**Genomic Selection Workshop**

Björn Usadel, Alisandra K. Denton

[Finding Data 3](#_Toc440636144)

[Basics 4](#_Toc440636145)

[Unix: Command-line basics 4](#_Toc440636146)

[Unix: Setting permissions 6](#_Toc440636147)

[Unix: Advanced command-line tools 7](#_Toc440636148)

[Quality Control 9](#_Toc440636149)

[Formatting reads: adapter trimming and quality filtering 11](#_Toc440636150)

[Demultiplexing 12](#_Toc440636151)

[SNP Calling 13](#_Toc440636152)

[Read mapping: Bowtie2 13](#_Toc440636153)

[The Sam Format 15](#_Toc440636154)

[Output visualization 16](#_Toc440636155)

[Samtools from alignment to SNPs 17](#_Toc440636156)

[The vcf format 19](#_Toc440636157)

[Filtering SNPs 19](#_Toc440636158)

[Checking our SNPs 20](#_Toc440636159)

[SNPs on Gramene 20](#_Toc440636160)

[SNPs on dbSNP 22](#_Toc440636161)

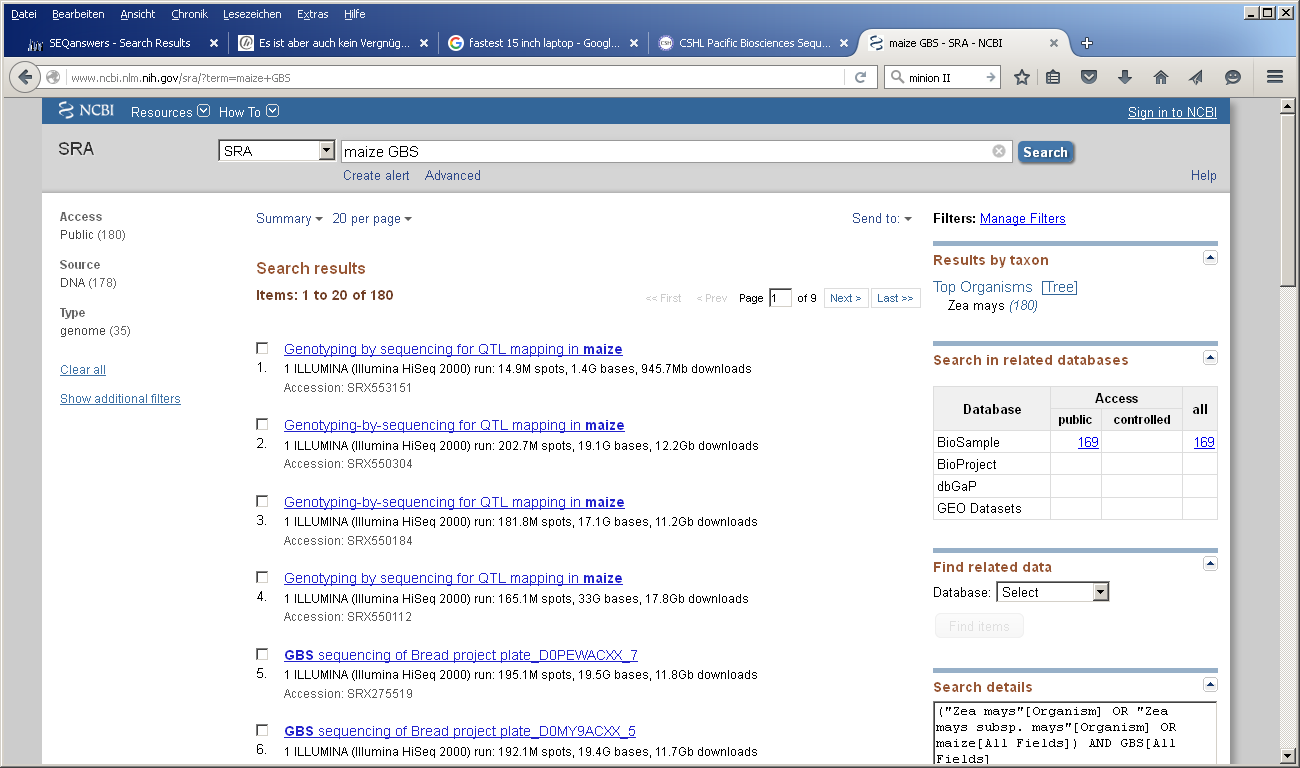
[Phasing Example Analysis 23](#_Toc440636162)

[Phasing 23](#_Toc440636163)

[Copyright 23](#_Toc440636164)

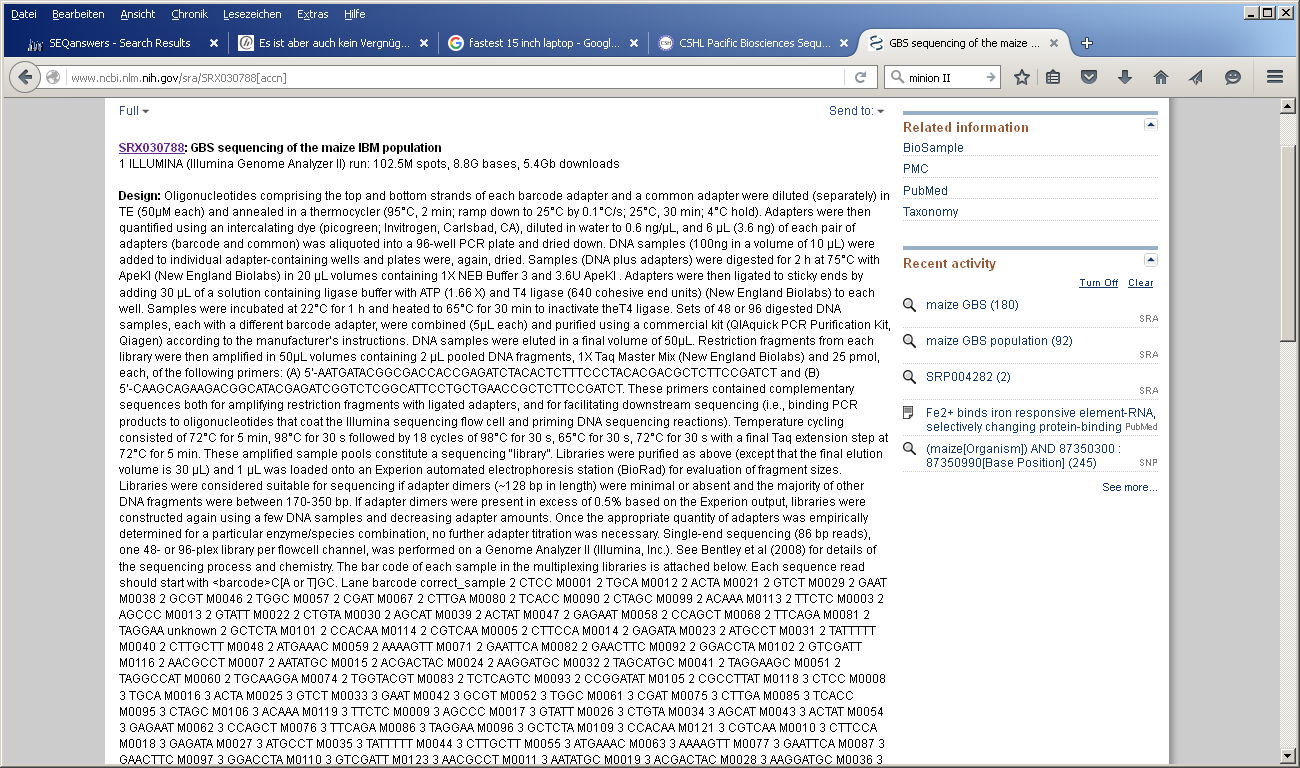
Finding Data

We navigate to the SRA (short read archive of the NCBI). There we search for maize GBS.

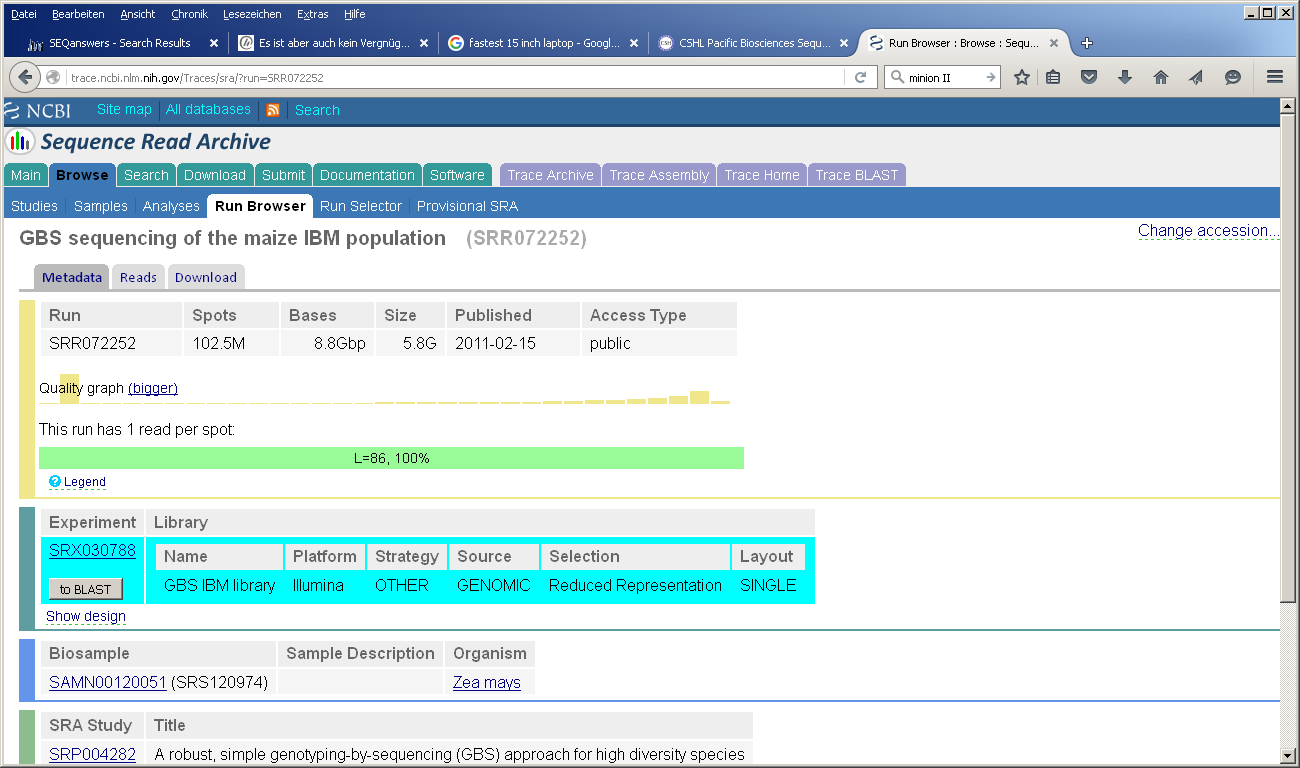


The very last result is the one we will be working with. This is a GBS study for the maize IBM population (you can also search for maize GBS IBM),

This is actually described well.



