

# **Critérios de escolha de linhagens e de metodologias de sequenciamento**

Desirrê Petters-Vandresen

Módulo I – Genômica no Estudo de Microrganismos

# Por que sequenciar um genoma de um microrganismo?

- Entender melhor um sistema biológico ou um organismo modelo
- Compreender as bases genéticas e genômicas da interação entre parasitas e hospedeiros
- Compreender processos evolutivos
- Bioprospecção e aplicação agrícola, médica ou industrial

**Perguntas a serem respondidas e/ou hipóteses a serem testadas!**

# Como escolher uma linhagem para sequenciamento?

- O sequenciamento é apenas o começo de um projeto de genômica
- Embora os custos de sequenciamento tenham reduzido ao longo dos anos, ainda não é uma técnica extremamente barata e trivial (dependendo do contexto)
- Escolher uma linhagem adequadamente é importante para maximizar os resultados em termos econômicos e científicos, principalmente quando os recursos são limitados!

# Como escolher uma linhagem para sequenciamento?

- A seleção da linhagem a ser sequenciada deve ser cuidadosa. Antes de escolher, é necessário pensar em perguntas como:
  - Quais questões científicas e práticas podem ser respondidas ao sequenciar esta linhagem?
  - A espécie possui importância agrícola, médica, industrial ou ecológica?
  - Quais recursos genéticos já estão disponíveis para a espécie ou espécies próximas?
  - Existe algum conhecimento genômico ou genético sobre a espécie em questão? Há estimativas quanto ao tamanho do genoma e nível de ploidia do organismo?

# Como escolher uma linhagem para sequenciamento?

- A seleção da linhagem a ser sequenciada deve ser cuidadosa. Antes de escolher, é necessário pensar em perguntas como:
  - Existem genomas disponíveis de outras linhagens da mesma espécie? Se sim, por que ainda é necessário sequenciar outra linhagem?
  - É possível obter DNA puro e em boa qualidade? Quão difícil será obter DNA suficiente?
  - Existe algum transcriptoma disponível para a espécie ou espécies próximas?



# Como escolher uma linhagem para sequenciamento?

- A seleção da linhagem a ser sequenciada deve ser cuidadosa. Antes de escolher, é necessário pensar em perguntas como:
  - Se ainda não existe transcriptoma, será possível desenvolver o projeto sem estes dados?
  - Caso o transcriptoma seja necessário, é possível obter RNA puro, em boa qualidade e quantidade suficiente?
  - Sequenciar apenas uma linhagem será suficiente? Quantas linhagens são necessárias?

# Como escolher uma linhagem para sequenciamento?

- A seleção da linhagem a ser sequenciada deve ser cuidadosa. Antes de escolher, é necessário pensar em perguntas como:
  - Há conhecimento sobre as diferentes etapas do projeto que serão necessárias (e. g. montagem, anotação gênica, anotação funcional...)?
  - Há recursos humanos, biológicos e financeiros disponíveis para as diferentes etapas do projeto?
    - Material biológico para sequenciamento e especialistas para sua obtenção
    - Equipamentos para sequenciamento e especialistas para sequenciamento
    - Recursos computacionais para montagem e anotação e especialistas para realização destes processos

# Qual estratégia de sequenciamento devo utilizar?

- Para escolher a estratégia de sequenciamento, é preciso levar em consideração:
  - Quais perguntas serão respondidas com estes dados?
  - Qual a cobertura de sequenciamento necessária?
  - Qual a contiguidade que montagem deve apresentar?



# Quais perguntas serão respondidas com estes dados?

- As perguntas de um projeto podem ser respondidas com diferentes tipos de “montagem”:
  - Genoma completo
  - Genoma rascunho
  - Somente genes
- Montagens mais completas permitem com que mais perguntas sejam respondidas, mas aumentam os custos envolvidos e o tempo de execução do projeto

# Genoma completo

- Um genoma é considerado completo quando a qualidade da sequência atende aos "Bermuda standards":
  - Taxa de erro de nucleotídeo: 1 a cada 10.000 bases ou menos para a maior parte da sequência (99,99% de acurácia)
  - Ausência de gaps na montagem
- Muitos genomas eucarióticos apresentam regiões complexas (e. g. regiões repetitivas) que são difíceis de sequenciar e montar: dificilmente um genoma completo (de acordo com a definição) é obtido
- Termos como "*working draft*" e "*essencialmente completo*" são usados para descrever genomas de qualidade superior à genomas rascunho, mas que ainda não são genomas completos


# Genoma rascunho

- Não há um padrão absoluto para reconhecer genomas rascunho (como no caso de genoma completo)
- Em geral, um bom genoma rascunho apresenta sequências com boa qualidade, cobertura e completude, e baixo nível de fragmentação
- Não é incomum encontrar genomas rascunho altamente fragmentados, com má qualidade e incompletos

