

A SEQUENCIAÇÃO EM ANÁLISES CLÍNICAS

Prof. Doutor José Cabeda

Índice

1.	Métodos para a sequenciação de fragmentos de DNA	2
1.1.	O método de sequenciação química	2
1.2.	O método de sequenciação de Sanger	2
1.2.1.	Clonagem dos fragmentos a sequenciar	4
1.2.2.	Os terminadores	5
1.3.	A resolução dos produtos da reacção de sequenciação	6
1.4.	A sequenciação directa por PCR	8
1.5.	O sequenciador ABI 310	9
1.6.	Vantagens e limitações da sequenciação em análises clínicas	15
2.	A sequenciação do genoma Humano	16
2.1.	Estratégias para a sequenciação do genoma Humano	16
2.1.1.	O consórcio Público: Estratégia do mapeamento	16
2.1.2.	A celera genomics: Estratégia do “Whole genome shot-gun”	19
2.2.	Estratégias para a identificação de genes na sequência do genoma Humano .	20
2.2.1.	recurso à identificação de genes ortólogos	20
2.2.2.	Identificação de sequências conservadas entre espécies	20
2.2.3.	Identificação de zonas de elevada densidade com sequências consenso para caixas promotoras/enhancers	21
2.2.4.	O mapa de EST.....	21

1. Métodos para a sequenciação de fragmentos de DNA

1.1. O método de sequenciação química

O método de sequenciação de Maxam-Gilbert utiliza um processo de degradação química para cortar o DNA em pontos específicos, produzindo fragmentos de DNA de diversos tamanhos. Os fragmentos assim produzidos são então separados em gel, o que permite pela determinação do seu tamanho determinar o nucleótido que se encontra em cada posição da sequência a analisar. Apesar de totalmente automatizado, e da modificação introduzida com a sequenciação multiplex (que permite analisar cerca de 40 clones por gel), este método é muito menos utilizado que o método enzimático (sequenciação de Sanger ou método de sequenciação dideoxi).

1.2. O método de sequenciação de Sanger

O método original de Sanger para a sequenciação de ácidos nucleicos é um método enzimático, utilizando uma reacção simples de polimerização. Por este motivo, e porque nenhuma polimerase inicia o seu trabalho se não tiver disponível uma extremidade 3'

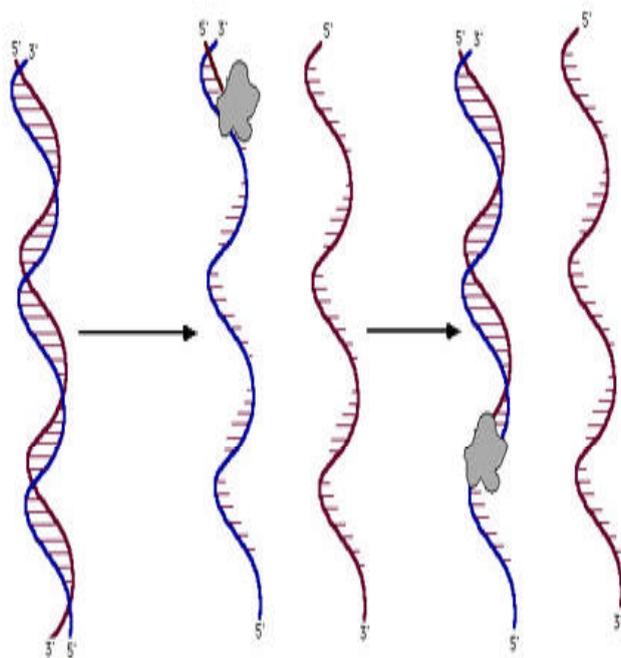


Fig. 1

livre, é necessário para a utilização deste método ter algum conhecimento da sequência da extremidade 5' da zona a sequenciar. Com este conhecimento é assim possível desenhar um primer complementar da cadeia a sequenciar, que por hibridização com esta vai fornecer a necessária extremidade 3' livre. De seguida a reacção prossegue, incorporando nucleótidos até que o nucleótido incorporado seja um terminador (ver secção à frente).

DNA Polymerase reads the template strand and synthesizes a new second strand to match:



IF 5% of the T nucleotides are actually dideoxy T, then each strand will terminate when it gets a ddT on its growing end:



Fig. 2 -

Como na reacção original de Sanger, era efectuado apenas um ciclo de polimerização, era necessário garantir uma elevada quantidade de DNA molde, bem como uma grande sensibilidade de detecção dos produtos da reacção. Assim, os nucleótidos utilizados eram frequentemente radioactivos, o que aumentava a sensibilidade de detecção dos produtos da reacção. Por outro lado, o aumento da concentração do DNA molde era assegurado com um passo inicial de clonagem dos fragmentos a sequenciar.

1.2.1. Clonagem dos fragmentos a sequenciar

A clonagem dos fragmentos a sequenciar era no período pré-PCR a única forma de garantir, por um lado a especificidade do material a sequenciar, e por outro a existência de uma amplificação do material genético antes da reacção de sequenciação. Para clonar o material genético, este era inicialmente fragmentado (por sonicação, ou mais frequentemente com recurso a reacções de restrição), separado em gel, purificado, e inserido em vectores de sequenciação. Estes não eram mais que plasmídeos ou DNA de fagos, alterados geneticamente, por forma a possuírem locais de ligação, e genes de selecção. Os locais de ligação serviam para efectuar o posicionamento do material a sequenciar de forma ordenada e precisa, habitualmente por recurso a reacções de restrição do vector, seguida de incubação com o DNA a sequenciar (insert), seguido de uma reacção de ligase. O vector assim produzido era então inserido em bactérias. Estas eram cultivadas, habitualmente na presença de antibióticos, o que apenas permitia o crescimento das bactérias que possuíssem o plasmídeo, já que um dos genes de selecção deste era um gene de resistência ao antibiótico utilizado. Um segundo gene de selecção frequentemente utilizado permitia às colónias bacterianas que tinham incorporado plasmídeo com o insert apresentar cor diferente das colónias cujo plasmídeo não possuía insert.

O conjunto de colónias obtido era chamado de biblioteca genómica, e destas colónias eram escolhidas algumas para serem sequenciadas. As colónias escolhidas eram então colocadas em meios líquidos de cultura bacteriana contendo o antibiótico anteriormente referido, o que permitia o crescimento exponencial das bactérias, e conseqüentemente a multiplicação do plasmídeo com o insert. No final da cultura, as bactérias eram utilizadas para purificar o plasmídeo, e deste retirar o insert.

