

Genomic effects of domestication in *Gossypium hirsutum*

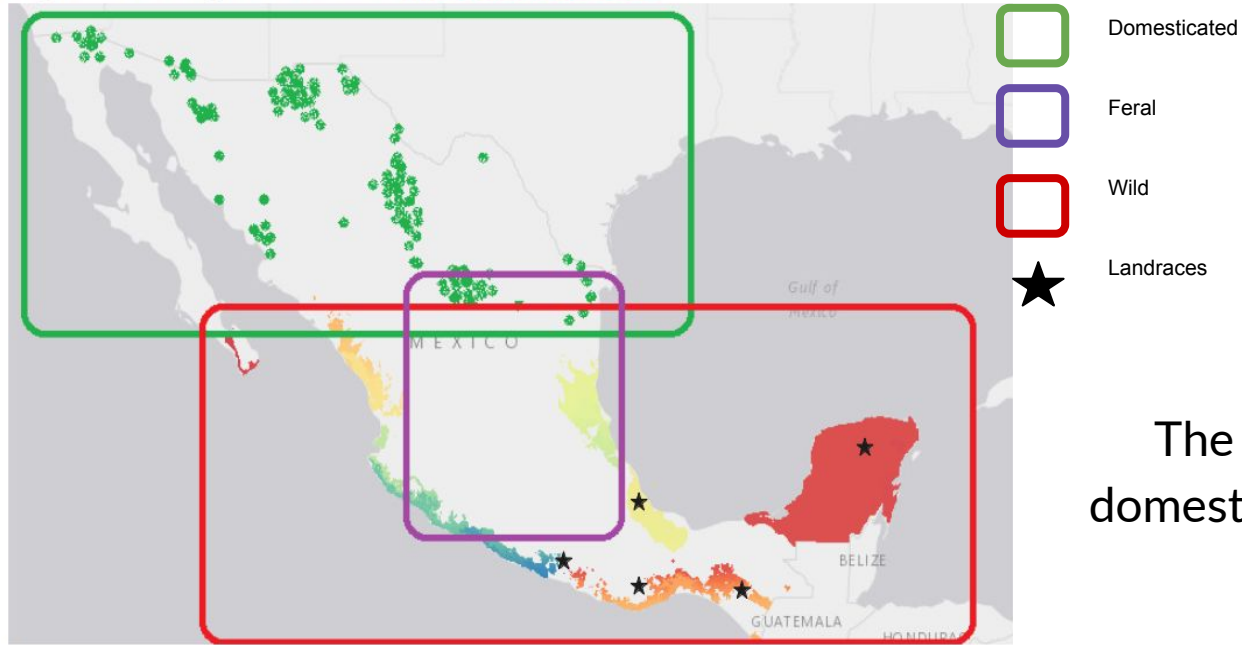


***Gossypium hirsutum*: specie with economic importance**

- It is cultivated for fiber and seed oil
- Generates ~ US\$ 50 billion
- The mexican codices are evidence that the cotton was used since prehispanic time
- The begining of domestications of cotton 5000 years ago



Distribution of mexican cotton

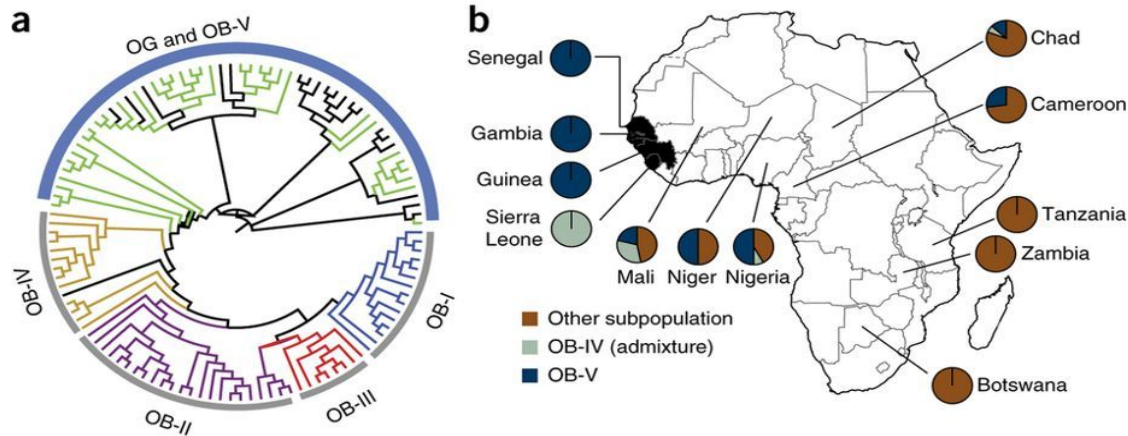


The center of origin, center of domestication and center of genetic diversity is Mexico.