By clicking on run at the bottom you get additional information



Basics

During this part we will address, how to navigate through Unix and how to invoke the main programs used.

Unix: Command-line basics

Unix is an operating system that is running on Linux machines and Macs. Nowadays the graphical user interface (GUI) of Linux basically looks like any other operating system that you are used to and we recommend that you largely use this GUI and only switch to the terminal if you have to. We will teach you now how to access the system through the terminal/shell, how to move through folders and a suite of little commands that are helpful in working with large datasets.

In terminal/shell, you can access directories just like you can through a GUI. Modern unix-based operating systems also have convenient GUIs, but you need to know how to work in terminal to run programs.

The “#” character signifies a comment or an instruction for you

1. # open terminal by finding the black box icon on the home screen
2. # list the contents of the directory you are in
3. ls
4. # move to your home directory
5. cd
6. # list the contents of the home directory
7. ls
8. # move to the examples directory
9. cd Documents/examples
10. # move the directory above the directory you are in
11. cd ..

Unix is case sensitive so cd RNASEQ is different from cd rnaseq

Also when you type **cd** it **has to be followed by a space**

cd RNASEQ but also

cd .. not cd..

Many of the programs require that the file you are working on and the program you are working with to be in the same folder. If you do not really know what you are doing, it is a safe idea to use small helper programs by copying them into the folder where the data files are. You can move the programs through the GUI just as you do with other operating systems. (If you are familiar with \*nix you can use symlinks as well or do it properly if you have the time)

1. Create two folders in your home directory, the folder /results and the folder /stuff by using the GUI.
2. Look again into your home folder to see what has changed
3. # open terminal
4. # move to your home folder
5. cd
6. # list the contents
7. ls
8. # now you can see the two folders you have added to your home folder – see terminal and GUI are really not that different ☺

If a program does not run, you might not have the permissions to run a program. If you are not the admin on the computer, make sure that you are in future or make your admin install and test all programs that you might want to use. In that case, the admin might give you authority to run the programs. If you need to do it yourself, it works like this

Unix: Setting permissions

There are several cases in which you might want to change the permissions of a file or even a folder.

If you try to read/write/execute a file you own, you can just add the permission you need. Therefore you need to understand the permissions given to files in UNIX. A file knows three entities it can be accessed from: u) The file owner (user) g) the usergroup the file belongs to o) any other person (also known as: the world). For each of the entities a file has separate permissions for reading, writing and executing it. First try to understand the permissions of a few files.

1. # print a detailed list of the current directory
2. ls -al

What you see now is a list of the files in the directory you’re currently working in. In the output you might recognize a part that looks something like: -rwxr-x--- or something similar. This is the attribute that controls file access. It always consists of 10 characters. The first character can be either “d” or “-“ and tells the system, whether the file is a directory or not. The residing 9 characters are in groups of three the permissions for user (which is owner) group (which is usually the owner’s primary group) and others. The permissions are read, write and execute. To change that you need the command “chmod” and a three character parameter, that states 1) whose permission to change, 2) add with “+” or remove with “-“, 3) the permission to “r” or “w” or “x”.

Create a New Document in the text editor containing the following two lines:  
#!/bin/bash  
date

Save the Document as file\_permissions.sh into your Home Directory.

1. # After changing to your home directory, try to execute the script
2. cd
3. ./file\_permissions.sh
4. # That won’t work, check the permissions, add executable to it and try again.
5. ls -al
6. chmod u+x file\_permissions.sh
7. ls -al
8. # Note the change. Now start the script again
9. ./file\_permissions.sh

For the following, IT people would bang their heads against a wall or jump off a bridge, but if you work on your own computer, where no one else has access, you might also give all permissions to everyone (even permission to delete the file)

1. # Check the current permissions. Make the file grant all permissions to everyone and check again
2. ls -al
3. chmod 777 file\_permissions.sh
4. ls -al

Unix: Advanced command-line tools

During our work with large datasets we have come across a few helpful commands that can be used to easily extract information. We will show you a few.

Look at the ReadFile1.fa and at the ReadFile2.fa

1. # View a file page by page
2. cd Documents/examples
3. less ReadFile1.fa
4. # press the Enter key, the file will move forward one line
5. # press the space key, the file will move forward one page
6. # press the Q key to leave
7. # Sometimes you only want to look at the first or last few lines
8. head ReadFile1.fa
9. tail ReadFile1.fa
10. # View a whole file in the shell
11. cat ReadFile1.fa

If you want to look at a file using the GUI on a Windows system, do **NOT** use the text editor. We recommend Notepad++ (<http://notepad-plus-plus.org/>). This will open even huge files with no to little problems. Notepad++ is also powerful enough to process translations of strings, something that can be really useful to reformat data. We will revisit Notepad++ when we look at extracting biological data.

Now, we will introduce a number of useful commands that are built into the operating system by showing examples how we use them to analyze files. We need to explain the | (“pipe”) character before. If you work on the command line as we do, you can tell the system to first do this, then do the next thing and the next by piping commands:

**Do this | with the result do that | and with the result of this do the next thing.**

We can write the results into a file in the same folder by ending the chain with > write to this file.txt. If you make Linux do something stupid, such as get stuck in a loop, you can always get out by typing **Ctrl+C**. We will not make you do that in the course but you should remember **Ctrl+C**.

1. # Count the number of sequences in a fasta file
2. # Take every line that has a ‘>’ character | count the number lines
3. grep “>” ReadFile1.fa | wc –l
4. # grep looks for lines (!) which have the ‘>’ character, then we move on and use wc –l to count the number of lines we have
5. # wc can also be used with the options –w (number of words) and –m (number of characters)
6. # Estimate the number of bases in a fasta file
7. # Take every line that does not have a ‘>’ character by using –v | count the number of characters
8. grep “>” ReadFile1.fa -v |wc -m
9. grep takes all lines that do not have the ‘>’ which in a fasta file are all lines except the headers, then wc –m counts all characters in the file
10. # Join two or more files together (files that are so large, that copy/paste operations no longer work since they require too much memory)
11. cat ReadFile1.fa ReadFile2.fa > Readfile1\_2.fa

Quality Control

First let’s get a data file. Common ways to download sequence files are shown below. The first command works with SRAtoolkit to download data from the Sequence Read Archives. Another general purpose downloading tool is wget. Finally you can download most files via your browser.

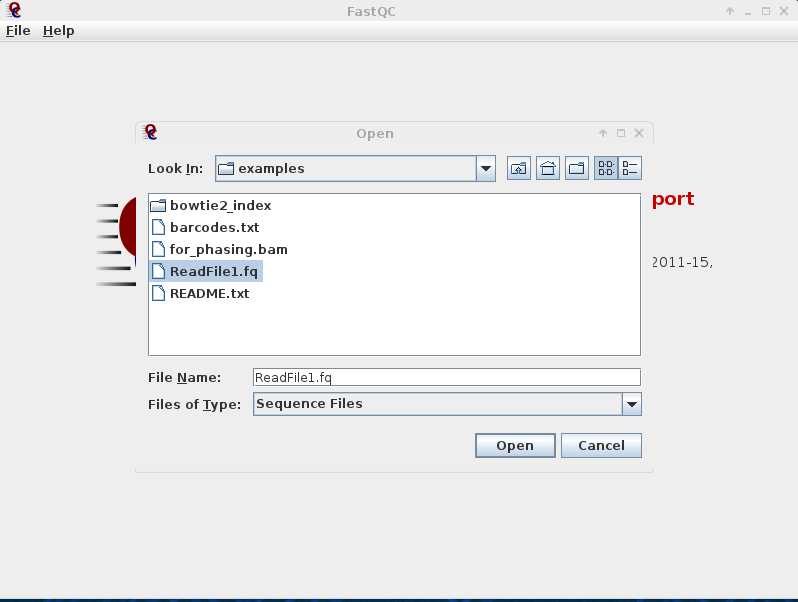
1. **# here are two ways to download a sequence from SRA**
2. **fastq-dump SRR072252**
3. **# this one will download all the files in the study SRP004282**
4. **wget -r** ftp://ftp.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByStudy/sra/SRP/SRP004/SRP004282/
5. **# note: don’t wait on these commands but instead continue with provided test data**

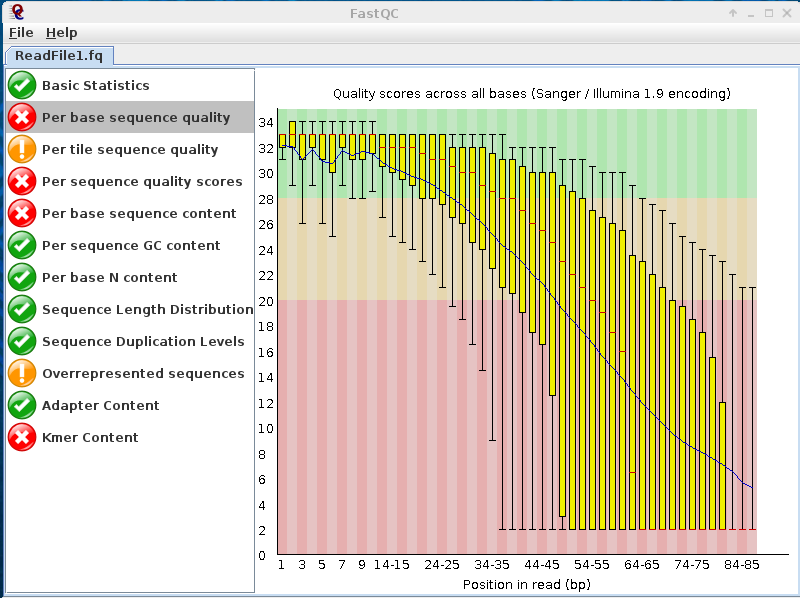
(The browser option also works well in Windows or MacOS so if you prefer processing the data on your own laptop, feel free to do so)

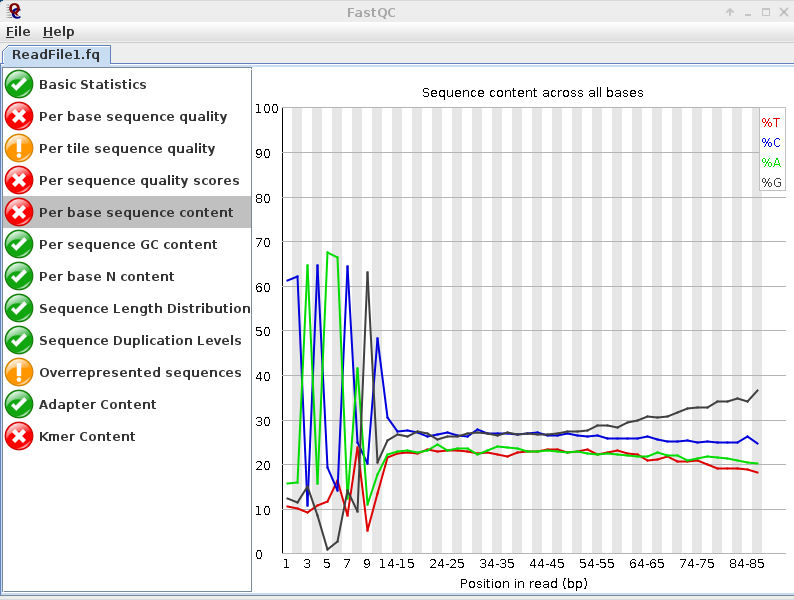
We use FastQC to quickly assess the quality of the raw sequencing data. It is a relatively fast program which offers a user friendly interface which allows one to quickly evaluate the datasets.

