

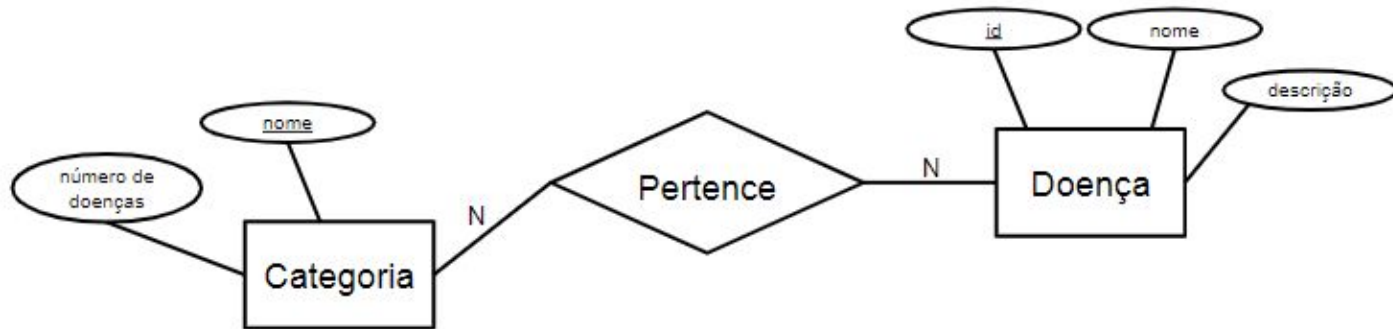
# Trabalho final

Eduardo Augusto Simão Vasconcellos	196240
Victória Pedrazzoli Ferreira	206664
Arthur Biscaino Fruch	164373
José Carlos Cieni Júnior	170859