Research of domestication process with molecular data

Category	Upstream	Exonic	Intronic	Downstream	Upstream/Downstream	Intergenic
SNP	149,469	132,271	283,859	123,958	7,742	5,931,601
InDel	43,047	3,952	51,547	32,258	2,066	589,730

Upstream: 1kb upstream region of a given gene; Downstream: 1kb downstream region of a given gene; Upstream/Downstream: 1kb upstream region of one gene, and meanwhile 1kb downstream region of another gene.



New sequencing technologies = more information

- Gain or loss of long-fragment sequences
- Chromosomal rearrangements
- Copy Number Variations
- Deleterious variants
- Differentiation in gene expression
- Frequency of specific alleles

Alteration of the
phenotype

Ecological
effects

Objective

Identify the genomic structural variant that define each component of the complex wild-to-crop of *G. hirsutum* to describe the domestication process of the specie.

Particular objectives

- Analyze introgressed genomic regions between parts of wild-crop complex to describe introgression events and their possible structural effects.
 - Characterize the structural variant of genomes in the neighborhood of transgene insertion to analyze their possible effects on transcription.
-

Plant materials



Nayarit
BCS
Yucatán
yucatanense

Verde
Coyuchi

TM-1 USA
China

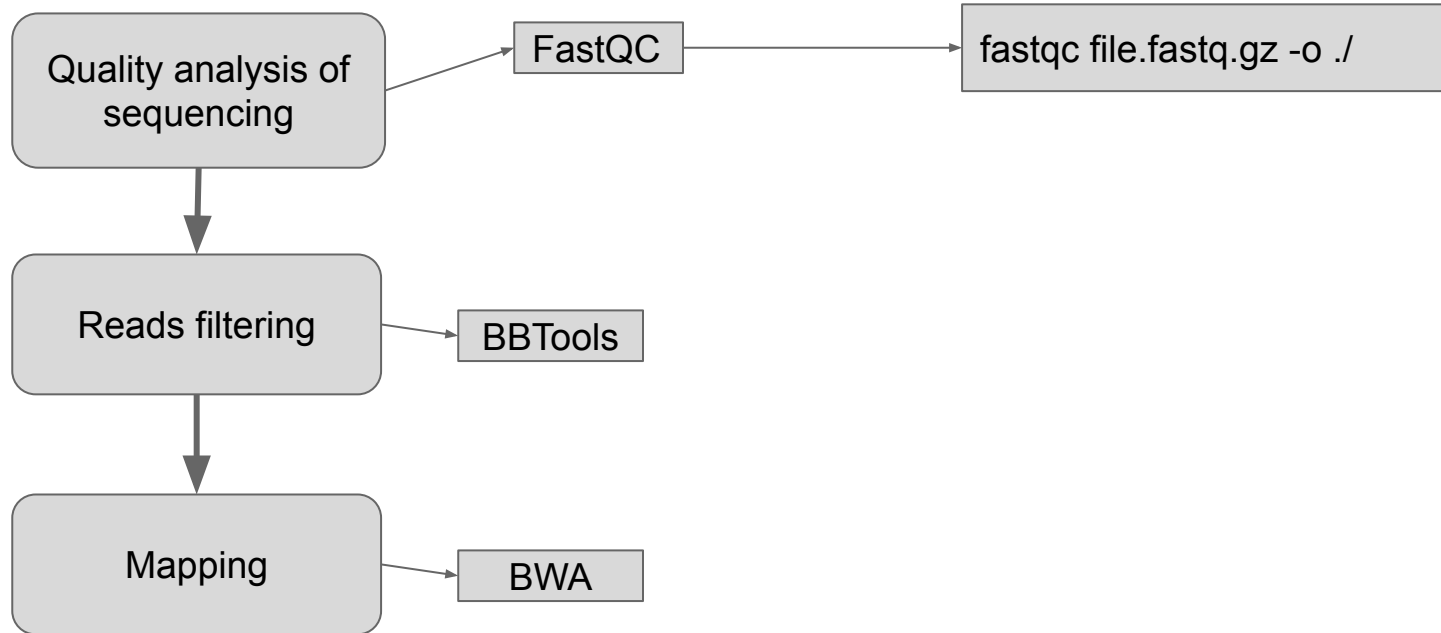
N= 11 genomes

Methods

Characteristics of data

- Sequencing with Illumina NovaSeq 6000
 - Library TruSeq DNA PCR free
 - Paired-end
-

