The 1st International Workshop on Conceptual Modeling for Life Sciences (CMLS) in conjunction with the 39th International Conference on Conceptual Modeling (ER 2020)

Towards the Generation of a Species-Independent Conceptual Schema of the Genome

Alberto García S. and Juan Carlos Casamayor

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

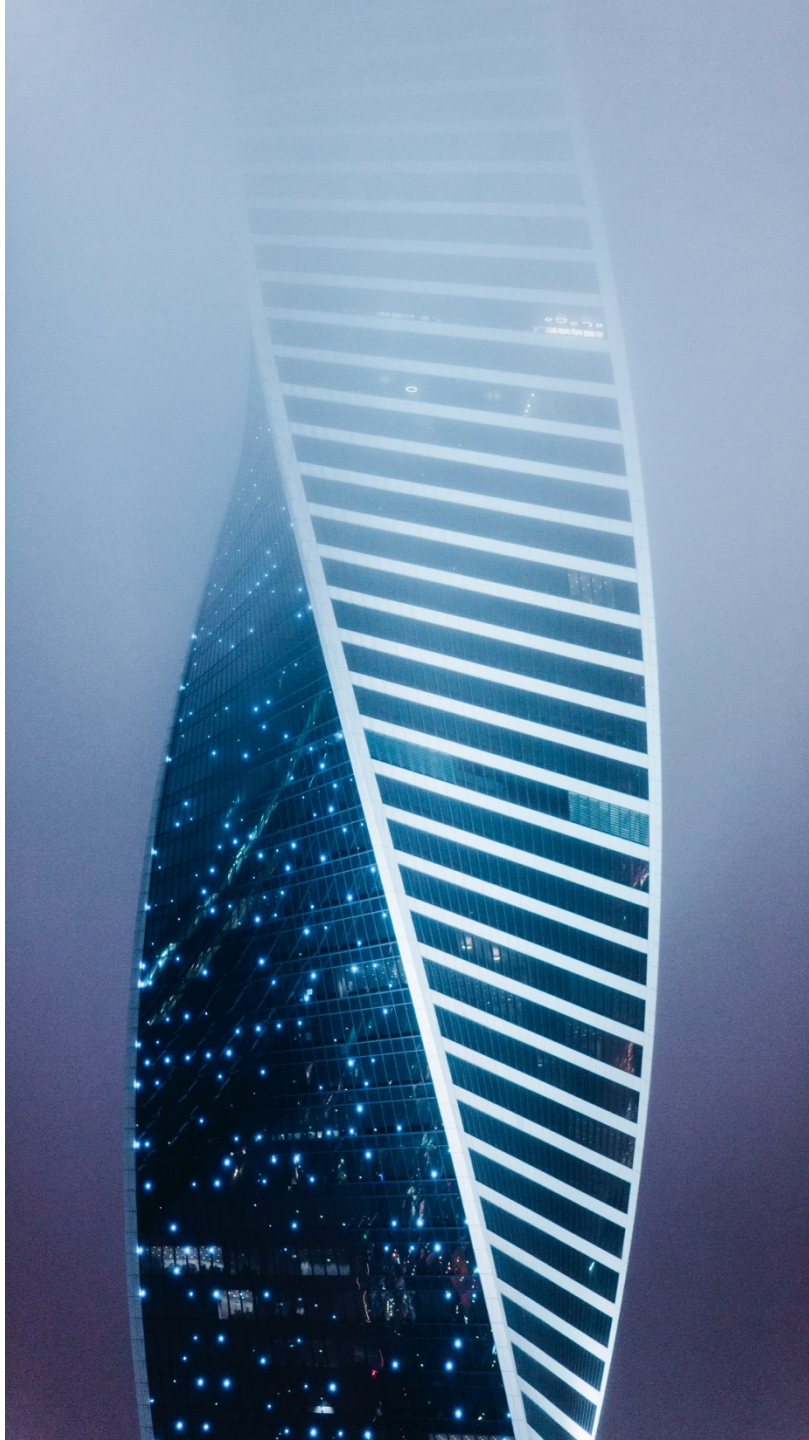
PROS
Centro de Investigación en Métodos de Producción de Software

INDEX

INTRODUCTION

What is the Genomic Domain?

- It is an interdisciplinary field of biology.
- It studies of genomes as a whole and other intragenomic interactions.
- It focuses on understanding the internal processes that drive life.
- It uses multiple approaches: structure, functionality, evolution, edition …



INTRODUCTION

! Why is Conceptual modeling **specially necessary** in the Genomic Domain?

- Relevant concepts are not clearly defined.
- The ever-changing nature of the domain, with new knowledge emerging continuously.
- Data-intensive domain where the information is exceedingly heterogeneous, dispersed, and isolated.

INTRODUCTION

What is our **previous work** in the Genomic Domain?

