

The LAILAPS plant science search engine: technology and implementation details



Jinbo Chen, Daniel Arend, Christian Colmsee, Uwe Scholz, Matthias Lange

Research Group Bioinformatics and Information Technology (BIT)
Leibniz Institute of Plant Genetics and Crop Plant Research (IPK),
OT Gatersleben, Corrensstr. 3, 06466 Stadt Seeland, Germany

Contact: chenj@ipk-gatersleben.de
Website: http://lailaps.ipk-gatersleben.de

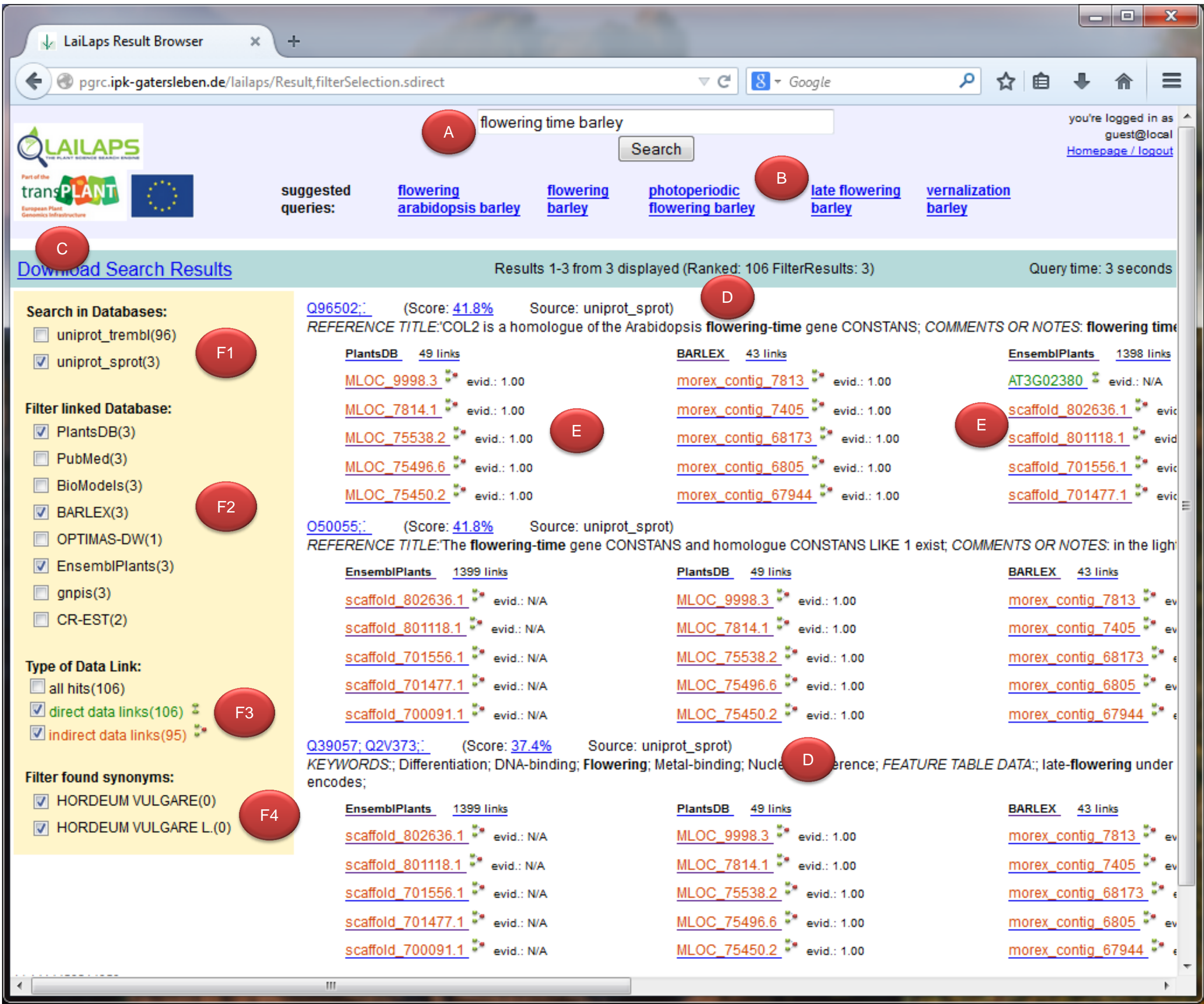
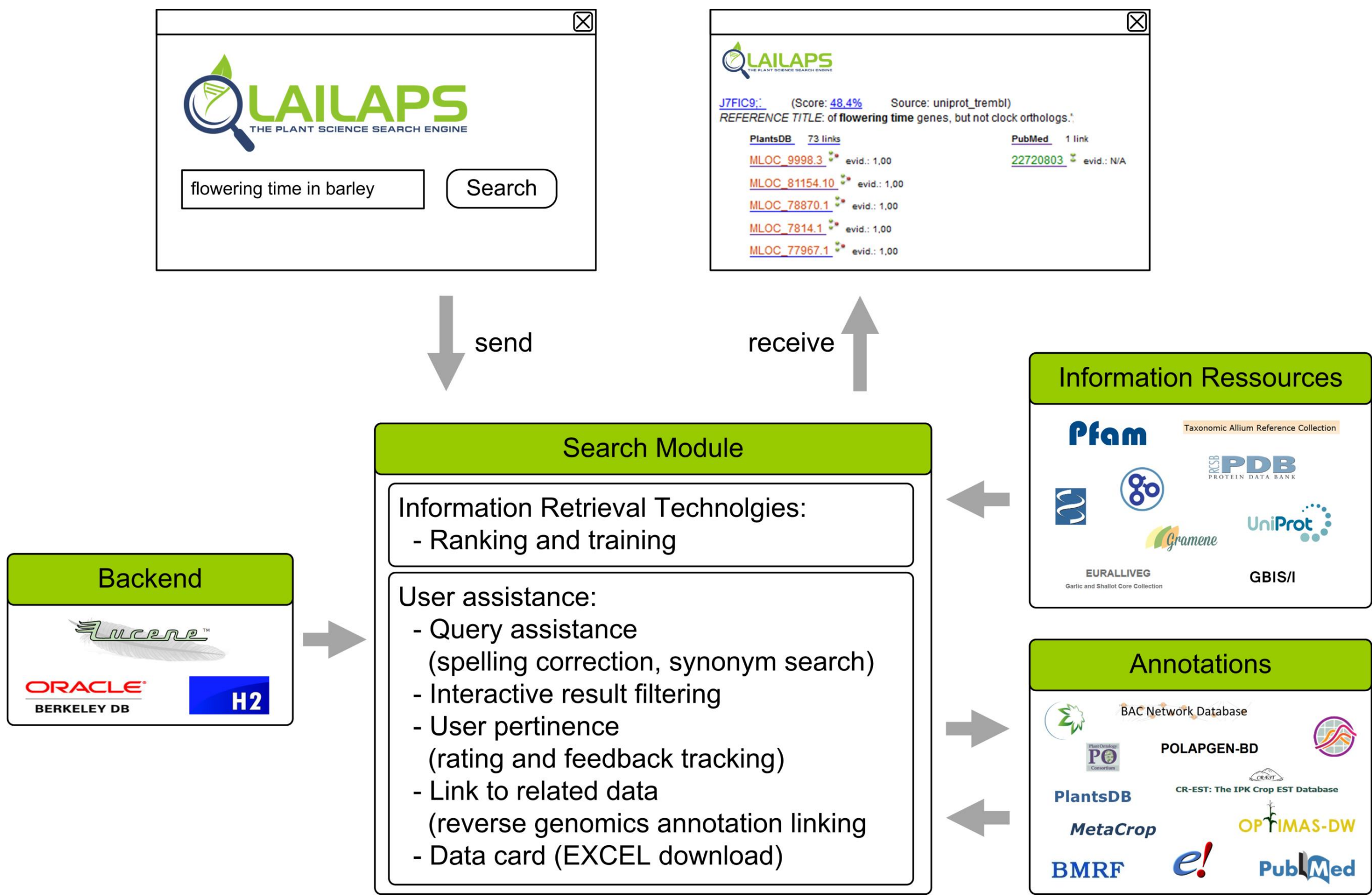
INTRODUCTION