# **Etapas 3: Análise SQL**

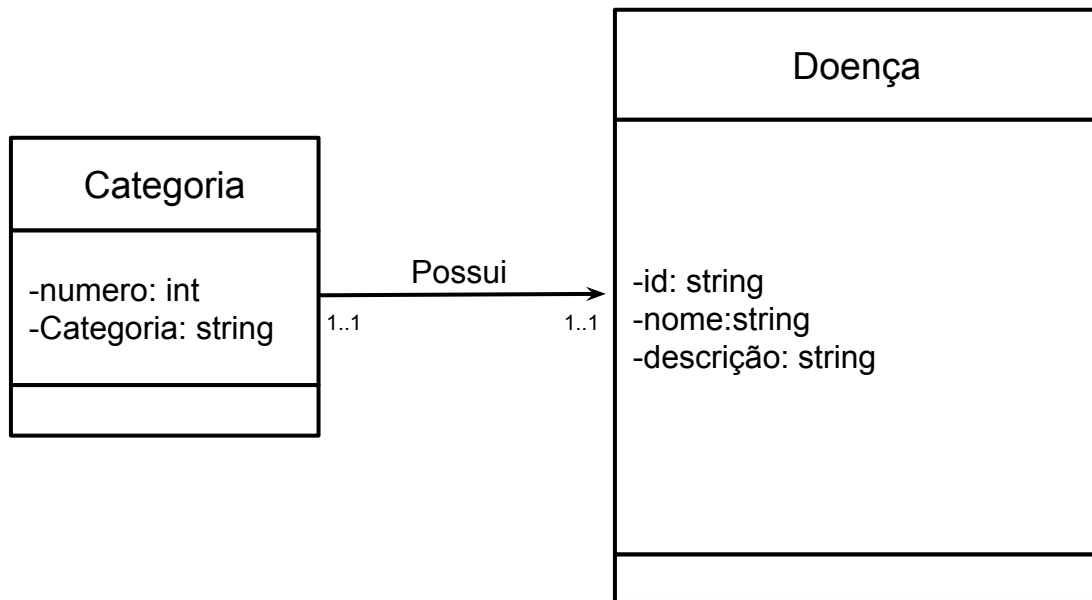
***UniProt - Human Diseases***

# Modelo conceitual - ER



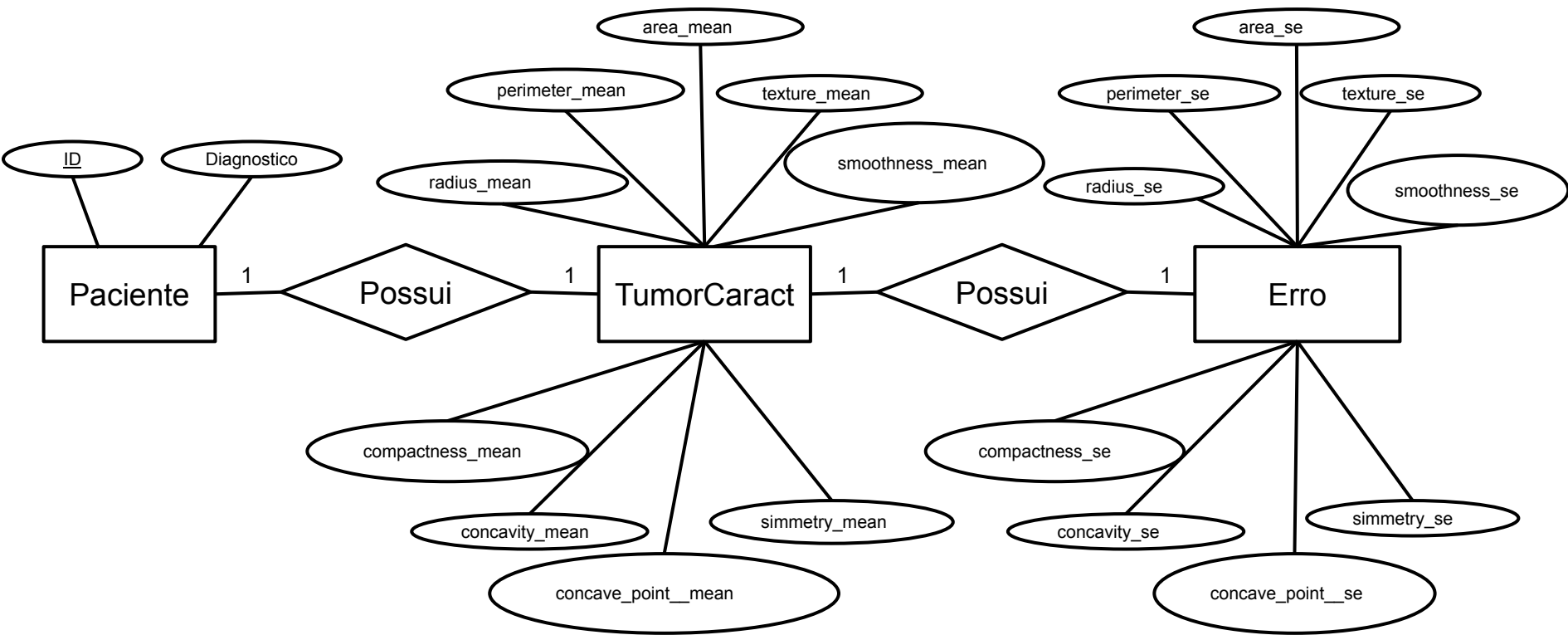


# Modelo Lógico



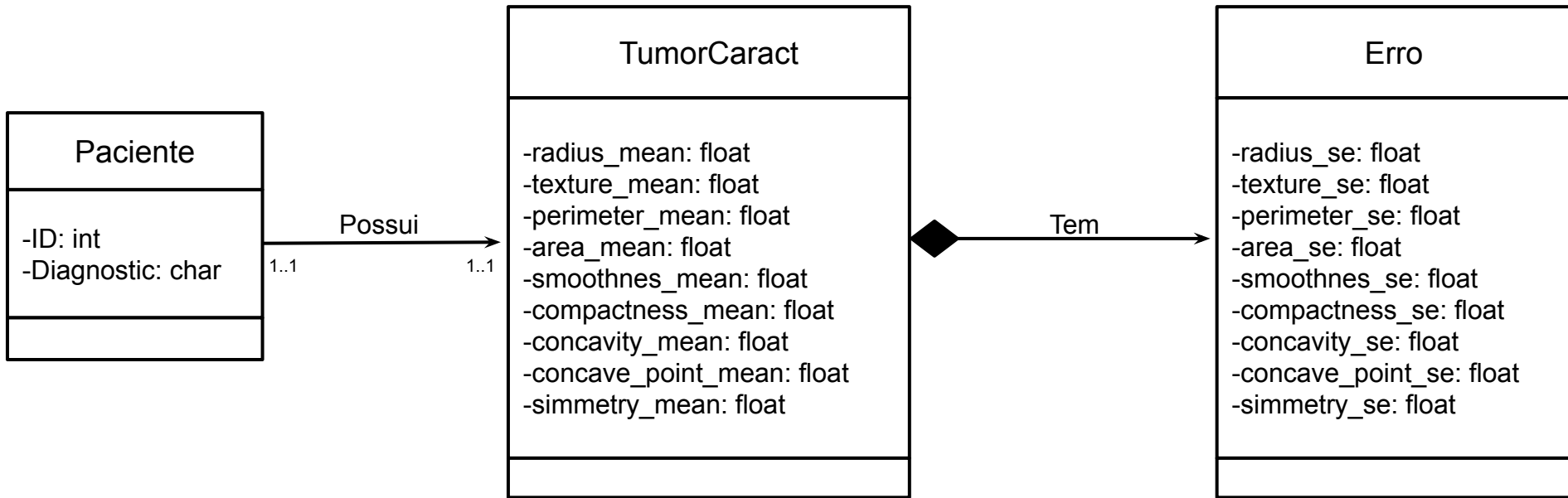


# Modelo conceitual





# Modelo Lógico- ER





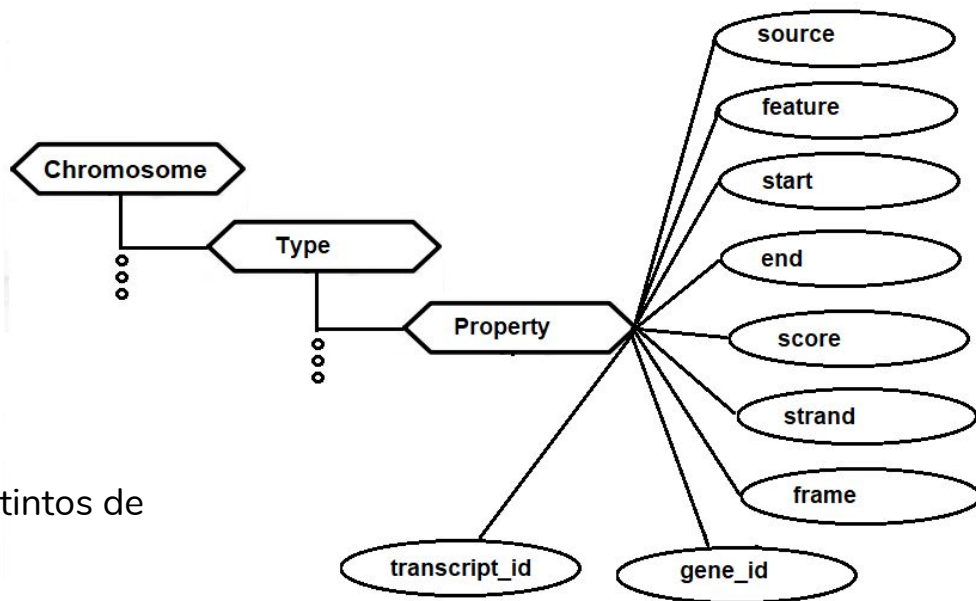
# **Etapas 4: XQuery**

***TB Database***





# Modelo



Pelo menos 6 tipos distintos de cromossomos

Incontáveis propriedades por tipo





# Propostas

- Contar o número de propriedades relacionadas a um tipo de cromossomo
- Contar o número de códons de parada(feature = 'stop\_codon')
- Retornar de todas as propriedades de um certo tipo que começam depois de uma certa posição (start >1200)
- Encontrar as propriedades de um certo tipo relativas a um gene específico
- Retornar propriedades com tamanho pequeno (start - end <10)
- Retornar todos os genes de um certo tipo que são códons de começo e parada



## Conclusão

As pesquisas propostas para essa base de dados se aproveitam da organização hierárquica do modelo. Um cromossomo pode possuir diversos tipos, que estarão em um nível hierárquico abaixo, que por sua vez podem possuir diversas características e propriedades, que estarão em níveis hierárquicos mais baixos.

