# Introduction

# Methods

## Simulated ageing

## Accession selection

We identified 340 candidate accessions in the Naturalis herbarium collections. Of these, 248 were annotated with known collection dates (ranging between 1800 and 2012) and 264 with known collection localities (42 distinct geographical units) including 106 with latitude/longitude coordinates. 318 of the herbarium sheets had previously been photographed under uniform conditions through Naturalis’s digitization activity. For these we computed the amount of available leaf material by counting the proportion of pixels in the green spectrum. This yielded a relative index ranging between 0.98 (an accession comprised of a single sprout) and 41.23 (more than 40% of the image occupied by putative leaf material).

## DNA extraction

## DNA sequencing

## Data analysis

Following basic quality control analysis (ref FastQC) we mapped our reads against the ITAG 2.3 reference genome (ref) using BWA (ref). To assess the extent of fragmentation and nucleotide misincorporation (ref) we analyzed our data using mapDamage (ref) and using a sliding window analysis of the base composition one base upstream from where our reads map against the reference.

For a selected set of candidate genes of commercial interest we constructed consensus sequences for each accession and multiple sequence alignments of our accessions with the reference genome and the nearest BLAST (ref) hit on GenBank that contains an annotated CDS for the locus. We then recomputed these sets of sequences as codon alignments using MACSE (ref).

For each codon alignment we inferred the maximum likelihood tree under a GTR substitution model using PHYML (ref). We then performed a BranchSiteREL test (ref) on the codon alignments and the resulting maximum likelihood trees using HYPHY (ref).

# Results

## Extraction results

## Raw reads

## Assembly

## Fragmentation

## Indels

## Phylogenetic analysis

## Analysis of selection

# Discussion

* We can extract and sequence nuclear genomes from 150+ year old herbarium sheets using normal extraction protocols
* The data show fragmentation but relatively little misincorporation
* The data show variation among accessions, including frame shifts and non-synonymous substitutions in coding regions of interest
* The data show divergent directional selection on traits of interest, e.g. in disease resistance and carbohydrate transport, both between species and between cultivars
* Some of the old cultivar genotypes are for wild-type phenotypes (e.g. non-uniform ripening)
* Museomics is doable and can shed light both on fundamental and applied questions
* To do: broader sampling to study origins, sequencing of virii

# Acknowledgements

# References

# Figures

# Tables

## Table 1. Herbarium accessions

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Accession** | **Species** | **Year** | **Green pixels before** | **Green pixels after** | **Coverage** | **Mapping quality** |
| U0015717 | *Solanum lycopersicon* L. | 1830 | 12.87 | 7.59 | 2.94x | 54.49 |
| U0015716 | *Solanum lycopersicon* L. | 1891 | 16.99 | 14.56 | 3.33x | 54.10 |
| WAG0463703 | *Solanum peruvianum*L. | 1958 | 16.7 | 12.2 | 2.97x | 44.62 |

## Table 2. Candidate genes

|  |  |  |  |
| --- | --- | --- | --- |
| **Locus ID** | **Location** | **Nearest GenBank record** | **Notes** |
| Solyc02g085500.2 | SL2.40ch02: 42945331- 42947065 | NM\_001247292 | Ovate protein |
| Solyc09g010080.2 | SL2.40ch09: 3475480- 3479343 | NM\_001247864 | Beta-fructofuranosidase |
| Solyc09g018220.1 | SL2.40ch09: 13621396- 13623981 | XM\_004246222 | Disease resistance protein RPP13-like |
| Solyc10g083290 | SL2.40ch10: 62435058- 62440444 | AF506005 | Cell-wall invertase |
| Solyc09g010090 | SL2.40ch09: 3480545- 3484159 | NM\_001247772 | Cell-wall invertase |
| Solyc10g083300 | SL2.40ch10: 62447632- 62451825 | XM\_004249387 | Beta-fructofuranosidase, insoluble isoenzyme 1 |
| Solyc06g008300 | SL2.40ch06: 2161344- 2164746 | AF053995 | Similar to *Lycopersicon pimpinellifolium* disease resistance protein Cf-2.2 encoded by the sequence presented in GenBank Accession Number U42445 |
| Solyc01g009690.1 | SL2.40ch01: 3972616- 3975213 | XM\_004228395 | Receptor-like protein 12-like isoform 2 |
| Solyc01g006550.2 | SL2.40ch01: 1124417- 1127114 | XM\_004228575 | Receptor-like protein 12-like |
| Solyc03g082780.1.1 | SL2.40ch03: 46190396- 46194193 | XM\_004228569 | Receptor-like protein 12-like |
| Solyc05g013300 | SL2.40ch05: 6400973- 6401944 | XM\_004239070 | Receptor-like protein kinase HERK 1-like |
| Solyc07g053630.2 | SL2.40ch07: 59386217- 59390311 | XM\_004243333.1 | Transcription activator GLK1 |
| Solyc10g008160.2 | SL2.40ch10: 2293088- 2295945 | XM\_004248130 | Transcription activator GLK2 |