Telecon 1

Visual Analytics for Plant Pangenomes (VAPP)

Astrid van den Brandt, Michel Westenberg & Sandra Smit

30 November 2020

EINDHOVEN UNIVERSITY OF TECHNOLOGY

WAGENINGEN UNIVERSITY & RESEARCH

Agenda

1. **Project Aim**
2. **Approach**
3. **Timeline**
4. **Preliminary results**
5. **Demonstration**
6. **PanTools update (Sandra Smit)**

If you were a crop, which one would you be?

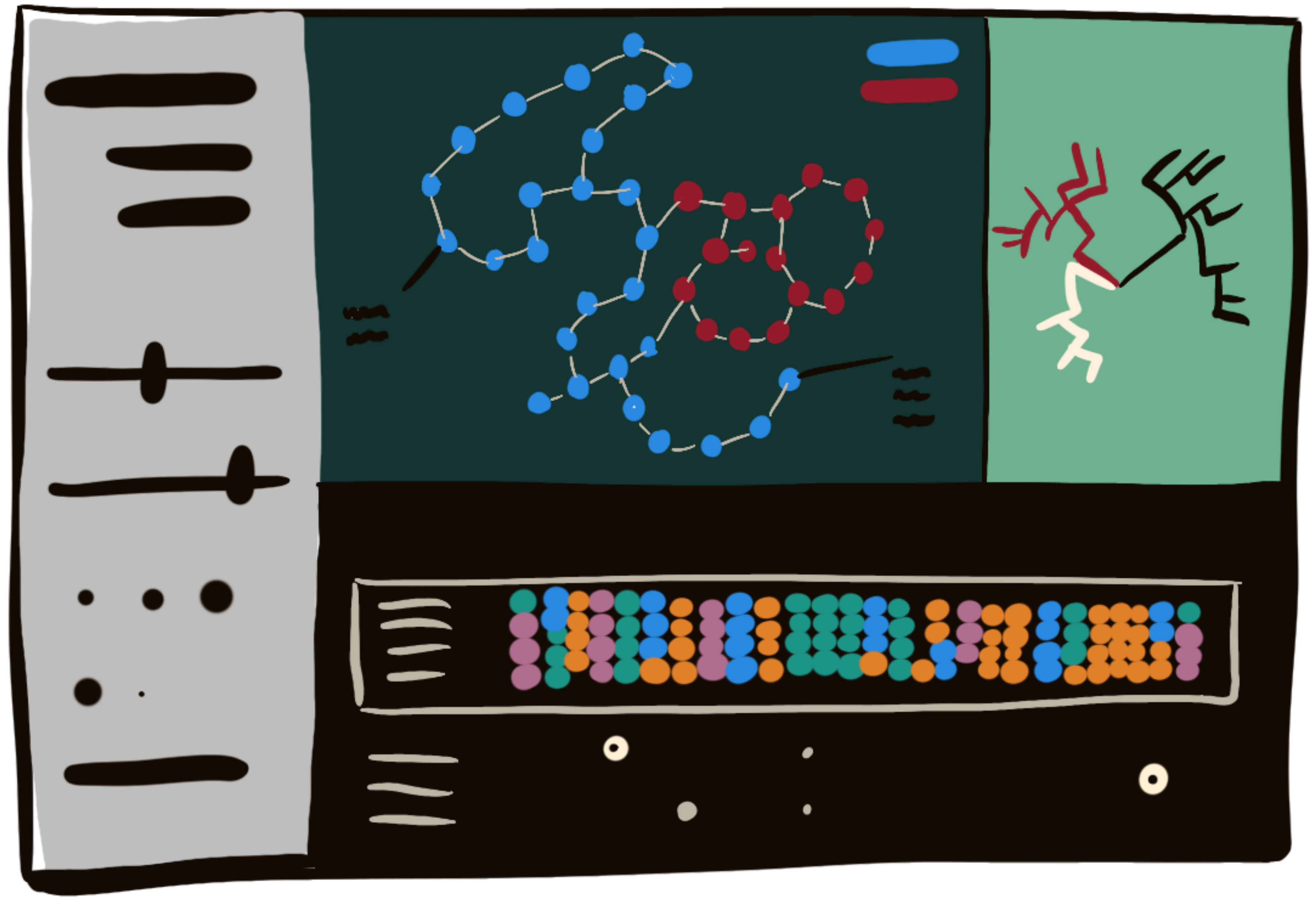
<please go to www.slido.com and enter code #92505>