1. fastqc

After having started the program we use the graphical user interface to load our dataset. FastQC supports many different FastQ formats including GZip compressed FastQ files. After the data has been loaded and analyzed by fastQC, we can go through the different tabs and visualize the data quality. FastQC provides an evaluation for each module and indicates whether the results seem normal (green tick), slightly abnormal (orange triangle) or very unusual (red cross).



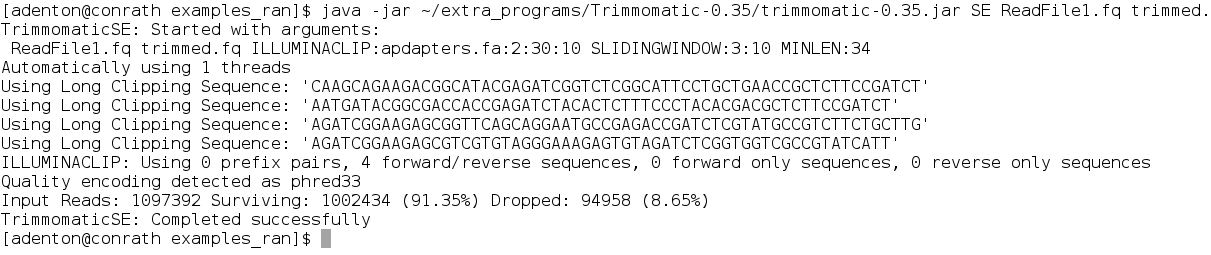




Formatting reads: adapter trimming and quality filtering

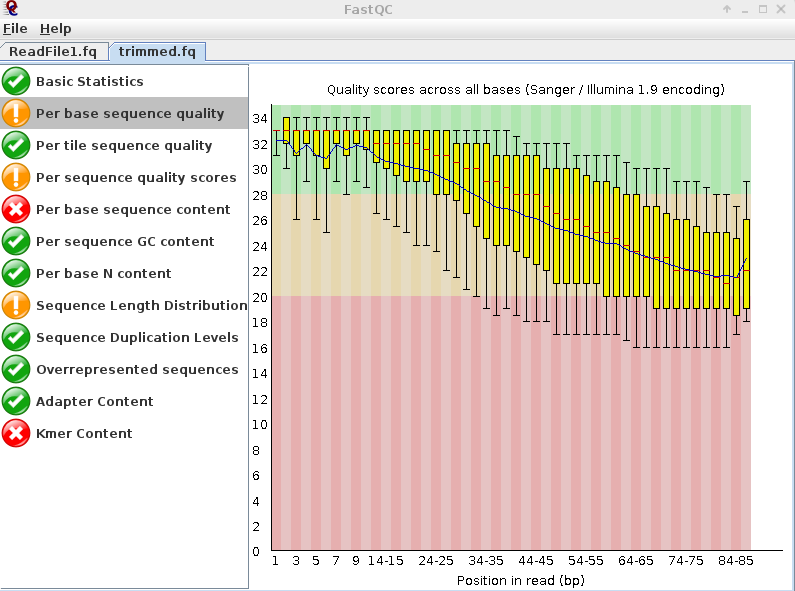
For trimming the adapters and PCR primers from the reads we will use trimmomatic. Trimmomatic has several different functions which include (i) adapter clipping (ii) quality clipping and (iii) discarding reads below a given quality or length. Trimmomatic is a modular tool and the user can perform only the functions they require. We perform adapter clipping, removal of lower quality bases (using a sliding window of size 3 and average quality 10) and reads that are too short to have a chance of being useful (below 24 bases). Before running Trimmomatic, we first need to know which library preparation kit was used and whether the library was run in paired-end (or single ended mode, the present data is single ended data so we indicate this by the SE switch

1. java -jar ~/Documents/software/Trimmomatic-0.35/trimmomatic-0.35.jar SE ReadFile1.fq trimmed.fq ILLUMINACLIP:adapters.fa:2:30:10 SLIDINGWINDOW:3:10 MINLEN:34



As you see ca 9% of the reads did not pass this lenient filtering criterion.

If you trim too stringently FASTQC might not get the quality encoding right! Then all data looks awkward and of low sequence quality



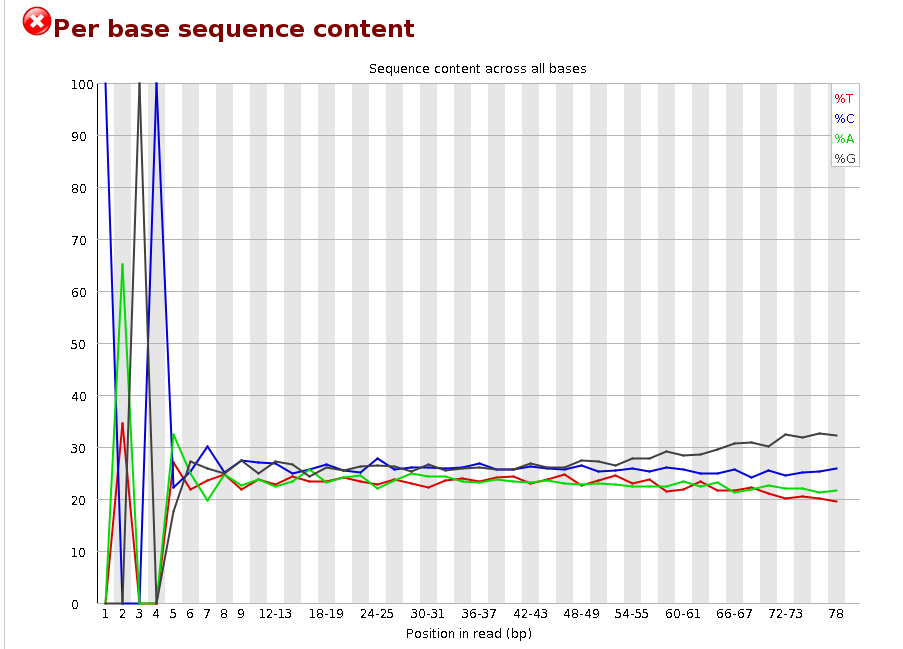
Now let’s have a look at the trimmed file in fastqc again. As you can see the quality got better but we did lose some reads.

Demultiplexing

1. # make a directory for the split files
2. mkdir demultiplexed
3. process\_radtags -f trimmed.fq -e apeKI -o demultiplexed -b barcodes.txt
4. # take a look at the results
5. ls demultiplexed/
6. # now every sequence will start with the site left behind by ApeKI
7. # as you can see in the fastqc report
8. fastqc demultiplexed/sample\_CCACAA.fq

In most cases one sequences highly multiplexed DNA when performing genotyping by sequencing, and for cases like RADseq the adapters are frequently customized, making the standard demultiplexing for Illumina adapters unhelpful. For the samples used here, the format is <barcode>[C(A or T)GC]<remaining sequence>. The barcode is between 4 and 8 bp long. The [C(A or T)GC] is the remainder of the ApeKI restriction site, which was used to enrich non-methylated DNA regions. A RAD-specific tool like process\_radtags from Stacks can be used to split the file according to barcode and remove

Note we just snuck in a different way to run fastqc. By giving it the filename in the command line, you can skip using the gui if you want to run a larger batch of files. Now if you **open the resulting html file in your browser** and look at per base sequence content, you can see all the reads start with the motif left by ApeKI, meaning the barcodes were successfully identified, called, and trimmed.



SNP Calling

Read mapping: Bowtie2

First, we will download chromosome #1 from the maize genome (Of course you would want to download the whole genome but then the run time would be way too long)

1. wget ftp://ftp.ensemblgenomes.org/pub/plants/release-30/fasta/zea\_mays/dna/Zea\_mays.AGPv3.30.dna.chromosome.1.fa.gz
2. gunzip Zea\_mays.AGPv3.30.dna.chromosome.1.fa.gz

You should now have a file Zea\_mays.AGPv3.21.dna.chromosome.1.fa in your directory

Bowtie2 is a very fast and powerful short read aligner. But it can tolerate only very few mismatches as do all fast aligners.

1. To align the reads with bowtie you have to build an index of the reference first
2. # Build a bowtie2 index, optional here as results are provided
3. bowtie2-build Zea\_mays.AGPv3.30.dna.chromosome.1.fa MAIZE1test

(You can try executing the above command but it will take about 10-20 minutes, the finished index is provided under examples/bowtie2\_index/MAIZE1)

Now align your read library with bowtie to the index

* 1. –p2 tells bowtie to run on two computer cores (~twice as fast)
  2. -x points to the index file
  3. -U points to the [U]npaired read file (with paired reads use -1 and -2)
  4. -S tells bowtie where to output the alignment file in [S]am format

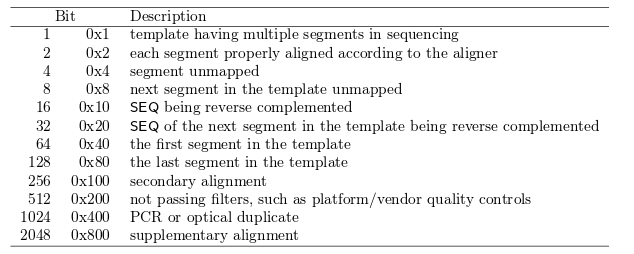
1. # map one of the files
2. bowtie2 –p2 -x bowtie2\_index/MAIZE1 -U demultiplexed/sample\_CCACAA.fq -S sample\_CCACAA.sam

Of course, while we are just taking one example for the workshop, you won’t want to map the files one by one in practice, but would rather loop through them. An example of how to do this is shown in the script examples/map\_all\_the\_samples.sh

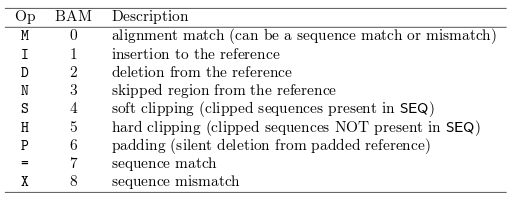
The Sam Format

The sam format contains alignment information by sequence. It has the following columns:

Probably the two most confusing, and yet important parts of the sam format are the FLAG and the CIGAR string.

The **FLAG** is a bitwise field, this makes a lot of sense when thinking about its encodings in 0's and 1's but for us it means we have to think of each of the elements in base 2 as an individual FLAG, hidden in in a base 10 number. Here's what they all mean:

Thus the number 272 means 256 + 16, means that it's a secondary alignment and that the SEQ is mapped as its reverse complement.

