

GCATGCATGCATGCATGCATGCATGCATGCA **N**extGenerationSequencing Design

NTGCATG CATGCATO JUATOLA 'GCATGCA ATGCATGCATGCA CAATG CATGON **GCATGCATGCATGCATGCA** *TGCATGCATGCATGCATGCATGCA TATGCATGCATGCAT? ATGC ATGCATGCATGCATG GCATGCAATGCATG SCAGTGCATGCAT ATGCA" SCATEL TATGCAN **CATGCA** 'ATCGCA' TGCATGC ATGCATGC ATGCATGC TGCATG **"GCAT** CATG `*TGC*

CATGCATGC®

ATGCATGCATGC

ATGCATGC CATGCATGC JGCATGC/ ATGCATG

ATGCATGO ATGCAAT ΓGCATGC ΓGCAGG1

- LAIGCATGCATGCATGCATGCA. I & CATGCATCGCATGCATCGCATGCATGCATG

> **Niklaus Zemp** June 2020

Bioinformatics Genetic Diversity Centre (GDC) **ETH Zurich**

JCATGCATG LATGCATGCAIGCATG STGC. GCAATG NGCATGCATGCATGCATGCATA 76CATGCATAAAGCADECAT JUATGCATGCATTCCATCC LATGCATGCATGCATGCA TGCATGCATGCAATGCATTT GCATGCAGTSCATGCATGC CATGCATGCA[®] JCATGCATGCA ATGCATGCATGCAT CATGCATGU TECATGCAT THERITGETA ATGCATG JAT G CAT



GD Genetic Diversity Centre Zurich



~500 Mio pages -> 54 km

Big data

Genetic









The second secon

- 1) Define research question and formulate hypotheses
- 2) Select application/methode
- 3) Select NGS platform
- 4) Define experimental design
- 5) Valdiate your results



C = LN/G

C coverage; L read length ; N inumber of reads; G genome length <u>https://support.illumina.com/downloads/sequencing_coverage_calculator.html</u>

What theoretical coverage do I need?

Genome assembly: 60-100 X Re-sequencing: 20 X RAD-seq: 20 X Pool-seq: 100 X RNA-seq: ~15-20 Mio reads for Eukaryotes Amplicon: >10 per OUT Low coverage re-sequencing: 0.5-2X



- Formulate your research questions
- There is no ideal application/method to answer all questions