Project Aim

To design a visual analytics system **that supports plant scientists** in using and assessing genetic variation in their crop data

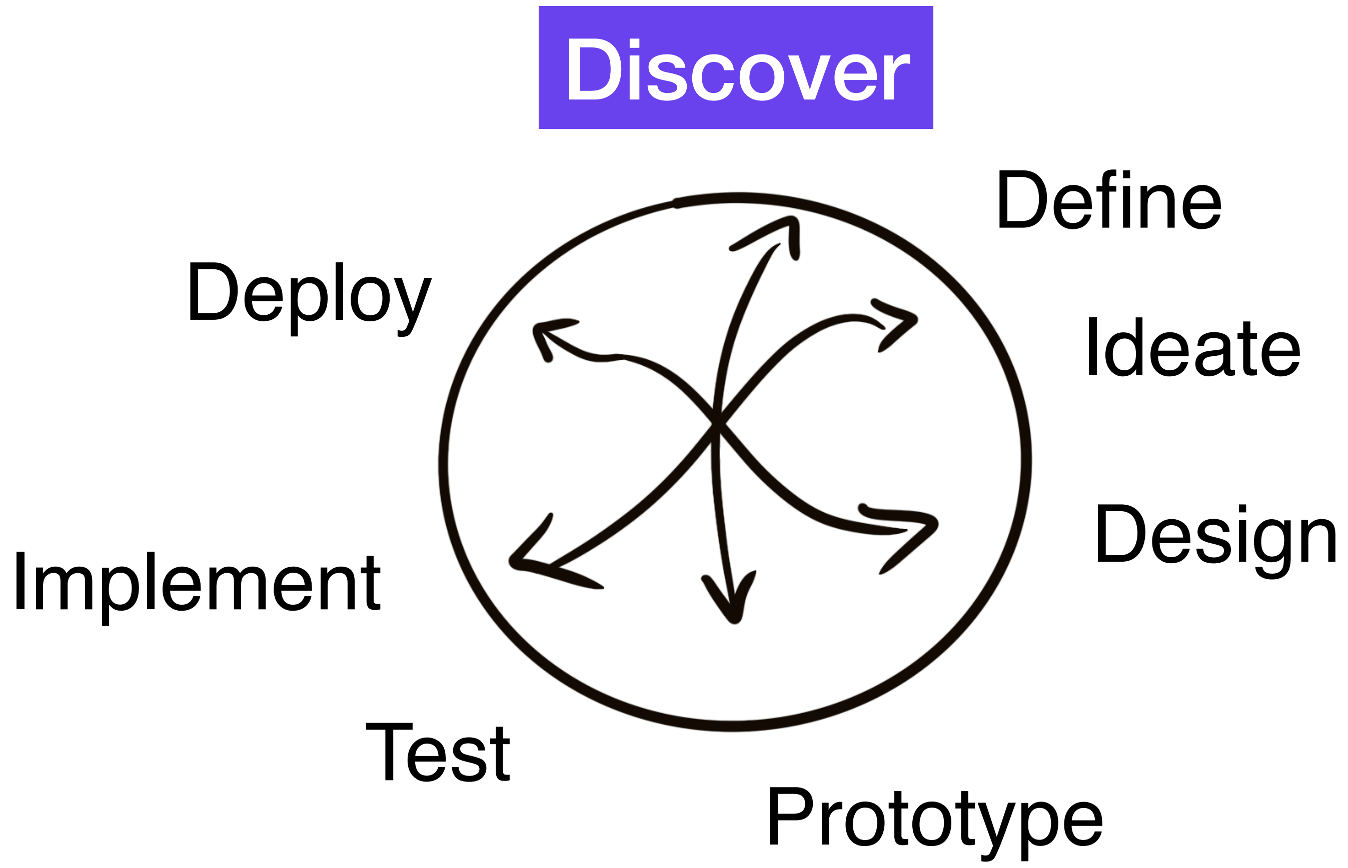
