



Metabarcoding projects

“From single individuals to biological communities”

Kessy Abarenkov
University of Tartu



What is DNA metabarcoding?

Taberlet et al. 2012. Towards next-generation biodiversity assessment using DNA metabarcoding.

“... Automated identification of multiple species from a single bulk sample containing entire organisms or from a single environmental sample containing degraded DNA (soil, water, faeces, etc.)”



What is DNA metabarcoding?

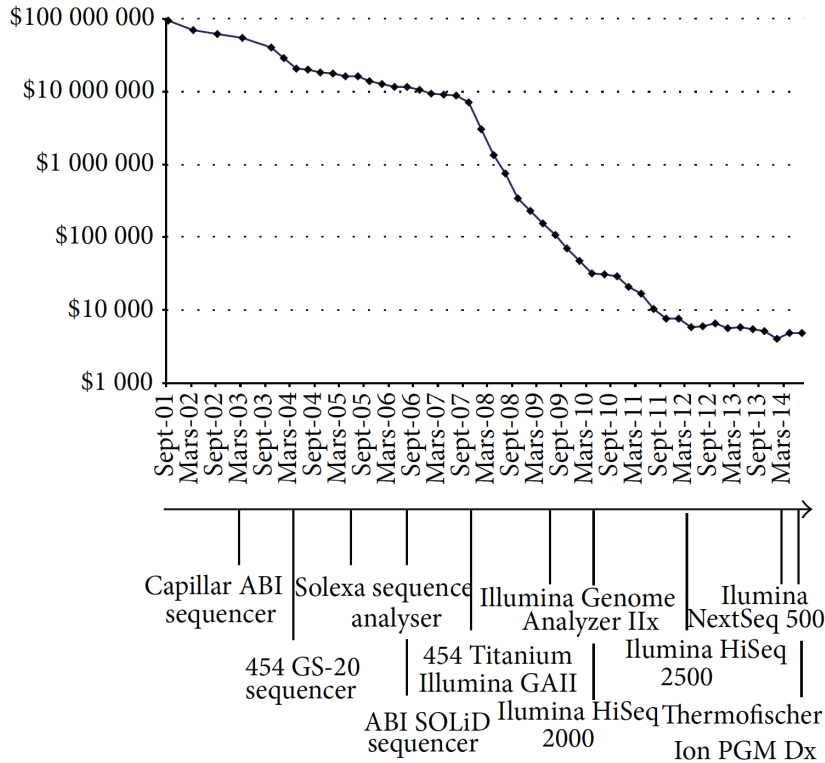
Taberlet et al. 2012. Towards next-generation biodiversity assessment using DNA metabarcoding.

“... Automated identification of multiple species from a single bulk sample containing entire organisms or from a single environmental sample containing degraded DNA (soil, water, faeces, etc.)”

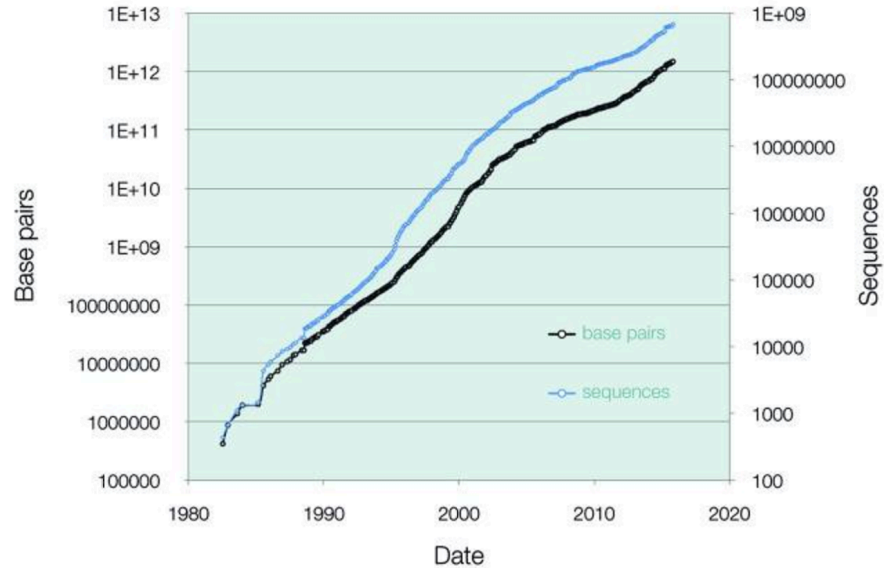
Metabarcoding largely owes to the revolution of next-generation sequencing which enables to, with moderately low cost and high throughput, measure and monitor biodiversity – its patterns and trends.