Como já foi referido, a sequenciação pelo método de Sanger exige a utilização de um primer (iniciador), o qual só é passível de ser desenhado quando se conhece a sequência do DNA alvo. Obviamente que esta situação cria um dilema tipo ovo-galinha (só se pode sequenciar aquilo de que já se conhece a sequência), a qual é no entanto ultrapassável com recurso ao vector de clonagem. Com efeito, próximo do local de clonagem do insert, o plasmídeo foi alterado por forma a possuir uma sequência conservada contra a qual pode ser desenhado um primer. No entanto, para utilizar este primer, é necessário que a remoção do insert se faça por forma a que esta sequência seja

conservada junto ao insert, o que se faz recorrendo a uma enzima de restrição diferente da inicialmente utilizada. Deste modo, a sequenciação precedida de clonagem tem ainda a vantagem de poder ser realizada, mesmo quando o DNA alvo não possui uma sequência conhecida.

1.2.2. Os terminadores

Os nucleótidos habitualmente utilizados na polimerização de DNA são desoxirribonucleótidos, enquanto os utilizados na polimerização do RNA são ribonucleótidos. O que distingue estes dois tipos de nucleótidos é a existência no carbono 2 da ribose de um grupo OH que não existe na desoxirribose (Fig.3). No entanto em ambos os nucleótidos existe um grupo OH no carbono 3 (fig. 3). Este grupo hidroxilo (OH) é essencial para que os nucleótidos possam formar uma ponte fosfodiéster com o nucleótido seguinte (fig 4), pelo que a remoção deste grupo torna o respectivo nucleótido (um didesoxirribonucleótido ou ddNTP) um terminador muito eficaz, já que, na incorporação de nucleótidos, a polimerase não detecta esta subtil mas decisiva alteração química.

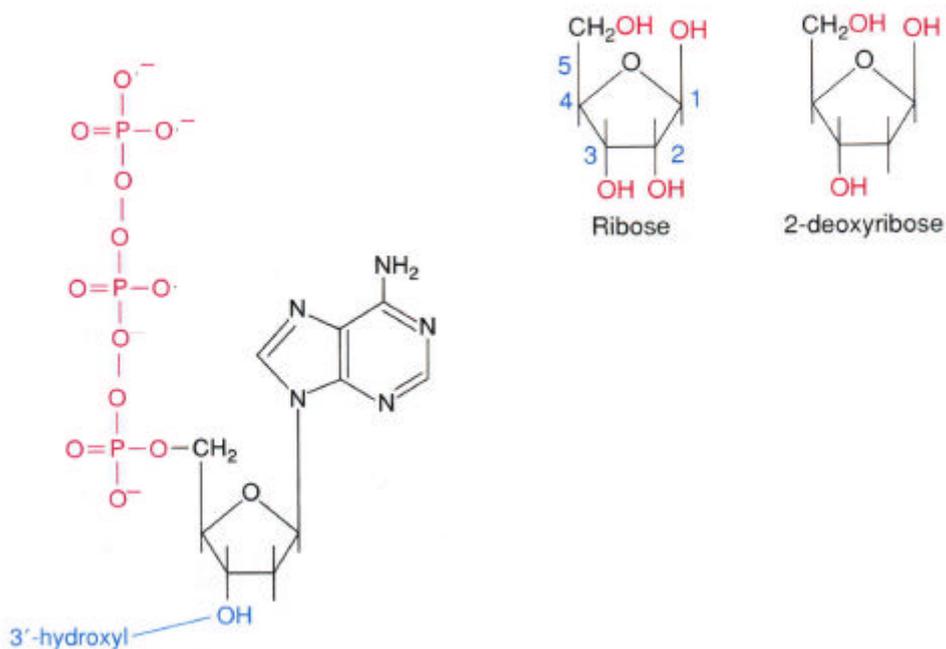


Fig.3 -

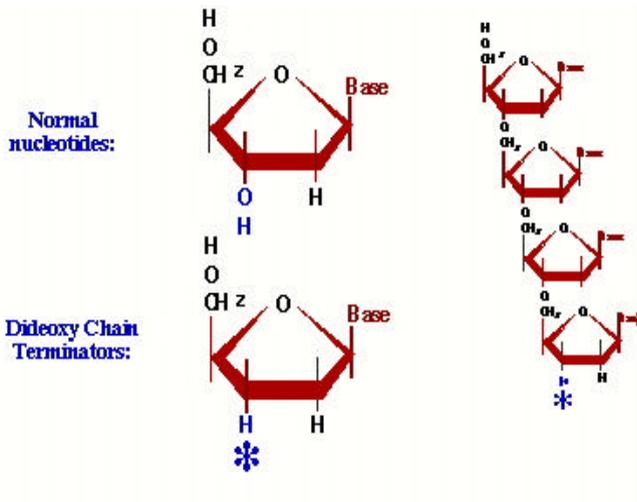


Fig. 4 -

1.3. A resolução dos produtos da reacção de sequenciação

Os produtos da reacção de sequenciação de Sanger necessitam então de ser resolvidos segundo o seu peso molecular em geis de grande resolução (habitualmente geis de acrilamida). Se em cada reacção de sequenciação for apenas incorporado uma espécie de terminador, cada reacção pode servir como revelador da posição em que aparece o respectivo nucleótido. A electroforese em paralelo das 4 reacções irmãs de sequenciação permite então obter um gel de sequenciação clássico, que se lê seguindo os produtos com peso molecular sucessivamente crescente (Fig. 5).

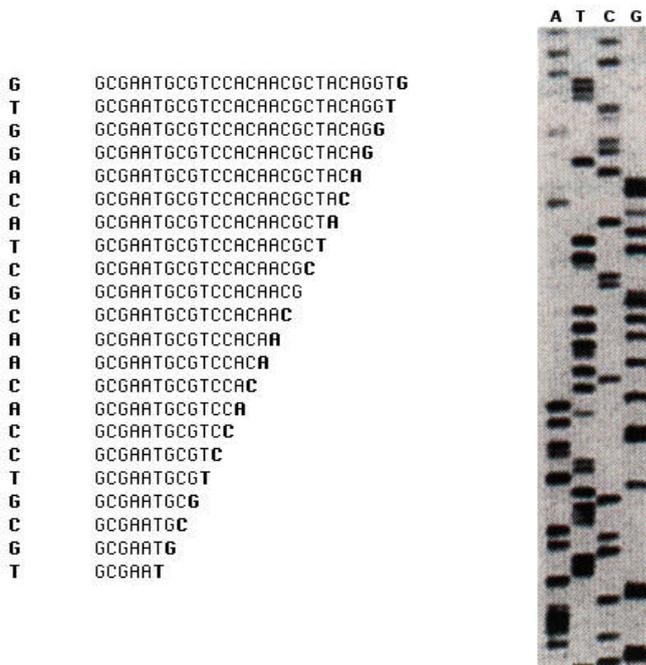


Fig. 5 -

Com o intuito de facilitar a leitura dos geis de sequenciação, tornando ao mesmo tempo mais precisas as diferenças de migração em gel, pretendeu-se incorporar todas as quatro reacções de sequenciação numa só, utilizando para isso terminadores modificados. A modificação adicional consiste na marcação de cada espécie de terminador com um fluorocromo, o que permite que cada produto de PCR específico emita luz num comprimento de onda bem determinado, e diferente dos demais. Foi assim possível efectuar todas as reacções num só tubo, correndo todos os produtos numa só “lane” do gel (Fig. 6), ainda que a leitura dos geis necessite agora de equipamentos específicos capazes de analisar os comprimentos de onda emitidos.

