POMBASE COMMUNITY LITERATURE CURATION

Antonia Lock[1], Kim Rutherford[2], Midori Harris[2], Jacqueline Hayles[3], Steve Oliver[2], Jürg Bähler[1], Valerie Wood[2]

[1]Department of Genetics, Evolution and Environment, and UCL Cancer Institute, University College London, London WC1E 6BT, UK; [2]Department of Biochemistry, University of Cambridge, Sanger Building, 80 Tennis Court Road, Cambridge CB2 1GA, UK; [3] The Francis Crick Institute , 1 Midland Rd, London NW1 1AT, UK

The model organism database PomBase obtains its highest-quality data by manual curation of the fission yeast literature. To supplement the work of its small staff of professional curators, PomBase has developed a community curation model that enables fission yeast researchers to participate directly in curating data from new publications. Community curators are incentivized by increased visibility of their recent publications. **To date (May 2017), the community has contributed 10,000 annotations for 440 publications, with a response rate of over 40% to community curation invitations.** Since PomBase is one of only three databases that provide manual literature curation for fungal species, electronic data transfer of high-confidence *S. pombe* annotations to other fungal species is an essential source of function-related data for the latter. Community contributions to PomBase therefore support research not only within the fission yeast community, but also throughout the broader community studying all fungi.

WHY COMMUNITY CURATION?

Rapidly increasing amounts of data

The information content in papers is increasing: more annotations can be extracted and more genes are described in recent publications
- The "paper format" is not an appropriate long-term storage format for data.
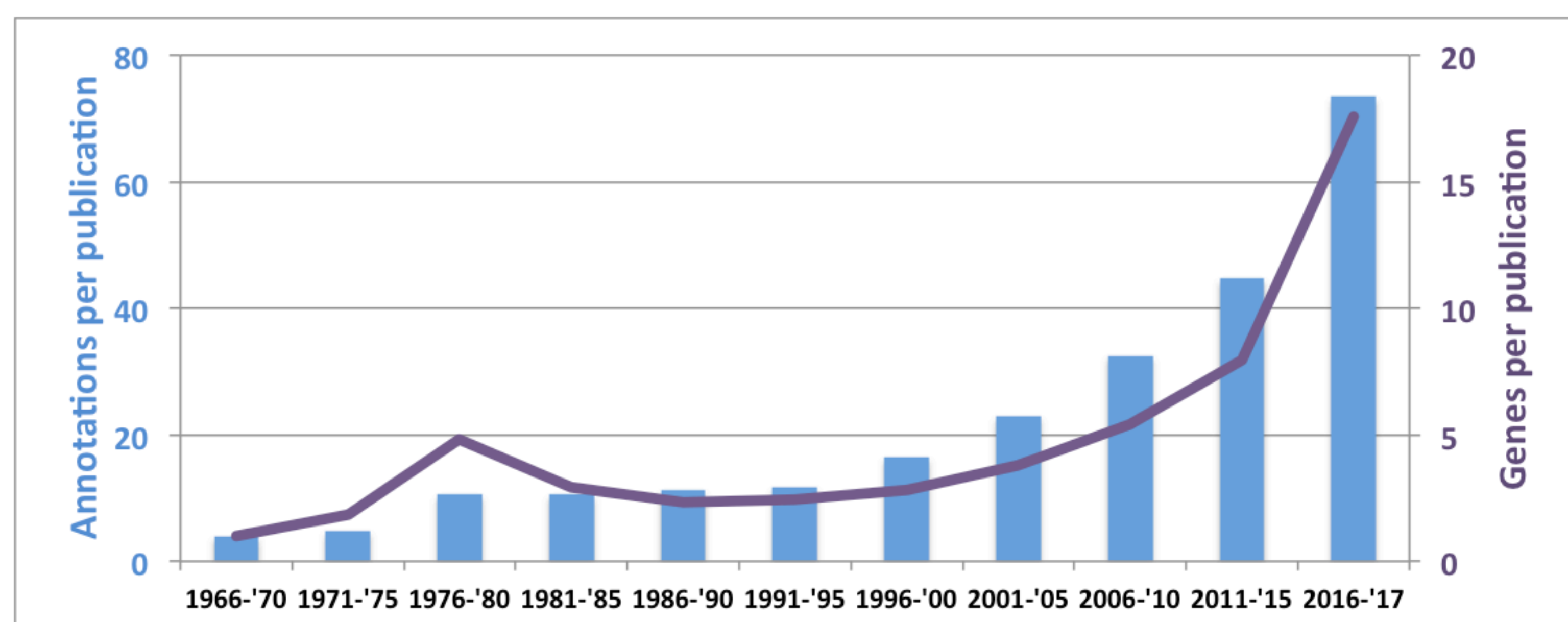
Data transfer