Reads Filtering and mapping



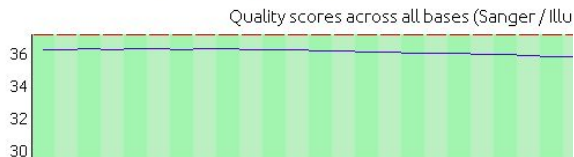
Summary

- ✓ [Basic Statistics](#)
- ✓ [Per base sequence quality](#)
- ✓ [Per tile sequence quality](#)
- ✓ [Per sequence quality scores](#)
- ✓ [Per base sequence content](#)
- ✓ [Per sequence GC content](#)
- ✓ [Per base N content](#)
- ✓ [Sequence Length Distribution](#)
- ✓ [Sequence Duplication Levels](#)
- ⚠ [Overrepresented sequences](#)
- ✓ [Adapter Content](#)

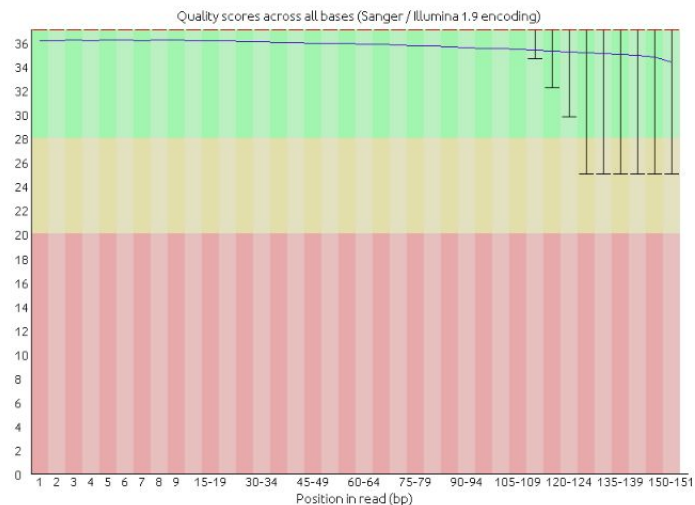
✓ Basic Statistics

Measure	Value
Filename	GhY3-1_R2.fastq.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	335981690
Sequences flagged as poor quality	0
Sequence length	151
%GC	34