- Improving precision medicine. The Conceptual Schema of the Human Genome (CSHG) is used in this context.

- Collaborating with agro-food researchers. The Conceptual Schema of the Citrus Genome (CSCG) is used in this context.

INTRODUCTION

? What **lessons** have we **learned**?

- Having different, independent conceptual schemes depending on the species increases complexity, and it is time-consuming.
- Can a holistic conceptual schema, valid to work with different species, be obtained?
- Can species-specific conceptual views be instantiated from such global schema?

A HOLISTIC CONCEPTUAL SCHEMA

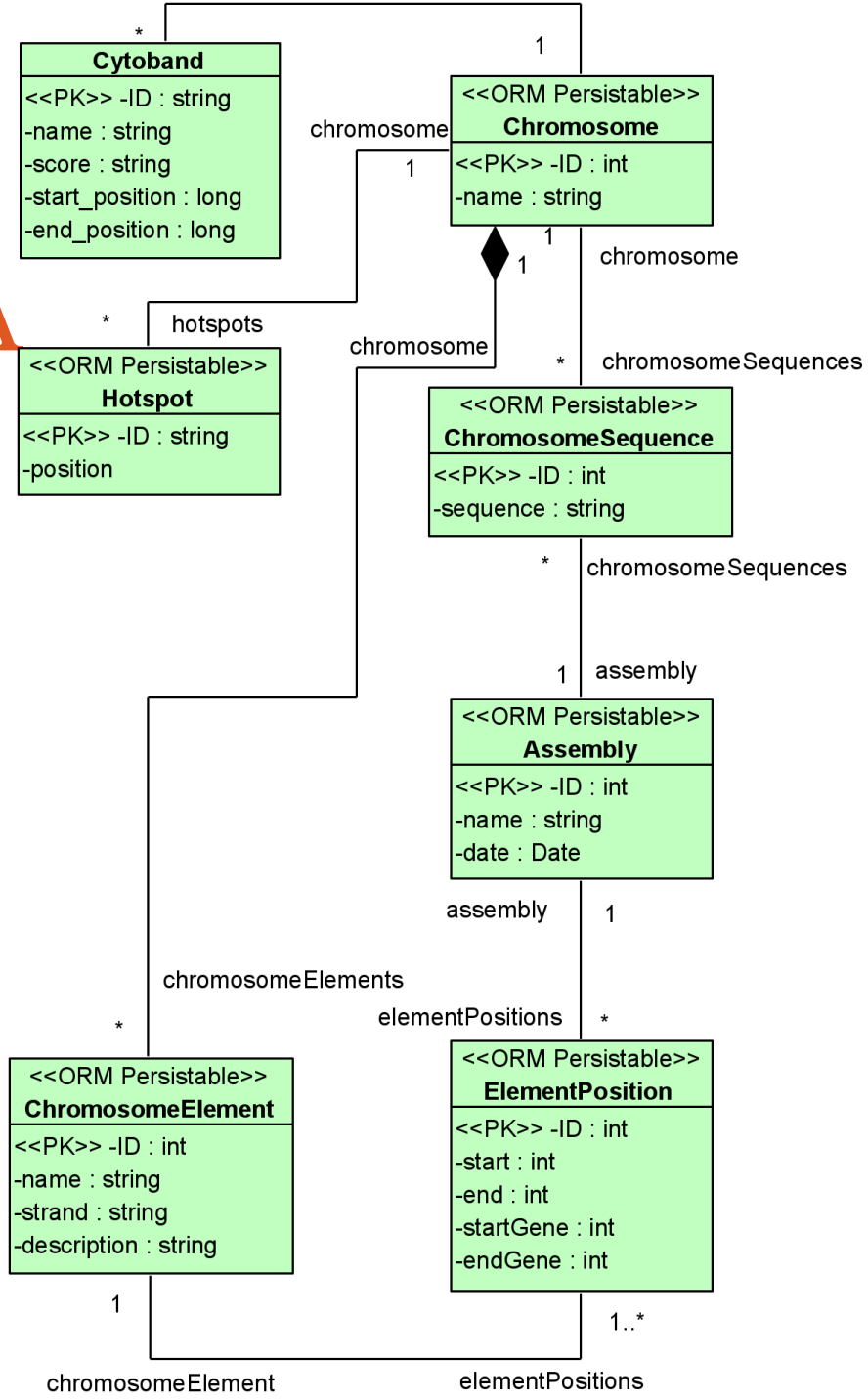
How to **achieve** our **goal**?

- We have compared both conceptual schemes to identify the needed changes to provide an expanded CS.
- Conceptual views from this CS are instantiated for both use cases.
- The CSHG has been used as a basis.
- The CSCG has been used as a complement to expand the CSHG.