Human Genome Sequencing cost vs accumulating data volumes in INSD

de Brevern et al. 2015.



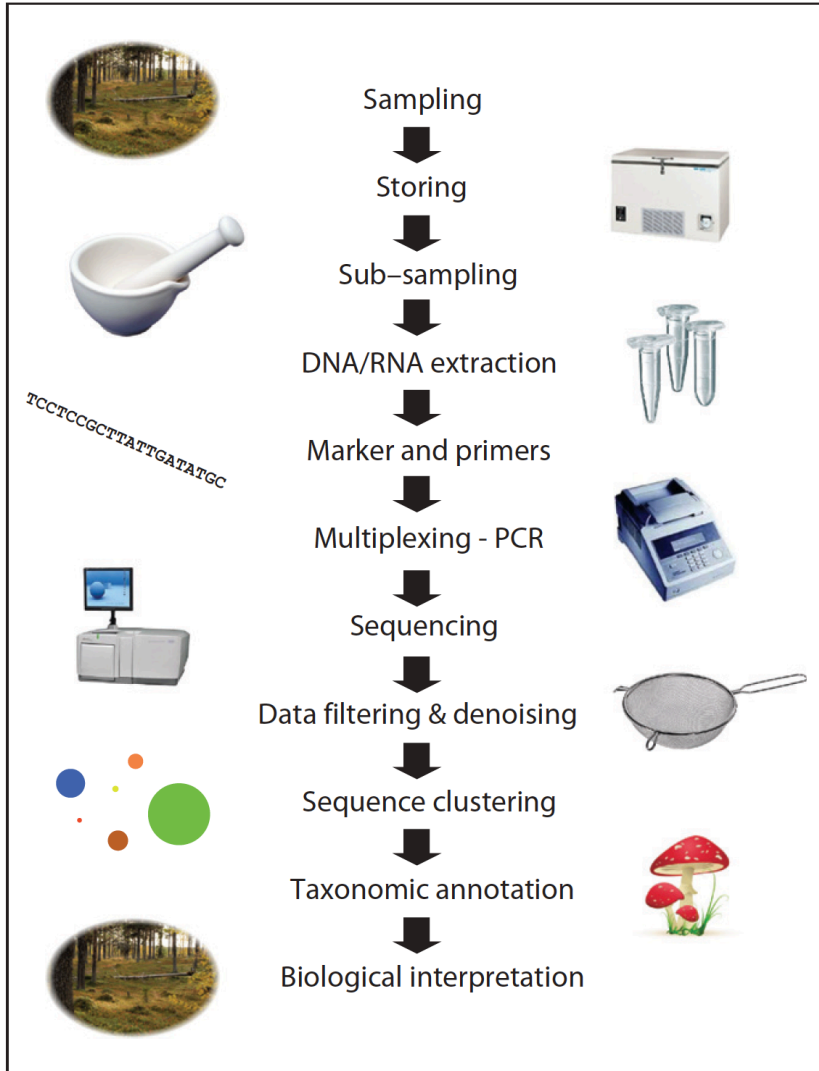
B



Cochrane et al. 2016.

