Biodiversity data sources
"European Biodiversity Archive"

Urmas Kõljalg

University of Tartu

EU BON Stakeholder Round Table "Requirements for Policy", Leibniz Association, Brussels Office, Rue du Trône 98 – 4°, 18 June 2013

Task 1.1 Assessment and evaluation of biodiversity data sources

The evaluation will assess important data characteristics, such as coverage, accessibility, quality, and format. Targeted data types 1) remote sensing data (incl. derived products e.g. vegetation and habitat maps, habitat classification schemes); 3) taxonomic backbone data; 4) ecological data; 5) specimen data from scientific collections; 6) species profile data, including descriptions and functional traits, conservation status, distribution and abundance data; 7) DNA sequence data.

Task 1.3 Gap analysis of available biodiversity information sources and identifying priorities

Based on the assessments of data sources undertaken for task 1.1. and the review of policy requirements in WP6 (task 6.1.), identified gaps in data coverage and quality for different information layers will be evaluated against scientific interests and capacities in the biodiversity information stakeholder communities, focusing at the European level. Taxonomic, geographic, thematic, and other areas of bias will be analyzed in assessed datasets and priority levels for closing gaps will be provided, also comparing European to global level coverage.

The biases and shortcomings of the actual biodiversity data that is available for European policy and how a portal and process like EU BON could contribute to fill the existing gaps.

How biodiversity data are collected?
What data are collected?
Who collects?
Where stored?
How accessible?
Best working examples?

GBIF

INSDC (NCBI "GenBank", ENA & DDBJ)

(a)