A HOLISTIC CONCEPTUAL SCHEMA

What are our **findings**?

1. **Structural view:**
   - both schemes follow a similar approach.
   - The CSCG is more technologically oriented.
   - The CSCG offers a mechanism to investigate evolutionary relationships among citrus.

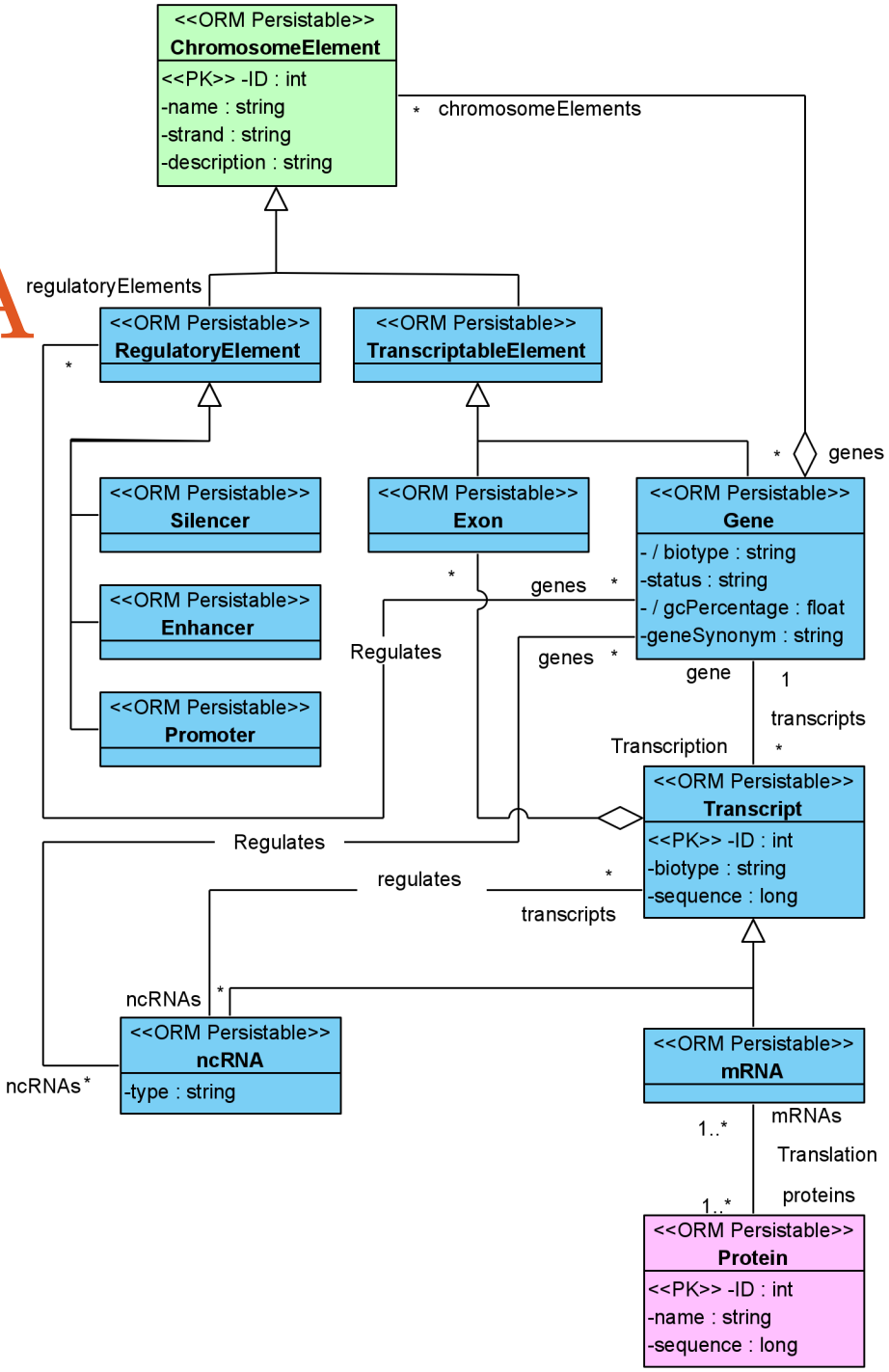


A HOLISTIC CONCEPTUAL SCHEMA

What are our **findings**?

