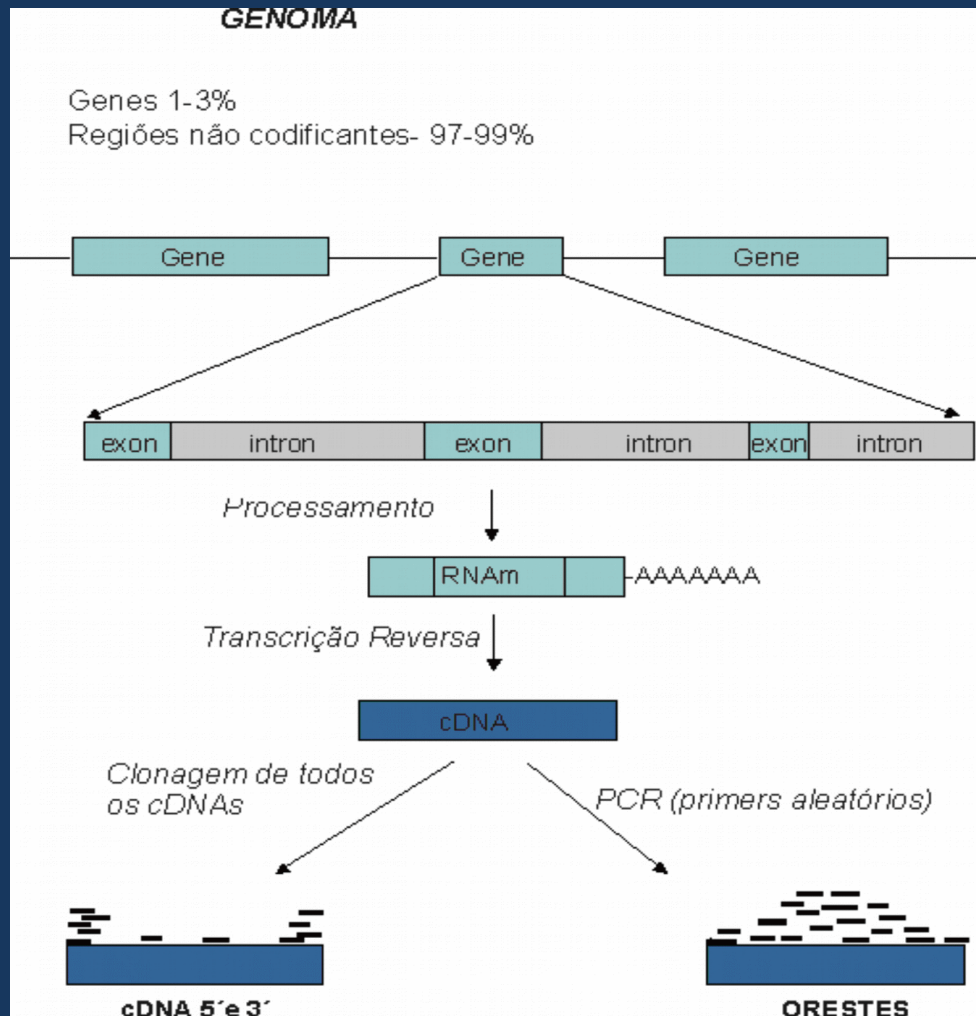


- É possível notar um tendência de organismos mais complexos possuírem genes com um maior numero de exons.

Seqüenciamento de transcriptomas

- Conforme mostrado no slide anterior organismos mais complexos tendem a possuir genes com um alto numero de exons. Além disso, o genoma destes organismos possuem uma alta quantidade de seqüências não-codificantes e portanto a predição da estrutura de genes não é trivial.
- Deste modo o seqüenciamento direto das moléculas de mRNA pode fornecer informações a respeito da estrutura de um gene, pois representa a molécula madura formada após os eventos de *splicing*
- Além disso, o seqüenciamento de mRNA permite a amostragem direta das seqüências codificantes permitindo com que um menor volume de seqüenciamento se obtenha maior informações sobre as proteínas deste organismo

Seqüenciamento de transcriptomas



Após o isolamento das moléculas de mRNA é realizada a reação de transcriptase reversa que irá gerar um fita de cDNA a partir de um mRNA molde.

