Biological Database Task

Breast cancer, the most common cancer in UK, is a disease of both humans and animals. However, its incidence in other mammals is very low. This suggests that the difference in cancer susceptibility in different species may be due to genetic differences. In order to explore this, I looked for the species with the most copies of genes associated with breast cancer. My SQL query identifies the genes annotated to breast cancer susceptibility by looking at the "gene_products" with a name similar to "breast cancer susceptibility". Since the genes related to breast cancer were stored in the GO database under the name of breast cancer or breast cancer types 1 and 2, I also looked at the "gene_products" with a name similar to "breast cancer type 1 susceptibility" and "breast cancer type 2 susceptibility". The query then groups the result by species and gives the number of genes in each species, ranking them from those with the most number of genes to those with the least.

The result showed human, with 3818 genes, as the species with the most copies of breast cancer-associated genes. European shrew, with 1408 genes, was ranked second. Chimpanzee, Rhesus monkey, dog and domestic cat were ranked 3rd to 6th. Although breast cancer is a disease of mammals, the genes associated with it were found in species other than mammals, such as Drosophila and ants. This shows that the genes related to cancer are responsible for the production of other proteins. It also suggests that the presence of cancer-associated genes does not necessarily result in developing cancer.

```
SELECT
 term.*,
 association.*,
 gene_product.*,
species.*,
count(*)
FROM species
INNER JOIN gene_product ON (gene_product.species_id=species.id )
INNER JOIN association ON (association.gene_product_id=gene_product.id)
INNER JOIN graph_path ON (graph_path.term2_id=association.term_id)
INNER JOIN term ON (term.id=graph_path.term1_id)

WHERE
gene_product.full_name LIKE 'breast cancer susceptibility%' OR
gene_product.full_name LIKE 'breast cancer type 1 susceptibility%' OR
gene_product.full_name LIKE 'breast cancer type 2 susceptibility%'
GROUP BY species.id
ORDER BY count(*) DESC;
```