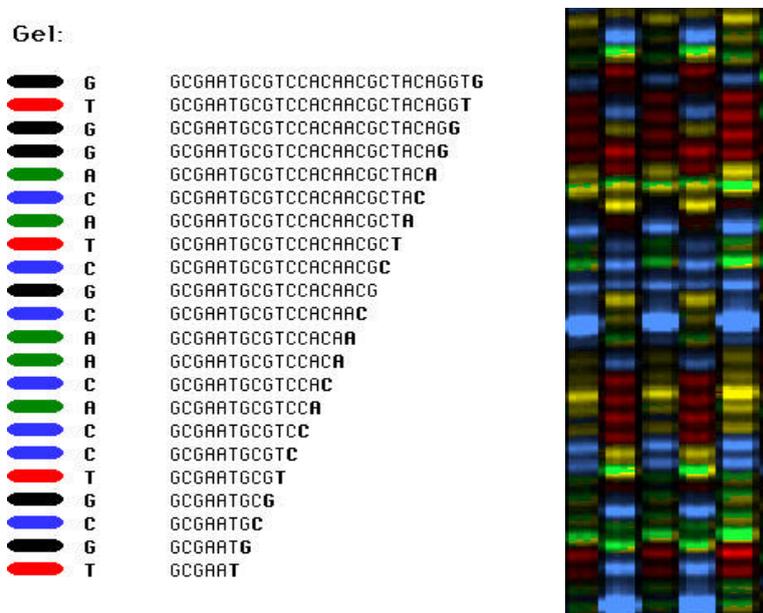


Fig. 6 -

1.4. A sequenciação directa por PCR

O advento do PCR veio permitir a simplificação substancial da sequenciação habitualmente utilizada num laboratório de rotina. Com efeito, ao contrário do que ocorre num laboratório de investigação, a rotina diagnóstica não lida frequentemente com sequências totalmente desconhecidas, já que habitualmente procura apenas detectar mutações/polimorfismos em genes ou loci previamente bem estudados. Assim, é possível na esmagadora maioria dos casos saltar todos os passos da clonagem, e estabelecimento de bibliotecas genómicas, já que quer a amplificação do material genético quer a especificidade do produto a sequenciar podem ser conseguidas com recurso a uma reacção de PCR convencional. Desta forma, a sequenciação directa inicia-se com um PCR que amplifica o gene/loci alvo com recurso a primers específicos e previamente escolhidos. A este PCR chamamos PCR simétrico, porque amplifica de igual modo ambas as cadeias do dsDNA original.

De seguida, o DNA é de novo submetido a 2 novas reacções de PCR em paralelo. Em cada uma destas reacções é utilizado apenas um dos primers iniciais, o que faz com que em cada reacção seja amplificado apenas uma das duas cadeias do DNA alvo. Por este motivo este PCR é denominado de PCR assimétrico.

O PCR assimétrico consiste ele próprio na reacção de sequenciação de Sanger. Por este motivo, a mistura de reacção necessita de possuir os terminadores (ddNTP) ainda que em menor concentração que os nucleótidos normais.

Uma vantagem adicional para a utilização do método da sequenciação directa consiste no facto de a reacção de sequenciação ser ela própria também uma reacção de PCR, o que aumenta muito a quantidade de produto final a ser detectado por electroforese.

1.5. O sequenciador ABI 310

Os produtos da reacção de sequenciação devem então ser resolvidos em função do seu tamanho por electroforese. Se tradicionalmente a electroforese vertical em geis de acrilamida era o método de eleição, na última década, a electroforese capilar tem-se tornado gradualmente técnica de eleição. As suas vantagens consistem na maior resolução, reprodutibilidade, capacidade de automatização, bem como numa menor exigência tecnológica ao nível dos detectores. O principal problema consiste num habitualmente menor nível de paralelismo, já que os equipamentos de electroforese

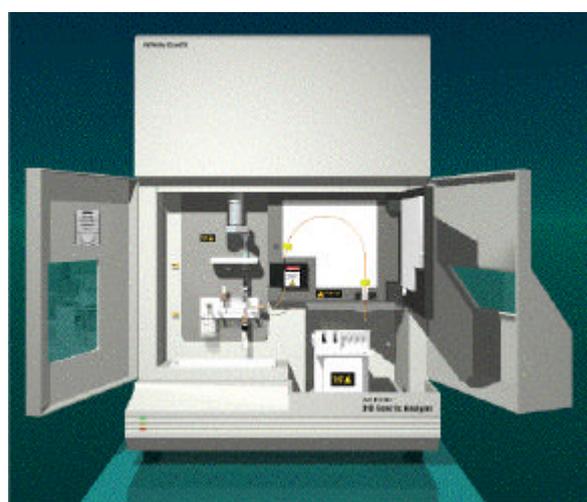


Fig. 7 -

capilar com capacidade para processarem mais que uma amostra em simultâneo são consideravelmente mais caros (mas existem, e são utilizados em alguns laboratórios de análises clínicas de rotina como os de medicina forense!).

De longe, o equipamento de electroforese capilar mais utilizado para resolver os produtos de uma reacção de sequenciação é o ABI 310 (Fig. 7), pelo que vamos deter-nos um pouco sobre o seu princípio de funcionamento.

O ABI 310 é um aparelho de electroforese capilar cujos componentes por questões de segurança (as diferenças de potencial aplicadas são da ordem das dezenas de milhar de volts) se encontram encerrados no interior de 2 portas com janelas (Fig. 7). A abertura das portas (Fig 7 B), permite encontrar vários componentes. Do lado direito em baixo, podemos visualizar o suporte das amostras (Fig. 8). Por cima deste fica o suporte