The CIGAR string encodes how the reads is aligned with any mismatches and gaps encoded as follows.

Most frequently we have perfect matches denoted as '[length]M', while a mismatch at position 12 of a 80bp alignment would look like 11M1X68M.

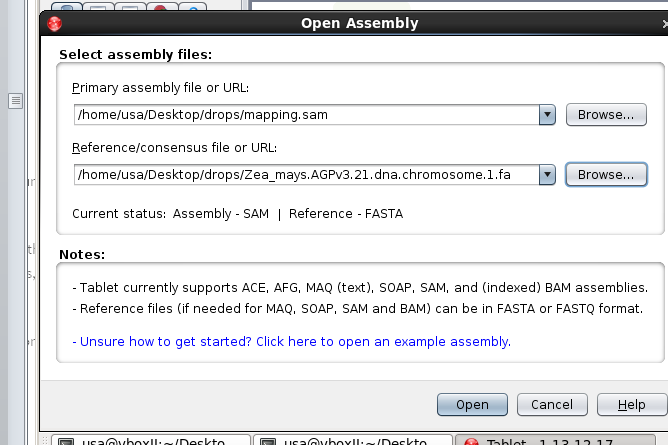
A good resource for further information on the sam format can be found at: <https://samtools.github.io/hts-specs/SAMv1.pdf>

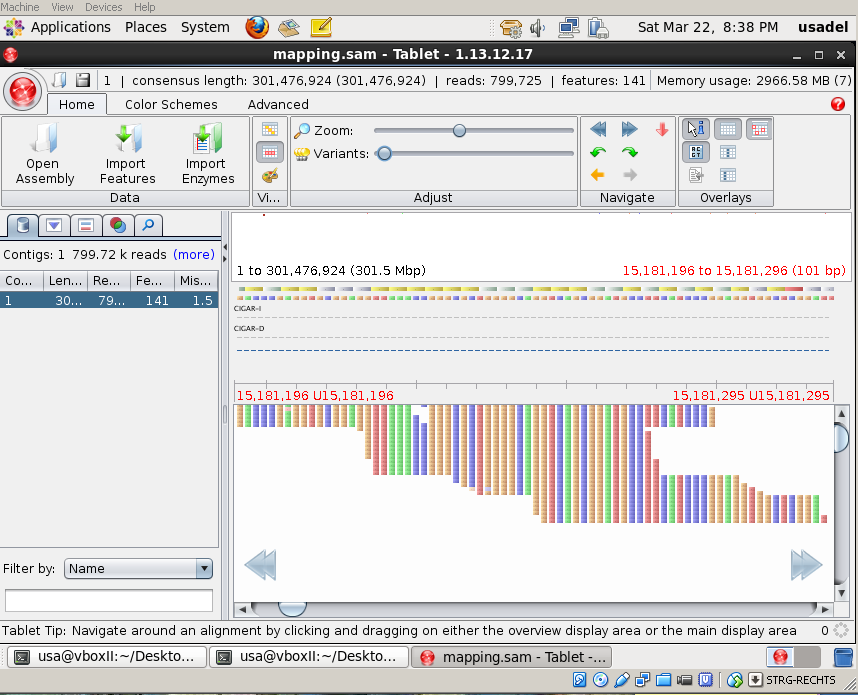
Output visualization

Bowtie output can be directly visualized with Tablet, a freeware available at <http://bioinf.scri.ac.uk/tablet/index.shtml>. You can use Bowtie not only for mapping to complete databases but also for looking at a single reference gene or contig by altering the reference in analogy to the process described above. If you use it that way, you can directly look at the output in a GUI using tablet.

Here we use the sam file as “assembly” and Chromosome #1 as reference. Efficient visualization requires indexes to be created for both the mapping file and the reference fasta file. So please continue through the ‘samtools faidx’ and ‘samtools index’ commands in the next section before trying visualization.

Tablet needs LOADS of memory to run. You might not be able to run Tablet with the settings shown here on the workshop laptops.





Samtools from alignment to SNPs

We now convert our ***sam*** file into a binary file these files are named ***bam*** files to speed things up a bit and to save space. This is of no concern here, but generally space is always an issue with next generation sequencing data.

1. samtools view -bS sample\_CCACAA.sam > mapping.bam
2. ls -altrh

You see that we save about 60% space now. Now we sort this binary file to allow better access

1. samtools sort mapping.bam > sorted.bam
2. # the next step isn’t necessary for SNP calling, but for visualization or other downstream processes
3. # the index allows more data rapid access in the bam file for some programs
4. samtools index sorted.bam

Now we need to create an index on our chromosome for samtools and SNP calling

1. samtools faidx Zea\_mays.AGPv3.30.dna.chromosome.1.fa
2. ls

(Check with ls if you have a file Zea\_mays.AGPv3.30.dna.chromosome.1.fa.fai now; this is the index)

Now we are finally ready to call some SNPs.

Samtools mpileup calculates the statistics for us, and the next command, bcftools view, does the actual SNP calling.

Options used for samtools mpileup

-u tells it the format to output (uncompressed bcf), which we won’t actually see as we ‘pipe’ it to the next command

-f points to the reference fasta

Options used for bcftools

-c tells bcftools to call SNPs

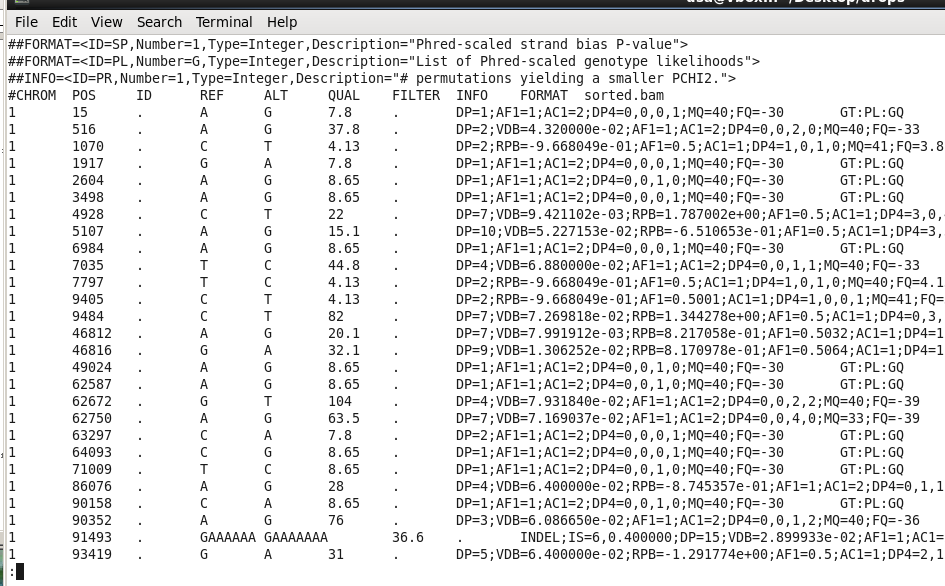
-v only reports potential variants

-g calls genotypes at variant sites

1. samtools mpileup -uf Zea\_mays.AGPv3.30.dna.chromosome.1.fa sorted.bam |bcftools call -vc - >snps.vcf

Note, there needs to be a space between ‘-‘ and ‘>’ in ‘- >snps.bcf ‘

Now we can have a look at the file snps.txt which is a human readable text file



The vcf format

What we get here is the so called “vcf” format

|  |  |  |
| --- | --- | --- |
| Col | Field | Description |
| 1 | CHROM | Chromosome name |
| 2 | POS | 1-based position. For an indel, this is the position preceding the indel. |
| 3 | ID | Variant identifier. Usually the dbSNP rsID. |
| 4 | REF | Reference sequence at POS involved in the variant. For a SNP, it is a single base. |
| 5 | ALT | Comma delimited list of alternative seuqence(s). |
| 6 | QUAL | Phred-scaled probability of all samples being homozygous reference. |
| 7 | FILTER | Semicolon delimited list of filters that the variant fails to pass. |
| 8 | INFO | Semicolon delimited list of variant information. |
| 9 | FORMAT | Colon delimited list of the format of individual genotypes in the following fields. |
| 10+ | Sample(s) | Individual genotype information defined by FORMAT. |

In the Info column we can see the number of reads covering a certain position indicated by the DP flag

Open the vcf file in a text editor as it also explains the info data sets

Filtering SNPs

As we see from the DP Info there are still quite some SNPs only supported by one read (DP=1)

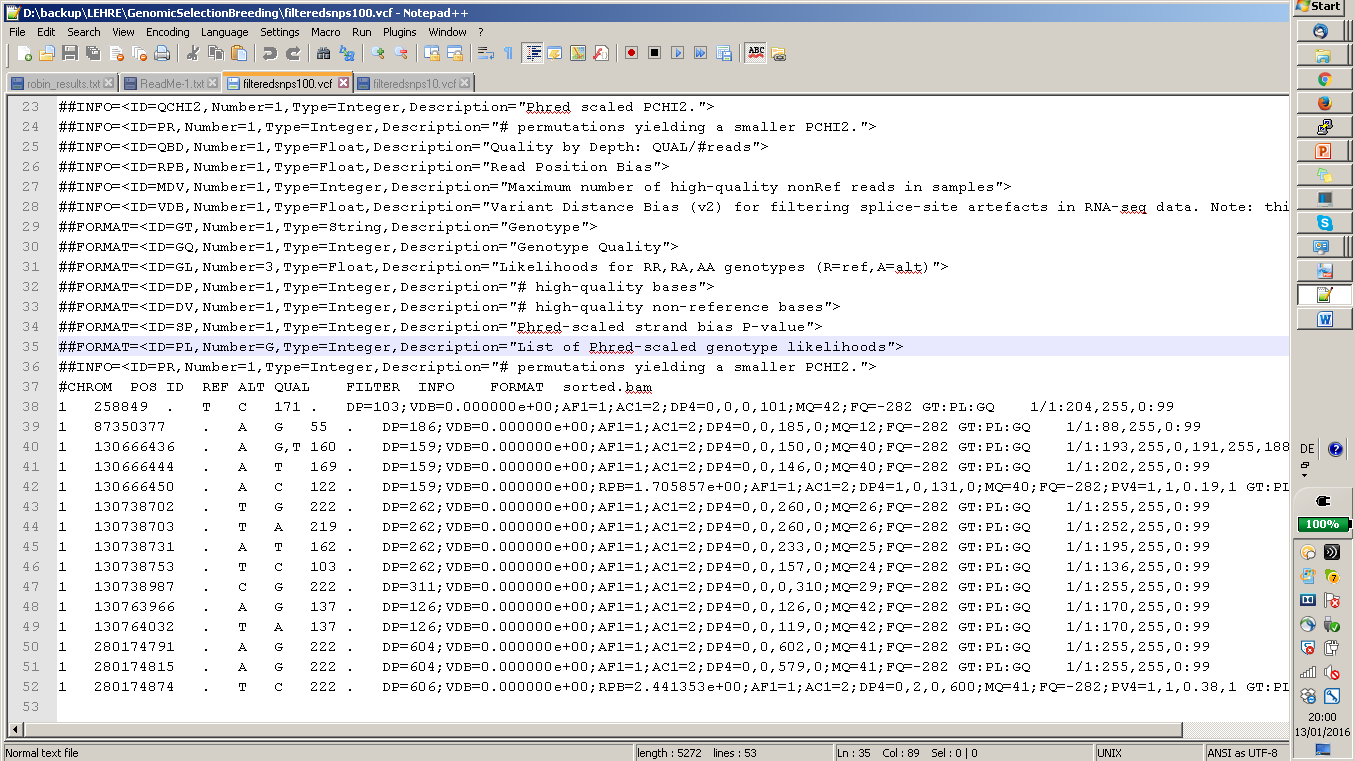
We can fix this by applying a filter

1. cat snps.vcf |vcfutils.pl varFilter -d 20 > filteredsnps.vcf

cat snps.txt just shows the file and vcfutils.pl varFilter only extracts those snps which have a coverage of at least 20