Pangenomes approach

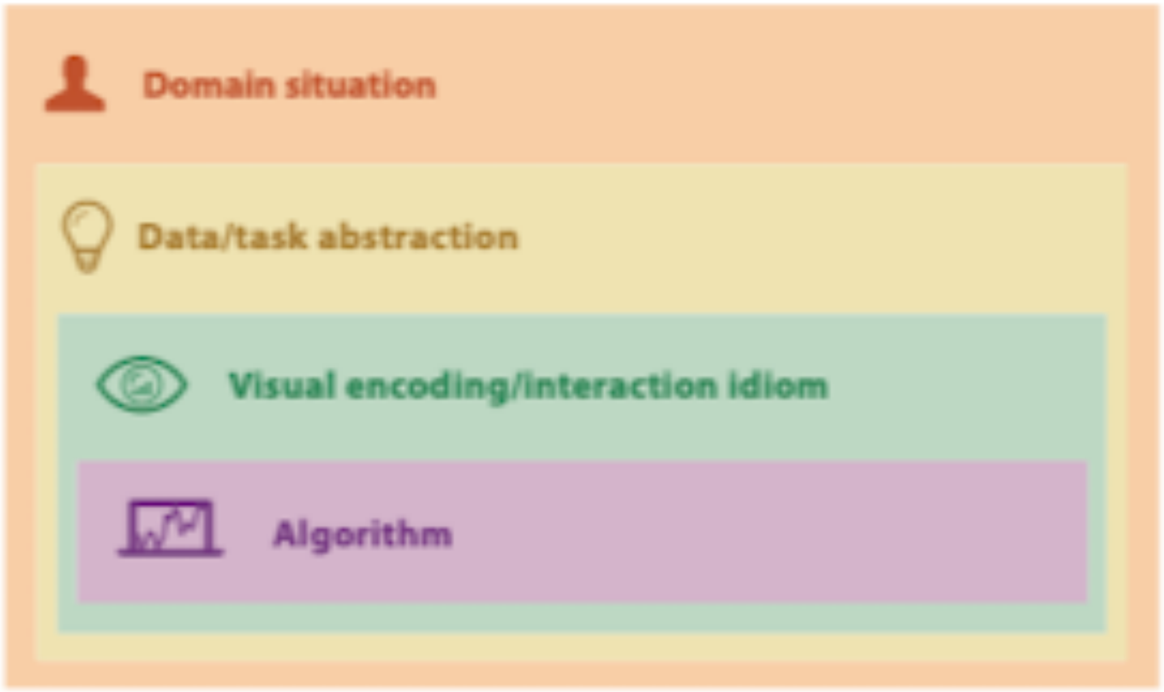
PanTools



Approach

Visualization Design Process





Munzner (2014). Four nested levels of vis design.