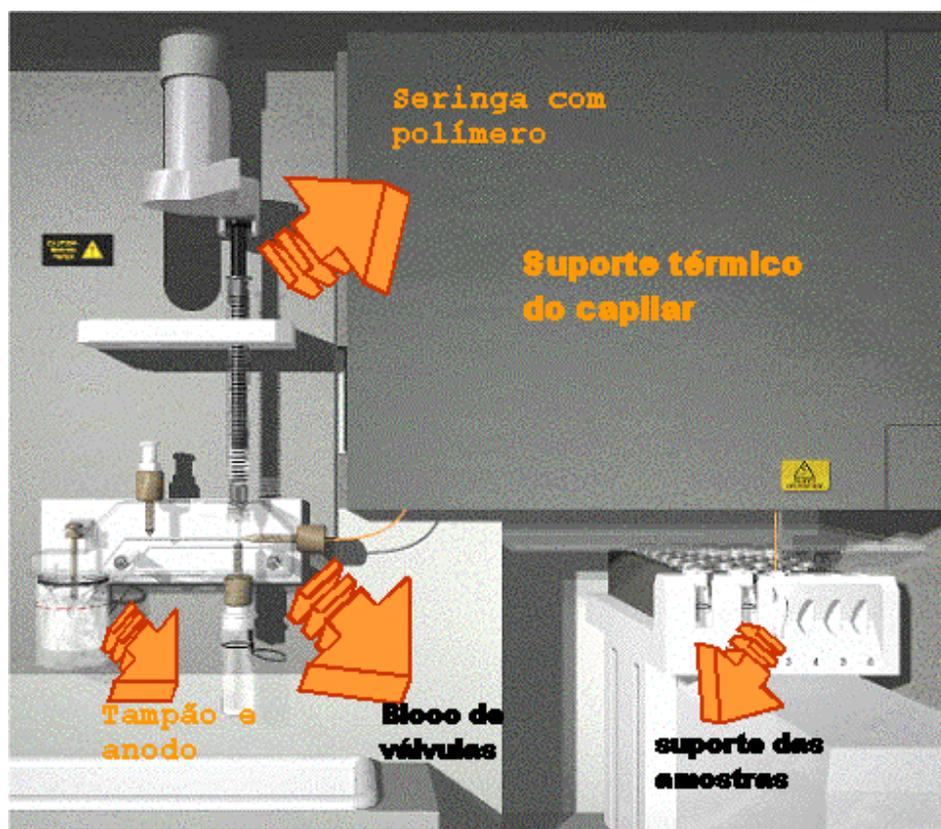


Fig. 8 -

térmico para o capilar (Fig. 8), o qual possui também uma porta que quando se abre expõe o capilar encostado a uma placa de cerâmica (Fig. 9) que assegurará durante a electroforese a manutenção da temperatura nos níveis programados. À esquerda do suporte térmico do capilar encontra-se o suporte da seringa com polímero (Fig. 8), a qual será utilizada para encher o capilar antes e durante a electroforese. Por baixo da seringa encontramos o bloco de válvulas, o qual controla o sentido do enchimento com o polímero e liga e desliga a aplicação da diferença de potencial nos momentos programados. O ânodo encontra-se aplicado ao tubo com tampão de electroforese, o qual se encontra na extremidade esquerda inferior do bloco de válvulas (Fig. 8).

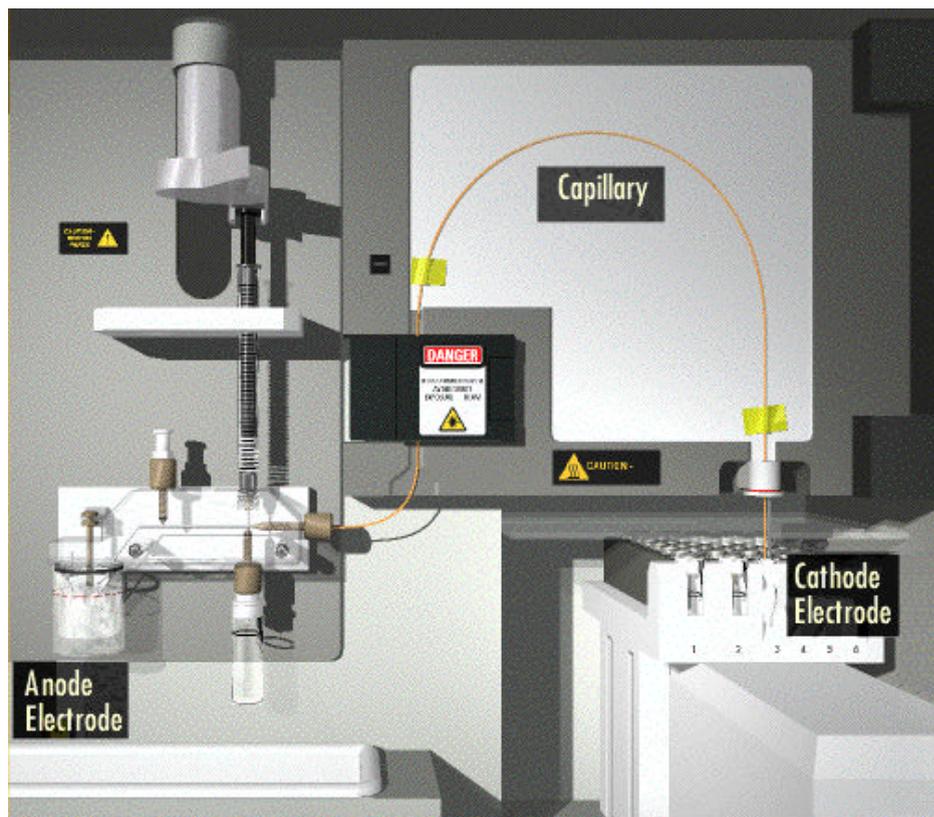


Fig. 9 -

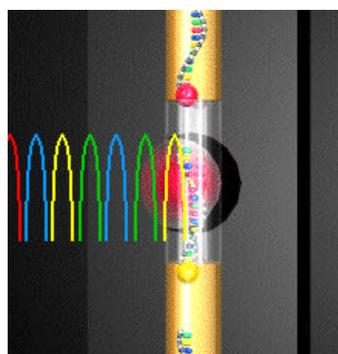


Fig. 10 -

Na extremidade inferior esquerda do suporte térmico do capilar encontra-se a janela de detecção (Fig.9), cuja porta abrindo-se expõe a janela transparente do capilar, alinhada com a lente que foca neste a fonte de luz de excitação, e recolhe a emissão de fluorescência (Fig. 10).

O funcionamento do aparelho consiste então em :

- 1) Fazer a montagem do equipamento, nomeadamente do capilar, amostras, tubos suplementares com água e tampão de electroforese, encher o capilar com polímero, assegurar que não existem bolhas de ar no sistema, e colocar o tampão de electroforese no ânodo.
- 2) Programar num computador a posição (no suporte) de cada amostra, o tipo de polímero e de terminadores utilizados, bem como as condições de electroforese a aplicar. Uma vez terminada a programação, o software controla de modo automático todo o processo, o qual consiste em:
 - a. O aparelho estabiliza a temperatura do capilar ao valor programado
 - b. Injecção da amostra: consiste na colocação de uma amostra na extremidade do cátodo do capilar, após o que é aplicada uma diferença de potencial, que impele o DNA a entrar no capilar.
 - c. Electroforese: O aparelho interrompe a diferença de potencial, retira a extremidade do capilar do tubo da amostra e lava-a num tubo com água. De seguida coloca esta extremidade do capilar num tubo com tampão de electroforese e reaplica uma diferença de potencial programada. Durante toda a corrida, o aparelho indica o que está a realizar (Figura 11),