Checking our SNPs

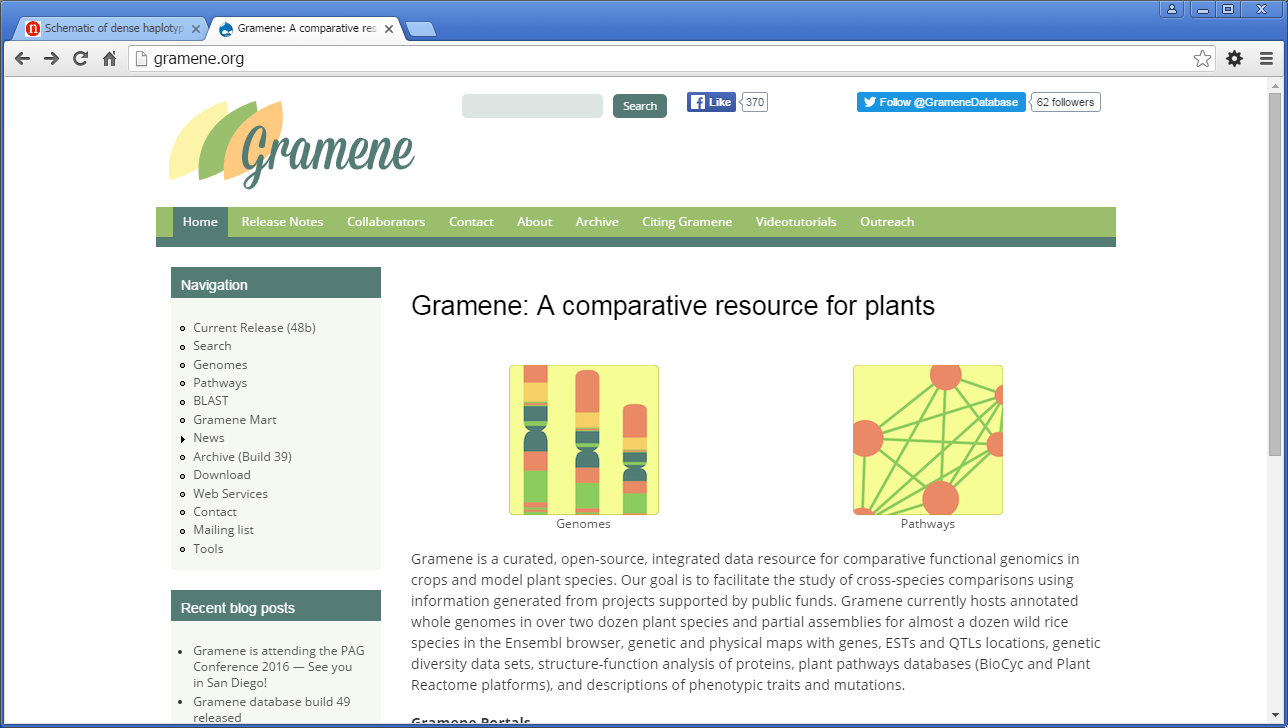
Our highly filtered SNPs for the genotype M0114 (barcoded CCACAA) looks something like this



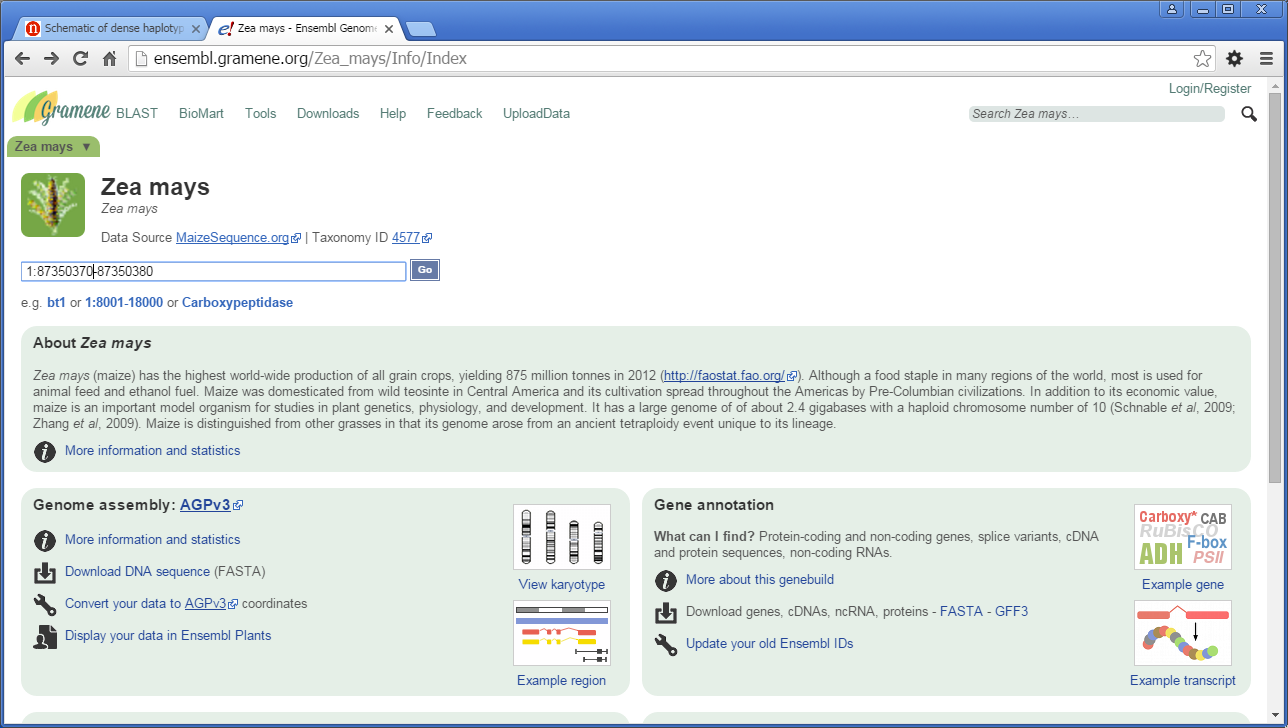
As we can see all of these SNPs have a high coverage now. First we will focus on the SNP at position 87350377 on Chromosome 1. We will use an ensembl style genome browser.

SNPs on Gramene

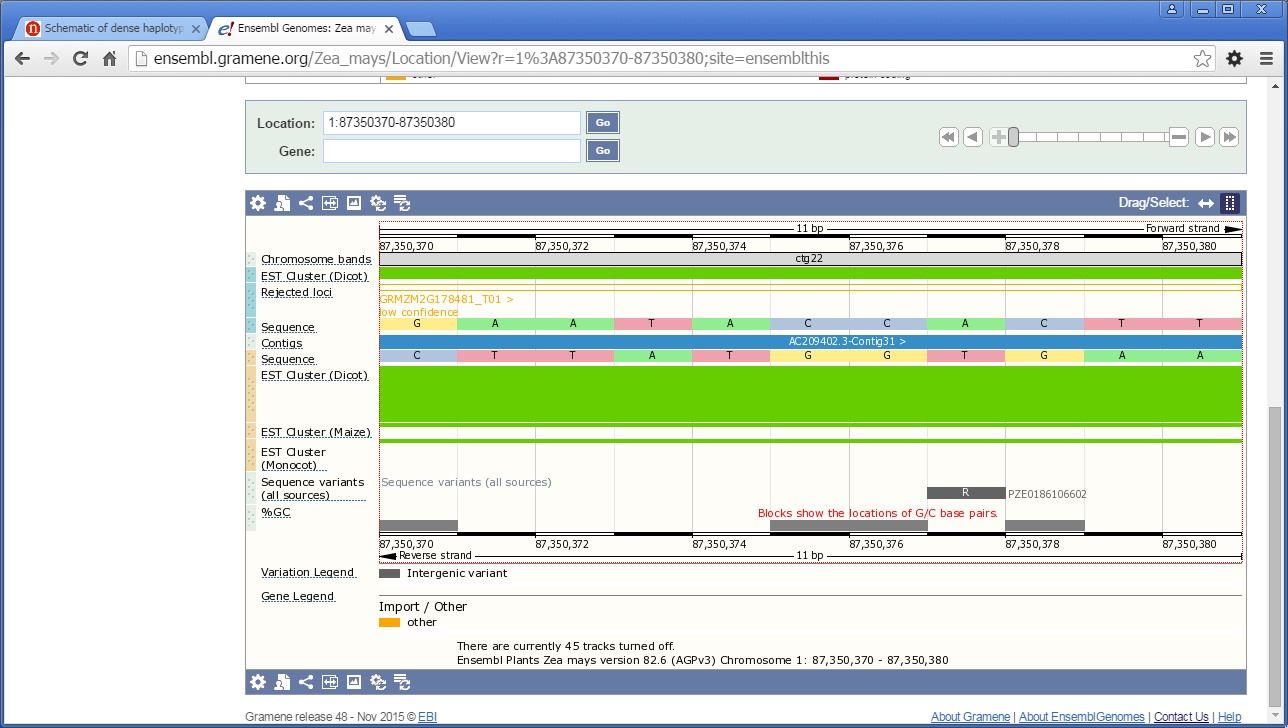
Got to Gramene.org and select Genomes in the menu on the left.