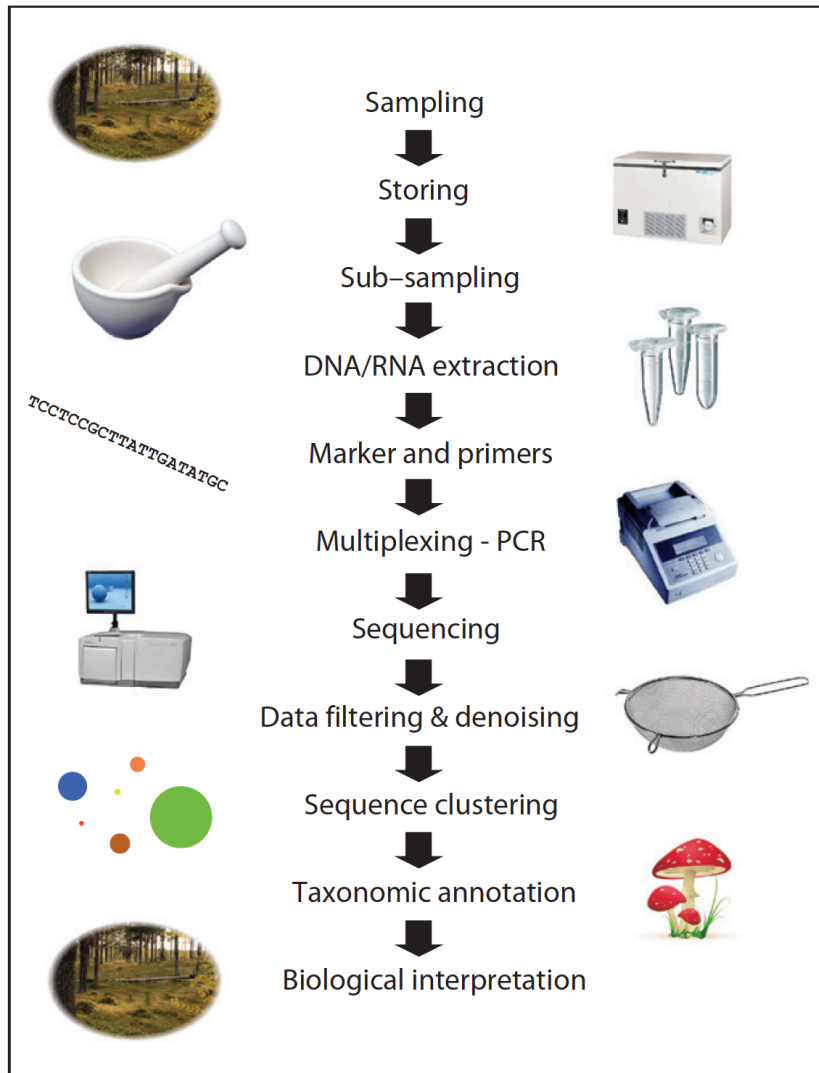
→ huge demands on bioinformatics tools for data analysis and management

Steps involved in HTS of fungal communities



Lindahl et al. 2013. Fungal community analysis by high-throughput sequencing of amplified markers – a user’s guide.

Steps involved in HTS of fungal communities



Lindahl et al. 2013. Fungal community analysis by high-throughput sequencing of amplified markers – a user's guide.

} Data management, storage and publishing?

Metabarcoding challenges

- Intermediate PCR step – generated amplification biases and contributes to errors
- Errors - caused by next-generation sequencing platforms
- Data analysis - need to build appropriate bioinformatics tools and infrastructure
- Data volumes - need for appropriate, centralized storage
- Lack of standardization and comparability (marker resolution, primer bias, data quality, bioinformatics pipelines, data standards)
- Lack of exhaustive and high-quality reference libraries (sequences generated with traditional barcoding approach and based on morphologically verified and curated specimens)
- Lack of communicating results (OTUs not easily recognizable across studies and time)



Metabarcoding challenges

Tedersoo et al. 2015. Standardizing metadata and taxonomic identification in metabarcoding studies.

27 fungal metabarcoding studies in total

37% missing from public repositories (Short Read Archive, DataDryad)

19% locked

15% lacking essential metadata for re-use

* Environmental metadata typically located in user-defined data fields with different units or no units at all

Metabarcoding data management I

Data storage and publishing:

- Local storage and data management
 - Short-term storage (performance and throughput, data compressed at all times)
 - Long-term storage (storage reliability, only raw data and final results should be archived, total cost vs regenerating the data)
 - Complete analysis workflow with all parameters and software versions (analysis transparency and reproducibility)
- Public data repositories ([INSD/SRA](#), [DataOne](#), [Dryad Digital Repository](#), [PlutoF platform](#))
- Availability & Accessibility vs Understandability & Usability?



Metabarcoding data management II

Heterogeneity of data and data formats – how to extract relevant and reliable information?

Standards and formats:

- Spreadsheets (CSV, TSV)
- FASTA
- OTU tables
- BIOM format - <http://biom-format.org/>
- GSC (MIXS, MIMARKS) - <http://gensc.org/>, <http://gensc.org/mixs/>



Metabarcoding data management in PlutoF

Current PlutoF platform allows to store and manage sample data according to GSC standard and upload representative sequences.

Example project –

[Mark et al. 2016. Barcoding lichen-forming fungi using 454 pyrosequencing is challenged by artifactual and biological sequence variation.](#)

Hands on time!