Annotations can be transferred from one species to another
-> existing resources benefit the whole fungal research community
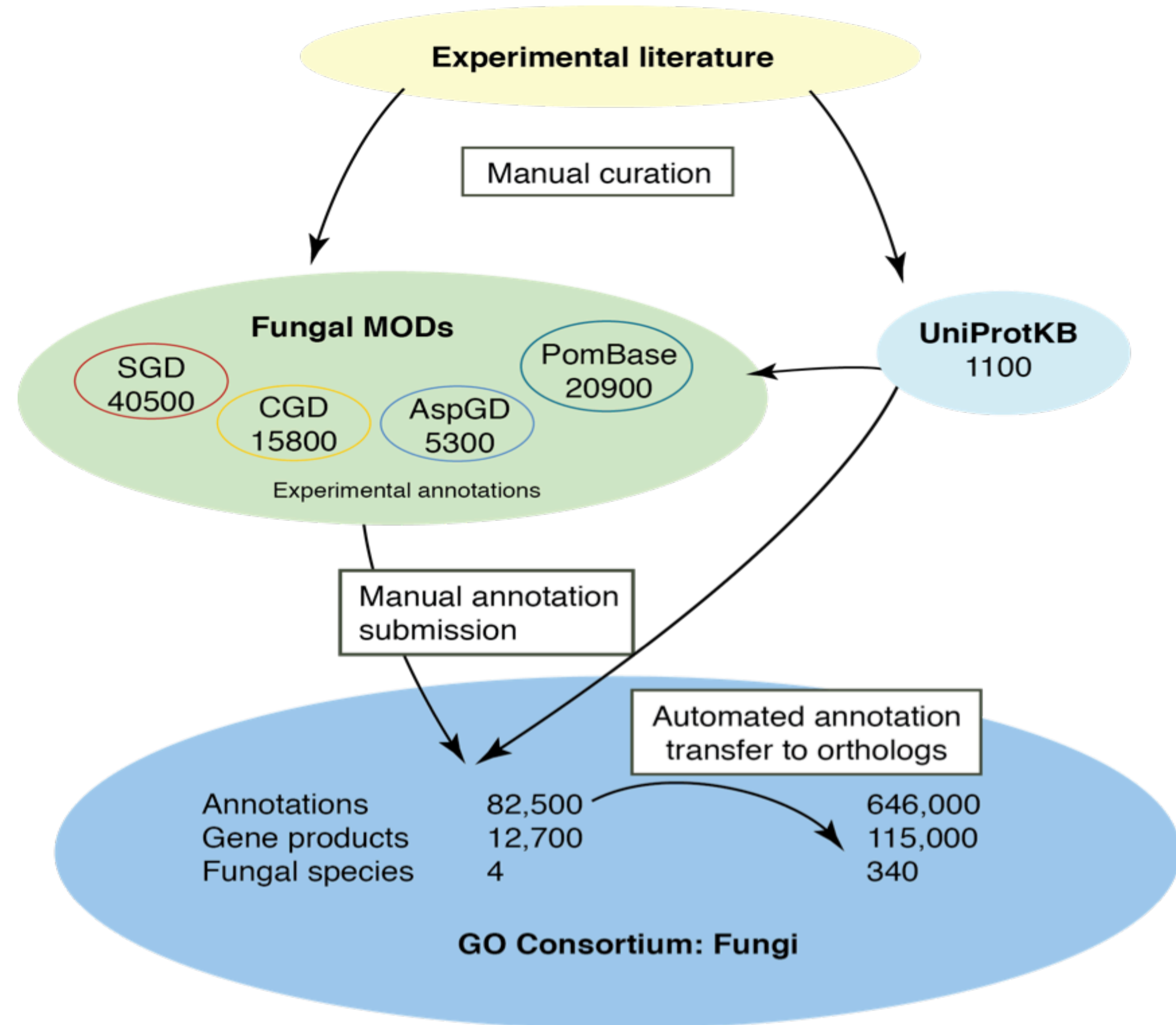
Data gets 'buried' in publications