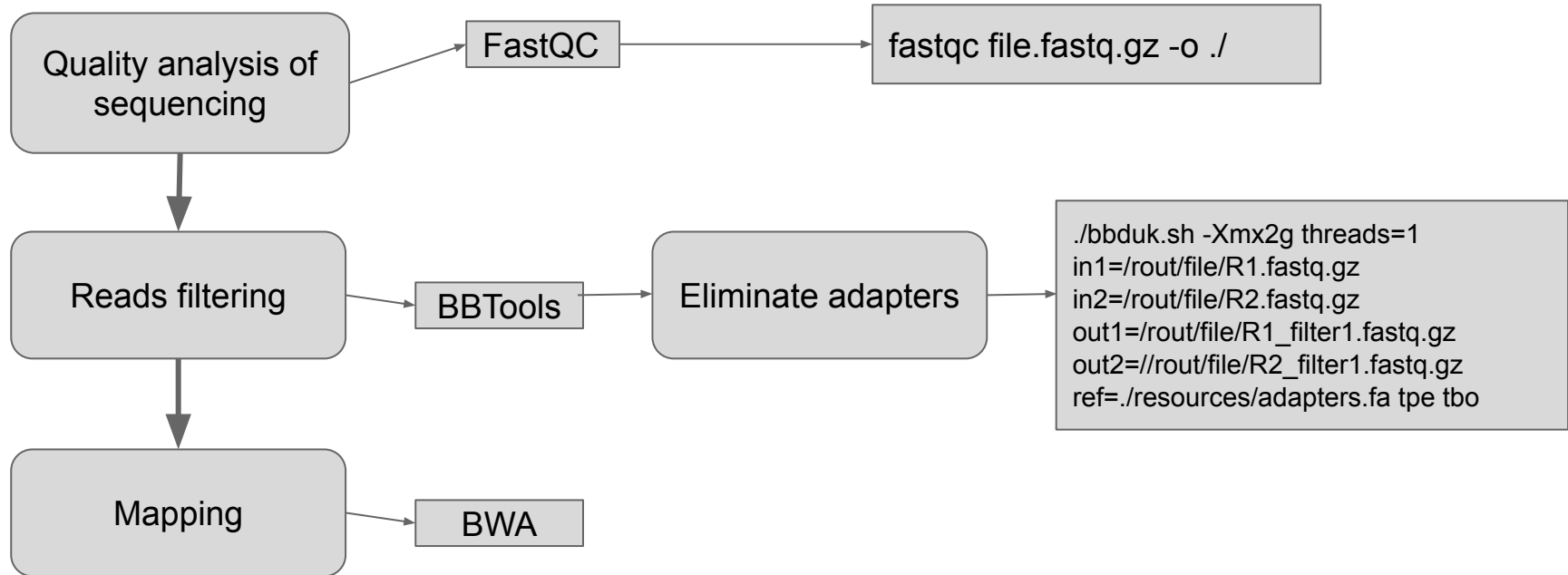
✓ Per base sequence quality



✓ Per base sequence quality

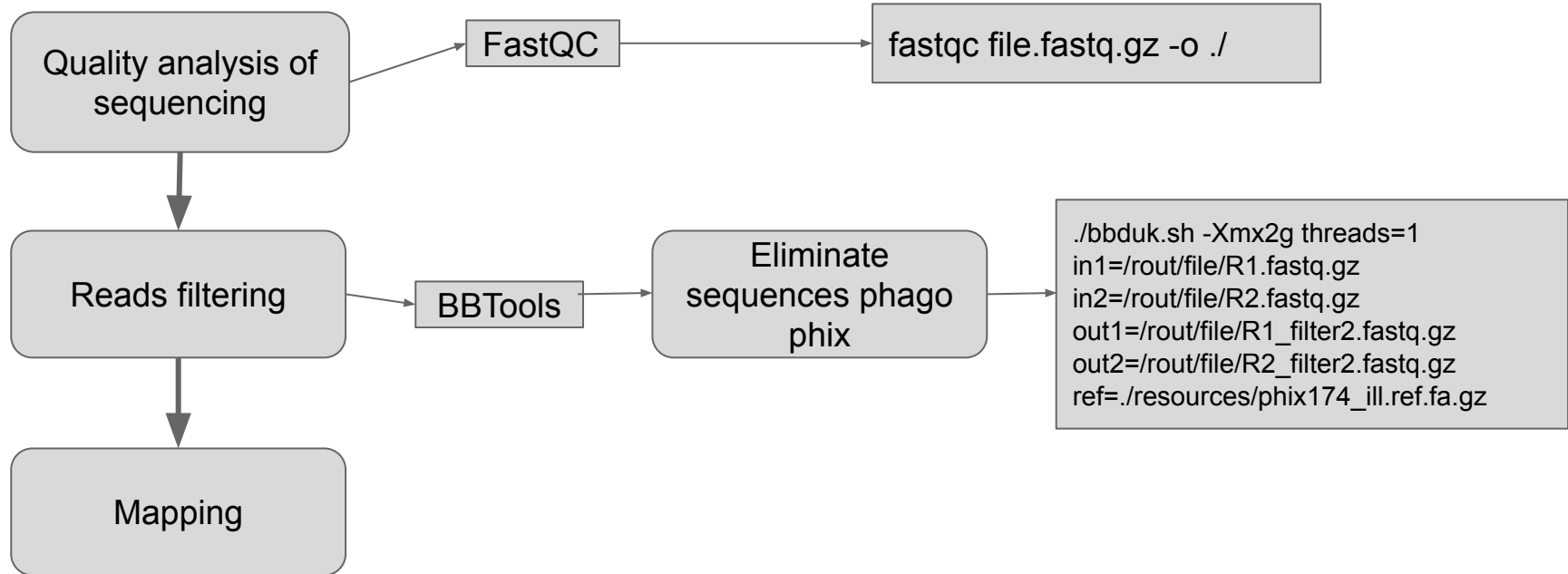


Reads Filtering and mapping

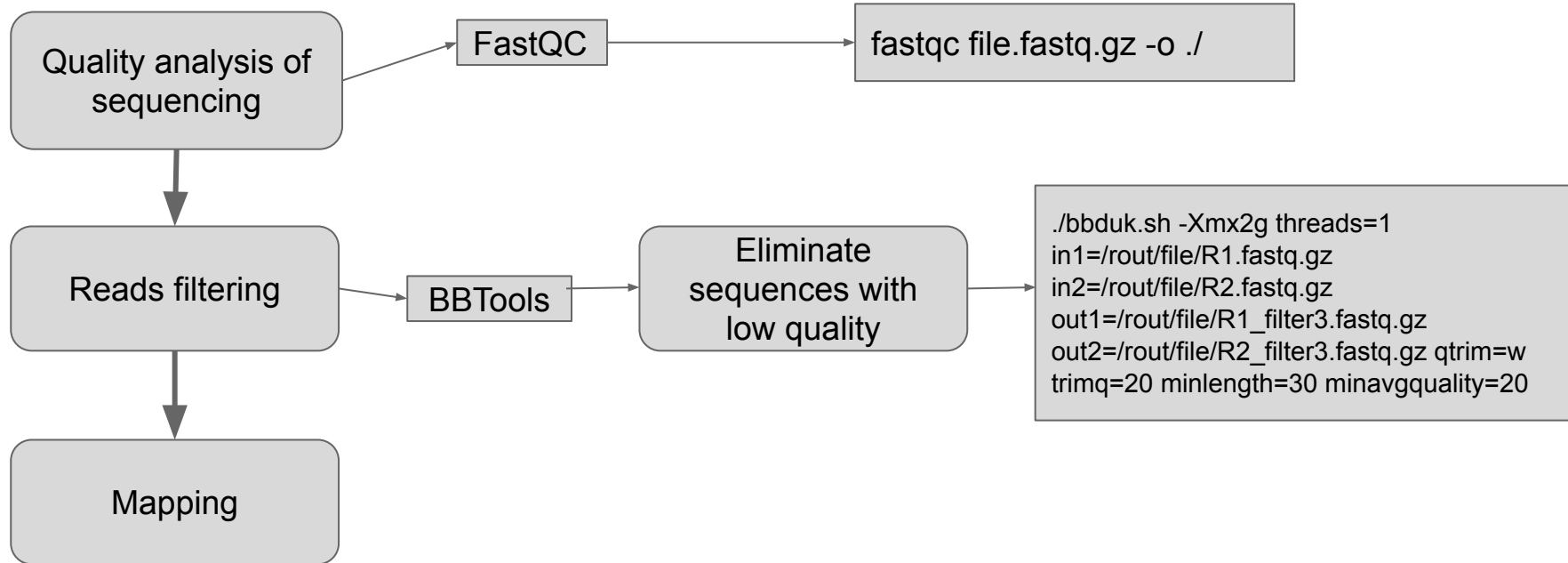



```
>TruSeq_Adapter_Index_13
GATCGGAAGAGCACACGTCTGAACTCCAGTCACAGTCAACAATCTCGTATGCCGTCTTCTGCTTG
>TruSeq_Adapter_Index_14
GATCGGAAGAGCACACGTCTGAACTCCAGTCACAGTTCCGTATCTCGTATGCCGTCTTCTGCTTG
>TruSeq_Adapter_Index_15
GATCGGAAGAGCACACGTCTGAACTCCAGTCACATGTCAGAATCTCGTATGCCGTCTTCTGCTTG
>TruSeq_Adapter_Index_16
GATCGGAAGAGCACACGTCTGAACTCCAGTCACCCGTCCTCGATCTCGTATGCCGTCTTCTGCTTG
>TruSeq_Adapter_Index_18_7