Dessa forma, se procurássemos alguma condição específica dentro das propriedades de um determinado tipo de cromossomo em um modelo relacional, teríamos uma pesquisa muito complicada de ser feita, a qual envolveria diversos joins e possivelmente selects aninhados, além de com certeza ser demorada, visto que necessitaria avaliar o campo “tipo” de todos os dados. O modelo hierárquico permite a simplificação desse tipo de consulta, de forma que conseguimos, através da hierarquia, acessar diretamente um tipo de cromossomo e as propriedades de níveis inferiores desse tipo, simplificando e otimizando as consultas.

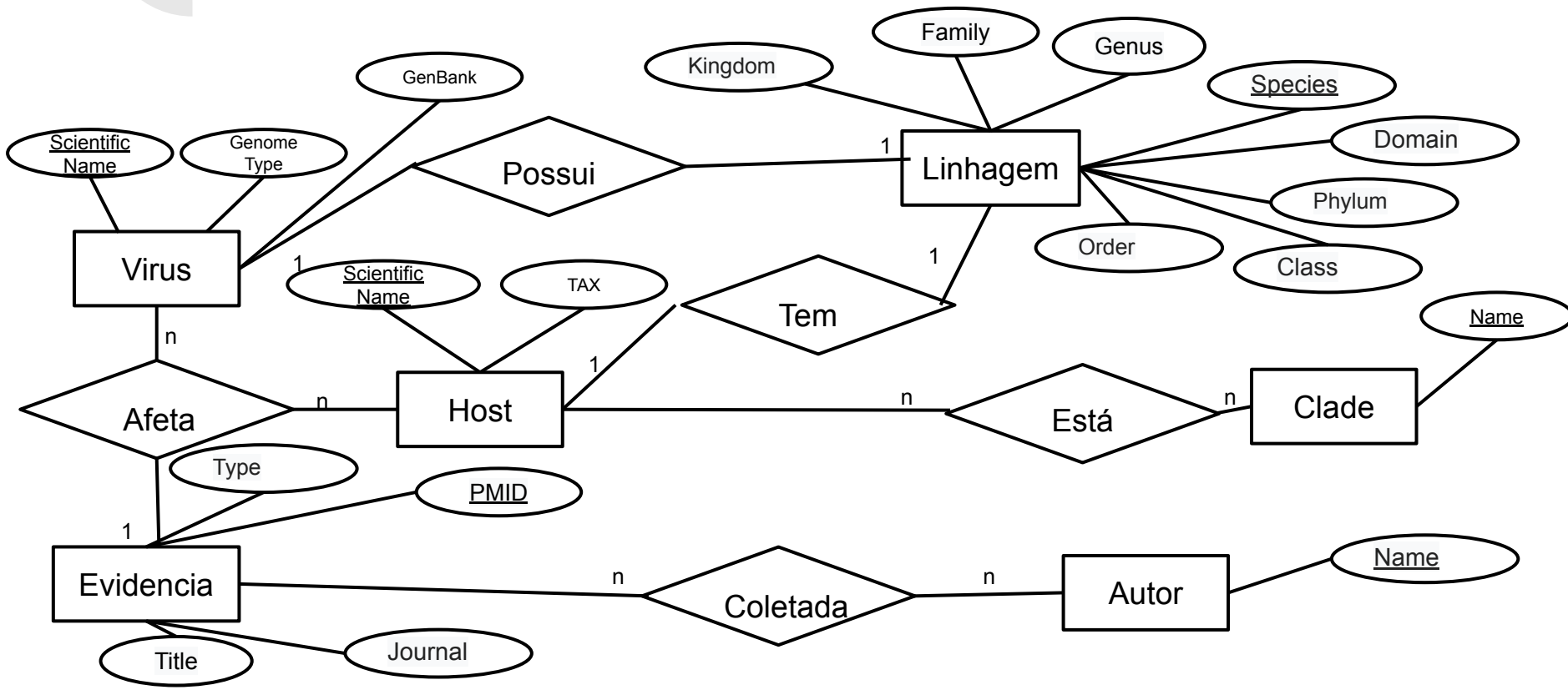
Vale a pena ressaltar que a fonte da dados é redundante, isso porque o mesmo teste em um cromossomo pode ser repetido várias vezes, e é mais fácil controlar isso em um arquivo XML

