# Tutorial for the 'Genotype' module

This module provides functions to extract and mine large-scale cucurbit variant datasets. The complete list of variant projects, reference genomes used to call the variants and sample list in each variant project is provided here (http://cucurbitgenomics.org/v2/geno_project).

## Retrieve variants within a gene

This function takes a gene ID from the reference genome used for variant calling for the select project and outputs a list of variants within the queried gene (**Figure 1**). Each variant ID is linked to its detailed page that provides the annotation of the variant, 500-bp up- and downstream flanking sequences of the variant (**Figure 2**), and allele frequencies in different groups of the population in a bar graph (**Figure 3**). Each bar is linked to a page that displays the genotype information of the variant in individual accessions from the corresponding group (**Figure 4**).

## Retrieve variants within a genome range

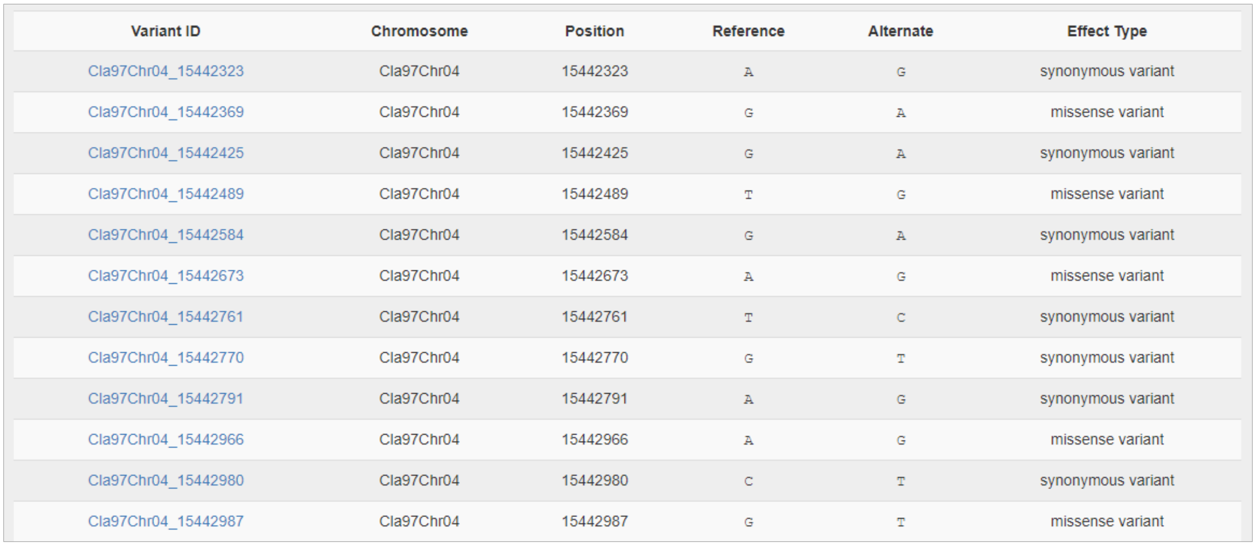
This function outputs a list of variants within a specific region of the reference genome that was used for variant calling of the selected project (**Figure 1**). After selecting a specific project, the corresponding list of chromosomes are automatically populated. For genome resequencing projects, the queried region should be <=30 kb in size, while for GBS projects there is no size limit.

## Variant genotype in a list of samples

This function outputs the detailed information of a specific variant including its annotation, 500-bp up- and downstream flanking sequences (**Figure 2**), and allele frequencies in different groups of the input samples (**Figure 3**). The input samples must be the sample IDs from the selected project (e.g., XG0001, XG0002… from project P01). Leaving the 'Sample list' field empty will output for all samples in the selected project.