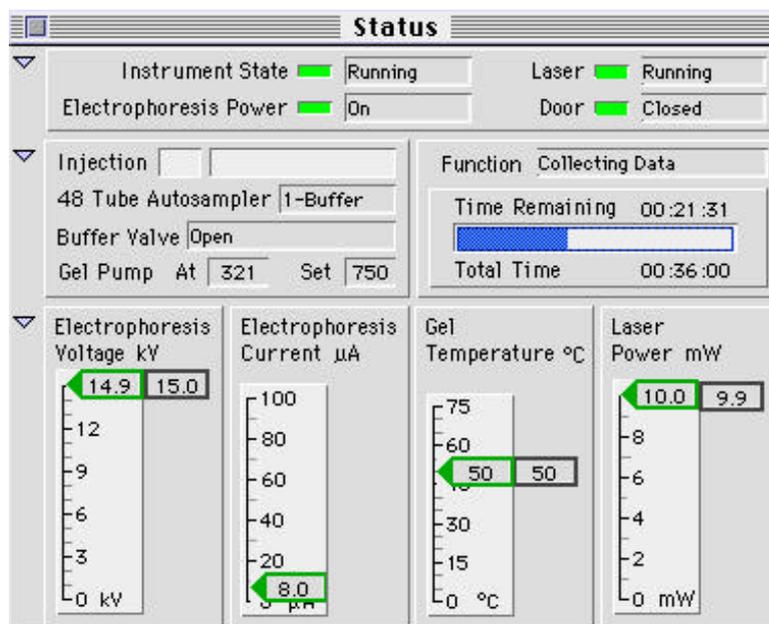


Fig.11 -

permitindo ao operador acompanhar as várias fases da electroforese.

- d. Leitura: Durante a electroforese, o aparelho regista a intensidade de fluorescência que é detectada na janela de leitura para cada um dos canais correspondentes aos terminadores utilizados (figura 12).

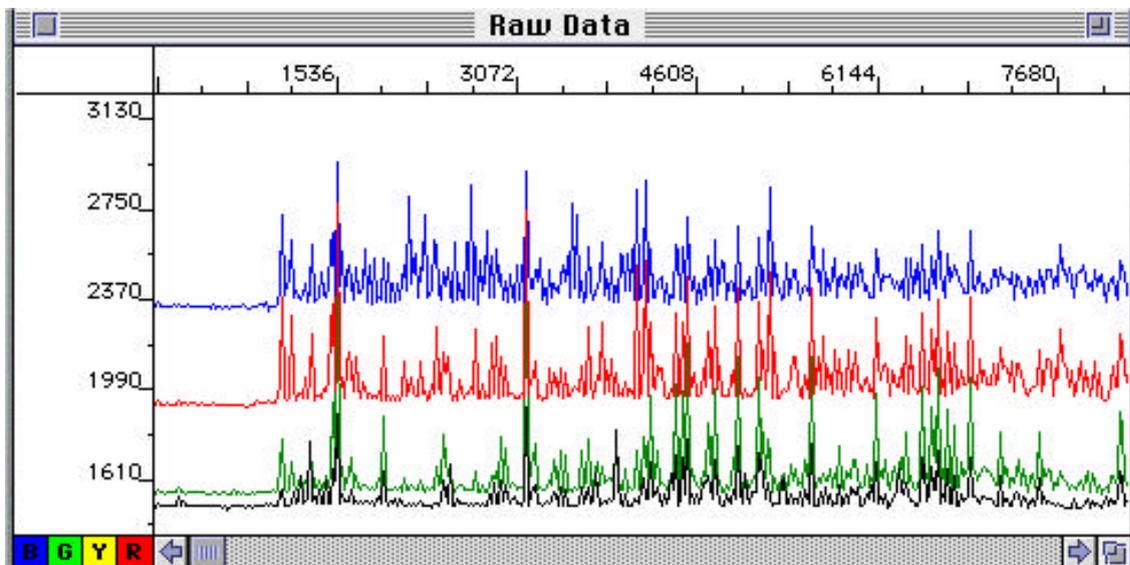


Fig. 12 -

- e. No final da electroforese de cada amostra, o processo é repetido até não existirem mais amostras a analisar no suporte.
- f. Análise: No final do processo, os ficheiros com os dados recolhidos são analisados por um programa que:
- Corrige as diferenças de mobilidade apresentadas devido aos terminadores utilizados
 - Corrige a sobreposição de espectros referente aos fluorocromos acoplados a cada terminador
 - Compara as curvas obtidas, inferindo a sequência de nucleótidos
 - O operador pode (e deve) ainda observar as curvas resultantes (cromatogramas; figura 13), e corrigir eventuais erros na atribuição da sequência. Por vezes é necessário pedir ao programa para reanalisar os dados utilizando configurações diferentes, que o operador considere mais adequadas à amostra em questão

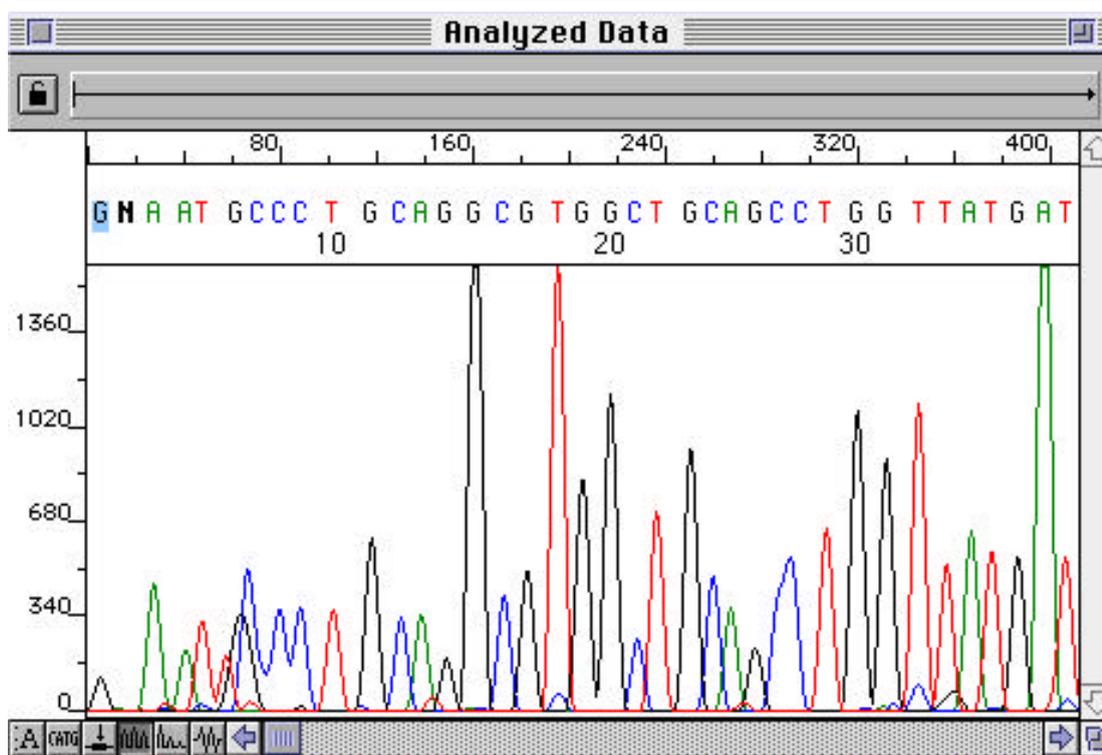


Fig.13 -

A utilização do ABI 310 (ou de qualquer um dos modelos posteriores) para sequenciação, é ainda facilitada pela disponibilidade comercial de kits (Fig. 14) e consumíveis (Fig. 15) prontos a usar, com protocolos padronizados, e de utilização relativamente fácil.