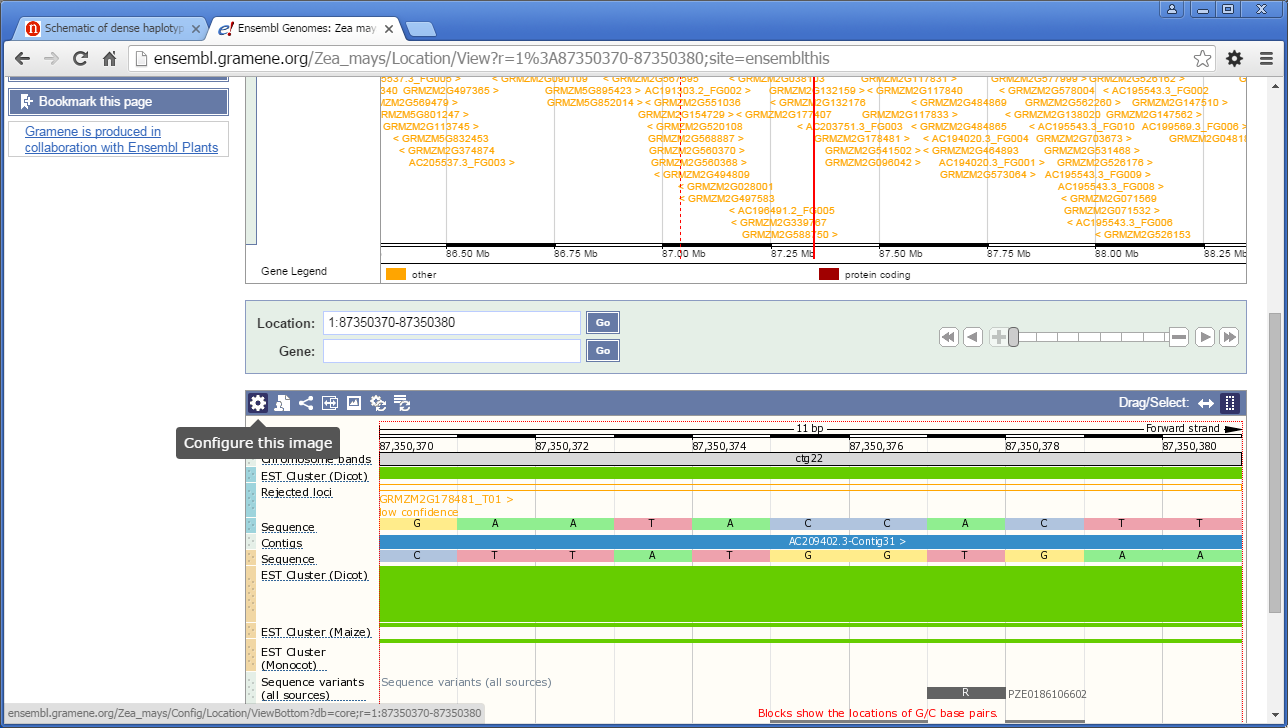
For the genomes select corn (Zea mays). On the following page we zoom in on the region where our SNP is supposed to be type into the box 1:8735070-8735080 (our SNP was supposed to be at 87350377)



In the resulting view scroll down

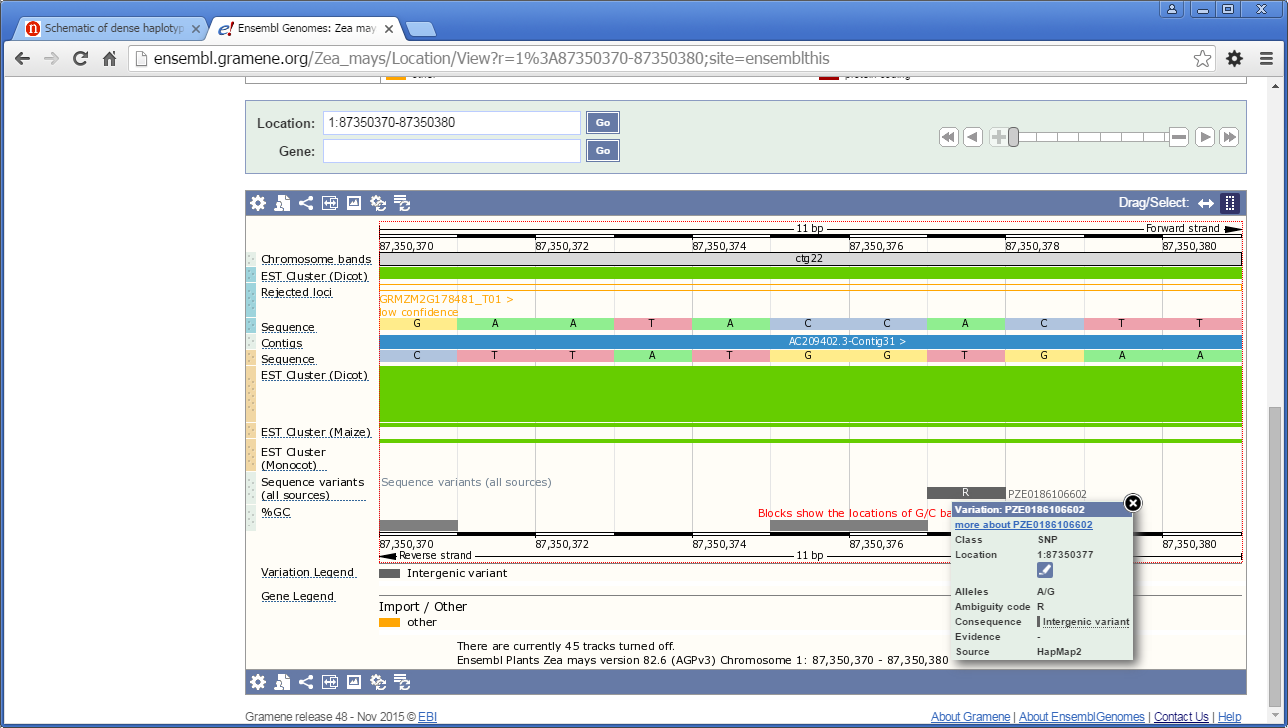


Here you should see a track called “sequence variants” if not click on this icon in the upper right corner



And turn the variant track on.

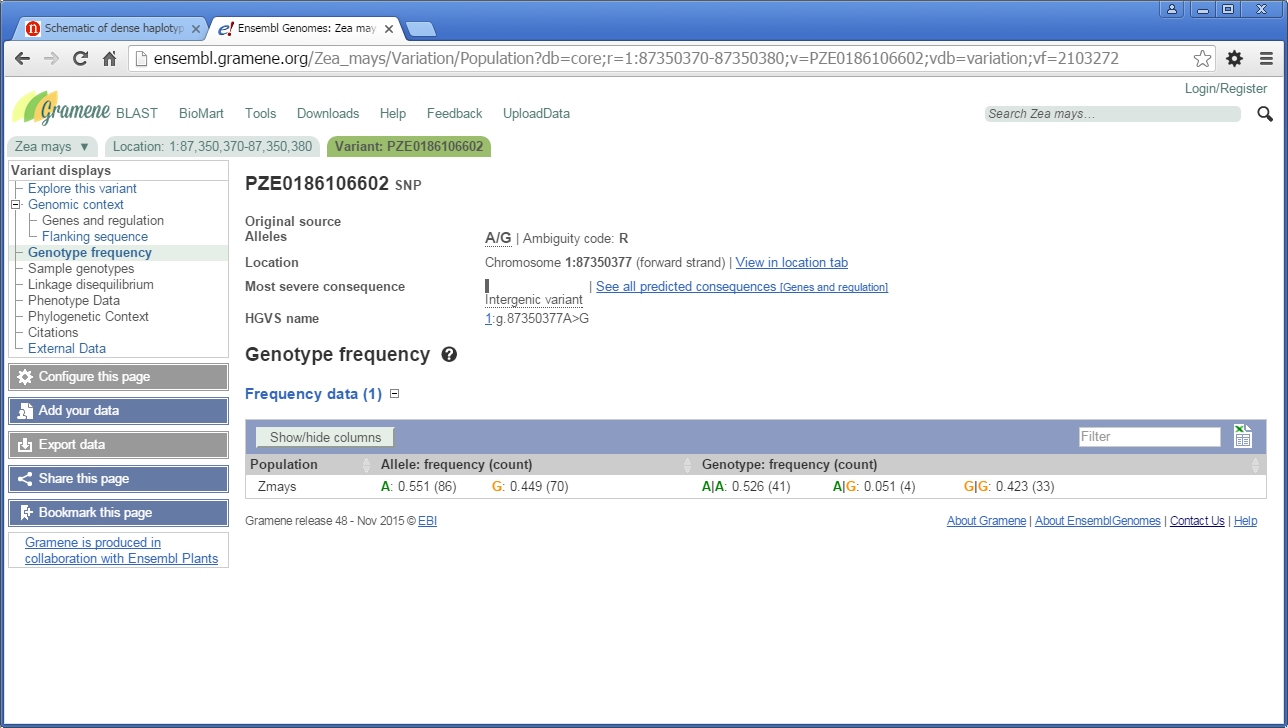
We see one variant labeled with R clicking on it yields this information window:



We can explore more info now. (more about PZE0186106602).