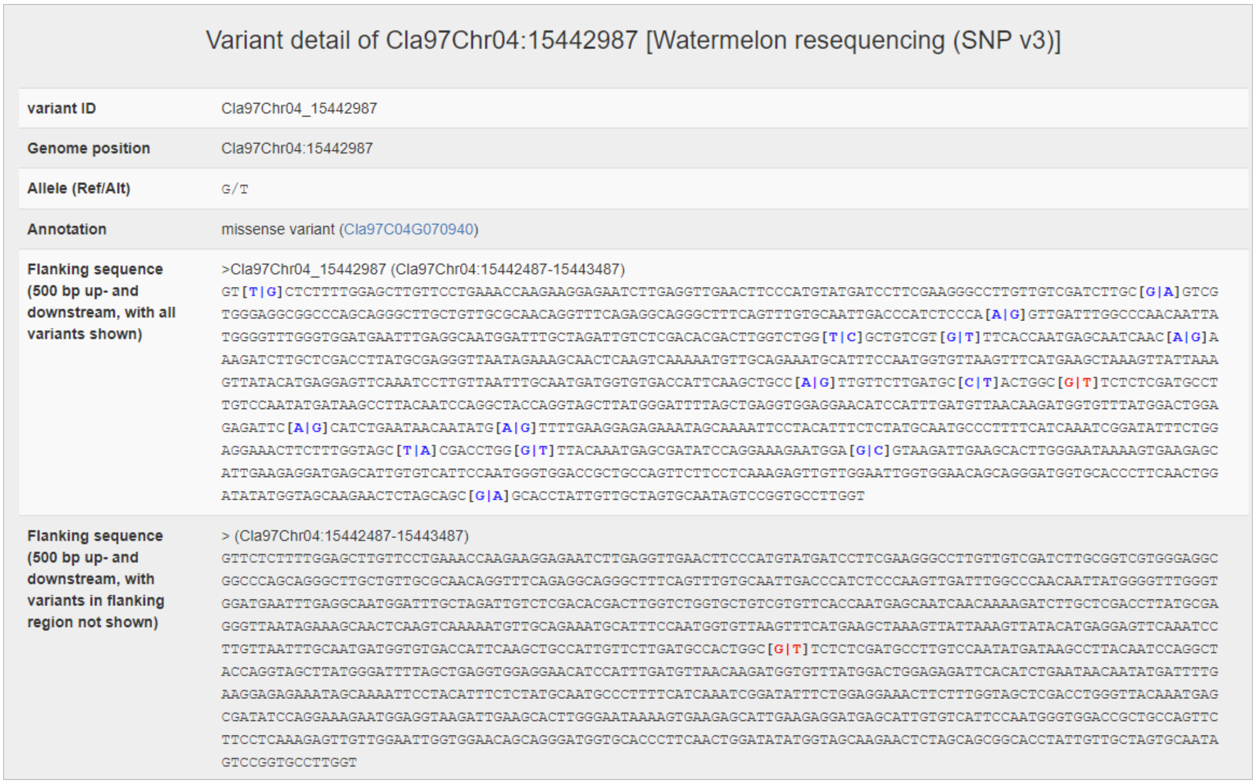
## Download variant data for a list of samples

This function allows to download a list of variants within a specific region of the reference genome in a subset of (or all) samples from the selected project. The output file can be in genotype or vcf format.



| Variant ID | Chromosome | Position | Reference | Alternate | Effect Type |
|---|---|---|---|---|---|
| Cla97Chr04_15442323 | Cla97Chr04 | 15442323 | A | G | synonymous variant |
| Cla97Chr04_15442369 | Cla97Chr04 | 15442369 | G | A | missense variant |
| Cla97Chr04_15442425 | Cla97Chr04 | 15442425 | G | A | synonymous variant |
| Cla97Chr04_15442489 | Cla97Chr04 | 15442489 | T | G | missense variant |
| Cla97Chr04_15442584 | Cla97Chr04 | 15442584 | G | A | synonymous variant |
| Cla97Chr04_15442673 | Cla97Chr04 | 15442673 | A | G | missense variant |
| Cla97Chr04_15442761 | Cla97Chr04 | 15442761 | T | C | synonymous variant |
| Cla97Chr04_15442770 | Cla97Chr04 | 15442770 | G | T | synonymous variant |
| Cla97Chr04_15442791 | Cla97Chr04 | 15442791 | A | G | synonymous variant |
| Cla97Chr04_15442966 | Cla97Chr04 | 15442966 | A | G | missense variant |
| Cla97Chr04_15442980 | Cla97Chr04 | 15442980 | C | T | synonymous variant |
| Cla97Chr04_15442987 | Cla97Chr04 | 15442987 | G | T | missense variant |