GATCGGAAGAGCACACGTCTGAACTCCAGTCACGTCCGCACATCTCGTATGCCGTCTTCTGCTTG
>TruSeq_Adapter_Index_19
GATCGGAAGAGCACACGTCTGAACTCCAGTCACGTGAAACGATCTCGTATGCCGTCTTCTGCTTG
>TruSeq_Adapter_Index_20
GATCGGAAGAGCACACGTCTGAACTCCAGTCACGTGGCCTTATCTCGTATGCCGTCTTCTGCTTG
>TruSeq_Adapter_Index_21
GATCGGAAGAGCACACGTCTGAACTCCAGTCACGTTTCGGAATCTCGTATGCCGTCTTCTGCTTG
>TruSeq_Adapter_Index_22
GATCGGAAGAGCACACGTCTGAACTCCAGTCACCGTACGTAATCTCGTATGCCGTCTTCTGCTTG
>TruSeq_Adapter_Index_23
GATCGGAAGAGCACACGTCTGAACTCCAGTCACGAGTGGATATCTCGTATGCCGTCTTCTGCTTG
>TruSeq_Adapter_Index_25
GATCGGAAGAGCACACGTCTGAACTCCAGTCACACTGATATATCTCGTATGCCGTCTTCTGCTTG
>TruSeq_Adapter_Index_27
GATCGGAAGAGCACACGTCTGAACTCCAGTCACATTCTTTATCTCGTATGCCGTCTTCTGCTTG
>I5_Nextera_Transposase_1
CTGTCTCTTATACACATCTGACGCTGCCGACGA
>I7_Nextera_Transposase_1
CTGTCTCTTATACACATCTCCGAGCCCACGAGAC
>I5_Nextera_Transposase_2
CTGTCTCTTATACACATCTCTGATGGCGCGAGGGAGGC
>I7_Nextera_Transposase_2
CTGTCTCTTATACACATCTCTGAGCGGGCTGGCAAGGC
>I5_Primer_Nextera_XT_and_Nextera_Enrichment_[N/S/E]501
GACGCTGCCGACGAGCGATCTAGTGTAGATCTCGGTGGTCGCCGTATCATT
>I5_Primer_Nextera_XT_and_Nextera_Enrichment_[N/S/E]502
GACGCTGCCGACGAATAGAGAGGTGTAGATCTCGGTGGTCGCCGTATCATT
>I5_Primer_Nextera_XT_and_Nextera_Enrichment_[N/S/E]503
:
```