# Somente genes

- Sequenciamento de RNA mensageiro (regiões transcritas) e produção de transcriptoma
- Custo mais baixo se comparado à genoma completo
- Melhor aproveitado se já existe um genoma de referência para a espécie
- Limitações:
  - Somente genes transcritos na condição avaliada serão sequenciados
  - Ausência de informação sobre regiões e sequências repetitivas, arquitetura genômica, organização cromossômica ou coordenadas dos genes

# Quais perguntas serão respondidas com estes dados?



Genoma essencialmente completo	<ul style="list-style-type: none"><li>• Regiões repetitivas</li><li>• Elementos transponíveis</li><li>• Organização cromossômica</li><li>• Arquitetura genômica</li><li>• Sintonia e rearranjos cromossômicos</li></ul>
Genoma rascunho	<ul style="list-style-type: none"><li>• Anotação gênica</li><li>• Anotação funcional</li><li>• Variantes populacionais (em regiões não codificantes)</li></ul>
Somente genes	<ul style="list-style-type: none"><li>• Estudos funcionais de genes específicos</li><li>• Variantes populacionais (em regiões codificantes)</li></ul>



# Qual a cobertura de sequenciamento necessária?

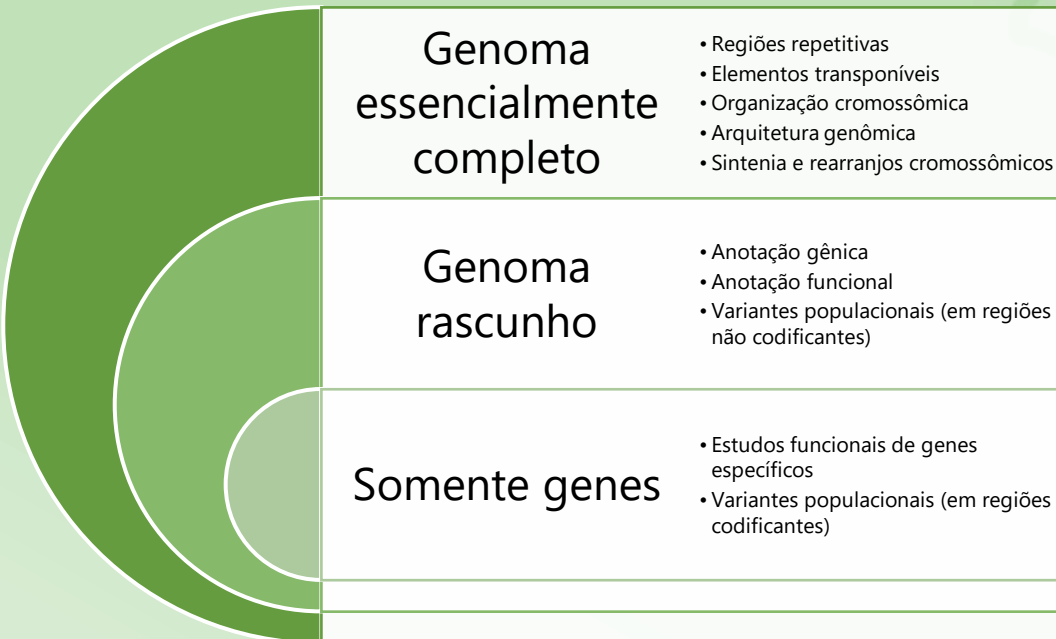
- A cobertura se refere à quantidade de vezes que o genoma foi sequenciado
- Alta cobertura: maior precisão e redução de erros nas montagens
- $Cobertura = \frac{Tamanho\ dos\ reads \times quantidade\ de\ reads}{Tamanho\ total\ do\ genoma}$

# Qual a contiguidade que montagem deve apresentar?

- A contiguidade está relacionada ao grau de fragmentação do genoma. Quanto maior a fragmentação, menor a contiguidade
- Para maior contiguidade, é necessária uma boa cobertura e uma técnica de sequenciamento com reads longos e capaz de resolver regiões repetitivas

# Qual estratégia de sequenciamento devo utilizar?

- **Somente genes e/ou genoma rascunho:** Illumina
- **Genoma essencialmente completo:** PacBio ou Nanopore, ou combinações entre dois ou três métodos



Método	Illumina	PacBio	Nanopore
Comprimento dos reads	100 – 300 pb	10 – 100 kb	Variável (até 1000 kb)
Taxa de erro	0.1%	5 – 15%*	5 – 20%*
Eficiência (bases por corrida)	200 – 600 Gb	10 – 20 Gb	5 – 10 Gb
Prós	Alta confiabilidade	Reads longos, velocidade e alta eficiência	Reads longos, velocidade e alta eficiência
Contras	Reads curtos, velocidade	Taxa de erro elevada	Taxa de erro elevada

\* Em baixa cobertura