Figure 1



Variant detail of Cla97Chr04:15442987 [Watermelon resequencing (SNP v3)]

| | |
|---|---|
| **variant ID** | Cla97Chr04_15442987 |
| **Genome position** | Cla97Chr04:15442987 |
| **Allele (Ref/Alt)** | G/T |
| **Annotation** | missense variant (Cla97C04G070940) |

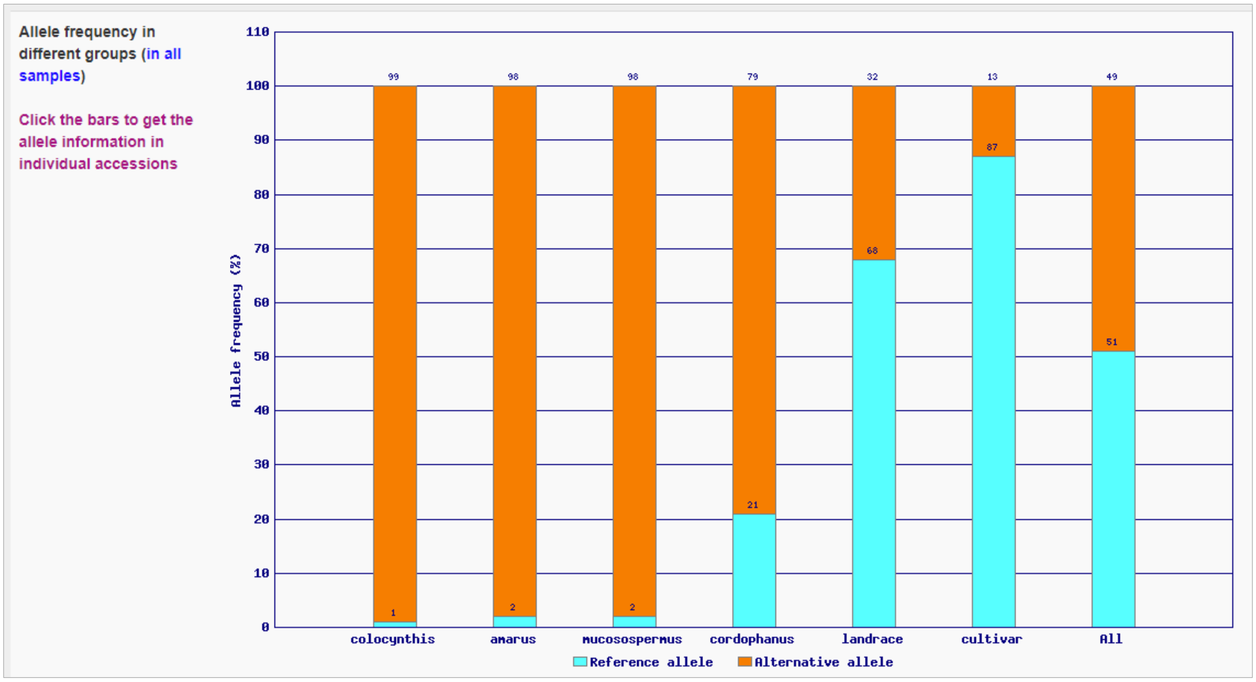
**Flanking sequence (500 bp up- and downstream, with all variants shown)**

```
>Cla97Chr04_15442987 (Cla97Chr04:15442487-15443487)
GT[T|G]CTCTTTTGGAGCTTGTTCCTGAAACCAAGAAGGAGAATCTTGAGGTTGAACTTCCCATGTATGATCCTTCGAAGGGCCTTGTTGTCGATCTTGC[G|A]GTCG
TGGGAGGCGGCCCAGCAGGGCTTGCTGTTGCGCAACAGGTTTCAGAGGCAGGGCTTTCAGTTTGTGCAATTGACCCATCTCCCA[A|G]GTTGATTTGGCCCAACAATTA
TGGGGTTTGGGTGGATGAATTTGAGGCAATGGATTTGCTAGATTGTCTCGACACGACTTGGTCTGG[T|C]GCTGTCGT[G|T]TTCACCAATGAGCAATCAAC[A|G]A
AAGATCTTGCTCGACCTTATGCGAGGGTTAATAGAAAGCAACTCAAGTCAAAAATGTTGCAGAAATGCATTTCCAATGGTGTTAAGTTTCATGAAGCTAAAGTTATTAAA
GTTATACATGAGGAGTTCAAATCCTTGTTAATTTGCAATGATGGTGTGACCATTCAAGCTGCC[A|G]TTGTTCTTGATGC[C|T]ACTGGC[G|T]TCTCTCGATGCCT
TGTCCAATATGATAAGCCTTACAATCCAGGCTACCAGGTAGCTTATGGGATTTTAGCTGAGGTGGAGGAACATCCATTTGATGTTAACAAGATGGTGTTTATGGACTGGA
GAGATTC[A|G]CATCTGAATAACAATATG[A|G]TTTTGAAGGAGAGAAATAGCAAAATTCCTACATTTCTCTATGCAATGCCCTTTTCATCAAATCGGATATTTCTGG
AGGAAACTTCTTTGGTAGC[T|A]CGACCTGG[G|T]TTACAAATGAGCGATATCCAGGAAAGAATGGA[G|C]GTAAGATTGAAGCACTTGGGAATAAAAGTGAAGAGC
ATTGAAGAGGATGAGCATTGTGTCATTCCAATGGGTGGACCGCTGCCAGTTCTTCCTCAAAGAGTTGTTGGAATTGGTGGAACAGCAGGGATGGTGCACCCTTCAACTGG
ATATATGGTAGCAAGAACTCTAGCAGC[G|A]GCACCTATTGTTGCTAGTGCAATAGTCCGGTGCCTTGGT
```