Reads Filtering and mapping



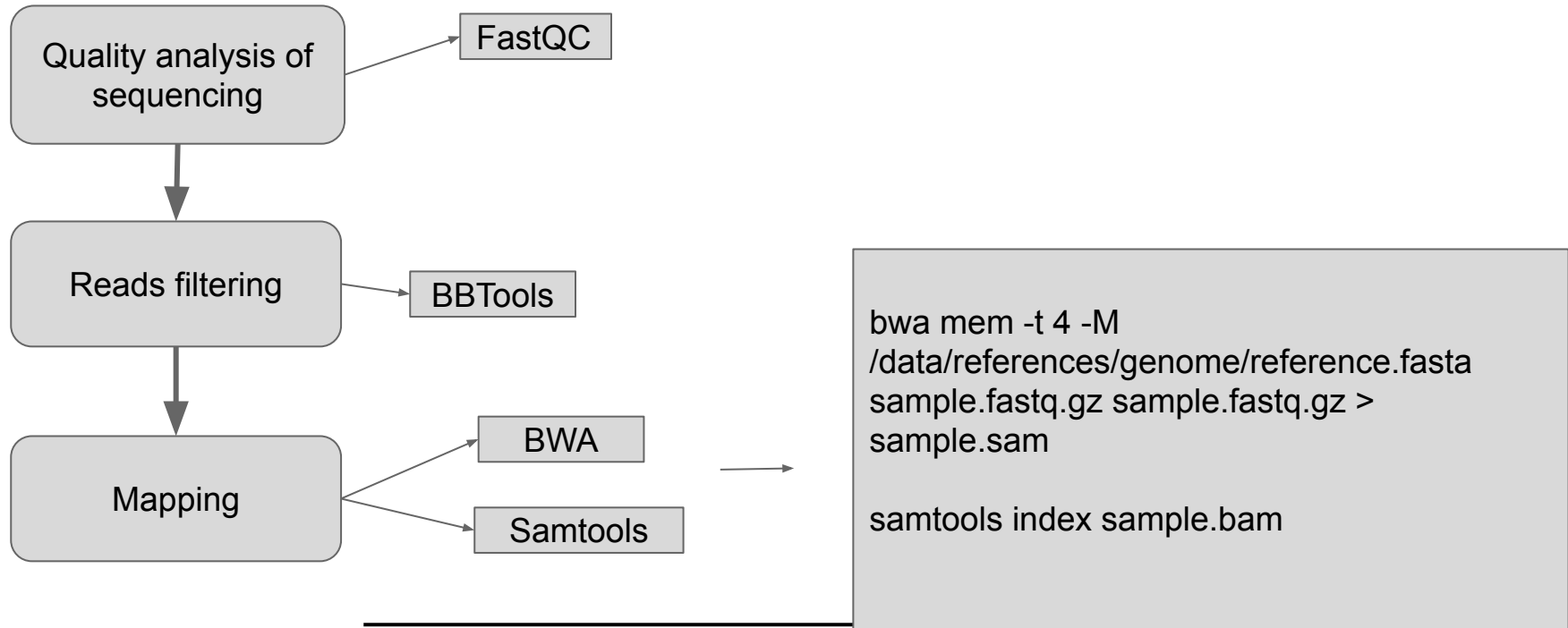
Reads Filtering and mapping



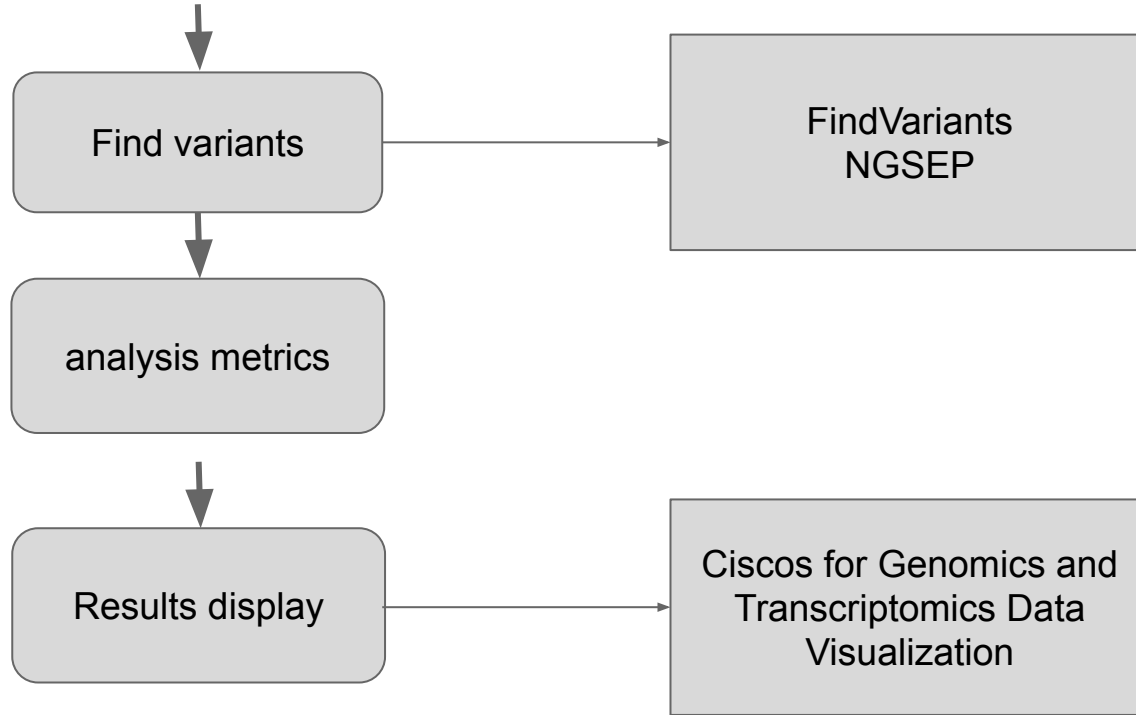
Sequence	Count	Percentage	Possible Source
GGG	375479	0.11175579240642547	No Hit

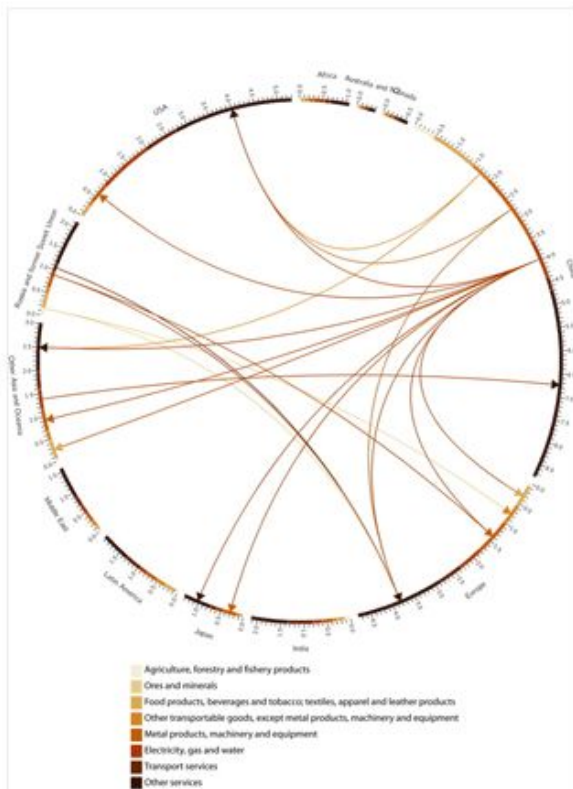
How to remove those sequences?