User-centered Design

User Studies

Task Analysis

Domain situation

Questions & Needs → Qualitative assessment

Interviews & shadowing

Task abstraction

Approach

Task Analysis

Domain situation

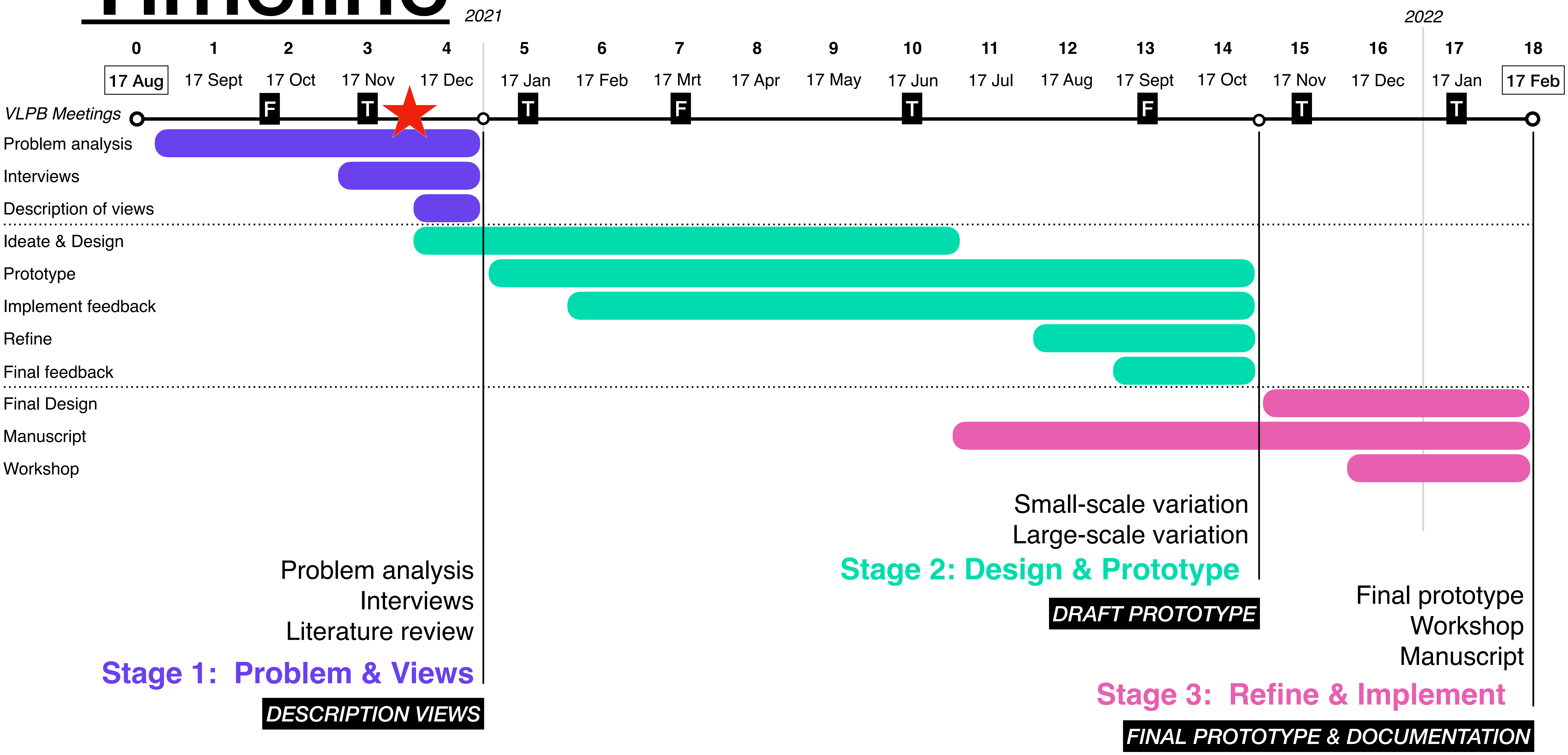
Questions & Needs → Qualitative assessment

Interview and shadowing: 3 rounds and 3 steps

1. **Introduction meeting**: discussing preliminaries and to scope the questions for the online interview
2. **Online interview**: semi-structured session of approximately 60 minutes duration
3. **Shadowing session**: for understanding the analysis context and environment

Timeline

We are here!



Problem analysis
Interviews
Literature review

Stage 1: Problem & Views

DESCRIPTION VIEWS

Small-scale variation
Large-scale variation

Stage 2: Design & Prototype

DRAFT PROTOTYPE

Final prototype
Workshop
Manuscript

Stage 3: Refine & Implement

FINAL PROTOTYPE & DOCUMENTATION

Preliminary results

1. Interview and shadowing sessions

Round 1	**Introduction**	**Interview**	**Shadowing**
HZPC	✅	1-12-2020	TBD
Genetwister	TBD	TBD	TBD