Statistics based on 98% threshold value	Total	Europe	Asia	N-America	S-America	Australasia	Antarctica	Africa	Unspecified
No of SHs*	52481	13779	8819	12941	2929	3417	313	2425	20294
No of ITS sequences	261521	50963	31560	53931	9635	8211	930	6183	100108
SH shared between continents		2832	1956	2633	686	745	130	507 N/A	
Percent of unique SHs		79.5%	78%	79.6%	76.6%	78.2%	58.5%	79.1%	
No of singletons (out. clusters)	4130	694	569	742	208	266	13	182	1456
No of singletons (inside clusters)	28395	5854	3730	5825	1183	1587	97	1006	9113
* Including all singletons									
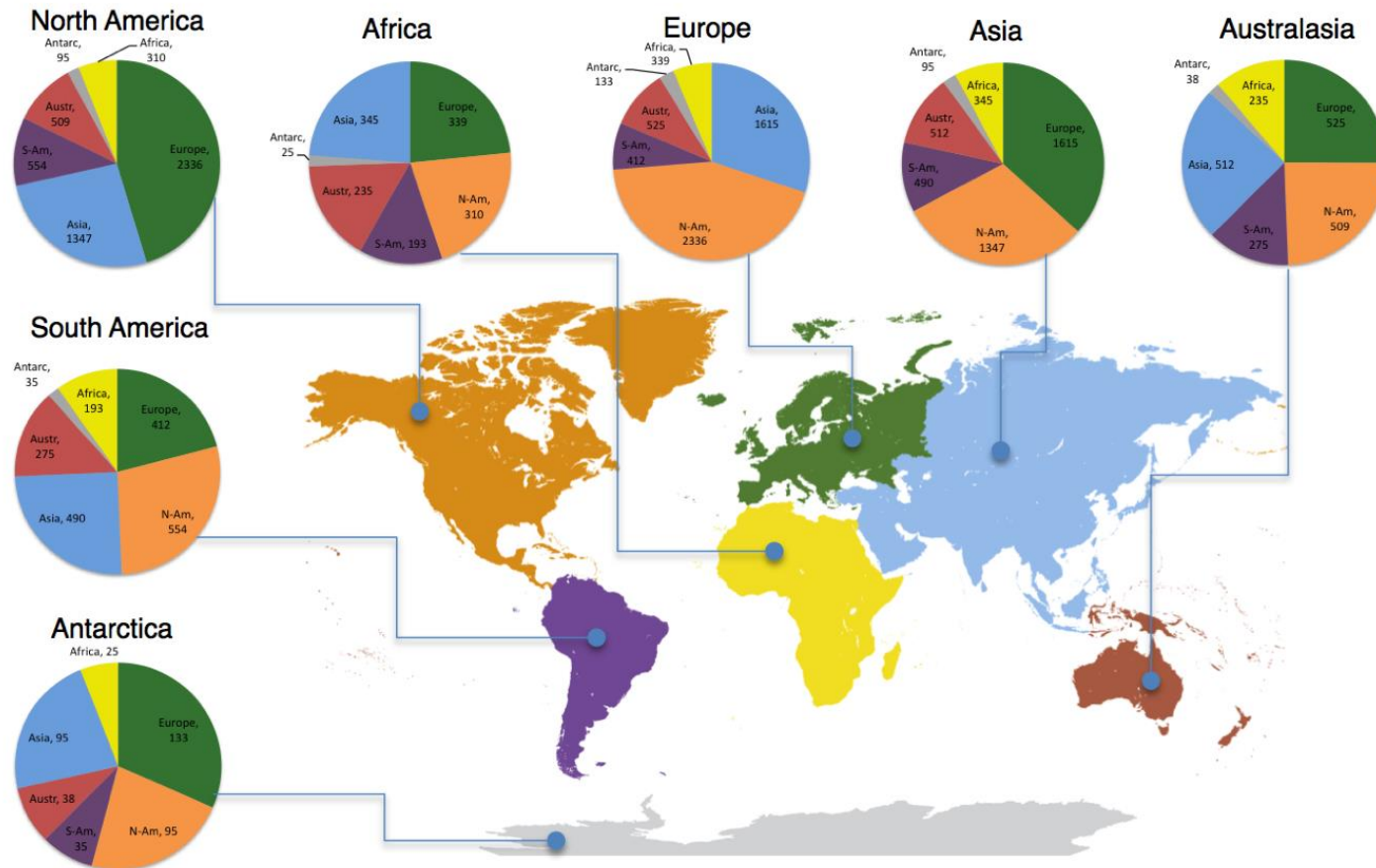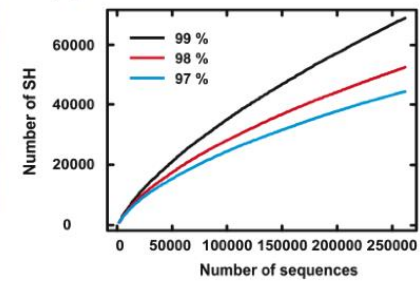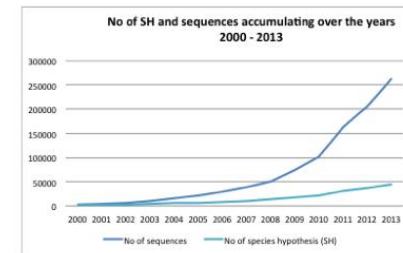
(e)

Phylum	No of SHs*	No of RefSeq
Basidiomycota	20804	1476
Ascomycota	20754	287
Chytridiomycota	285	0
Glomeromycota	1658	77
Zygomycota	602	0
Neocallimastigomycota	85	2
Blastocladiomycota	31	0
Incertae sedis	9	0
Unspecified	11387	0
*Based on 98% threshold value		

(f)

Threshold value	Number of SHs
97%	44537
97.5%	48007
98%	52481
98.5%	58594
99%	68938

Kõljalg et al. Molecular Ecology (accepted)

GBIF

INSDC (NCBI "GenBank", ENA & DDBJ)

"European Biodiversity Archive"
In collaboration with ENA (INSDC)?

Primary goals: data storage and making data available for downloading

Thank you!