**Flanking sequence (500 bp up- and downstream, with variants in flanking region not shown)**

```
> (Cla97Chr04:15442487-15443487)
GTTCTCTTTTGGAGCTTGTTCCTGAAACCAAGAAGGAGAATCTTGAGGTTGAACTTCCCATGTATGATCCTTCGAAGGGCCTTGTTGTCGATCTTGCGGTCGTGGGAGGC
GGCCCAGCAGGGCTTGCTGTTGCGCAACAGGTTTCAGAGGCAGGGCTTTCAGTTTGTGCAATTGACCCATCTCCCAAGTTGATTTGGCCCAACAATTATGGGGTTTGGGT
GGATGAATTTGAGGCAATGGATTTGCTAGATTGTCTCGACACGACTTGGTCTGGTGCTGTCGTGTTCACCAATGAGCAATCAACAAAAGATCTTGCTCGACCTTATGCGA
GGGTTAATAGAAAGCAACTCAAGTCAAAAATGTTGCAGAAATGCATTTCCAATGGTGTTAAGTTTCATGAAGCTAAAGTTATTAAAGTTATACATGAGGAGTTCAAATCC
TTGTTAATTTGCAATGATGGTGTGACCATTCAAGCTGCCATTGTTCTTGATGCCACTGGC[G|T]TCTCTCGATGCCTTGTCCAATATGATAAGCCTTACAATCCAGGCT
ACCAGGTAGCTTATGGGATTTTAGCTGAGGTGGAGGAACATCCATTTGATGTTAACAAGATGGTGTTTATGGACTGGAGAGATTCACATCTGAATAACAATATGATTTTG
AAGGAGAGAAATAGCAAAATTCCTACATTTCTCTATGCAATGCCCTTTTCATCAAATCGGATATTTCTGGAGGAAACTTCTTTGGTAGCTCGACCTGGGTTACAAATGAG
CGATATCCAGGAAAGAATGGAGGTAAGATTGAAGCACTTGGGAATAAAAGTGAAGAGCATTGAAGAGGATGAGCATTGTGTCATTCCAATGGGTGGACCGCTGCCAGTTC
TTCCTCAAAGAGTTGTTGGAATTGGTGGAACAGCAGGGATGGTGCACCCTTCAACTGGATATATGGTAGCAAGAACTCTAGCAGCGGCACCTATTGTTGCTAGTGCAATA
GTCCGGTGCCTTGGT
```