Normalmente esta transcrição é realizada com um oligo-dT como primer o que permite com que o mRNA inteiro seja transcrito

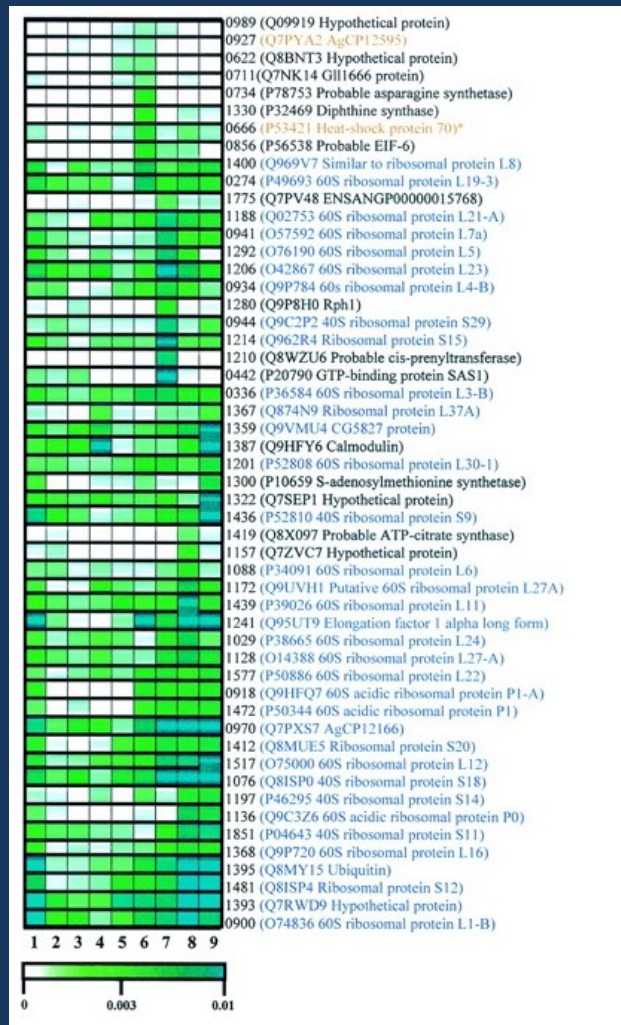
Os cDNA produzidos são clonados e o conjunto de plasmídeos produzidos é denominado biblioteca

Seqüenciamento de transcriptomas

Abundance distribution	
Total copies	55,172 (100)
> 1000	5
501 to 1000	6
101 to 500	22
51 to 100	26
11 to 50	153
6 to 10	235
2 to 5	3,084
1	9,936