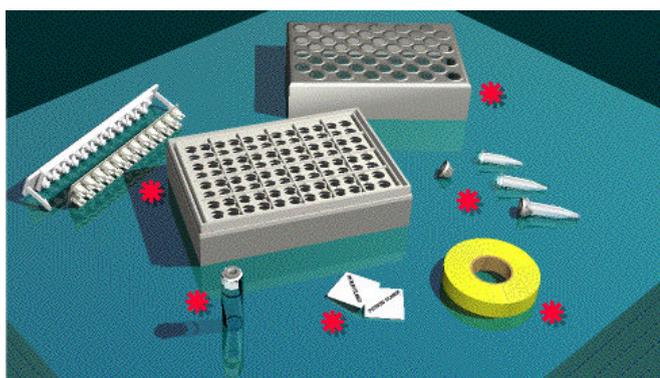


Fig. 14 -

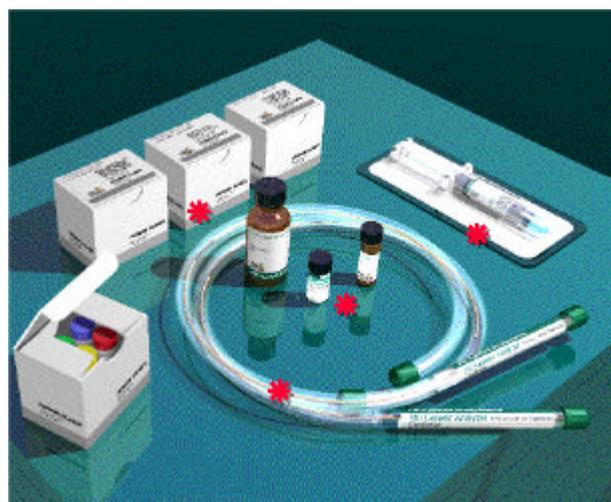


Fig. 15 -

1.6. Vantagens e limitações da sequenciação em análises clínicas

A utilização da sequenciação em análises clínicas apresenta como principal vantagem a identificação inequívoca da existência de uma mutação/polimorfismo associada a um dado perfil clínico (fenótipo). No entanto, é necessário ter presente que não é habitualmente razoável (em análises clínicas) utilizar a sequenciação nas seguintes situações:

- 1) Situações em que o diagnóstico tem carácter urgente
- 2) Situações em que a informação clínica não permite ter uma ideia precisa de qual o(s) gene(s) em que é provável encontrar a mutação
- 3) Situações em que o número de genes candidatos é muito elevado
- 4) Situações em que o(s) gene(s) candidato possui um locus de elevado tamanho e não há uma concentração habitual das mutações numa zona do gene.

Por outro lado convém ter sempre presente que o tempo de execução desta técnica, apesar de já muito reduzido em comparação com o que era possível há 10 anos, é consideravelmente superior ao necessário para um diagnóstico baseado em Real-Time-PCR.

Convém ainda ter presente, que a identificação de uma alteração genética num indivíduo com uma dada patologia, não associa de imediato essa alteração como causadora do fenótipo clínico. Em todos os casos, é necessário estabelecer a natureza da mutação (polimorfismo ou mutação; causadora de alteração proteica ou não; causadora de alteração reguladora do gene ou não; perturbadora dos mecanismos de splicing ou não), bem como estudar a sua ocorrência ou não numa população “normal”, sem os mesmos sinais clínicos. Em suma, a identificação de uma alteração genética é apenas o início, servindo para justificar a existência de estudos funcionais. Ressalva-se obviamente desta situação a descoberta de alterações que já tenham sido previamente identificadas e caracterizadas noutros doentes.

2. A sequenciação do genoma Humano

2.1. Estratégias para a sequenciação do genoma Humano

2.1.1. O consórcio Público: Estratégia do mapeamento

Quando se fala de um mapa do genoma humano, fala-se de realidades que podem ser bastante diversas. A definição original de um mapa do genoma refere-se à identificação da posição de marcadores genéticos em intervalos tão pequenos quanto possível, ao longo de todo o genoma. Este tipo de mapa tem o intuito de facilitar a identificação da zona do genoma envolvida numa dada doença em estudo. A razão para a utilização de um mapa deste tipo prende-se com o facto de por crossing-over os marcadores tenderem a recombinar entre si. No entanto, como o crossing-over entre dois loci é tão mais provável quanto mais distantes eles se encontrarem um do outro, a frequência de recombinação entre dois marcadores pode ser utilizada para inferir a sua distância. Assim, se considerarmos um fenótipo clínico como estando associado a um locus genético desconhecido, a determinação da frequência com que este hipotético locus se recombina com loci de posição conhecida permite delimitar uma zona do genoma onde existe maior probabilidade de se encontrar um gene responsável pela característica clínica em estudo. Estes mapas foram os primeiros a ser desenvolvidos, chamam-se mapas genéticos, e procuram ter uma resolução de aproximadamente 2 Mb (milhões de pares de bases) ou 2 cM (centimorgan; 1 cM corresponde à distância a que se encontram dois loci que recombinam em 1% dos cruzamentos e equivale aproximadamente a 1 Mb). Note-se que num mapa genético, a localização do marcador num cromossoma é muitas vezes desconhecida, apenas se conhecendo a posição relativa (distância) entre os vários marcadores.

O mapa genético pode ser refinado, utilizando tecnologia do DNA recombinante (técnicas de engenharia genética). Para esse efeito são produzidos mutantes contendo cromossomas fragmentados por radiação, os quais são depois mantidos por fusão com outras células. A determinação de que marcadores permanecem juntos após a fragmentação permitiu acelerar o processo de mapeamento, bem como obter maior resolução no mapa genético.