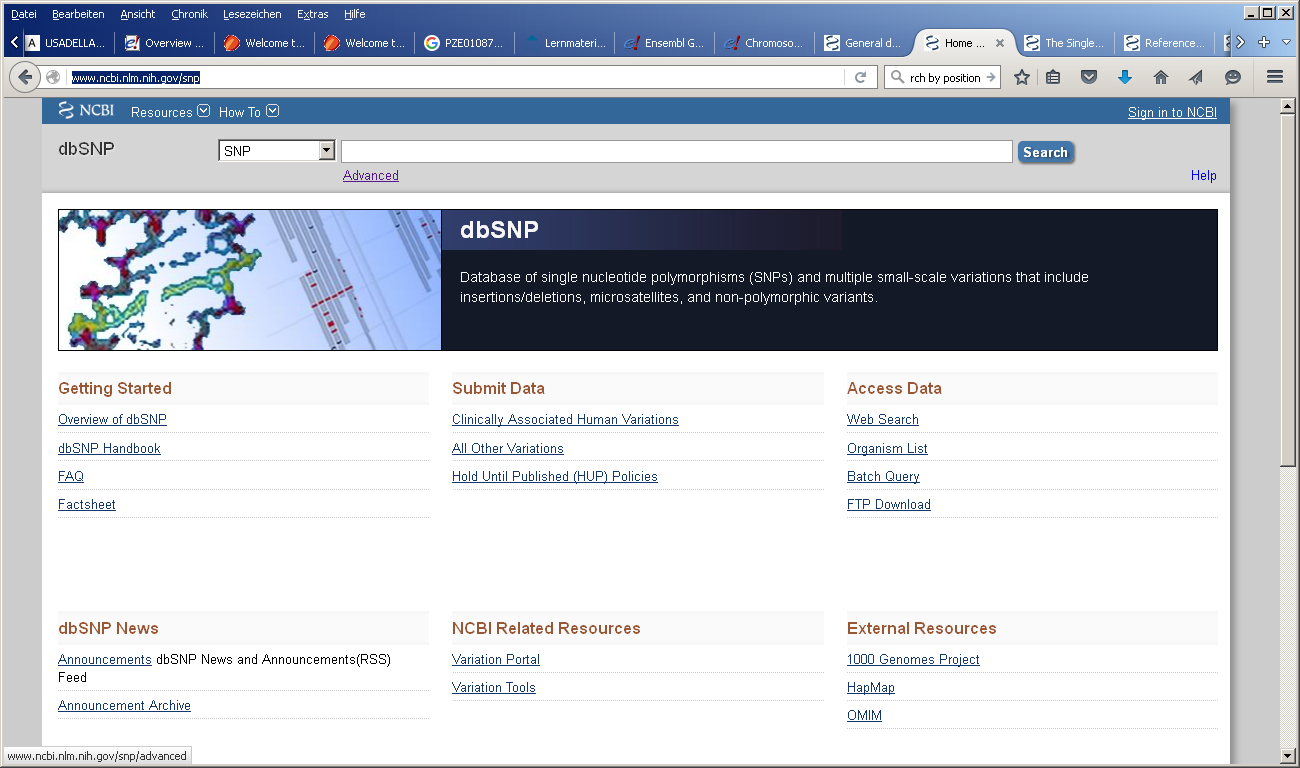


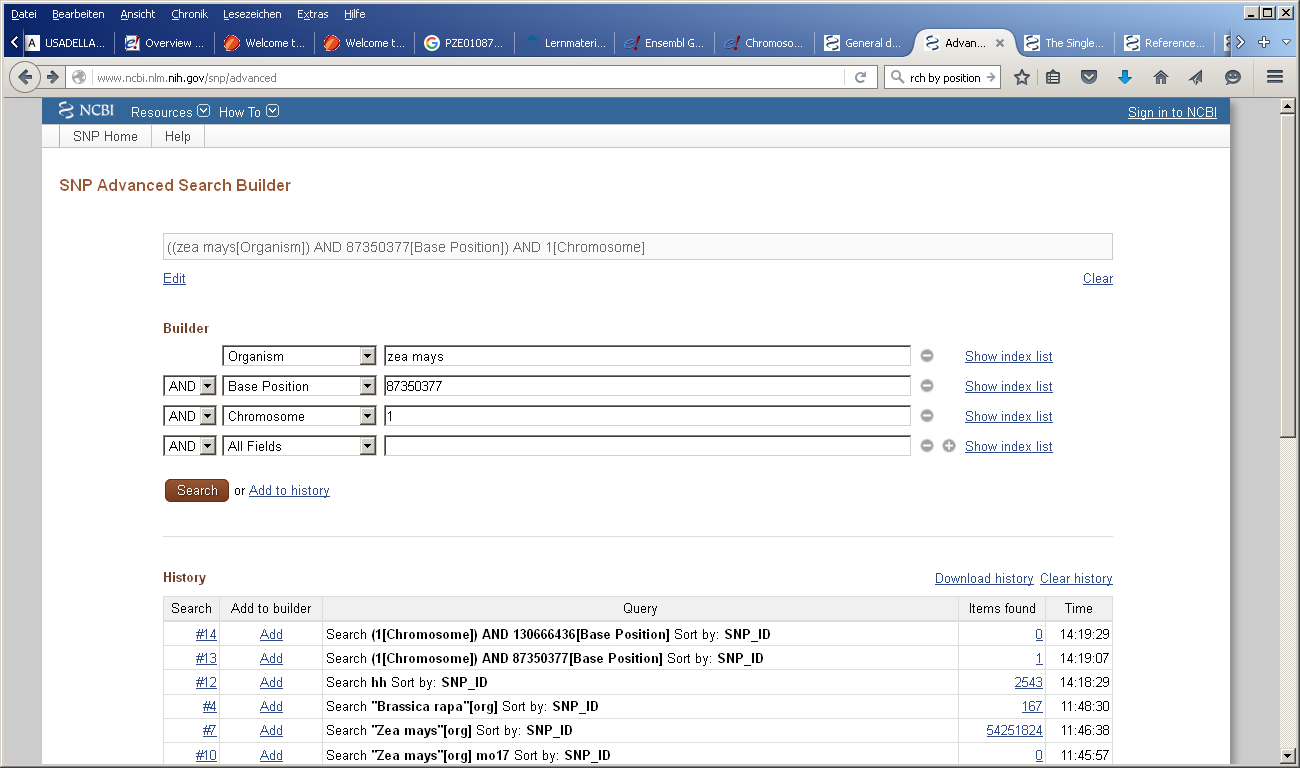
Genomic context shows us SNPs around and genotype frequency shows us the frequency. AA and AG are both highly probably whereas A|G is less likely.



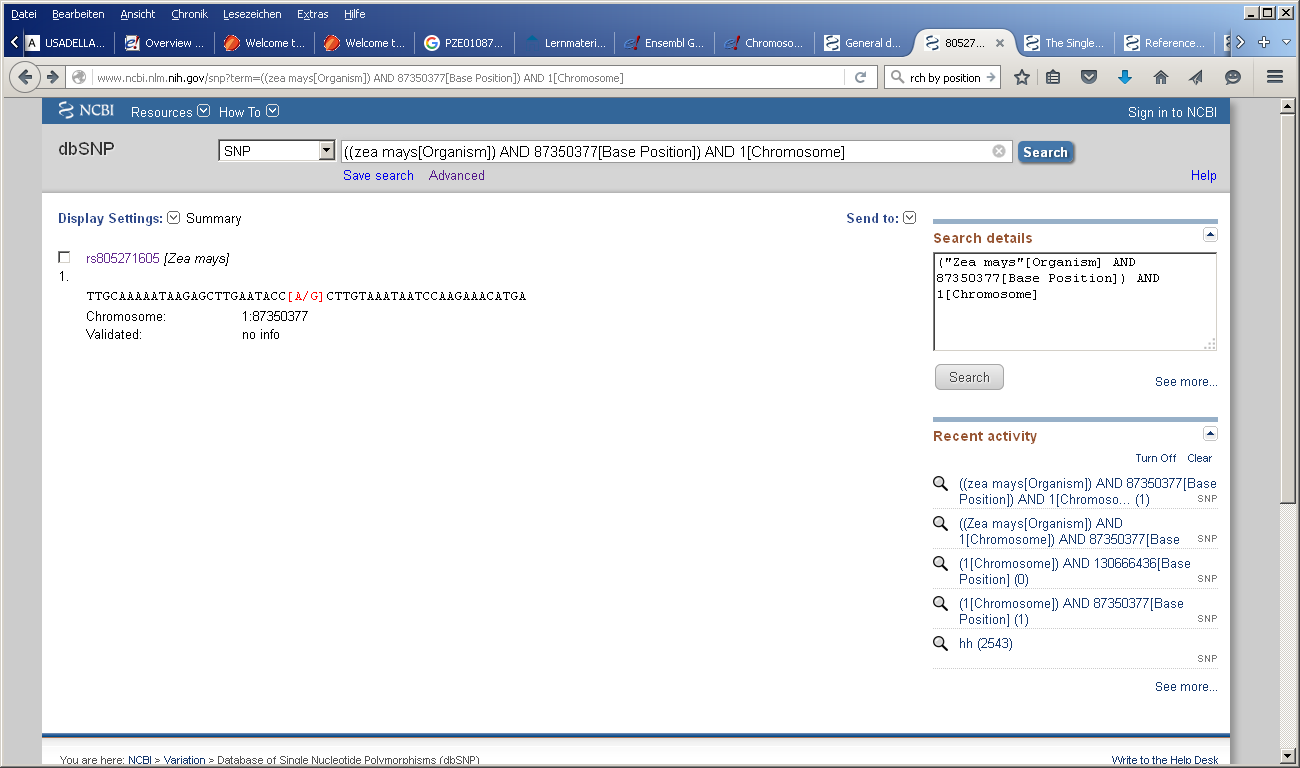
SNPs on dbSNP

Let’s use dbSNP for the same SNP go to <http://www.ncbi.nlm.nih.gov/snp> and go for advanced



In the resulting screen choose Organism as Zea mays, Base position as 87350377 and chromosome as 1 

And we see that our SNP is also in dbSNP.



Phasing Example Analysis

Phasing

One application of the SNPs derived from (RAD)seq is inferring haplotypes. This can use statistical inference from the population data and/or reads that overlap mutiple SNPs and while many tools are available, it can also be performed with samtools. However since we use (connected) reads your phasing blocks in RADseq would be very small. (What would be a better application than RADSeq)

1. # for\_phasing.bam has longer, paired reads that we mapped previously
2. samtools phase for\_phasing.bam > phased.txt
3. # view results
4. less phased.txt

Copyright

This workshop script is copyrighted under terms of the creative commons license for attribution 3.0. (<http://creativecommons.org/licenses/by/3.0/>)  
That means you are free to share, copy and modify this manuscript for any purpose, commercially as well as non-commercially, under the condition that you attribute the authors:   
Björn Usadel & Alisandra Denton

Parts of this script (linux comammand line are based on RNA seq data written by Simon Schliesky  
& Andrea Bräutigam) and variant calling input from Marie and Tony Bolger.  
  
Legal code of the license:

THE WORK (AS DEFINED BELOW) IS PROVIDED UNDER THE TERMS OF THIS CREATIVE COMMONS PUBLIC LICENSE ("CCPL" OR "LICENSE"). THE WORK IS PROTECTED BY COPYRIGHT AND/OR OTHER APPLICABLE LAW. ANY USE OF THE WORK OTHER THAN AS AUTHORIZED UNDER THIS LICENSE OR COPYRIGHT LAW IS PROHIBITED.

BY EXERCISING ANY RIGHTS TO THE WORK PROVIDED HERE, YOU ACCEPT AND AGREE TO BE BOUND BY THE TERMS OF THIS LICENSE. TO THE EXTENT THIS LICENSE MAY BE CONSIDERED TO BE A CONTRACT, THE LICENSOR GRANTS YOU THE RIGHTS CONTAINED HERE IN CONSIDERATION OF YOUR ACCEPTANCE OF SUCH TERMS AND CONDITIONS.

