

Coordinating and sharing genomic data in pigs (1/2)

SABRE-TP Workshop 2020



Andreas Hofer

19.08.2020



Outline

(1/2) Andreas Hofer (SUISAG)

- Introduction & aim
- Concept for sharing data
 - Type of data and processor, storage
 - Outlook

(2/2) Claudia Kasper-Völkl (Agroscope)

- Requirements and rules for pre-publication data sharing
 - Data structure, access to and use of data
- General discussion



Introduction & Aim

- Animal breeders and researchers increasingly rely on genomic data of animals for their work.
- Several players: Breeding organisation (BO), research groups (RG)
 (ETHZ, VetSuisse, HAFL, Agroscope), funding agencies (e.g. BLW)
- Ressources should be shared to:
 - Avoid repeating generating same genomic data (safe costs)
 - Increase power of statistical analyses

Aim of initiative:

- Coordinating and sharing genomic data in pigs
- So far 2 discussion rounds of research groups (ETHZ, UniBE, HAFL, Agroscope) with SUISAG
- Model for other species?

19.08.2020 SABRE-TP 2020 3



Concept of coordination and sharing data

| Type of data | Processor / Storage | Remarks |
|--|--|---|
| Meta data Animals (Pedigree), Samples, type of genomic analyses, key to raw genomic data | Kept and updated in data bases of breeding organizations | Information flow: RG→BO on newly generated genomic data BO→RG on available ressources |
| Raw genomic data provided by lab (SNPs, sequence (FASTQ)) | Should end up in predefined shared storage (public as far as needed for publications, otherwise private) | Predefined data formats ID = anonymous key should facilitate public storage |
| Processed/condenced genomic data (SNPs after QC, gVCF) | Produced by BO and RG Local and/or shared storage | To be shared if analysis is expensive and used parameters are well documented |

19.08.2020 SABRE-TP 2020 4



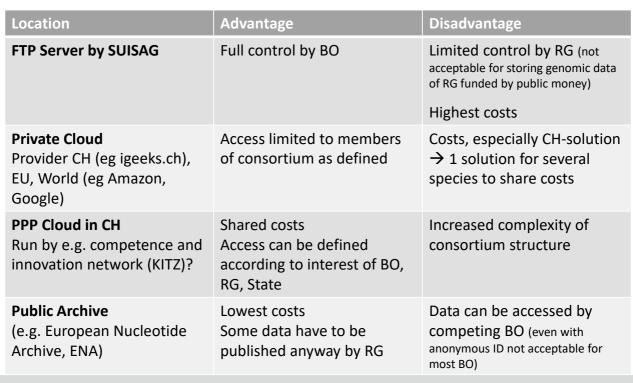
Storage of shared data: Requirements

- Easy access for BO and RG
 - Upload & download of large data volumes (not very often, speed of transfer not very critical, but transfer must be reliable)
 - Data structure to allow easy access to data of interest
- Data security
 - Backup, limited access in case of private data
- Low costs
 - Storage space, data transfer, maintenance

19.08.2020 SABRE-TP 2020 S



Storage of shared data: Potential locations



19.08.2020 SABRE-TP 2020 6



Outlook

- Short term
 - BO extend(ed) their data bases to accommodate metadata
 - BO keep track of storage and locations of raw genomic data
 - RG keep BO informed about newly generated genomic data
- Medimum term
 - Agreement among BO and RG on pre-publication data sharing (independent on questions on storage solutions/locations)
 - Start discussion on storage solutions/locations
 - Visions of BO (SUISAG, ASR/Qualitas, others)
 - Potential role of KITZ (BLW)

19.08.2020 SABRE-TP 2020 7