**Inferência de Haplótipos**

**1. Introdução**

Um **haplótipo** sequência de DNA que foi herdada de um dos pais. Portanto, cada ser humano possui dois haplótipos para a maioria das regiões do genoma. Sabe-se que as bases genéticas por trás de importantes traços, como por exemplo as doenças, podem ser melhor entendidas por meio da associação entre a ocorrência de determinados haplótipos e determinados traços. Mas, antes de entendermos melhor o que significa um haplótipo e sua importância, faz-se necessário a definição de algums conceitos biológicos básicos para o melhor entendimento.

**1.1 Conceitos Básicos**

**//** DNA, cromossomo, divisão celular

**Alelo**

Um alelo é uma das muitas formas de uma sequência de DNA de um determinado gene, portanto cada gene pode ter diferentes alelos. Às vezes, diferentes sequências de DNA (alelos) podem resultar em diferentes traços (ex.: cor), mas às vezes diferentes alelos podem ter o mesmo resultado na expressão gênica.

A maioria dos organismos têm dois conjuntos de cromossomos (diplóides). Esses organismos têm uma cópia de cada gene (um alelo) em cada cromossomo. Se ambos os alelos forem iguais, eles são homozigotos, caso contrário, são classificados como heterozigotos

**Locus**

É a localização específica de um gene ou sequência de DNA em um cromossomo. Uma variante de determinada sequência de DNA em um dado locus é chamada de alelo. Em alguns parágrafos aparecerá o nome *loci*, que nada mais é do que o plural de *locus.*

**Variação**

Os processos de seleção natural, mutação, recombinação, conversão de genes, rearranjos genômicos, transferência lateral de genes, mistura de populações e deriva genética aleatória têm misturado e remisturado alelos em muitos *loci* de modo a criar a vasta variedade de genótipos encontrados em muitas populações. O desafio é encontrar aqueles genótipos que tenham associações biologicamente significativas com importantes traços de interesse.

Uma população de espécies de organismos geralmente inclui múltiplos alelos em cada locus entre vários indivíduos. Variação alélica em um locus é medida como o número de alelos (polimorfismo) presentes ou a proporção de heterozigotos na população.

**Single Nucleotide polymorphisms (SNP)**

Um polimorfismo de único nucleotídeo é uma variação em uma sequência de DNA que ocorre quando um único nucleotídeo (A, T, C ou G) difere entre os membros de determinada espécie ou entre o par de cromossomos em um indivíduo. Por exemplo, os fragmentos de sequências ACCTAGT e ACCGAGT são diferentes por causa de um único nucleotídeo. Vale ressaltar que, como nesse exemplo, a maioria dos SNPs têm 2 nucleotídeos.

SNP é o tipo de polimorfismo que ocorre com mais frequência (responsáveis por 90% de toda a variação genética humana). Variações nas sequências de DNA podem afetar o modo como as pessoas desenvolvem doenças e respondem a um agente infeccioso, drogas, vacinas e outros agentes. Isto dá aos SNPs um grande valor para a investigação biomédica e o desenvolvimento de produtos farmacêuticos ou diagnósticos médicos.

**Genótipo e Haplótipo**

O conjunto de informações genéticas de um dos dois conjuntos de cromossomos é chamado de haplótipo, enquanto o conjunto de informações da união dos dois tipos de conjuntos corresponde ao genótipo. Quanto à avaliação da contribuição genética para um determinado traço, é muito mais informativo ter dados de haplótipos que dados de genótipos.

Um haplótipo pode ser um locus, muitos loci ou um cromossomo inteiro, vai depender do número de recombinações que ocorreram em um dado conjunto de loci. Porém, a forma de representação de haplótipos mais comum é através de SNPs. Portanto, podemos considerar um haplótipo como um conjunto de SNPs.

**1.2 O Problema Biológico**

**1.3 O Problema Computacional**

**2. Algoritmos**

**2.1 Métodos Baseados em Regras e o Algoritmo de Clark**

**2.2 O Critério de Parcimônia**

**2.3 Perfect Phylogeny Haplotyping**