**1. Definitions**

1. **"Adaptation"** means a work based upon the Work, or upon the Work and other pre-existing works, such as a translation, adaptation, derivative work, arrangement of music or other alterations of a literary or artistic work, or phonogram or performance and includes cinematographic adaptations or any other form in which the Work may be recast, transformed, or adapted including in any form recognizably derived from the original, except that a work that constitutes a Collection will not be considered an Adaptation for the purpose of this License. For the avoidance of doubt, where the Work is a musical work, performance or phonogram, the synchronization of the Work in timed-relation with a moving image ("synching") will be considered an Adaptation for the purpose of this License.
2. **"Collection"** means a collection of literary or artistic works, such as encyclopedias and anthologies, or performances, phonograms or broadcasts, or other works or subject matter other than works listed in Section 1(f) below, which, by reason of the selection and arrangement of their contents, constitute intellectual creations, in which the Work is included in its entirety in unmodified form along with one or more other contributions, each constituting separate and independent works in themselves, which together are assembled into a collective whole. A work that constitutes a Collection will not be considered an Adaptation (as defined above) for the purposes of this License.
3. **"Distribute"** means to make available to the public the original and copies of the Work or Adaptation, as appropriate, through sale or other transfer of ownership.
4. **"Licensor"** means the individual, individuals, entity or entities that offer(s) the Work under the terms of this License.
5. **"Original Author"** means, in the case of a literary or artistic work, the individual, individuals, entity or entities who created the Work or if no individual or entity can be identified, the publisher; and in addition (i) in the case of a performance the actors, singers, musicians, dancers, and other persons who act, sing, deliver, declaim, play in, interpret or otherwise perform literary or artistic works or expressions of folklore; (ii) in the case of a phonogram the producer being the person or legal entity who first fixes the sounds of a performance or other sounds; and, (iii) in the case of broadcasts, the organization that transmits the broadcast.
6. **"Work"** means the literary and/or artistic work offered under the terms of this License including without limitation any production in the literary, scientific and artistic domain, whatever may be the mode or form of its expression including digital form, such as a book, pamphlet and other writing; a lecture, address, sermon or other work of the same nature; a dramatic or dramatico-musical work; a choreographic work or entertainment in dumb show; a musical composition with or without words; a cinematographic work to which are assimilated works expressed by a process analogous to cinematography; a work of drawing, painting, architecture, sculpture, engraving or lithography; a photographic work to which are assimilated works expressed by a process analogous to photography; a work of applied art; an illustration, map, plan, sketch or three-dimensional work relative to geography, topography, architecture or science; a performance; a broadcast; a phonogram; a compilation of data to the extent it is protected as a copyrightable work; or a work performed by a variety or circus performer to the extent it is not otherwise considered a literary or artistic work.
7. **"You"** means an individual or entity exercising rights under this License who has not previously violated the terms of this License with respect to the Work, or who has received express permission from the Licensor to exercise rights under this License despite a previous violation.
8. **"Publicly Perform"** means to perform public recitations of the Work and to communicate to the public those public recitations, by any means or process, including by wire or wireless means or public digital performances; to make available to the public Works in such a way that members of the public may access these Works from a place and at a place individually chosen by them; to perform the Work to the public by any means or process and the communication to the public of the performances of the Work, including by public digital performance; to broadcast and rebroadcast the Work by any means including signs, sounds or images.
9. **"Reproduce"** means to make copies of the Work by any means including without limitation by sound or visual recordings and the right of fixation and reproducing fixations of the Work, including storage of a protected performance or phonogram in digital form or other electronic medium.

**2. Fair Dealing Rights.** Nothing in this License is intended to reduce, limit, or restrict any uses free from copyright or rights arising from limitations or exceptions that are provided for in connection with the copyright protection under copyright law or other applicable laws.

**3. License Grant.** Subject to the terms and conditions of this License, Licensor hereby grants You a worldwide, royalty-free, non-exclusive, perpetual (for the duration of the applicable copyright) license to exercise the rights in the Work as stated below:

1. to Reproduce the Work, to incorporate the Work into one or more Collections, and to Reproduce the Work as incorporated in the Collections;
2. to create and Reproduce Adaptations provided that any such Adaptation, including any translation in any medium, takes reasonable steps to clearly label, demarcate or otherwise identify that changes were made to the original Work. For example, a translation could be marked "The original work was translated from English to Spanish," or a modification could indicate "The original work has been modified.";
3. to Distribute and Publicly Perform the Work including as incorporated in Collections; and,
4. to Distribute and Publicly Perform Adaptations.
5. For the avoidance of doubt:
   1. **Non-waivable Compulsory License Schemes**. In those jurisdictions in which the right to collect royalties through any statutory or compulsory licensing scheme cannot be waived, the Licensor reserves the exclusive right to collect such royalties for any exercise by You of the rights granted under this License;
   2. **Waivable Compulsory License Schemes**. In those jurisdictions in which the right to collect royalties through any statutory or compulsory licensing scheme can be waived, the Licensor waives the exclusive right to collect such royalties for any exercise by You of the rights granted under this License; and,
   3. **Voluntary License Schemes**. The Licensor waives the right to collect royalties, whether individually or, in the event that the Licensor is a member of a collecting society that administers voluntary licensing schemes, via that society, from any exercise by You of the rights granted under this License.

The above rights may be exercised in all media and formats whether now known or hereafter devised. The above rights include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. Subject to Section 8(f), all rights not expressly granted by Licensor are hereby reserved.

**4. Restrictions.** The license granted in Section 3 above is expressly made subject to and limited by the following restrictions:

1. You may Distribute or Publicly Perform the Work only under the terms of this License. You must include a copy of, or the Uniform Resource Identifier (URI) for, this License with every copy of the Work You Distribute or Publicly Perform. You may not offer or impose any terms on the Work that restrict the terms of this License or the ability of the recipient of the Work to exercise the rights granted to that recipient under the terms of the License. You may not sublicense the Work. You must keep intact all notices that refer to this License and to the disclaimer of warranties with every copy of the Work You Distribute or Publicly Perform. When You Distribute or Publicly Perform the Work, You may not impose any effective technological measures on the Work that restrict the ability of a recipient of the Work from You to exercise the rights granted to that recipient under the terms of the License. This Section 4(a) applies to the Work as incorporated in a Collection, but this does not require the Collection apart from the Work itself to be made subject to the terms of this License. If You create a Collection, upon notice from any Licensor You must, to the extent practicable, remove from the Collection any credit as required by Section 4(b), as requested. If You create an Adaptation, upon notice from any Licensor You must, to the extent practicable, remove from the Adaptation any credit as required by Section 4(b), as requested.
2. If You Distribute, or Publicly Perform the Work or any Adaptations or Collections, You must, unless a request has been made pursuant to Section 4(a), keep intact all copyright notices for the Work and provide, reasonable to the medium or means You are utilizing: (i) the name of the Original Author (or pseudonym, if applicable) if supplied, and/or if the Original Author and/or Licensor designate another party or parties (e.g., a sponsor institute, publishing entity, journal) for attribution ("Attribution Parties") in Licensor's copyright notice, terms of service or by other reasonable means, the name of such party or parties; (ii) the title of the Work if supplied; (iii) to the extent reasonably practicable, the URI, if any, that Licensor specifies to be associated with the Work, unless such URI does not refer to the copyright notice or licensing information for the Work; and (iv) , consistent with Section 3(b), in the case of an Adaptation, a credit identifying the use of the Work in the Adaptation (e.g., "French translation of the Work by Original Author," or "Screenplay based on original Work by Original Author"). The credit required by this Section 4 (b) may be implemented in any reasonable manner; provided, however, that in the case of a Adaptation or Collection, at a minimum such credit will appear, if a credit for all contributing authors of the Adaptation or Collection appears, then as part of these credits and in a manner at least as prominent as the credits for the other contributing authors. For the avoidance of doubt, You may only use the credit required by this Section for the purpose of attribution in the manner set out above and, by exercising Your rights under this License, You may not implicitly or explicitly assert or imply any connection with, sponsorship or endorsement by the Original Author, Licensor and/or Attribution Parties, as appropriate, of You or Your use of the Work, without the separate, express prior written permission of the Original Author, Licensor and/or Attribution Parties.
3. Except as otherwise agreed in writing by the Licensor or as may be otherwise permitted by applicable law, if You Reproduce, Distribute or Publicly Perform the Work either by itself or as part of any Adaptations or Collections, You must not distort, mutilate, modify or take other derogatory action in relation to the Work which would be prejudicial to the Original Author's honor or reputation. Licensor agrees that in those jurisdictions (e.g. Japan), in which any exercise of the right granted in Section 3(b) of this License (the right to make Adaptations) would be deemed to be a distortion, mutilation, modification or other derogatory action prejudicial to the Original Author's honor and reputation, the Licensor will waive or not assert, as appropriate, this Section, to the fullest extent permitted by the applicable national law, to enable You to reasonably exercise Your right under Section 3(b) of this License (right to make Adaptations) but not otherwise.

**5. Representations, Warranties and Disclaimer**

UNLESS OTHERWISE MUTUALLY AGREED TO BY THE PARTIES IN WRITING, LICENSOR OFFERS THE WORK AS-IS AND MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND CONCERNING THE WORK, EXPRESS, IMPLIED, STATUTORY OR OTHERWISE, INCLUDING, WITHOUT LIMITATION, WARRANTIES OF TITLE, MERCHANTIBILITY, FITNESS FOR A PARTICULAR PURPOSE, NONINFRINGEMENT, OR THE ABSENCE OF LATENT OR OTHER DEFECTS, ACCURACY, OR THE PRESENCE OF ABSENCE OF ERRORS, WHETHER OR NOT DISCOVERABLE. SOME JURISDICTIONS DO NOT ALLOW THE EXCLUSION OF IMPLIED WARRANTIES, SO SUCH EXCLUSION MAY NOT APPLY TO YOU.

**6. Limitation on Liability.** EXCEPT TO THE EXTENT REQUIRED BY APPLICABLE LAW, IN NO EVENT WILL LICENSOR BE LIABLE TO YOU ON ANY LEGAL THEORY FOR ANY SPECIAL, INCIDENTAL, CONSEQUENTIAL, PUNITIVE OR EXEMPLARY DAMAGES ARISING OUT OF THIS LICENSE OR THE USE OF THE WORK, EVEN IF LICENSOR HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

**7. Termination**

1. This License and the rights granted hereunder will terminate automatically upon any breach by You of the terms of this License. Individuals or entities who have received Adaptations or Collections from You under this License, however, will not have their licenses terminated provided such individuals or entities remain in full compliance with those licenses. Sections 1, 2, 5, 6, 7, and 8 will survive any termination of this License.
2. Subject to the above terms and conditions, the license granted here is perpetual (for the duration of the applicable copyright in the Work). Notwithstanding the above, Licensor reserves the right to release the Work under different license terms or to stop distributing the Work at any time; provided, however that any such election will not serve to withdraw this License (or any other license that has been, or is required to be, granted under the terms of this License), and this License will continue in full force and effect unless terminated as stated above.

**8. Miscellaneous**

1. Each time You Distribute or Publicly Perform the Work or a Collection, the Licensor offers to the recipient a license to the Work on the same terms and conditions as the license granted to You under this License.
2. Each time You Distribute or Publicly Perform an Adaptation, Licensor offers to the recipient a license to the original Work on the same terms and conditions as the license granted to You under this License.
3. If any provision of this License is invalid or unenforceable under applicable law, it shall not affect the validity or enforceability of the remainder of the terms of this License, and without further action by the parties to this agreement, such provision shall be reformed to the minimum extent necessary to make such provision valid and enforceable.
4. No term or provision of this License shall be deemed waived and no breach consented to unless such waiver or consent shall be in writing and signed by the party to be charged with such waiver or consent.
5. This License constitutes the entire agreement between the parties with respect to the Work licensed here. There are no understandings, agreements or representations with respect to the Work not specified here. Licensor shall not be bound by any additional provisions that may appear in any communication from You. This License may not be modified without the mutual written agreement of the Licensor and You.
6. The rights granted under, and the subject matter referenced, in this License were drafted utilizing the terminology of the Berne Convention for the Protection of Literary and Artistic Works (as amended on September 28, 1979), the Rome Convention of 1961, the WIPO Copyright Treaty of 1996, the WIPO Performances and Phonograms Treaty of 1996 and the Universal Copyright Convention (as revised on July 24, 1971). These rights and subject matter take effect in the relevant jurisdiction in which the License terms are sought to be enforced according to the corresponding provisions of the implementation of those treaty provisions in the applicable national law. If the standard suite of rights granted under applicable copyright law includes additional rights not granted under this License, such additional rights are deemed to be included in the License; this License is not intended to restrict the license of any rights under applicable law.