Research data is scattered heterogeneously in distributed databases. Its dedicated use as homogeneous knowledge resource requires information retrieval technology to find and extract genome to phenome associations. Here, we present LAILAPS^{1,2} as a comprehensive information retrieval system to mine plant genomic data in the context of phenotypic attributes. The system has been developed in frame of national and European research projects for information retrieval in plant genomics with focus to support forward genetics studies of phenotype-genotype associations. The concept is to link knowledge of gene functions and phenotypes to genomics data. This enables effective information extraction and homogenizes databases in a non-invasive way.

Integrated Search In Genome Knowledge Resources

LAILAPS features:

AI-based relevance ranking, semantic query suggestion, linking phenotypes to genotypes, integrated genome databases, evidence and relevance result sorting, customizable in-house installation



Legend to frontend screenshot:

The search box (A) features an interactive spelling correction. A suggested list of semantically related keywords (B) may be used for an interactive query refinement. All queried and relevant ranked functional descriptions are shown as text excerpt (D). Corresponding links to relevant data (i.e. by genome annotation) (E), are provided. The links can be direct (green) or indirect (red). Filter enable to choose preferred knowledge sources (F1), linked fact databases (F2), direct or indirect references (F3) or synonyms (F4). The search results can be downloaded as a Microsoft Excel sheet (C).