# **Etapas Final: Neo4j**

***Virus-Host DB***

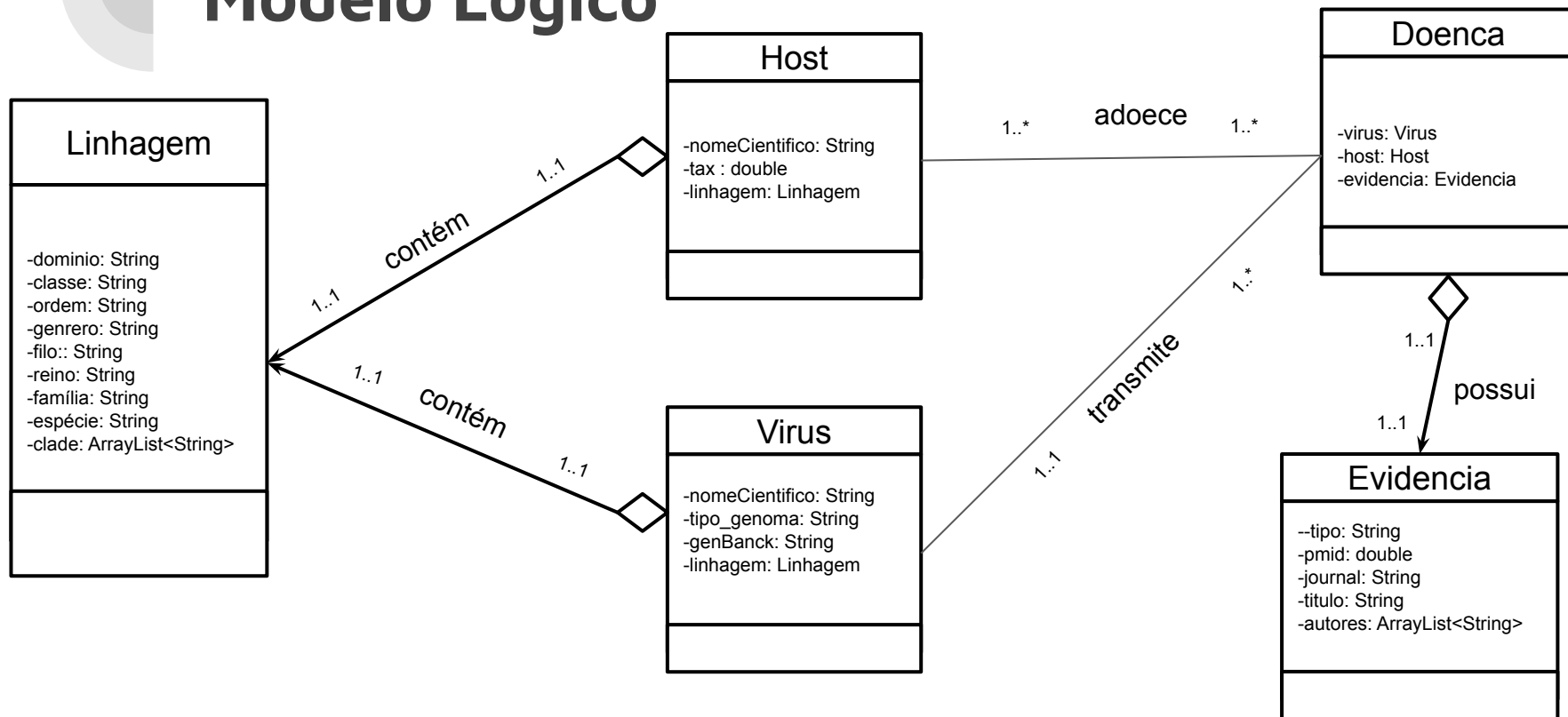


# Modelo Conceitual - ER





# Modelo Lógico





## Conclusão

O uso de uma linguagem mais ilustrativa é interessante de ser aplicado pois estamos interessados em analisar apenas algumas informações específicas da base, e dessa maneira é mais fácil observar se o grafo construído apresenta alguma característica marcante como bipartição