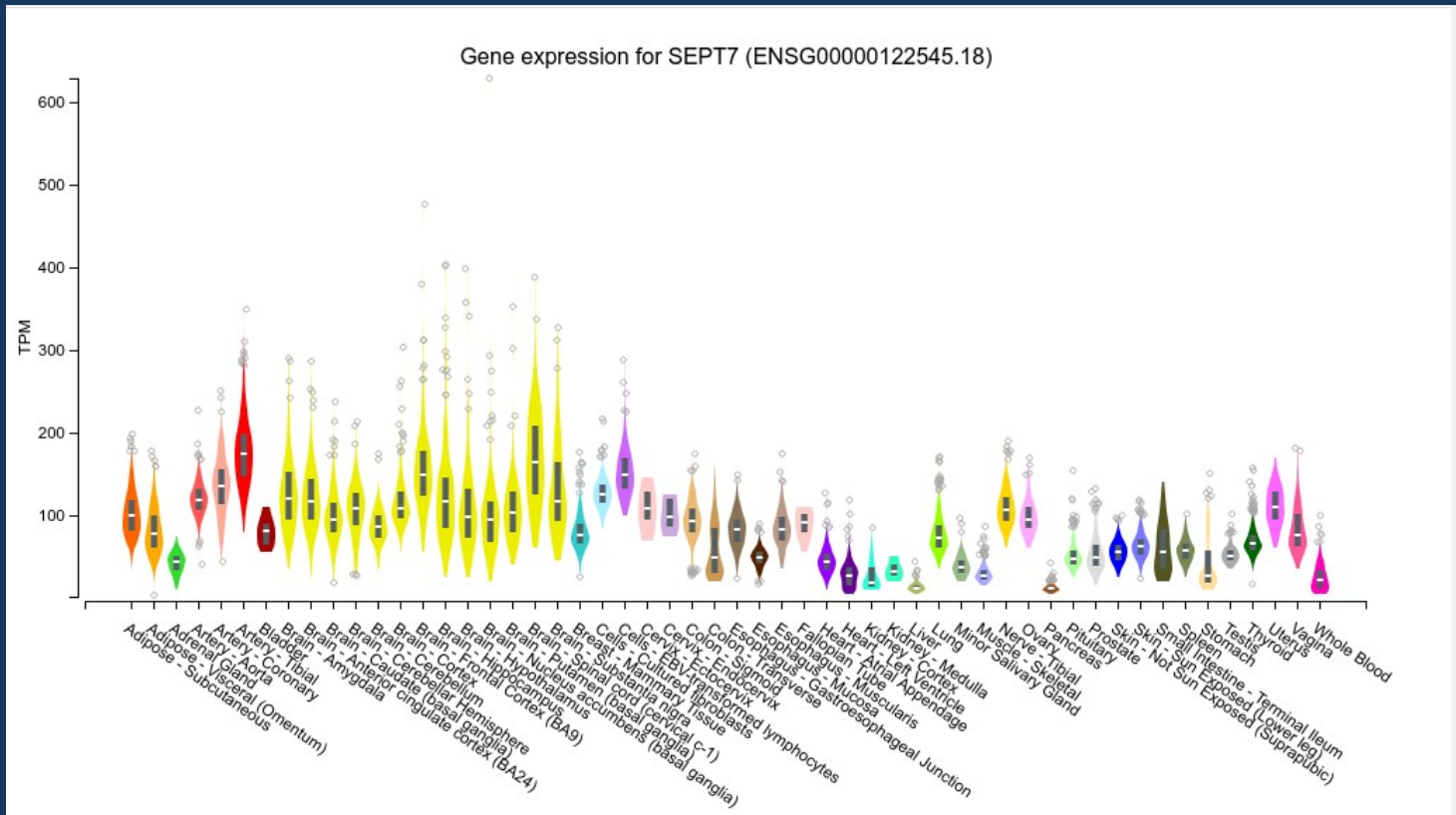
- Entretanto a abundancia de diferente mRNAs em uma célula varia muito. Existem alguns poucos mRNAs que possuem um numero de moléculas até 1000 X maior que a maioria dos mRNAs.
- Deste modo, seqüenciamentos de bibliotecas de mRNAs tendem a amostrar muito umas poucas moléculas e pouco um conjunto grande
- Além disso, nem todos os mRNA vão estar sendo expressos em um único tecido ou fase de vida do organismo e por isso para obter uma descrição completa dos mRNAs de um organismos vários destes deverão se amostrados

RNA-Seq



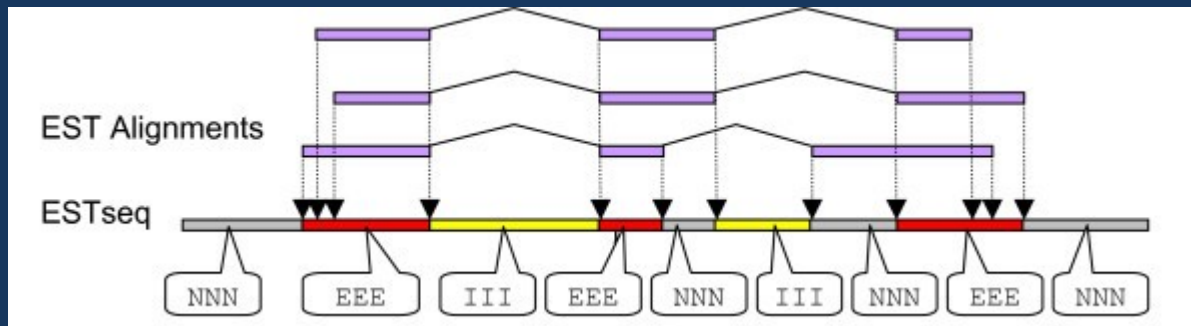
- Devido ao fato do numero de clones seqüenciados em uma biblioteca ser proporcional a sua abundancia é possível realizar uma comparação de diferentes bibliotecas para obter uma medida da expressão diferencial de um gene

RNA-Seq



Bancos de dados contém informações relativas a expressão de genes em diversos tecidos para milhares de amostras.

Splicing alternativo



Seqüenciamento de transcritos permite a dedução de eventos de splicing alternativo a partir do mapeamento deste nas seqüências de DNA genômico

