

## Uma nova ciência para alavancar a pesquisa agropecuária

A informação genética dos seres vivos é escrita na linguagem do DNA. Esta foi uma descoberta magnífica feita há mais de meio século e nos instiga até hoje a novas pesquisas. O desafio é decodificar a informação contida no DNA e transformar esse conhecimento em benefícios sustentáveis para a humanidade. Benefícios ligados à saúde, como diagnósticos mais apurados e precoces para prevenção de doenças, melhores tratamentos, etc, e também benefícios ligados à produção de alimentos e outros produtos de origem vegetal e animal de forma ambiental, social e economicamente sustentável. Estamos falando de estudar genomas. Isso só é possível hoje graças aos avanços no campo da biologia molecular e ao surgimento de métodos ultra avançados para acessar a informação presente nos genomas – isto é, no DNA – dos organismos.

O desafio e, por que não dizer, a beleza desse tipo de estudo é que os organismos são muito bem projetados. Cada organismo – uma planta, por exemplo – possui em seu genoma toda a informação necessária para a sua construção e replicação. Assim, toda a funcionalidade e as particularidades que diferenciam um organismo dos demais estão codificadas em seu genoma. Até as reações aos estímulos do ambiente são, de certa forma, produtos do genoma. E a variação é virtualmente infinita entre indivíduos. Mesmo dentro de uma espécie há genótipos diferentes que podem se manifestar em fenótipos específicos desejáveis ou não. Por isso é tão importante estudar os genomas.

Para simplificar, imaginemos que o genótipo é o projeto arquitetônico do organismo e o fenótipo é o resultado da execução desse projeto, considerando possíveis interferências do ambiente de produção. Isto é suficiente e podemos passar para o próximo parágrafo. Caso o leitor queira aprofundar o tema, consideramos nesse artigo que o genótipo é a constituição genética completa de um organismo e que o fenótipo são as manifestações do genótipo em características observáveis, como morfologia, desenvolvimento, fisiologia, propriedades bioquímicas, e também o modo como os fatores ambientais são recebidos e processados, modificando ou não essas características.

O custo de sequenciamento cai ano a ano e dezenas de genomas de plantas já foram sequenciados. O genoma da soja é um deles. A primeira versão, ligeiramente incompleta, foi

publicada em 2010. Sem entrar em detalhes, podemos dizer que é muito difícil sequenciar e reconstruir um genoma em sua total integridade. Isso ocorre por várias limitações dos métodos de sequenciamento – mesmo os mais avançados hoje existentes – e também por características do próprio genoma – ploidia e segmentos repetitivos, por exemplo – que dificultam o trabalho de montagem do quebra-cabeça. Isso mesmo: é um verdadeiro quebra-cabeça, visto que o DNA de um genoma precisa primeiro ser quebrado em milhões de pedacinhos para então ser lida a sequência de bases de cada um desses pedacinhos e só depois reconstruir a sequência completa do genoma.

Embora a linguagem do DNA seja a mesma nos genomas, digamos, de todos os seres vivos do planeta (os genomas são escritos num alfabeto de apenas 4 letras: as bases A, T, G e C), nota-se que há uma grande variação em complexidade e tamanho entre genomas de espécies diferentes. O genoma da soja, por exemplo, possui 978 milhões de bases; três vezes menor que o genoma do milho. Se impresso, o genoma da soja completaria 300 livros do tamanho da bíblia sagrada. Nota-se que são muitas as informações registradas no genoma de uma espécie. As mais bem estudadas são os milhares de pequenos segmentos que codificam as proteínas. Esses pequenos segmentos do genoma são os genes. E o número de genes varia enormemente entre as espécies. Para uma espécie como a soja, estima-se pouco mais de 50 mil genes codificando proteínas.

São as proteínas que fazem praticamente todo o trabalho numa célula viva. Germinação, crescimento, defesa contra invasores, fotossíntese, enchimento de grãos, enfim, tudo emerge da atividade coordenada de um conjunto de proteínas intercomunicando-se harmoniosamente. Mas nem todas as milhares de proteínas participam de cada evento. Isso acontece porque a informação contida no genoma de um organismo determina quando e onde cada proteína é provida de modo necessário e suficiente. Um organismo transgênico, por exemplo, traz em seu DNA pelo menos um gene com o código modificado, gerando proteínas diferenciadas em sua composição e/ou em seu modo de expressão.

Como a linguagem do DNA é universal, de posse de um genoma, pode-se compará-lo com outros genomas em busca de características que o diferenciam. Um ou mais genótipos podem ser estudados frente a diferentes estímulos, como a infecção por um patógeno ou uma condição

ambiental desafiante, para então detectar quais genes modulam-se no conjunto para, no final, produzir o fenótipo de interesse – uma característica de resistência a um patógeno, por exemplo. É possível também estudar centenas de genótipos distintos em conjunto, escaneando seus genomas e caracterizando as diferenças genotípicas ligadas às suas particularidades fenotípicas. O resultado é a geração massiva de marcadores moleculares úteis para seleção genômica. Nesse tipo de estudo é fundamental que os genótipos tenham sido fenotipados *a priori* para as características de interesse, nas diversas condições ambientais que se aplicam ao cultivo da espécie. Esta é a parte mais complicada porque, hoje, fenotipar apropriadamente é bem mais custoso do que sequenciar genótipos.

Com a informação gerada nesses estudos, os melhoristas podem selecionar os fenótipos de interesse com uma simples varredura no genoma, sem a necessidade de esperar pelo ciclo completo de desenvolvimento da planta. Outra vantagem importante é que a seleção genômica possibilita selecionar características poligênicas de baixa herdabilidade, ou seja, aquelas características, cujos fenótipos são muito influenciados pelo ambiente. Na seleção fenotípica tradicional isso é mais difícil.

Temos hoje, em laboratórios ao redor do mundo, máquinas sequenciando genomas de toda a vida do planeta. E os dados gerados são em grande parte de domínio público. Embora o que conseguimos captar hoje em informação biológica seja ínfima frente às montanhas de dados que estão sendo geradas, provavelmente em nenhum momento na história da humanidade houve tamanha abertura para conhecer as complexidades da vida. Em meio a essas novas demandas e necessidades, surgiu a bioinformática. A bioinformática é uma ciência que estuda como a informação é armazenada e processada nos organismos. É um novo campo de estudos que se posiciona na interseção entre biologia, ciência da computação e análise de dados. Dedicar-se ao processamento digital da informação genômica. A pesquisa agropecuária deverá beneficiar-se largamente com a aplicação dessa nova ciência.

A crescente demanda por alimentos, fibras e energia, as mudanças climáticas no planeta e os constantes surgimentos de novas pragas e doenças nos cultivos, são desafios importantes, mas podem ser enfrentados pela pesquisa agropecuária. Uma das maneiras mais sustentáveis de



enfrentá-los será explorar a vasta diversidade genética existente nos bancos de germoplasma. A boa notícia é que, com os avanços da ciência e o surgimento da bioinformática, já é possível imaginar um programa de melhoramento montando alelos ou genes desejáveis em cultivares promissoras, com performance otimizada de modo sustentável, usando uma abordagem multidisciplinar que integre vários campos do conhecimento.

**Autor:**

Eliseu Binneck - Pesquisador da Embrapa Soja