EXAMPLE

LAILAPS instance to explore wheat and barley genome (as of June 2015)

indexed knowledge resources

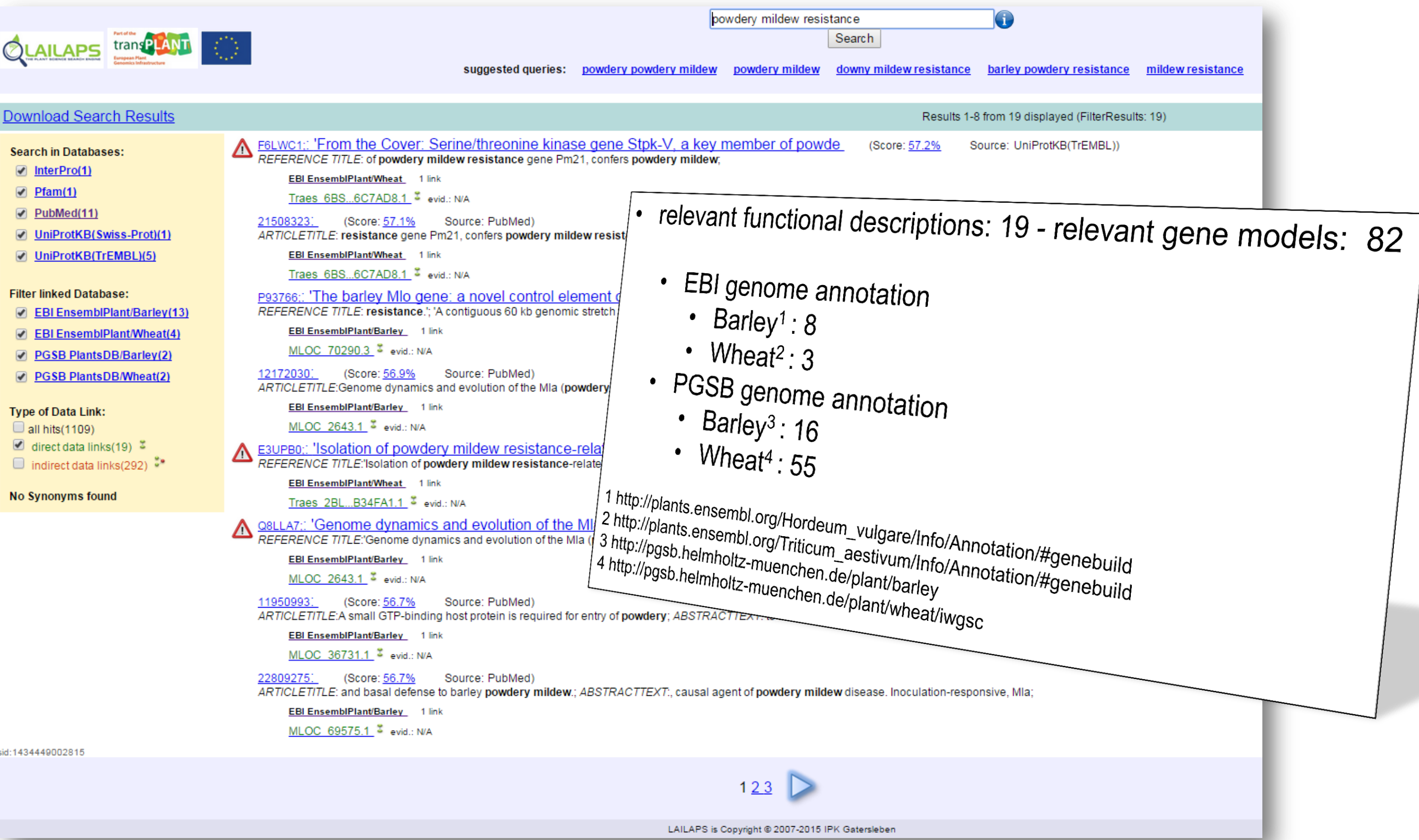
Gene Ontology Consortium 43,131
InterPro 27,460
Pfam 14,831
PubMed 12,602,333
UniProtKB(Swiss-Prot) 548,586
UniProtKB(TrEMBL) 48,744,721

loaded genome annotation

EBI EnsemblPlant/Barley 15,524
EBI EnsemblPlant/Wheat 157,533
PGSB PlantsDB/Barley 131,928
PGSB PlantsDB/Wheat 454,153