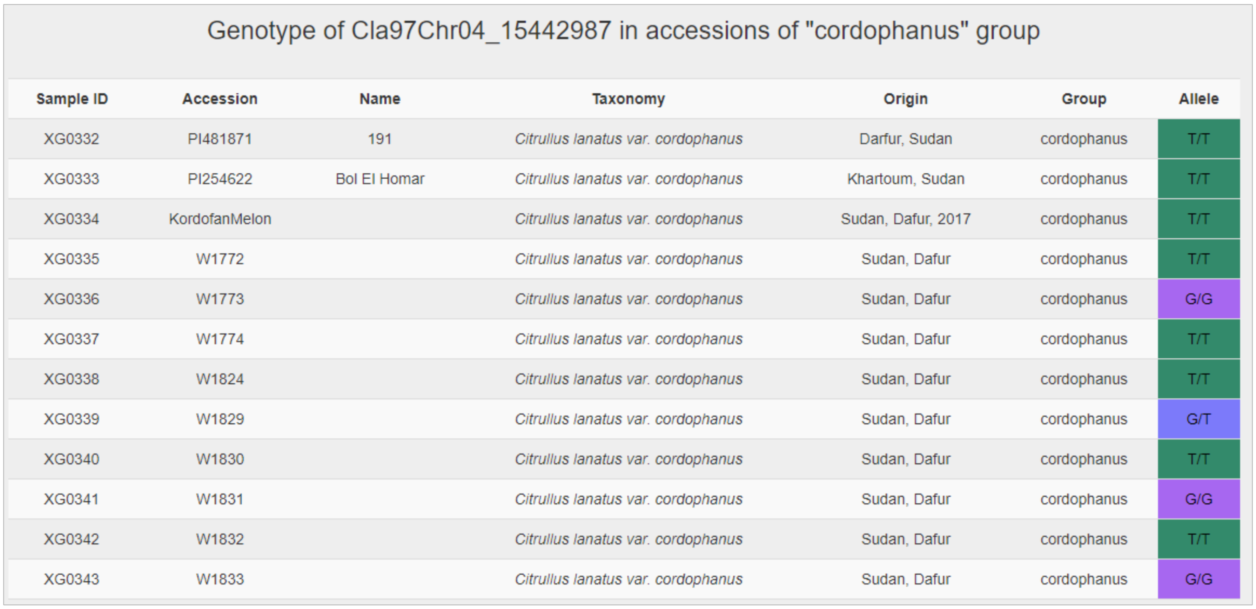
Figure 2



Allele frequency in different groups (in all samples)

Click the bars to get the allele information in individual accessions

Figure 3



Genotype of Cla97Chr04_15442987 in accessions of "cordophanus" group

| Sample ID | Accession | Name | Taxonomy | Origin | Group | Allele |
|---|---|---|---|---|---|---|
| XG0332 | PI481871 | 191 | *Citrullus lanatus var. cordophanus* | Darfur, Sudan | cordophanus | T/T |
| XG0333 | PI254622 | Bol El Homar | *Citrullus lanatus var. cordophanus* | Khartoum, Sudan | cordophanus | T/T |
| XG0334 | KordofanMelon | | *Citrullus lanatus var. cordophanus* | Sudan, Dafur, 2017 | cordophanus | T/T |
| XG0335 | W1772 | | *Citrullus lanatus var. cordophanus* | Sudan, Dafur | cordophanus | T/T |
| XG0336 | W1773 | | *Citrullus lanatus var. cordophanus* | Sudan, Dafur | cordophanus | G/G |
| XG0337 | W1774 | | *Citrullus lanatus var. cordophanus* | Sudan, Dafur | cordophanus | T/T |
| XG0338 | W1824 | | *Citrullus lanatus var. cordophanus* | Sudan, Dafur | cordophanus | T/T |
| XG0339 | W1829 | | *Citrullus lanatus var. cordophanus* | Sudan, Dafur | cordophanus | G/T |
| XG0340 | W1830 | | *Citrullus lanatus var. cordophanus* | Sudan, Dafur | cordophanus | T/T |
| XG0341 | W1831 | | *Citrullus lanatus var. cordophanus* | Sudan, Dafur | cordophanus | G/G |
| XG0342 | W1832 | | *Citrullus lanatus var. cordophanus* | Sudan, Dafur | cordophanus | T/T |
| XG0343 | W1833 | | *Citrullus lanatus var. cordophanus* | Sudan, Dafur | cordophanus | G/G |

Figure 4