Um outro tipo de mapa, com uma resolução pretendida da ordem dos 0,1 Mb (ou 100Kb) é o mapa físico. Os mapas físicos acrescentam relativamente ao mapa genético

a informação da localização do marcador num cromossoma, e em que zona do cromossoma. Obviamente que o mapa físico de menor resolução é o mapa citogenético ou cromossómico, o qual se baseia na identificação de um padrão específico de bandas cromossómicas após um tratamento com determinados agentes. Um mapa de cDNA

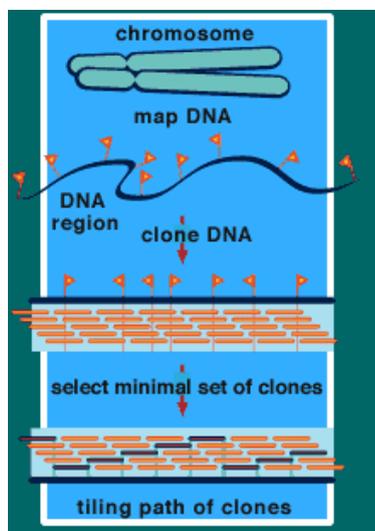


Fig. 16 -

indica a localização de exões no mapa cromossómico.

Um mapa de cosmídeos indica em que zona de que cromossoma se localiza cada um dos fragmentos genómicos clonados em cosmídeos, bem como a sua sobreposição relativa (fig.16). Um mapa de macrorestrição indica em que zonas de cada cromossoma uma dada enzima de restrição reconhece o DNA e o fragmenta (forçosamente enzimas com um baixo número de locais de corte, daí o chamar-se **macro**restrição).

Obviamente, o mapa físico de maior resolução é a sequência completa dos nucleótidos em cada cromossoma (resolução de 1bp).

A estratégia de sequenciação do consórcio público que na última década do século XX iniciou o programa de sequenciação do Genoma Humano consistiu em primeiro obter um mapa físico do genoma ao nível de cosmídeos, e de subclones destes cosmídeos. Desta forma, esperava-se que o processo de atribuição da ordem de encaixe das sequências parcelares obtidas fosse mais fácil e mais fidedigno. A estratégia consistiu então em definir um mapa físico para um conjunto cromossomas artificiais de levedura (YAC) depois de definir um conjunto mínimo destes clones, subdividir cada clone em sub-clones com sobreposições entre si, cloná-los em cromossomas artificiais de bactérias (BAC) contendo cerca de 200000 bp, e mapeá-los e assim sucessivamente utilizando cosmídeos (40000 bp) e plasmídeos (2-10000 bp) até ter um mapa de clones em plasmídeos, suficientemente pequenos para que pudessem ser sequenciados individualmente (Fig. 17).

Deste modo, o resultado da sequenciação de cada clone poderia ser utilizado para, através das sobreposições com os clones vizinhos definir uma sequência contínua que resulte do alinhamento ordenado das sequências parcelares (Fig. 18)

Os resultados revelaram que esta estratégia produz de facto sequências relativamente fáceis de trabalhar e ordenar numa única sequência, mas tem o problema de constituir

um processo moroso e que depende da estabilidade de cada fragmento em YAC's, BAC's e plasmídeos.

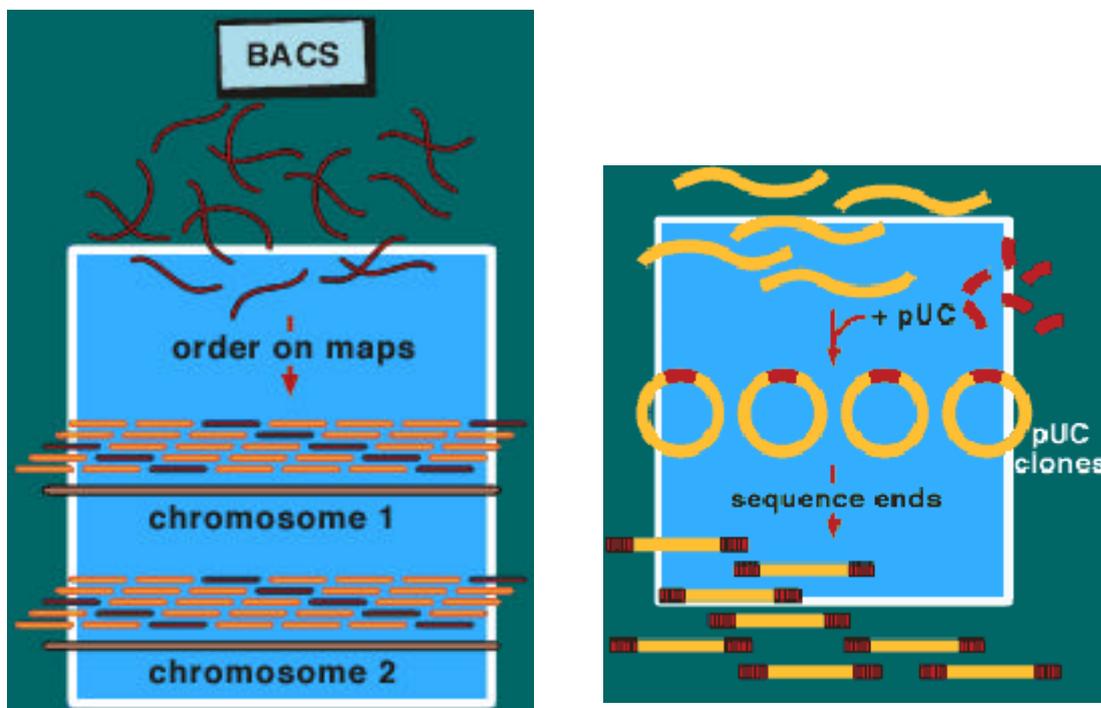


Fig. 17 -

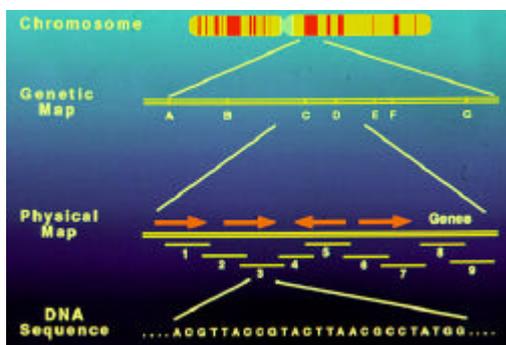


Fig.18 -

2.1.2. A celera genomics: Estratégia do “Whole genome shot-gun”

Como resposta à grande morosidade do processo de clonagem, mapeamento, e subclonagem, foi desenvolvido um processo alternativo para fragmentos de DNA, o qual consiste em produzir fragmentos aleatórios da sequência a estudar, e antes mesmo de os mapear, sequenciá-los (Fig. 19). Após a obtenção das sequências, o seu

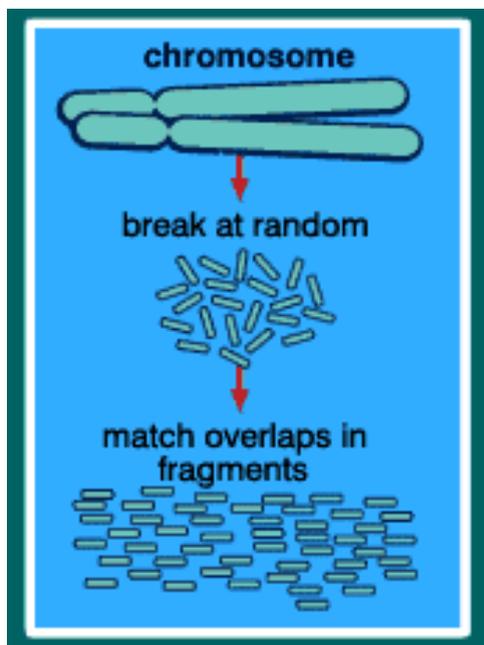


Fig. 19 -

alinhamento numa sequência contínua é feita recorrendo a massivo poder computacional.