2. **Transcription view:**

- The CSCG models the transcription process in a more simple way.
- Non-coding variants are more relevant in the CSCG.
- Protein functionality is studied in more detail in the CSCG.
- The structure of mRNAs is modeled in the CSCG

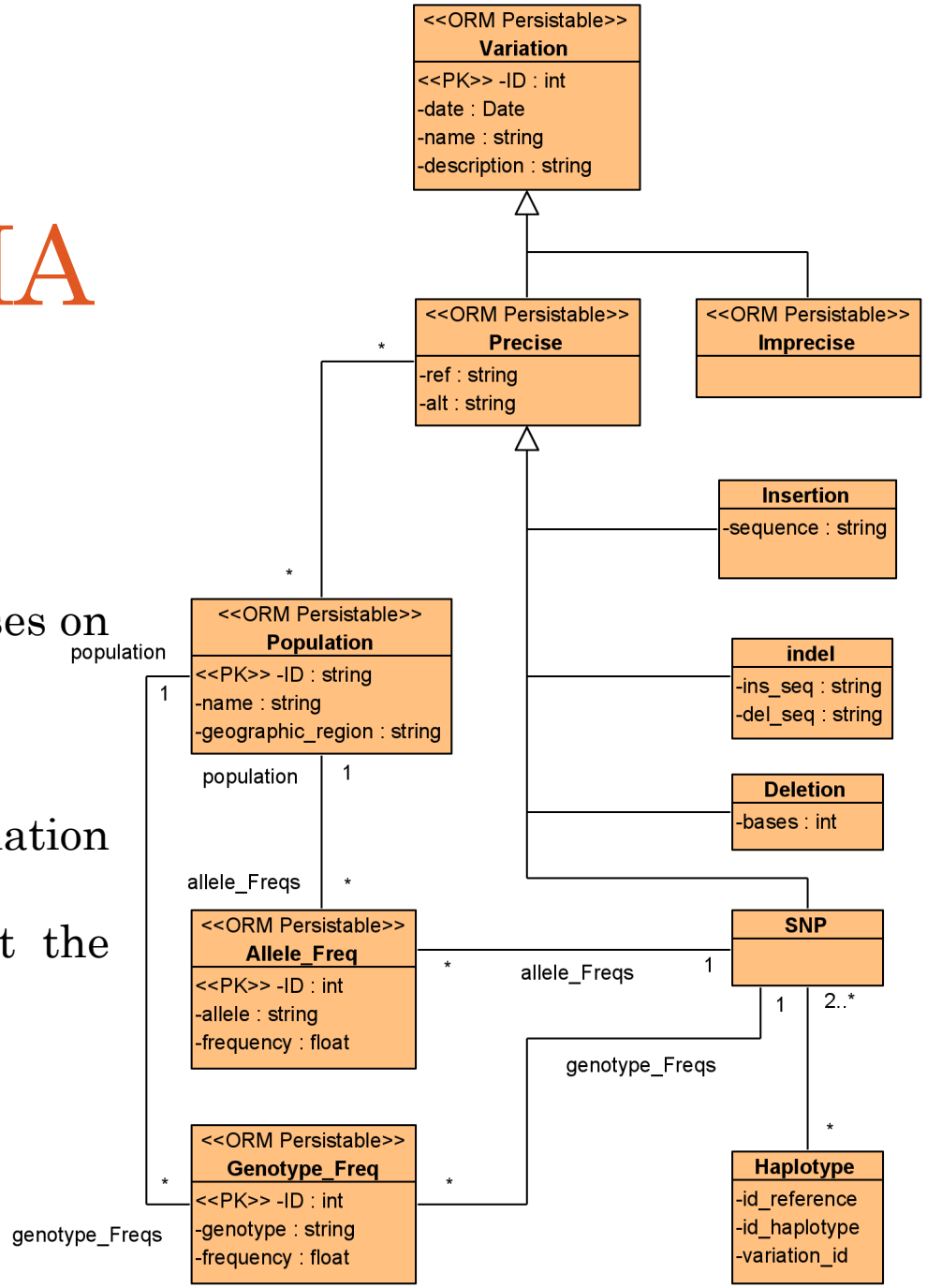


A HOLISTIC CONCEPTUAL SCHEMA

What are our **findings**?

3. **Variation view:**

- the CSHG focuses on SNPs while the CSCG focuses on SNPs and INDELs.
- The CSHG focuses on variations at the population level while the CSCG focuses on variations at the individual level.



CONCLUSIONS

- ✓ The specific particularities of genomic use cases result in too **diverse data analytics**.
- ✓ One of these particularities is the working species, which results in the existence of **multiple conceptual schemes**.
- ✓ It is feasible to have a **holistic CS** that is valid for different species.
- ✓ Specific **views** can be generated from such a holistic schema, **increasing modeling efficiency**, and **knowledge generation**.

The 1st International Workshop on Conceptual Modeling for Life Sciences (CMLS) in conjunction with the 39th International Conference on Conceptual Modeling (ER 2020)

Towards the Generation of a Species-Independent Conceptual Schema of the Genome

Alberto García S.

Phd Student

Email: **algarsi3@pros.upv.es**

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

PROS

Centro de Investigación en Métodos de Producción de Software

Alberto García S. and Juan Carlos Casamayor