- 650 fungal species have been sequenced, and are associated with ~300.000 publications.
- Only 4 fungal species – all Ascomycota – have dedicated databases.
  -> Much published knowledge is not catalogued in a computer-friendly manner
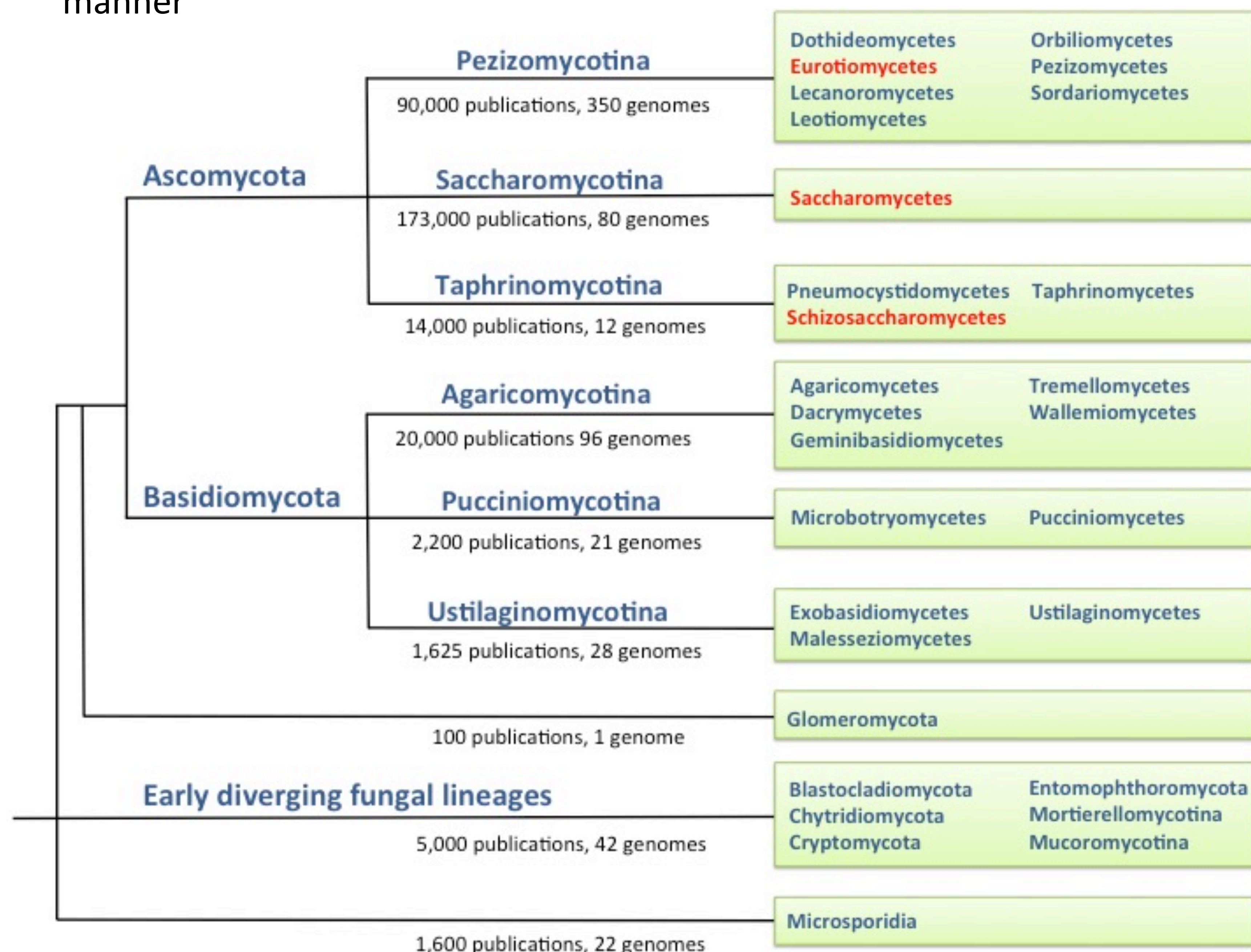
Researcher, database & community benefits

- Familiarizes community curators with formal data representation
- Increased visibility of the community curators' publications
- Co-curation between a professional curator and a biological expert result in higher quality annotations
- Formal language makes data from different sources comparable
- Facilitate locating data from different publications
- Support analysis of large gene sets

COMMUNITY CURATION USING THE ONLINE TOOL CANTO

Users can create different types of annotations

New annotations are shown in a table

Intuitive search interface

More specific terms are suggested