Esta estratégia permite encurtar substancialmente o tempo necessário para obter uma sequência, mas não assegura que:

- 1) que todo o DNA alvo tenha sido sequenciado
- 2) que seja possível alinhar as sequências parcelares numa única sequência consenso
- 3) Que a sequência consenso tenha sido montada de forma correcta

Para obviar a estas dificuldades, esta estratégia envolve a sequenciação de um número muito maior de fragmentos, procurando assegurar que todo o DNA a

estudar foi sequenciado 7 a 9 vezes. Deste modo aumenta-se a redundância das sequências parcelares, aumentando a probabilidade de que todo o DNA tenha sido sequenciado, e que a montagem final seja correcta.

Esta estratégia foi inicialmente desenvolvida com vista a apressar a sequenciação do genoma, aplicando-a aos clones intermédios (cosmídeos), mas uma nova empresa (Celera Genomics) entretanto fundada com o apoio da Applied Biosystems, pretendeu aplicar esta estratégia a todo o genoma, fundando um plano paralelo de sequenciação do genoma humano, em concorrência directa com o consórcio público. Este projecto, apesar do enorme desafio informático que acarretava (o passo decisivo consistia na capacidade de formar sequências contínuas com as sequências parcelares), acabou por ter um enorme sucesso, devido aos enormes recursos informáticos que o projecto congregou (a Celera Genomics possui hoje os

supercomputadores com maior poder de cálculo para uso não militar em todo o mundo).

2.2. Estratégias para a identificação de genes na sequência do genoma Humano

A descodificação de um genoma, não fica nunca completa com a sequência de nucleótidos que compõe cada cromossoma desse genoma, do mesmo modo que nenhum texto é compreensível se as palavras que o compõem não tiverem sido identificadas e o seu significado for perceptível. Na verdade, a sequência de um genoma, na ausência da identificação dos respectivos genes e elementos reguladores (sejam eles reguladores da transcrição genética, da replicação, etc) equivale a possuir um livro escrito numa língua completamente desconhecida. A informação existe, mas não é convertível em conhecimento, nem é utilizável para outros fins que não o armazenamento. Assim, o passo seguinte à elaboração da sequência de um genoma consiste em identificar os elementos funcionais desse genoma, e as suas funções respectivas. Para tal, existem também várias estratégias complementares.

2.2.1. Recurso à identificação de genes ortólogos

Uma vez que hoje em dia dispomos de informação parcelar sobre o genoma de um alargado leque de organismos biológicos, essa informação (sequência de genes com funções e elementos reguladores conhecidos) pode ser utilizada para procurar genes homólogos no meio da sopa de letras em que consiste um genoma sequenciado. Pode-se assim encontrar genes com elevada homologia entre espécies diferentes (genes ortólogos), e daí deduzir a localização de um novo gene no genoma sequenciado, bem como uma provável organização intrão/exão, localização de promotores e enhancers.

2.2.2. Identificação de sequências conservadas entre espécies

A disponibilidade de sequências genómicas quase completas para mais do que um organismo, e em alguns casos para organismos no mesmo ramo evolutivo (ex homem-

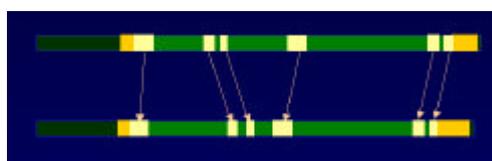


Fig. 20 -

ratinho), permite efectuar comparações entre os genomas, por blocos e encontrar zonas que ao longo da escala evolutiva foram conservadas com grande homologia. É assim

possível encontrar zonas genómicas com maior probabilidade de possuírem genes.

2.2.3. Identificação de zonas de elevada densidade com seqüências consenso para caixas promotoras/enhancers

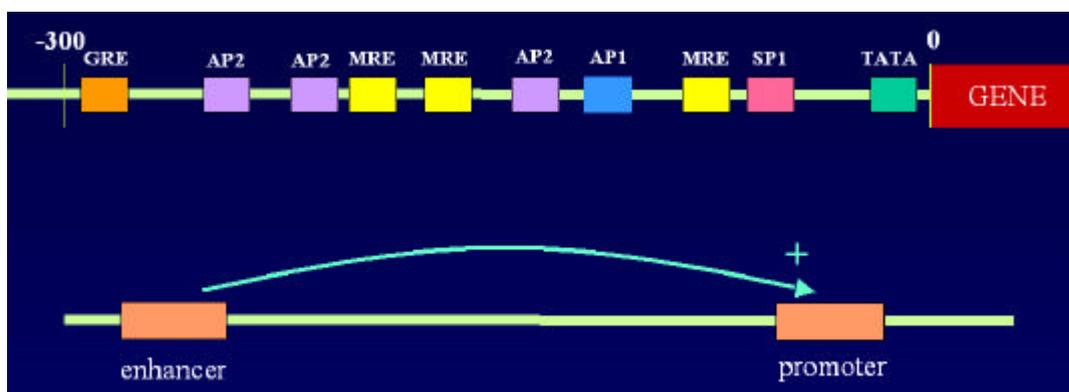


Fig. 21 -

Conhecem-se hoje seqüências consenso para vários domínios de ligação a factores de transcrição, bem como seqüências que funcionam como enhancers. Como tanto uns como outros são elementos apenas úteis na vizinhança de genes, a sua identificação permite também focar a atenção em zonas de maior probabilidade de ocorrência de genes. Em particular, zonas onde se identificam várias seqüências consenso para factores de transcrição são fortes candidatos à existência de genes.

2.2.4. O mapa de EST

Os EST são “Expressed sequence tags”. Tratam-se de pequenas seqüências derivadas de cDNA (ou de DNA complementares de mRNA), as quais podem ser sequenciadas, marcadas e hibridizadas com os clones, permitindo construir um mapa de zonas de expressão genómica que provavelmente contêm genes. De igual modo, as seqüências consenso derivadas dos EST podem ser utilizadas para procurar na seqüência total do genoma onde existem zonas para as quais foram encontrados EST, e que por isso devem conter genes.