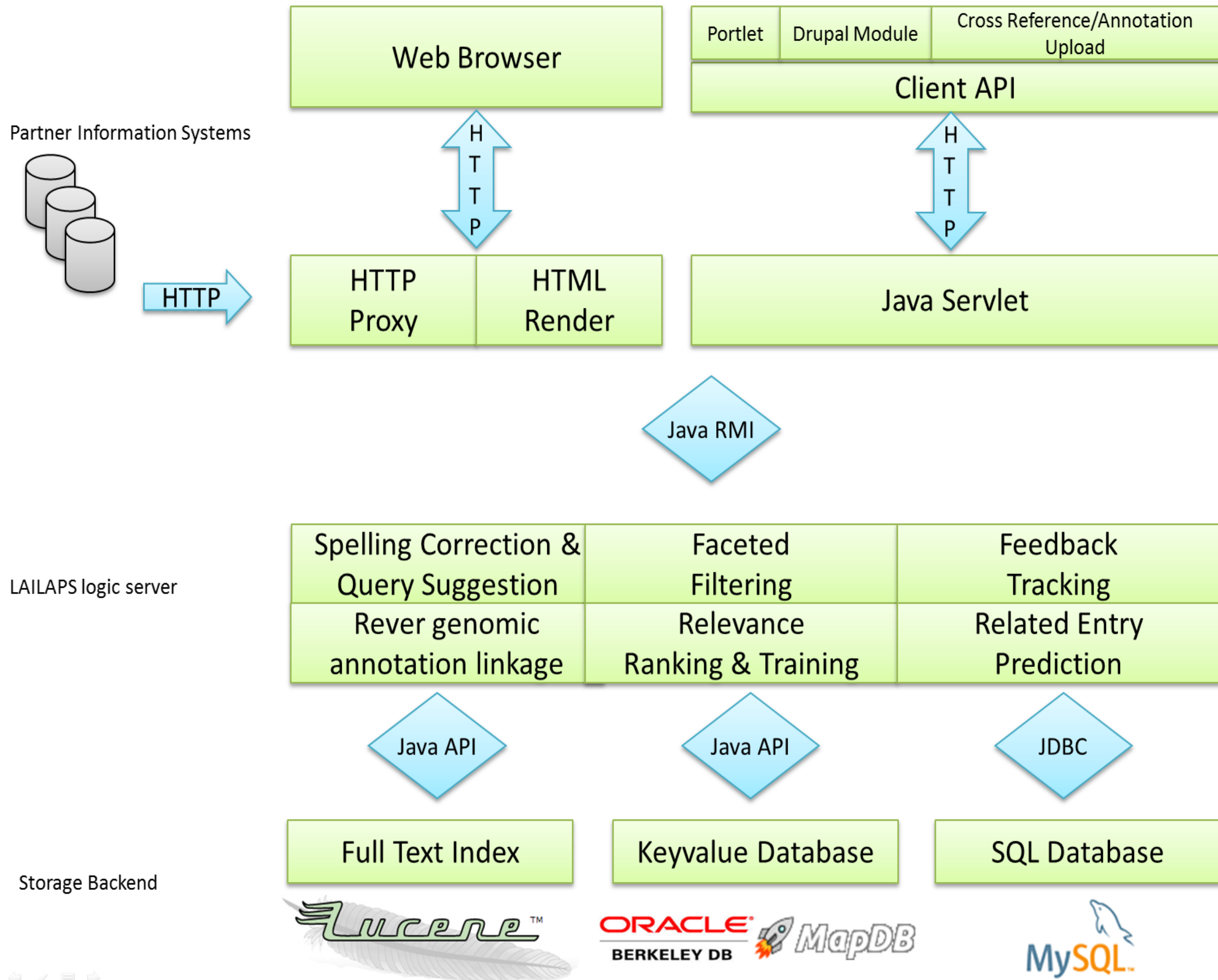
Sample trait: "powdery mildew resistance"



TECHNOLOGY

Hardware Requirements:

- RAM: 16GB
- CPU: 4 core, 4GHz
- storage 800GB SSD
- cost < \$1500
- performance:
 - max. response time: 20 sec. (query for: „gene“ etc.)
 - 25 parallel queries



REFERENCES

- [1] M. Esch, J. Chen, C. Colmsee, M. Klapperstück, E. Grafahrend-Belau, U. Scholz, M. Lange LAILAPS - the plant science search engine. Plant Cell Physiol. 56 (2015) e8.
[2] M. Lange, K. Spies, C. Colmsee, S. Flemming, M. Klapperstück and U. Scholz. The LAILAPS Search Engine: A Feature Model for Relevance Ranking in Life Science Databases. Journal of Integrative Bioinformatics, 7(3):e118, 2010.1

Acknowledgements

This work was supported by "de.NBI-German network for bioinformatics infrastructure", which is a project of the Federal Ministry of Education and Research (BMBF). We thank Jens Bauernfeind and Thomas Münch, Heiko Mieh as administrators of the project website, the LAILAPS server, the code and artifact repositories.