Reads Filtering and mapping

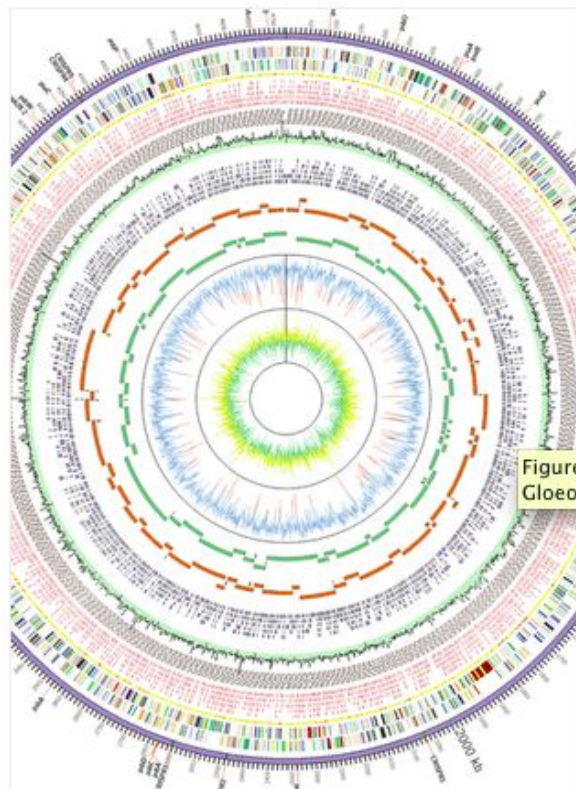


Reads Filtering and mapping

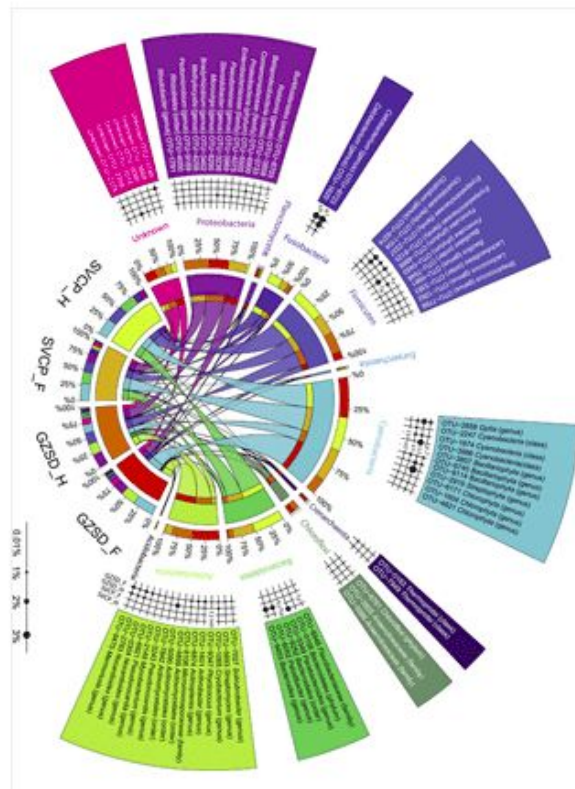




▲ 4 · 23 Oct 2013 | Kanemoto K, Moran D, Lenzen M *et al.* (2013) [International trade undermines national emission reduction targets: New evidence from air pollution](#) *Global Environmental Change*



▲ 5 · 23 October 2013 | Saw JHW, Schatz M, Brown MV *et al.* (2013) [Cultivation and Complete Genome Sequencing of Gloeobacter kilaeuensis sp. nov., from a Lava Cave in Kilauea Caldera, Hawaii](#) *PLoS One* 8:e76376.



▲ 6 · 17 Oct 2013 | Ye L, Amberg J, Chapman D *et al.* (2013) [Fish gut microbiota analysis differentiates physiology and behavior of invasive Asian carp and indigenous American fish](#) *The ISME journal*