Preliminary results

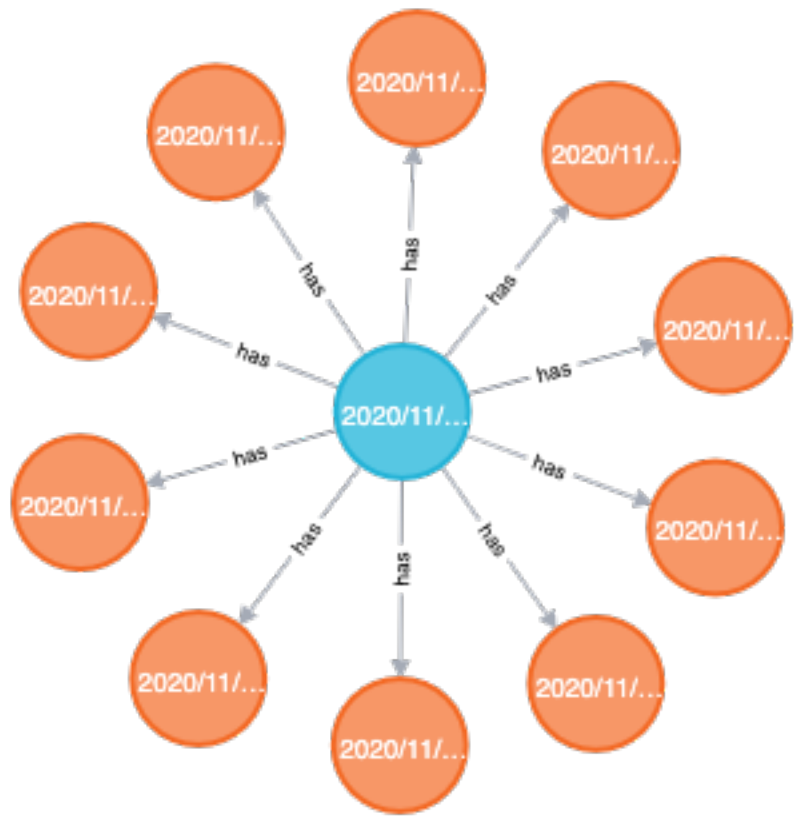
2. Development prototype:

2.1 Yeast data set (from NCBI genome database)

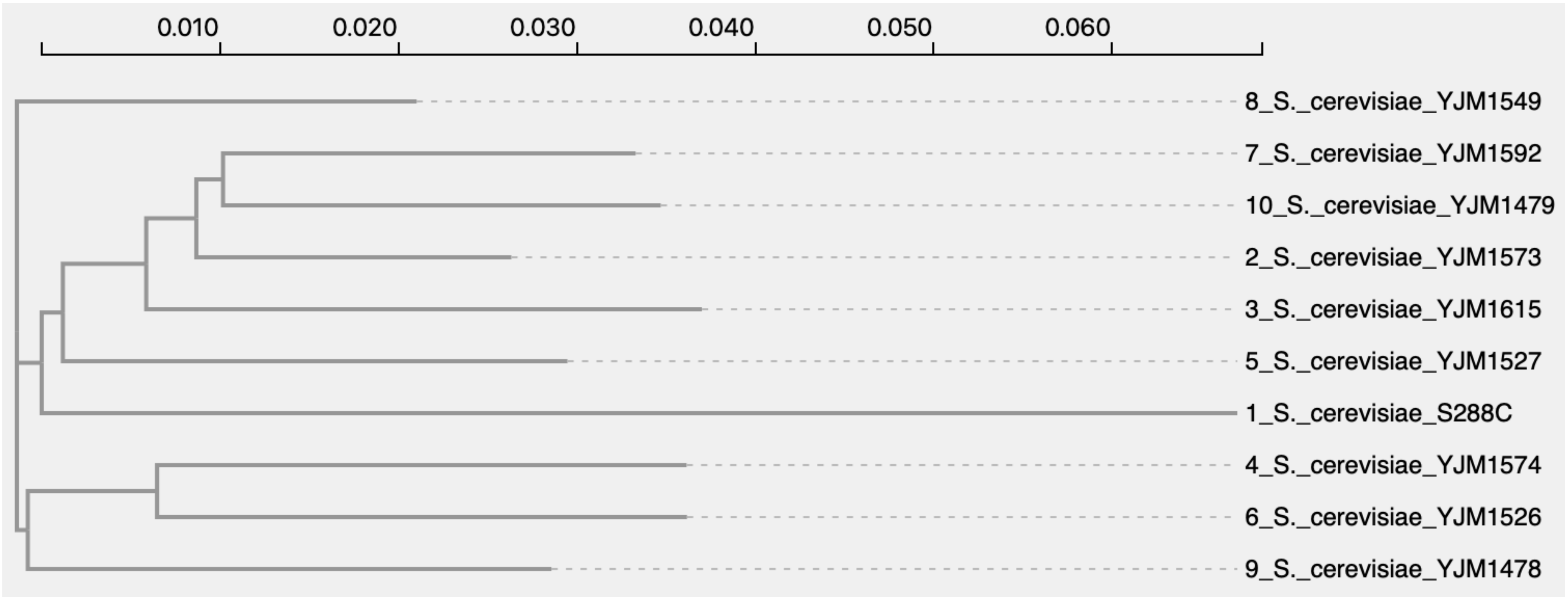
40 strains in 4 levels:

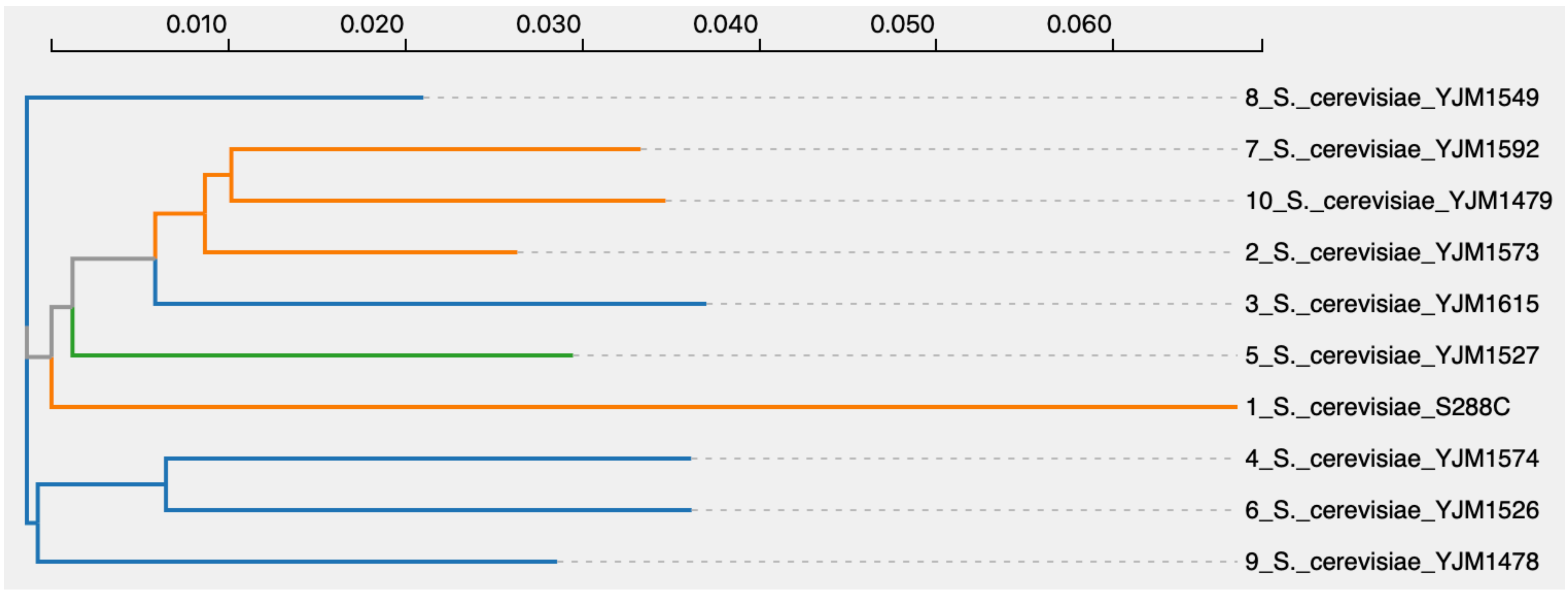
- 10x identical strain
- 10x same species
- 10x same family
- 10x same genus

Species pangenome



2.2 Building and showing phylogenetic trees



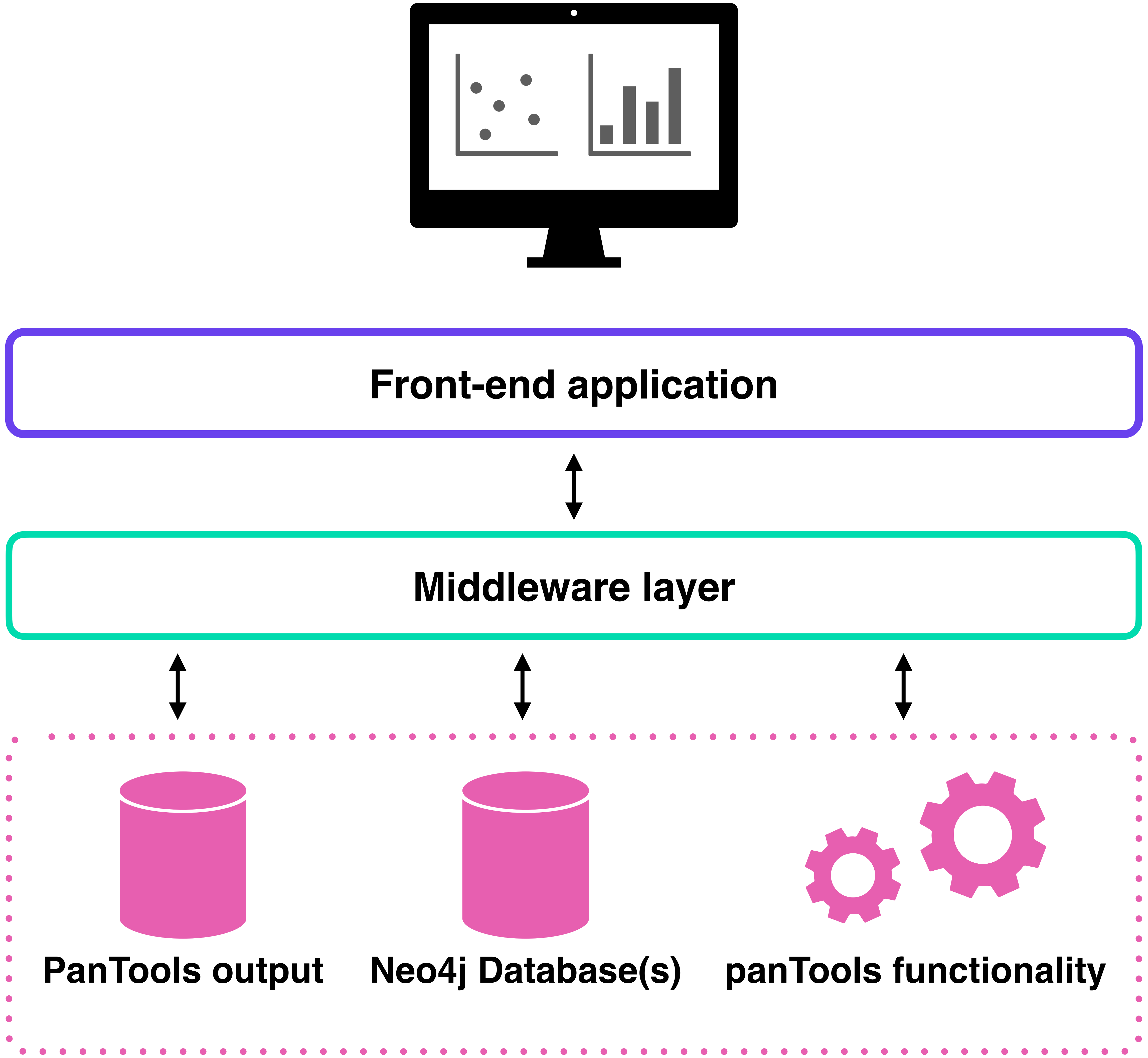


Preliminary results

2. Development prototype:

2.3 Exploring the system architecture

- ✅ Connect panTools output
- ✅ Connect Neo4j database
- ✅ Integrate database info
- ✅ Show phylogenetic trees
- 🤔 Connect PanTools functionality
- 🤔 Configure background task queue



Front-end application

Middleware layer

PanTools output

Neo4j Database(s)

panTools functionality

Questions?